USE OF MPEG-4 IN BROADBAND APPLICATIONS

Mukta L. Kar, Ph.D. and Yasser Syed, Ph.D.
Cable Television Laboratories, Inc.
Sam Narasimhan, Ph.D. and Ajay Luthra, Ph.D.
Motorola Broadband Communication Sector

ABSTRACT

The emerging MPEG-4 standard in development today will provide new opportunities to broadband environments. Streaming Media, enabled through MPEG-4 technologies, can create personal viewing and interactive applications, such as Video-on-Demand (VoD), targeted advertising, and hot-button interactive product purchasing. This paper will describe some of the newer capabilities added to the MPEG-4 standard that are advantageous to broadband environments. Furthermore, cable initiatives for digital local ad insertion (DVS 253 and DVS 380) will be discussed in light of digital compression technologies, such as rate-remultiplexing. We will look at how MPEG-4 can achieve this within an MPEG-2 structure and still allow for bandwidth reduction. Finally, this paper will discuss how a targeted advertising capability can result from MPEG-4 implementation.

Introduction

Since the early days of cable, the primary source of revenue has been the broadcast delivery of premium analog content. This type of delivery mechanism does not utilize resources and information in an optimized way. Resources and information are wasted in this approach because not everyone will watch programs in their entirety at a pre-scheduled time or consider the information (advertisements) as relevant. Many cable systems have been upgraded from one-way analog to the modern hybrid-fiber coax (HFC) two-way digital broadband networks of today. Digital delivery using advanced digital modulation and compression technology has expanded a limited analog RF (Radio Frequency) spectrum to a large digital bandwidth potentially increasing the channel capacity of a cable plant by a few folds [6]. This digital bandwidth can be used to offer more niche channels to capture a more targeted audience interested in the programs as well as the advertisements being shown. This is one of the few ways to utilize digital delivery optimally to increase the revenue of the cable industry by offering more niche programming/content. Additionally, the large digital bandwidth also can be used for delivering Internet protocol (IP)-based services [5][6].

Streaming Media adds a new dimension by allowing a personally designed viewing experience. The viewer's choice is not limited to local headend offerings. With streaming media, a viewer can request content at a time when he/she desires to see it, and from a source that may reside anywhere in the world. This content can be delivered through a headend-managed, DOCSIS-enabled IP transport stream or through a traditional MPEG-2 delivery system to a set-top box (STB) [6]. It also adds in the ability to target content to a specific type of viewer rather than broadcasting the same content to the entire audience. A streaming media session adds increased value by capturing a viewer who is definitely interested in the program. Revenue can come from the customer actually purchasing the streaming media content or through the use of targeted advertisements that are more relevant to the viewer [5].

Through Streaming Media, the customer can actually purchase content that he/she is willing to watch at his/her convenience. One of the most visible applications of this technology is video-on-demand (VoD) services where movies can be streamed to the customer at his request. The application can be extended to other types of content, such as "how to" programs, news programs, and weekly serials. This service can be enhanced by adding in VCR capabilities, such as pause, fast-forward, and rewind. However, there would be an increased amount of downstream data generated by each customer request [4], and some upstream traffic due to interactive VCR features. The additional traffic in both directions would be an added strain to a

two-way cable plant. For services like these to exist on a large scale in a two-way cable infrastructure, a technology like MPEG-4 needs to be implemented to permit high-quality interactive video to stream at low bit rates [4].

