FACTA UNIVERSITATIS (NIŠ)
Series: Electronics and Energetics vol.13, No.1, April 2000, 109–130

MPEG AND ITU-T VIDEO COMMUNICATION: STANDARDIZATION PROCESS

Zoran S. Bojković

Abstract. This paper seeks to provide MPEG (MPEG-2, MPEG-4, MPEG-7) and ITU-T (H.263, H.263+, H.263++, H.26L) standardization processes. These developed standards explore every possibility of the digital environment.

Key words: MPEG, ITU-T, bitstream, standardization, processing chain, applications.

1. Introduction

Modern video compression techniques offer the possibility to store or transmit the vast amount of data necessary to represent digital video in an efficient and robust way [1], [2]. New audio-visual applications in the fields of communication, multimedia and broadcasting became possible based on digital video technology. With the advances in VLSI technology, it has become possible to open more application fields to a large number of users and, therefore the necessity for digital video standards arose. Video standards have to rely on compromises between what is theoretically possible and what is technologically feasible. Commercially, international standardization of video communication systems and protocols aims to serve two important purposes: interoperability and economy of scale. Namely, interworking between video communication equipment from different vendors is a desirable feature for users and equipment manufacturers. It enables large scale international video data exchange via storage media or communication networks. An increased demand leads to economy of scale i.e, the mass production of VLSI systems and devices which in turn makes video equipment more affordable for a wide firld of applications and users [3]. Standards can

Manuscript received March 24, 2000. A version of this paper was presented at the fourth IEEE Conference on Telecommunications in Modern Satellite, Cables and Broadcasting Services, TELSIKS'99, October 1999, Niš, Serbia.

The author is with Faculty of Transport and Traffic Engineering, University of Belgrade, Vojvode Stepe 305, 11000 Belgrad, Serbia, e-mail zoranb@sf.sf.bg.ac.yu.

only be successful in the market place if the cost performance ratio is well balanced. This is specifically true in the field of video coding where a large variety of innovative coding algorithms exist, but may be too complex for implementation.

From the beginning of 1980, a number of international video standardization activities has started with CCITT (now ITU-T), followed by working to develop international standards for digital audio and video transmission and storage. Unlike H.261 which is optimized for teleconferencing where motion is naturally limited, the ISO/IEC MPEG video standard is devised for a wide range of video and motion pictures. The goal of MPEG is to define a standard for coded representation of moving pictures associated audio and the systems [2], [4], [5]. It defines the structure of coded video bit streams transmitted to the decoder but leaves the encoder architecture undefined, so that some encoders may produce higher quality compression, some will be real time and some will require manual intervention. MPEG compression is lossy and asymmetric with the encoding more complex than the decoding. To support the wide range of application profiles, a diversity of input parameters including flexible picture size and frame rate can be specified by the user.

There are different versions of MPEG, denoted by MPEG-x. The first version of MPEG-1 was targeted at CD-ROM and applications at a bitrate of about 1,5 Mb/s. MPEG-1 has also proved useful in other applications, such as for early direct broadcast satellite transmission and for computer-generated video, where transmission bandwidth and storage capacity are limited or expansive. MPEG-2 addresses high-quality coding for all digital transmission of broadcast-TV-quality video at data rates of 2-50 Mb/s. The major applications include digital storage media, digital television including HDTV, broadcasting over cable, satellite and other broadcast channels, as well as other communication applications [6]. MPEG-2 can produce the video quality needed for multimedia entertainment piped to the home and for more demanding business and scientific applications, too. MPEG-4 has a different set of objectives to provide video on low bandwidth transmission lines or for low capacity storage devices. Besides traditional block/based coding as in MPEG-1 and 2, H.261, H.263 etc, MPEG-4 addresses object based and content based coding with composition, manipulation interactivity etc. It also extends to both synthetic and natural video/audio, layered coding, multi-view, shape/texture/motion coding of objects, animation and all possible combination of these features. Its roles extend to Internet, Web TV, large databases (storage, retrieval, transmission), mobile networks, multimedia, electronic games, education, training, entertainment, defense, medicine, etc. The new number of the MPEG family called "Multimedia Content Description Interface", is MPEG-7 [3], [4]. It will extend the limited current search capabilities to include more information types, such as video, image, audio, graphics, animation, etc. In other words, MPEG-7 will specify a standardized description of various types of multimedia information. This description will be associated with the content itself, to allow fast and efficient searching for multimedia that is of a user's interest.

The H.263 was developed for video compression at rates below 64 kb/s. The key technical features of H.263 are available: block-size motion compensation, overlapped -block motion compensation, pixel extrapolating motion vectors, 3- dimensional run-level-last variable-length coding, median vector prediction and more efficient header information signalling. In this paper, MPEG and ITU-T video communication standardization processes will be presented. MPEG-2, MPEG-4, MPEG-7 as well as H.263, H.263+, H.263++, H.26L standards will be analyzed separately.

2. MPEG Bitstream Syntax

The MPEG bitstream syntax is more flexible than ITU-T Recommendation H.261, reflecting the wider variety of applications anticipated from MPEG. Table 1 shows the MPEG bitstream syntax. As can be seen, the MPEG bitstream syntax is constructed in several hierarchical layers with different functions. The MPEG syntax specifies that the variable length coding of video may start and end on slice boundaries. This arrangement prevents errors from propagating during the decoding process by forcing the decoder to start processing variable length video data from a known state at the beginning of every slice.

Table 1. MPEG bitstream syntax

Syntax	Function
Video sequence layer	Global context unit: video frame size, frame rate, bitrate, minimum decoder buffer size, constrained parameters information
Group-of-pictures layer	Video coding unit: random access, search, editing
Picture layer	Frame coding unit: type (I,B,P), displayorder position
Slice layer	Resynchronization unit: a string of macroblocks in rasterscan order used for resynchronization during decoding
Macroblock layer	Motion compensation unit: a 16×16 block
Block layer	DCT unit: an 8×8 block

3. MPEG-2 Video Standard

Both MPEG-1 and MPEG-2 are the video coding standards for television based applications in the 1990's. Since the MPEG-1 standard was intended for audio-visual coding for digital storage media (DMS) applications and since DSMs typically have very low or negligible bit error rates, the MPEG-1 system part was not designed to be highly robust to bit errors. Also, the MPEG-1 system was intended for software oriented processing, and thus large variable length packet was preferred to minimize software overhead. A key factor for world-wide success of MPEG-1 is the generic structure of the standard which supports a broad range of applications and application specific parameters. MPEG-1 is targeted for small picture resolutions (352-by-240 or 352-by-288 SIF resolution, with SIF standing for source input format). Emerging applications, such as digital cable TV distribution, networked database services via ATM, and satellite and terrestrial digital broadcasting distribution were seen to benefit from the increased quality resulting from the MPEG-2 standard. Work was carried out in collaboration with the ITU-T SG 15 Experts Group for ATM video coding and in 1994 the MPEG-2 Draft International Standard was released. It is identical to the ITU-T H.262 Recommendation [7,8]. The specification of the standard is intended to be generic. Hence, the standard aims to facilitate the bitstream interchange among different applications and transmission and storage media.

Fig. 1 depicts the general block of a multiscale video coding scheme. As can be seen, two layers are provided, each layer supporting video at a different scale. Multiresolution representation can be achieved by downscaling the input video signal into a lower resolution video. The downscaled version is encoded into a base layer bitstream with reduced bitrate. The upscaled reconstructed base layer video (upsampled spatially or temporally) is used as a prediction for the coding of the original input video signal. The prediction error is encoded into an enhancement layer bitstream. The display of video at highest resolution with reduced quality is possible by decoding the lower bitrate base layer. Thus, scalable coding can be used to encode video with a suitable bitrate allocated to each layer in order to meet specific bandwidth requirements of transmission channels or storage media.

MPEG-2 has been an international standard since November 1994. In fact, it is at the commercially applicable stage and has been adopted for a number of application areas. There is a strong commitment to use this standard. Implementation of MPEG-2 is possible in several ways: software-only, hardware-assisted, decoder-only and digital signal processing-based encoder/decoder implementation. A number of the MPEG-2 decoders became

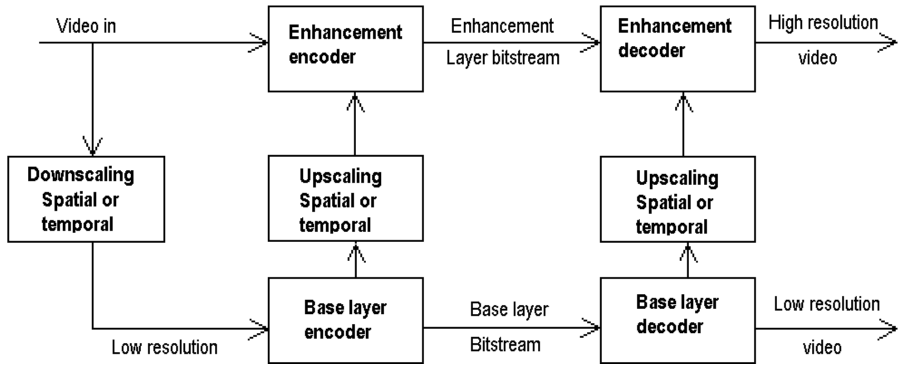

Fig.1 Block scheme of a multiscale video coding scheme.

commercially available towards the end of 1994. The world wide acceptance of MPEG-2 in consumer electronics has lead to large production scales making MPEG-2 decoder equipment affordable and therefore also attractive for other related areas like video communications and storage as well as multimedia applications in general. Because MPEG-2 video provides higher quality but is too complex for software-only playback, hardware-assisted codecs provide substantially better image quality and better coding and playback performance than software-only video codecs.

4. MPEG-4 Video Standardization

New types of networks are carrying audiovisual (AV) information from mobile narrowband to broadband while hardware and software technologies continue to progress. On the other hand, new developments taking place in production and handling of AV data have given rise to new demands on AV coding standards. The MPEG-1 and MPEG-2 AV coding standards are successful because they allow digital AV service with high performance, both concerning quality and compression efficiency. However, these standards are limited to a representation of AV information where video is a sequence of frames with a certain number of lines [9].

The MPEG-4 standard is fundamentally different in nature from its predecessors, and it makes the move towards representing the scene as a composition of objects, rather than just the pixels [10], [11]. ISO MPEG-4 started its standardization in July 1993 with the character to develop a generic video coding algorithm mainly targeted for a wide range of low bitrate multimedia applications. MPEG-4 is still a standard in the making, but its effects are already being felt. Starting from original goal of providing an audiovisual coding standard for very low bitrate channels, such as found in mobile applications or today's Internet, MPEG-4 is evaluated to a complex toolkit. The MPEG-4 defined composition technology is called binary format for scene description which achieves the important goal of saving bits for the purpose of transmitting scan-composition information. As MPEG-4 deals with objects, it becomes necessary to provide a mechanism enabling the author of a scene to indicate how the different audiovisual objects are composed and for a user to interact with the scene by changing its original composition, if this is allowed by the author. In general, MPEG-4 images as well as image sequences are considered to be arbitrarily shaped in contrast to the standard MPEG-1 and MPEG-2 definitions. The MPEG committee has defined the following set of functionalities in MPEG-4:

a) Content-based multimedia data access tools,

b) Content-based manipulation and bitstream editing,

c) Hybrid natural and synthetic data, both video and audio coding,

d) Improved temporal random access,

e) Improved coding efficiency,

f) Coding a multiple concurrent data stream,

g) Robustness in error-prone environments,

h) High content scalability.