Another revenue source opened up by streaming media is targeted advertising. In the existing method of broadcast delivery, national commercials are inserted into the program stream by the networks before delivery to the cable headends and broadcast affiliates. However, advertisements and short programs originated locally are being inserted (overwritten on top of selected national ads) per cue tones delivered by the networks. The revenue from the commercials is not only significant compared to the total revenue, but provides a good cash flow to the cable industry. The existing method of ad delivery is not very efficient. Technically, the existing approaches splice the local ad into the national distributed program stream using analog techniques by uncompressing and converting the digital streams into an analog format. The output is then delivered over the local cable network as an analog signal with a degradation in visual quality as well as a loss in bandwidth revenue. In terms of scope, these commercials are broadcast to the entire regional viewing audience (large area), but fail to be relevant to many in the viewing audience. With streaming combined with digital statistical remultiplexing techniques, these slots can contain different local digitally inserted commercials that can be targeted towards each customer [5]. For instance, a car company can put out a different car commercial for each of their product lines based upon a customer profile for the same commercial time slot. This way the most relevant commercial reaches the customer. As techniques become further developed, streaming also can enhance broadcast commercials by combining additional data to create a higher visual quality commercial instead of having the commercial constrained to the encoding bit rate of the channel. Lastly, the commercial can be made more interactive by providing a way for the customer to stream back his responses. An example of this would be the ability to immediately buy a product advertised in a commercial. Similar to VoD, these types of services will increase traffic loads in the cable network to accommodate the individualized information generated by the streaming application. A technology like MPEG-4 will be necessary to allow these services to exist on a large scale by reducing the bandwidth required

for low bit rate transmissions and allowing for interactivity [1][4].

Streaming Media in MPEG-4

MPEG-4 Background

The MPEG-1 standard was created to satisfy the need for storage applications, such as CD-ROMs and adopted frame-based video compression methods. Later MPEG-2 standards were developed to meet requirements of the broadcast industry. It included field/frame-based video coding to deal with interlaced video as well. In brief, the MPEG-2 standard was optimized primarily for one-way broadcast delivery of television content [1]. The MPEG-4 standard effort initially differed by focusing on low bitrate coding applications over IP connections. Later its scope increased to cover a wider range of multimedia applications including videophone, interactive TV, streaming video etc. [4]. Additionally, to include interactive capability at various levels, the coding paradigm changed from frame-based to content-based or object-based coding. One of the more important advantages of MPEG-4 is that it is well optimized for scalable low bit rate (LBR) transmissions (less than 2 Mbps) and the second one is the ability to selectively incorporate natural and synthetic objects into a scene. These advantages will allow streaming media to exist on a large scale by economizing the available bandwidth while allowing interactivity to be integrated [3].

Object-based coding has other advantages. It reduces bandwidth demands by allowing for object-based coding instead of frame-based coding that exists in MPEG-2 technologies. In this case, instead of a frame being coded in its entirety, separate audio/video (A/V) objects within the frame can be encoded at different quality and rates. Objects do not have to be re-transmitted each time a scene changes, rather only manipulation information (scale, translation) of the object could be sent. Each A/V element can be encoded in its own elementary or set of elementary streams in the form of video object planes (VOP). Different encoding parameters can be assigned based upon the nature of the object to allow for the most efficient type of encoding. For instance, a natural image can be encoded by a video codec while a text-based object can be encoded by a text-based codec, which would display a higher quality object
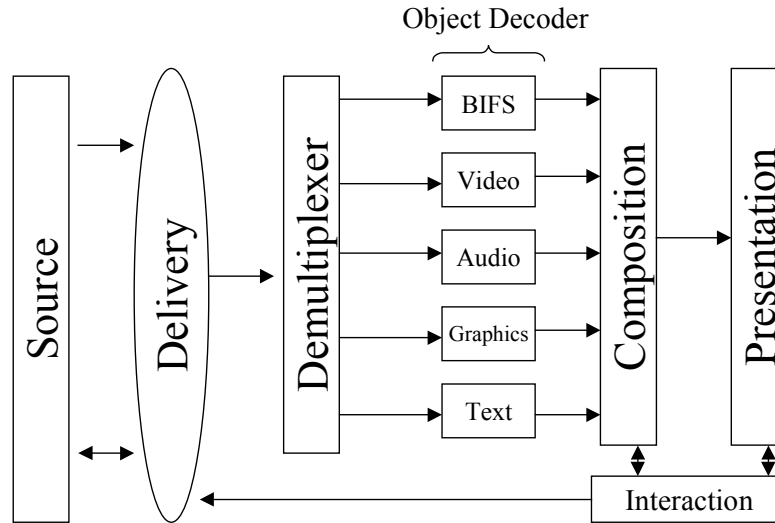
Figure 1. Block Diagram of MPEG-4 System

with a smaller amount of bits. By using the right codec type, bits can be conserved without sacrificing video quality. There are five general categories of coders (Figure 1): 1) video, 2) audio, 3) graphics, 4) text, and 5) scene. MPEG-4 adapts each of its codecs to conform to multiple profiles and levels of transmission to accommodate different delivery formats. The video and audio codecs are used on natural video and audio objects optimized for good quality at low bit rates. In addition, MPEG-4 adds in a spatial and temporal scalability factor to traditional A/V coding and provides graceful degradation of objects during times of congestion [3]. The graphics coder provides a means to animate and render synthetic objects. Synthetic objects can be computer-generated objects with interactive components. An example of this would be a "hot-button" to purchase an advertised product. The text coder provides an efficient way to code text. The scene coder, called BIFS (Binary Format for Scene), is responsible for scene omposition and rendering. This coder can manipulate (spatially, temporally), layer and even edit out objects from the scene [7]. It also can add or delete streams such that a directed channel change (DCC) can occur to allow for targeted advertising [12].

New Scalable Profiles in MPEG-4

The adoption of the advanced simple profile (ASP) and fine grain scalability (FGS) profile in the MPEG-4 visual standard allows for different layered levels of quality that are advantageous for streaming media over the Internet, which consists of many heterogeneous networks. ASP allows the highest possible quality within the MPEG-4 standard for a traditional video consisting of rectangular shape frames by allowing the use of B-frames, quarter-pel interpolation, global motion compensation (GMC) and interlaced coding format. It also allows pictures to be compressed at higher than common interchange format (CIF) resolution—half horizontal resolution (HHR) and full resolution—and at high compressed bit rates. It does not include shape coding tools and thus does not have the complexity associated with arbitrary shape coding [1][3]. The addition of FGS provides for transmission of base and enhancement layer streams. The base layer contains the lowest quality coded image of the object and is compliant with the ASP. The enhancement layer in the FGS profile adds to the base layer to increase the visual quality of the object. A spatial enhancement layer can be done by the FGS layer, and a temporal enhancement layer to the visual stream can be
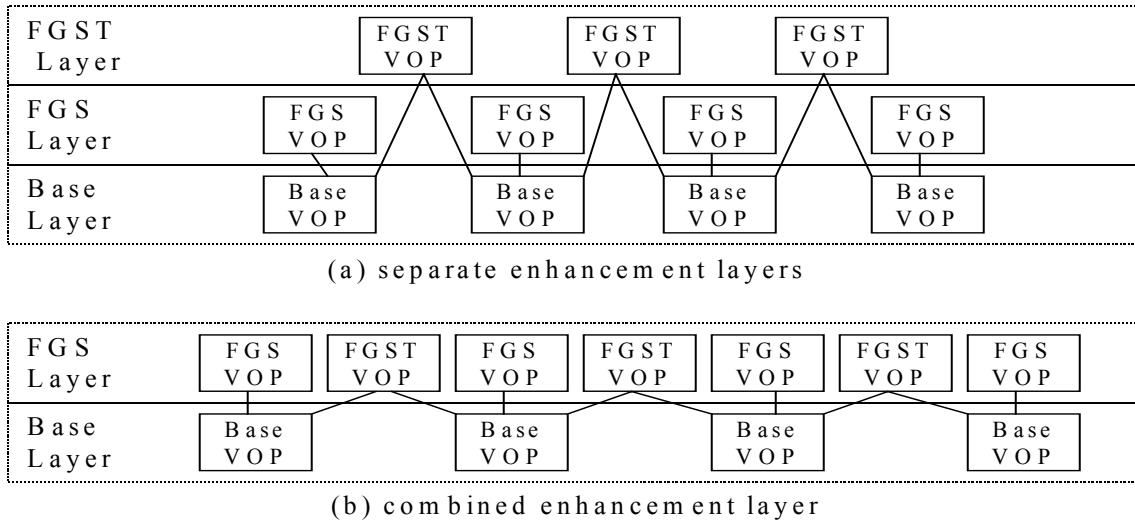
Figure 2. Base & Enhancement Layers Combinations for Streaming Video over the Internet