The objective of MPEG-4 is thus to provide an AV representation standard supporting new ways of communication, access and manipulation of digital AV data and offering a common technological solution to various communication paradigms among which the borders are disappearing. MPEG-4 will supply an answer to the emerging needs of application fields, including interactive AV services, games, AV home editing, advanced AV communication services, such as mobile AV terminals, teleshoping, AV on the Internet as well as remote monitoring and control [12]. A further contribution made by MPEG-4 is the delivery multimedia integration format (DMIF). DMIF's goals are twofold: hide the delivery technology details from the DMIF user and ensure the establishment of end-to-end connections. The former called the DMIF application interface is a unified interface that allows an application to access AVO's in a way that is transparent to the type of delivery: a local disk, a two-way channel, or even a broadcast channel. The latter is achieved through DMIF signaling messages.

MPEG-4 combines some of the typical features of other MPEG standards with the new ones coming from existing or anticipated manifestation of multimedia [13]:

- Independence of application from lower layer details, as in the Internet paradigm
- Technology awareness of lower layer characteristics (scalability, error robustness)
- Application software downloadibility as in the network computer paradigm
- Reusability of encoding tools and data
- Interactivity not just with an integral AV bitstream, but with the individual pieces of information

The structure of MPEG-4 is composed of four different elements: syntax, tools, algorithms and profiles [14], [15]. The syntax is an extensible description language that allows for selection, description and downloading of tools, algorithms and profiles. A tool is a technique (contour representation, motion compensation) that is accessible via syntax. An algorithm is organized collection of tools that provides one or more functionalities. A profile is a collection of algorithms constrained in a specific way to address a specific application. Hence, MPEG-4 is defined to offer a flexible syntax and an open set of tools supporting a range of both novel and conventional functionalities that extend beyond just better coding efficiency at very low bitrates. MPEG-4 has also developed a flexible and extensible syntactic object-oriented description language called MPEG-4 Systems and Description Language (MSDL) which not only allows for the description of the bitstream structure, but also configuration and programming of the decoder. MSDL supports the varied set of coding techniques that realize the new functionalities supported by MPEG-4.

For the storage and transmission of AVO data a high coding efficiency meaning a good quality of the reconstructed data is required [16]. Improved coding efficiency is particularly necessary for applications such as video transmission over mobile networks or the Internet.

4.1. Overview of MPEG-4 system

Fig. 2 gives a very general overview of MPEG-4 system. The objects that make up a scene, are sent or stored together with information about their spatio-temporal relationships i.e. composition information. The compositor uses this information to reconstruct the complete scene again. Composition information is used to synchronize different objects in time and to give them the right position in space. Separating this function from the pure decoding of objects introduces the possibility to influence the presentation of a scene, on a screen or through the loudspeaker. Coding different objects separately makes it possible to change the speed of a moving object in the same scene or make it rotate, to influence which objects are sent and with which quality and error protection. In addition, it permits composing a scene with objects that arrive from different locations.

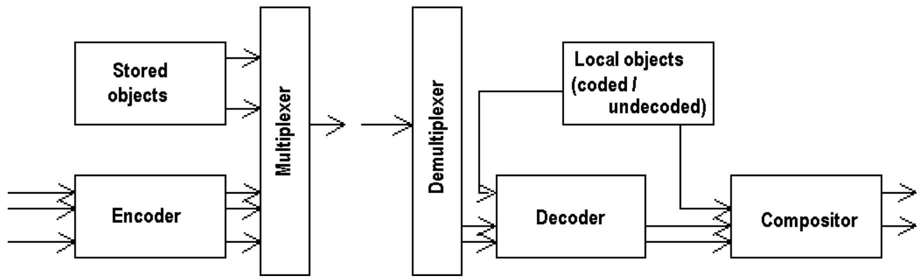

Fig.2 General overview of MPEG-4 system.

The MPEG-4 standard will not prescribe how any given scene is to be organized in objects i.e. segmented. The reason for this is that MPEG-4 decoding standard which does not specify the encoding. The segmentation is usually regarded to take place even before the encoding, as a preprocessor step, which is never a standardization issue. By not specifying preprocessing and encoding, the standard leaves room for manufacturers of systems to distinguish themselves from their competitors by providing a better quality or more options. It also allows use of different encoding strategies for different applications and leaves room for technological progress in analysis and decoding strategies. MPEG-4 is building on the proven success of three fields: digital television, interactive graphics, applications, synthetic content in the World Wide Web and provides the standardized technological elements enabling the integration of the production, distribution and content access paradigms of the three fields.

4.2. Video verification model

The current focus of MPEG-4 video is the development of Video Verification Models (VMs) which evolve through time by means of core experiments. The Verification Model is a common platform with a precise definition of encoding and decoding algorithms, which can be presented as tools addressing specific functionalities. New algorithms/tools are added to the VM and old algorithms/tools are replaced in the VM by successful core experiments [17].