done by the FGS temporal scalability (FGTS) layer. Both these layers only require the base layer information to be decoded and have a scalable embedded bitstream property (where the number of bits processed is proportional to the image quality). This is different from the discrete scalability profile put forth in MPEG-2 or earlier versions of the MPEG-4 profiles. Each type of enhancement layer can be layered on top of each other or combined into a single enhancement stream layer (Figure 2) The different profile levels (five for the ASP and five for FGS) adds in different capabilities, such as number of objects, visual image size, temporal and/or spatial scalability, buffer size, maximum packet length, and maximum bit rate [2].

This base/enhancement layer partition provides several advantages. In an IP-congested environment, a graceful degradation of the image (by sacrificing the quality of certain visual objects) can occur during times of net congestion through the use of an embedded bit stream property. Also, multiple CPE devices with different bandwidths, QoS parameters, or processing restrictions, can view the video stream at different resolutions or qualities due to the stream's scalability and embedded properties. For lower bandwidth connections, selective enhancements through bitplane shifting and coefficient weighting can improve visual quality by prioritizing enhancements of certain regions of the video first [2][3]. This approach also can allow for a picture-in-picture format without doubling the bandwidth. In VoD or PVR applications, the base layer can provide a quick continuous search capability within the bandwidth constraints of the connection. Since elementary streams do not have to come from the same source (this is determined by the BIFS information), enhancement layers can be added to increase the visual quality of an advertisement without a corresponding increase in bandwidth.

Current Streaming Media Technology

Existing media streaming codecs largely used on today's PCs do not take advantage of all the interactive services enabled by a sequence mix of synthetic and natural objects. Current implementation by several equally popular proprietary codecs limit network streaming application to video- and audio-focused material delivered to the PC. As coverage extends to other CPE devices, and implementation develops more personal interactive applications, the complexity in codec and system development will either lead to a limit in growth due to the support of too many proprietary formats or will drive the acceptance of a single format [4]. This single format may arise from a proprietary implementation or from the MPEG-4 standardization process. In either case, the support of the full scope of streaming media on a large scale will require the adoption of MPEG-4 features, if not the exact approaches to implement them.

AD INSERTION

Cable's Response to Existing Local Analog Ad Insertion

Existing ad insertion systems used in today's local cable headends are, at best, hybrid in nature. Locally generated commercials and short programs are stored in an ad server in a digitally compressed format. The digitally compressed commercial is then decoded and converted to analog before insertion into a network program using network cue-tones for timing and duration information of the avail [9], as well as analog splicing techniques [8] .

The hybrid method is fine if content is delivered from a local headend in analog format. But in an all digital delivery environment, a standardized digital-insertion method [9], where a compressed commercial is inserted into a compressed network program in the headend, is the most desirable. With this objective, the SCTE digital video subcommittee (DVS) developed a standard [11] entitled "Digital Program Insertion Cueing Message for Cable (DVS 253)." This standard defines just the cue-message and does not impose any constraints on insertion/splicing equipment. The cue message carries timing information using the coordinated universal time (GPS UTC) for scheduling and MPEG PTS time for frame-accurate insertion, which the splicing device may use to perform the splice. Cue messaging, if required, may be passed on to authorized downstream equipment, such as a pass-through via a remultiplexer to a set-top box. The timing correction needed after remultiplexing also may be transmitted to maintain timing accuracy of the cue signal.