General structure of the encoder in the MPEG-4 video VM is shown in Fig.3 [13]. The significant structure of this encoder is the representation base on video object (VO) defining a visual scene [18]. A user, or an intelligent algorithm may choose to encode different VOs composing a source data with different parameters, different encoding methods, or may even choose not to code some of them. In most applications, each VO represents a semantically meaningful object in the scene. To maintain a certain compatibility with available video materials, each uncompressed VO is represented as a set of Y, U and V components plus information about its shape, stored frame after frame in predefined temporal intervals. Another important feature of the video VM is that the encoder and decoder can therefore function in different frame rates, which do not even need to be constant through the video sequence. Interactivity between the user and the encoder or the decoder can be concluded in different ways. The user may decide to interact at the encoding level, either in coding control to distribute the available bitrate between different VOs or to influence the multiplexing to change parameters such as the composition script at the encoder. The user can also influence the decoding at the demultiplexer by requesting the processing of a portion of the bitstream only, such as the shape [19].

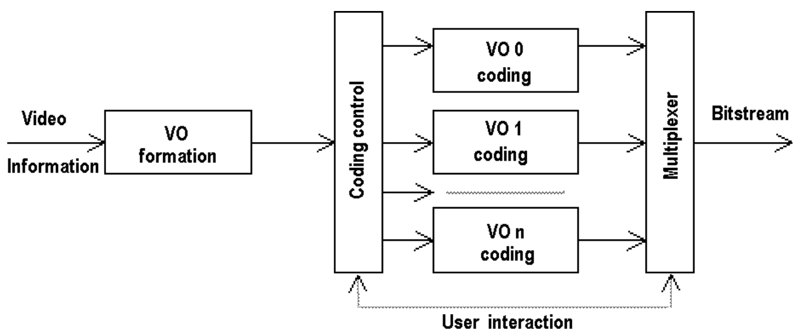

Fig.3 Block diagram of the encoder in the MPEG-4 video VM.

General structure of the decoder in the MPEG-4 video VM is presented in Fig.4. The structure of the decoder is basically similar to that of the encoder in reverse, except for the composition block end. The method of the decomposition of different VOs depends on the applications and the method of multiplexing used at system level. One important issue is that of synchronization among different VOs and other entities such as audio data. In order to fulfill the required functionalities of the video VM, the data structure used in the syntax of encoder/decoder should be designed carefully [20]. The following hierarchy of classes is used in the video VM syntax: Video Session, Video Object, Video Object Layer, and Video Object Plane.

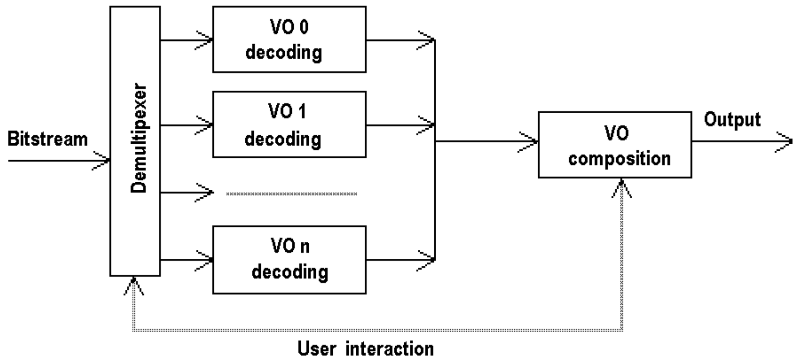

Fig.4 Block diagram of the decoder in the MPEG-4 video VM.

5. MPEG-7 Standardization

The objective of MPEG-7 standardization process is to provide an interoperable solution to extend, the capabilities of today's proprietary solutions in identifying multimedia content. It is expected that MPEG-7 will be a successful standard, which will transform demands from all fields converging to multimedia industry in technical specifications and at the right time. At the same time, this standard has to leave enough free space for further improvement of technical solutions through competition in the market. MPEG-7 work plan is shown in Table 2 [21].

MPEG-7 will specify a standard set of descriptors D that can be used to describe various types of multimedia information. MPEG-7 will also standardize description definition language DDL for definition of other descriptors as well as structures description scheme DS for the descriptors and their relationships. Combination of descriptors and schemes shall be associated with the multimedia content itself, to enable fast end efficient search for material of user's interest. Multimedia material may include: still images, graphics, 3D models, audio, speech, video (both natural and synthetic) and information about how these elements are combined in a multimedia environment. The special types of this general data may include facial expressions or personal characteristics. It is important to emphasize that MPEG descriptors will not depend on the way the content is coded or achieved. Moreover, it is possible to add MPEG-7 descriptor to analog video signal and an image printed on a paper. However, MPEG-7 is to a great extent build around MPEG-4 standard, which codes audio/video material as objects that have certain relation in time (synchronization) and space (location on the scene).

Table 2. MPEG-7 work plan

Requirements	October 1996
Call for proposals	October 1998
WD (Working Draft)	December 1999
CD (Committee Draft)	October 2000
FCD (Final Committee Draft)	February 2001
DIS (Draft International Standard)	July 2001
IS (International Standard)	September 2001

MPEG-7 will support the application of database searching ("pull" applications: on-line/off-line video databases, retrieval and search/location of multimedia information) and data distribution ("push" models on Internet, filtering according to user behavior, agents: broadcasting and Web casting). The most general block-diagram of MPEG-7 application is shown in Fig. 5: extraction of characteristics (analyze), standardized description and search engine (application) [21]. Analyze and searching are not standardized because it is not necessary for interoperability and it leaves a space for industry for continuous development and solution improvement.

Fig. 5 Processing chain and scope of MPEG-7 standard.

In Fig. 6 block diagram of an application that uses MPEG-7 standard for description of multimedia content is shown. Final performances of application crucially depend on a database structure: the indexing information will have to be structured, e.g. in hierarchical or associative way [21], [22], [23], [24]. The way the MPEG-7 data will be used to answer user queries is outside the scope of the standard. This means, for example, that video material may be queried using video, music, speech, etc. It is to the search engine to match the query data and the MPEG-7 AV description.

MPEG-7 search/query engine could freely access any complete or partial description associated with any AV object in any set of data, perform a

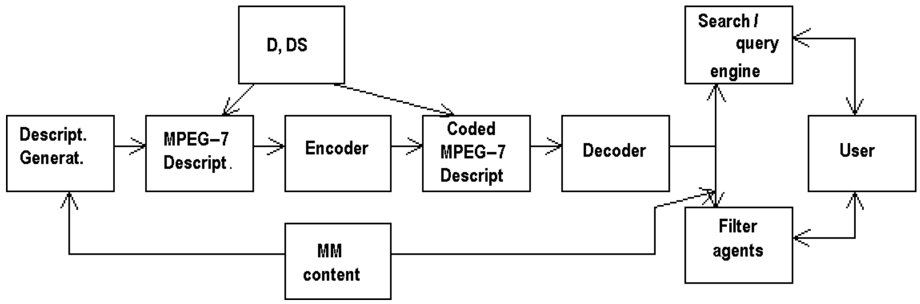

Fig.6 Applications using MPEG-7.

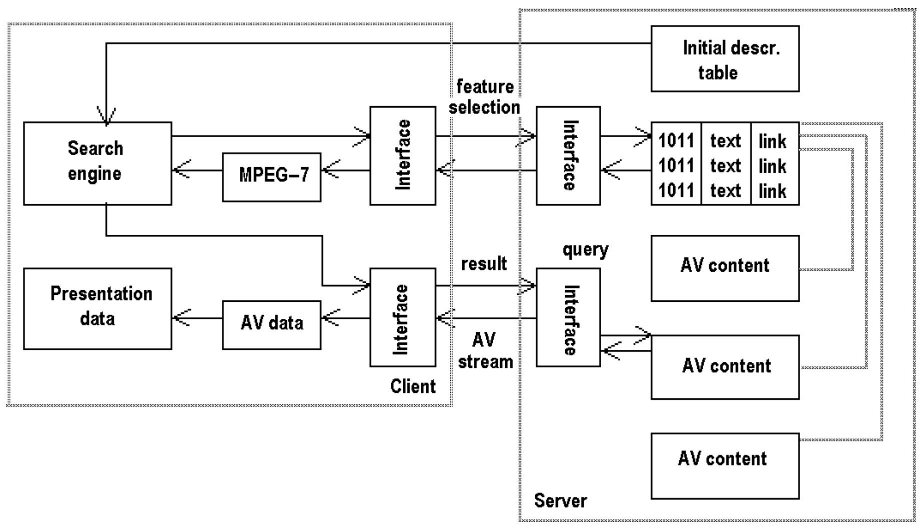

Fig. 7 Example of a client-server architecture in MPEG-7 based data search.

ranking and retrieve the data for display by using the link information. An example of architecture is illustrated in Fig.7 [21].

Access to multimedia material according to their content includes the following scientific/technical fields and research areas [25]:

- digital image processing (scene, face and gesture detection)
- processing of audio signals (phrase detection, word spotting ...)
- the problems of semantic and artificial intelligence (natural languages, understanding in general sense...)

- interface problems (query entering, results representation)
- computer networks (databases and distributed processing)
- servicing problems (delivery, copyrights, payment).

Theory and tools for browsing and image/video manipulation are at present in the initial state of their development and many of the problems are still unsolved. For decoders, image processing researcher has focused on problems related to medical and military images. Access to multimedia material according to its content, include a number of scientific fields: digital image/audio processing, solving problems semantically with artificial intelligence, interface problems, communication networks [26], [27].

6. ITU-T Recommendations

In the original ITU-T work plan, the goal was to define a "near term" recommendation in 1996, followed by a "long term" recommendation several years later. The near term recommendation is what is referred to as H.263. The original H.263 was developed for video compression at rates below 64 kbits per second and more specifically at rates below 28,8 kbits per second. This was the first international standard for video compression which would permit video communications at such low rates [28], [29]. The key technical features of H.263 are variable block-size motion compensation, overlapped-block motion compensation, pixel-extrapolating motion vectors, 3-dimensional run-level-last variable-length coding, median motion vector prediction, and more efficient header information signaling (and relative to H.261, arithmetic coding, half-pixel motion estimation, and bi-directional prediction). In the original H.263 Recommendation, there is a mode by which a bi-directionally predicted frame or B frame (as in MPEG-2) could be multiplexed with a subsequent P frame in the bitstream thus saving some of the overhead of having a separate frame in the bitstream.. The motion vectors for B frame would be interpreted from the P frame motion vectors and a small offset to the interpolated vectors could also be transmitted in the bitstream. Although in most cases, P-B pair of frames would be more efficient than a pair of P frames, in some situations this arrangement would perform worse.

6.1. H.263+

After H.263 was completed, it became apparent there were incremental changes that could be made to H.263 that can visibly improve its compression performance. Thus, it was decided in 1996 that a revision to H.263 would be created which incorporated these incremental improvements. This is H.263 "plus" with several new features. Hence, the name H.263+ (now called H.263 Version 2). H.263+ contains approximately 12 new features that do not exist in H.263. These include new coding modes that improve compression efficiency, support for scalable bitstreams, several new features to support packet networks and error-prone environments, added functionality and support for a wider variety of video formats. H.263+ falls into the family of video coders commonly referred to as hybrid DPCM/DCT-based video compression algorithms. Other video coders in this family include H.261 [30], MPEG-1, MPEG-2, FCC/HDTV and MPEG-4 video. These video coders typically employ 16x16 pixel block-based motion estimation and frame differencing to reduce temporal redundancy. Discrete cosine transform (DCT) is applied to 8x8 pixel block of the resulting residual frame. The transform coefficients are then quantized to reduce spatial redundancy. Lossless coding techniques are applied to the resulting symbols to further reduce statistical redundancies. A frame rate that has been temporally predicted is referred to as an INTER coded or P frame. The frame is used as a basis of prediction, which is usually the decompressed version of the previous frame, is called the reference frame. When a frame is coded with no prediction from a prior frame, it is referred to as an INTRA coded or I frame. An INTRA coded frame is, for all intents and purposes, simply the result of still image compression applied to a frame video [10]. A B (bi-directionally interpolated) frame is predicted based on motion estimation/compensation and of the previous and/or future frame. In H.263+ an improvement to the original mode allows a complete motion vector to be transmitted for the macroblock (MB) in B frame. Thus, a P/B frame pair can do no worse than a pair of P frames. H.263+ provides the means by which the reference frame can be resized, translated or generally warped before being used as prediction for the current frame. This is referred to as Reference Picture Resampling (RPR) mode. Several of the capabilities in RPR mode can positively impact compression performance. For example, it is possible to indicate a global motion parameter. This parameter may be able to describe the global motion, thus allowing most of the block based motion vectors to be zero. With RPR, an encoder could dynamically switch to quarter sized video until the large motion subsides, thus avoiding "jerky" video. The display would continue to be full size, however the video would have to be interpolated back to full size prior to display.

Among the new features of H.263+, one of several which corrects design inefficiencies of the original H.263 recommendation is Modified Quantization Mode. This mode has four key elements:

- Indication for larger quantizer changes from macroblock-to-macroblock to better react to rate control requirements

- The ability to use a finer chrominance quantizer to better preserve chrominance fidelity
- Capability to support the entire range of quantized coefficient values rather than having to clip values greater than 128
- Explicitly restricting the representation of quantized transform coefficients to those that can reasonably occur

The second modification of the original H.263 Recommendation is motion vector range, which was for the most part [- 16, 15,5]. When H.263+ mode is invoked, the range is generally larger and depends on the frame size as shown in Table 3 [31].

Table 3. Motion vector ranges in H.263+

Frame sizes up to	Motion vector range
352×288	[−32, 31.5]
704×576	[−64, 63.5]
1408×1152	[−128, 127.5]
Widths up to 2048	Hor. range [−256, 255.5]

H.263+ supports a wider variety of input video formats than H.263. In addition to five standard sizes, arbitrary frame sizes, in multiples of 4, from (32×32) to (2048×1152) can be supported, as well as other pixel aspect ratios besides 12:11, and other picture clock frequencies besides 29,27 Hz.

H.263+ video enhancement layers belong to one of three categories: temporal, SNR or spatial enhancement. Temporal enhancement is the process by which the frame rate can be increased over the base layer. This is accomplished via disposable bi- directionally predicted (B) frames as presented in Fig.8 [32]. Predicting from prior and subsequent frames usually improves compression performance. B frames are not used to predict (with or without motion compensation) the B and P frames. However, when added, they raise the frame rate of the video sequence. Hence, the name temporal enhancement layer.

SNR enhancement layer is shown in Fig.9 [32]. It is a refinement to the coded base layer frames. When a video frame is compressed, the decoded version is not an exact replica of the original frame. This is because H.263+ is a lossy compression. It selectively throws away information in order to improve compression performance without excessive degradation of visual quality. If there is enough bandwidth for a decoder to receive the SNR enhancement layer, the visual quality can usually be improved over the

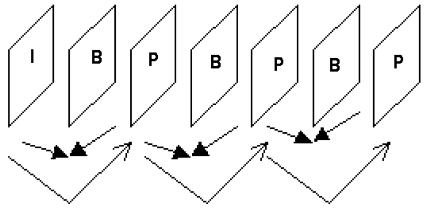

Fig.8 Bi-directionally predicted (B) frames.

base layer. Besides the prediction from the base layer, the H.263+ encoder has the choice of including prediction from the previous frame in the SNR enhancement layer as well. This is a modified form of a bi-directionally predicted frame, called EP frame. When an SNR enhancement frame is only predicted from the base layer and not a previous frame, it is called an EI frame.

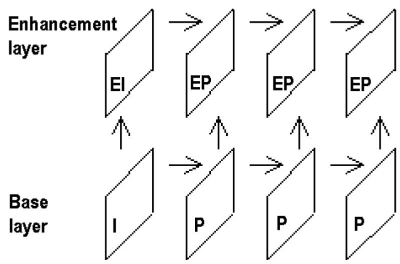

Fig.9 SNR enhancement layer.

Spatial enhancement is closely related to SNR enhancement. The only difference is that the enhancement layer is twice the size vertically and horizontally from the base layer. In this case, the input video is first down-sized both vertically and horizontally prior to encoding as the base layer. After that, the decompressed base layer frames are interpolated back to the original size before being used as prediction for the spatial enhancement layer frames. Different enhancement types can be combined to provide a very flexible layered bitstream architecture.

Part of the process of transporting video bitstream over packet networks is the fragmentation of the bitstream before being converted into a packet payload. H.263+ improves support for this fragmentation operation through the use of arbitrary resynchronization markers. The resynchronization markers are provided by slice headers in a sub-mode of the slice structured mode. Slice headers can be inserted at any macroblock boundary.

When packet loss becomes substantial, it is possible to define subpictures within each picture which do not depend on any information outside the subpictures boundaries within the current frame or in any referenced picture. In this way, error due to packet loss can be kept from propagation to other regions. This is referred to as independent segment decoding mode.

H.263+ also provides the ability to indicate that a prior frame is suitable to be used as a reference for compression, either positively or negatively. This is referred to as reference picture selection.

6.2. H.263++

The H.263++ development effort is intended for near-term standardization of enhancements to produce a third version of the H.263 video codec for real-time telecommunications and related non-conversional services. The work plan for H.263++ is shown in Table 4 [9].

Table 4. H.263++ work plan

September 1997	Adoption of work plan
November 1998	First formal draft adoptions
February 1999	Last formal draft adoptions
November 1999	Final draft for determination
February 2000	Determination
July 2000	Final draft for decision
November 2000	Decision

Key technical areas showing potential for performance gain of H.263++ are:

- Error resilient data partitioning,
- 4×4 block size motion compensation,
- Adaptive quantization,
- Enhanced reference picture selection,
- Inverse DCT mismatch reduction,
- Deblocking and deringing filters,
- Error concealment.

6.3. H.26L

The long term recommendation H.26L (previously called H.263 L) is scheduled for standardization in 2002 and may adopt a completely new compression algorithm. The H.26L is aimed at developing new video coding technology beyond the capabilities of incremental enhancements to H.263, for longer-term standardization. It is aimed at very low bitrates, real time, low end-to-end delay coding for a variety of source materials. It is designed to have low complexity permitting software implementation, enhanced error robustness (especially for mobile networks) and adaptable rate control mechanisms. The applications targeted by H.26L include: real-time conversational services, Internet video applications, sign language and lip- reading communication, video storage and retrieval service, video store (VoD) and forward services (video mail) and multi-pint communication over heterogeneous networks. The IMT-2000, future mobile communication environment, is a key application area for H.26L. H.26L should provide more error robustness capability as well as substantially improved compression performance for an application to future mobile multimedia systems. The work plan for H.26L is shown in Table 5 [10].

Table 5. H.26L work plan

January 1998	Call for proposals
November 1998	Evaluation of the proposals
April 1999	1st Test Model of H.26L (TML1)
1999-2001	Collaboration phase
October 2001	Determination
July 2002	Decision

The following technical proposals are evaluated in response to the call for proposals for H.26L [33], [34].

- Modified prediction/transform based method
- Vector quantization with blocks approximation either by reference to a codebook or by motion compensation from a previous frame
- Loop-filtering method for reducing the blocking artifacts, corner outliner and the ringing noise
- Adaptive scalar quantizer scheme using non-zero level codebooks
- DCT-based embeded video coder using rearrangement of DCT coefficients
- Low bitrate video coder: rough segmentation, affine motion compensation scheme, vector quantization and multi-shape DCT
- Data partitioning method using data reordering algorithm
- Video coding using long-term memory for multiple reference frames and affine motion compensated prediction

7. Conclusion

The standard developed over five years by the MPEG explores every possibility of the digital environment. Recorded images and sounds co-exist with their computer-generated counter-parts, a new language for sound promises compact-disk quality at extremely low data rates, while the multimedia content could even adjust itself to suit the transmission rate and quality.

The world wide acceptance of MPEG-2 in consumer electronics has lead to large production scales making MPEG-2 decoder equipment inexpensive and therefore also attractive for other related areas, such as video communications and storage and multimedia applications in general. Because MPEG-2 was designed as a transmission standard, it supports a variety of packet formats (including long and variable length packets of from 1 up to 64 kbits) and provides error correction capability that is suitable for transmission over cable TV and satellite links.

MPEG-4 allows the user to interact with objects within the scene, whether they derive from the so-called real sources, such as moving video, or from synthetic sources such as computer-aided design. Authors of content can give users the power to modify scenes by deleting, adding or repositioning objects or to alter the behavior of the objects. The video coding standard MPEG-4 is enabling content-based functionalities by introduction the concept of video object planes. MPEG-4 will integrate most of the capabilities and features of multimedia into one standard including live audio/video, synthetic objects and text. One of the outstanding features of the evolving MPEG-4 standard is the possibility of object-based image access to coded video data. The user has the possibility to interact with the audiovisual scene by selecting single objects, move the object within the virtual scene and change object attributes like size or orientation, The problems associated with rate control for coding multiple video objects were addressed, too. This type of algorithm is useful in supporting the object-based functionalities which are central to the emerging MPEG-4 standard.

MPEG-7 aims to standardize a set of multimedia description schemes as well as to support a number of audio and visual descriptions, including free text, N-dimensional spatio-temporal structure, statistical information, objective attributes, subjective attributes, production attributes and composition information. For visual information, descriptions will include color, visual objects, texture, sketch, shape, volume, spatial relations, motion and deformation. MPEG-7 also aims to support a means to describe multimedia material hierarchically according to abstraction levels of information in order to efficiently represent a user's information need at different levels.

The H.263 video standard is based on techniques common to many correct video coding standards. The key technical features of H.263 are available block-size motion compensation, overlapped block motion compensation, pixel extrapolating motion vectors, 3-dimensional run-level-variable length coding, medium vector prediction and more efficient header information signaling. H.263 supports five standardized picture formats: sub-QCIF, QCIF, CIF, 4 CIF and 16 CIF.

H.263+ offers many improvements over H.263. The added flexibility opens H.263+ to a broader range of video scenes and applications, such as wide format pictures, resizeable computer windows and higher refresh rates. Another major improvement of H.263+ over H.263 is the improvement of the delivery of video information in error-prone, packet-lossy or heterogeneous environments by allowing multiple display rates, bitrates and resolutions to be available at the decoder. Furthermore, picture segment dependencies may be limited likely reducing error propagation.

Finally, H.26L is designed to have low complexity permitting software implementation enhanced error robustness and adaptable rate control mechanisms.

REFERENCES

1. R.Shafer, T.Sikora: *Digital video coding standards and their role in video communications.* Proc. IEEE, Vol. 83, June 1995, pp. 907-924.
2. V.Bhaskaran, K.Konstantinides: *Image and Video Compression Standards, Algorithms and Architectures.* Norwell, MA: Kluwer Academic, 1997.
3. T.Sikora: *MPEG digital video-coding standards.* IEEE Signal Processing Magazine, vol. 14, September 1997, pp. 82-100.
4. T.Naveen, S.C.Fhu: *Compression of spatial and frequency keys.* ISO/IEC JTC1/SC 29/WG 11, Doc. 615, 1993.
5. A.Puri, B.G.Haskel: *Picture format scalable coding structures.* ISO/IEC JTC1/SC 29/WG 11, Doc. 673, 1993.
6. T.Chiang, D.Anastassiou: *Hierarchical coding of digital television.* IEEE Comm. Magazine, vol. 32, May 1994, pp. 38-45.
7. T.Sikora, T.K.Tan, K.N.Ngan: *A performance comparison of frequency domain pyramid scalable coding scheme within the MPEG framework.* Picture coding Symp., March 1993, pp 16.1-16.2.
8. Y.Wang, A.Puri: *Spatial-temporal adaptive interlace extraction.* ISO/IEC JTC1/SC 29/WG 11, Doc. 509, Sept. 1992.
9. ITU-T Study Group 16: *Video Coding Experts Group (Question 15).* Doc. Q15F09, Report of the ad hoc committee H.263++ development, Seoul, Korea, Nov. 1998.
10. M.Tekalp: *Digital Video Processing.* Upper Saddle River, NY: Prentice-Hall, 1995.

11. ISO/IEC JTC1/SC 29/WG 11: *Doc.N2172.* March1998.

12. L.Chiarglione: *MPEG-4 project description.* Document ISO/IEC JTC1/SC 29/WG 11, Doc. N 1177, Munich, Germany, MPEG Meeting, Jan.1996.

13. L.Chiarglione: *MPEG and multimedia communications.* IEEE Trans. CSVT, vol. 7, Feb. 1997, pp 5-18.

14. ISO/IEC JTC1/SC 29/WG 11: *Doc. N 1022.* July1995.

15. ISO/IEC JTC1/SC 29/WG 11: *Doc. N 1164.* Jan.1996.

16. R.Nell, A.Zakhor: *Very low bitrate video coding based on matching persuits.* IEEE Trans. CSVT, Vol. 7, Feb. 1997, pp. 158-171.

17. T.Sikora: *The MPEG-4 video standard verification model.* IEEE Trans. CSVT, vol. 7, Feb. 1997, pp 19-31.

18. Z.Bojkovic: *MPEG-4 video verification model: encoding process.* Proceedings ICOMT'98, Budapest, Hungary, Oct. 1998, pp 125-132.

19. Z.Bojkovic: *Multimedia communication system: modeling, standardization, requirements.* Proceedings ICOMT'96, Budapest, Hungary, Oct. 1996, pp 5-14.

20. Z.Bojkovic, D.Prokopljevic: *MPEG-4 applications.* Telecommunications, The Community of the Yugoslav PTT, Vol. XLVII, No.1, March 1999, pp 69-79.

21. ISO/IEC JTC1/SC 29/WG 11: *Doc. N 2460, MPEG-7: Context and objectives V.10.* Oct. 1998.

22. ISO/IEC JTC1/SC 29/WG 11: *Doc. N 2461, MPEG-7: Requirements V.7.* Oct. 1998.

23. ISO/IEC JTC1/SC 29/WG 11: *Doc. N 2462, MPEG-7: Applications V.7.* Oct. 1998.

24. ISO/IEC JTC1/SC 29/WG 11: *Doc. N 2571, MPEG-7: Experimental model (XM) Ver.0.5.* Dec. 1998.

25. R.Picard: *Light-years form Lena: video and image libraries of the future.* IEEE ICIP, Washington, D.C, Oct. 1995, pp. 310-313.

26. ISO/IEC JTC1/SC 29/WG 11: *Doc. N 2464, MPEG-7 proposal package description (PPD).* Oct. 1998.

27. ISO/IEC JTC1/SC 29/WG 11: *Doc. N 2469, Call for proposals for MPEG-7 technology.* Oct. 1998.

28. ITU-T Recommendation H.263: *Video coding for low bitrate communication.* March 1996.

29. K.Rijkse: *H.263: video coding for low bitrate communications.* IEEE Trans. Comm, vol. 34, No. 12, Dec. 1996, pp. 42-45.

30. ITU-T Recommendation H.261: *Video codec for audiovisual services at px64 kbit/s.* March 1993.

31. M.Ghanbari: *Two-layer coding of video signals for VBR networks.* IEEE J. Selected Areas in Comm, vol. 7, June 1989, pp. 771-781.

32. ITU-T Draft Recommendation H.263 Version 2 H.263+: *Video coding for low bitrate communication.* Sept. 1997.

33. ITU-T Study Group 16: *Video Coding Experts Group (Question 15).* Doc. Q15D62, Call for proposals for H.26L video coding, Geneva, Jan. 1998.

34. ITU-T Study Group 16: *Video Coding Experts Group (Question 15).* Doc. Q15F10, Report of the ad hoc committee H.26L development, Seoul, Korea, Nov. 1998.