In addition to these two message components, each splice command enables splicing of complete programs or individual components of a program (such as video or audio or data) through the use of component_tags enabled by the stream_id_descriptor. Today's systems use only program-level splicing, where all the components of a program are replaced at the splice point. In the future, programs and advertisements will be "enhanced" using data broadcast and interactive elements. Typical enhanced commercials could include delivery of discount coupons, prize drawings, or free software. These systems will use component-l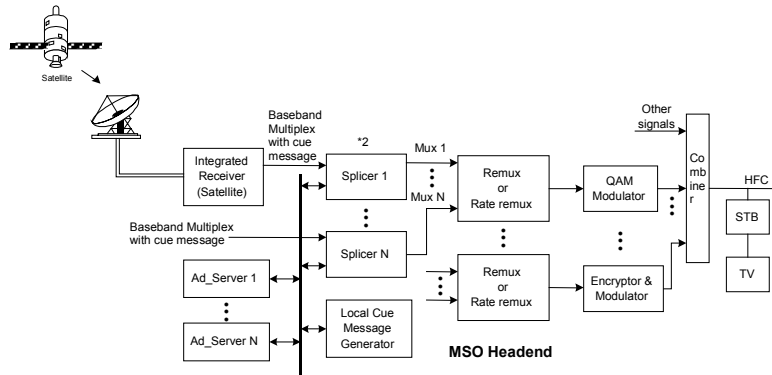evel splicing, where only selected components of a program are replaced at the splice point. This splicing enables pre-loading of data streams that are part of the same program by inserting the data component ahead of A/V content. This may be done to load and run the data enhancements in the receiver's application engine.

In addition, the SCTE sub-committee also has developed draft standard DVS 380 [12], which standardizes the API's between the ad-server and splicing equipment. DVS 253, along with standardized APIs defined in DVS 380, will allow splicers and commercial ad servers to interoperate with each other.

Figure 3 shows the functional block diagram of a digital headend where local commercials are inserted into a program utilizing the cue message multiplexed in the transmitted program. The satellite-integrated receiver demodulates the RF signal to its baseband MPEG-2 transport stream and decrypts it. The cue message, multiplexed in the transport stream is detected and is processed by the splicer and ad server. At the splice_insert time, the splicer switches from the network program to the local ad. At the end of the avail, the splicer switches back to the network program. A headend may have multiple splicers and ad servers interconnected among them so that the same commercial need not be stored in more than one server. The cue message standard, DVS 253, does not specify how to splice between two bitstreams. The techniques and resultant constraints for splicing compressed streams have been left to splicing equipment manufacturers. This adaptability also allows for incorporation of advancements in knowledge, such as MPEG-4-enabled technologies.

Digital Splicing into Stat-Mux Channels

In a digital environment, statistical multiplexing of compressed video streams helps to utilize digital channels better than constant bitrate encoded streams, assuming that the peak demand for bits from all video encoders does not coincide. The important constraint here is that all the videos need to be encoded while multiplexing. A previously compressed stream typically had to be decompressed and re-encoded before it could be inserted in a statistical multiplex. This is a detriment to local ad insertion due to the equipment cost of decompression and recompression of the stream as well as visual quality degradation. As an

Satellite

Baseband Multiplex with cue message

Integrated Receiver (Satellite)

*2

Splicer 1

Mux 1

Mux N

Remux or Rate remux

QAM Modulator

Other signals

Combiner

HFC

STB

TV

Baseband Multiplex with cue message

Splicer N

Ad_Server 1

Ad_Server N

Local Cue Message Generator

Remux or Rate remux

Encryptor & Modulator

MSO Headend

Figure 3. Typical Commercial Insertion System per DVS 253

alternative, a remultiplexer can be used to manipulate or "groom" the multiplexed stream in the compressed domain. By definition, a *remultiplexer* receives one or more multiplexed streams as input and creates a new output multiplexed stream from local operator-selected programs, such as local ads. Nominally, a remultiplexer does not alter bitrates while constructing a new multiplex out of the input streams. The technology that deals with multiplexing compressed video streams and trims the resulting multiplex to match an assigned constant total transmission channel bitrate, is known as *rate-remultiplexing*. Rate remultiplexing meets the latter constraint by transcoding individual video streams within the output multiplex. Transcoding is the technique by which a compressed video stream is translated to a lower bitrate strictly within the compressed domain. It can reduce the bitrate of MPEG-2-compressed video without fully decoding and re-encoding a bitstream. Thus, without cascaded compression, degradation in picture quality is not noticeable with occasional or moderate reductions in the average bitrate of individual video streams.

In addition through rate-remultiplexing techniques, storage requirements for commercials can be reduced by storing only one high-quality version in the server and using rate-remultiplexing to adjust its bitrate. Rate-remultiplexing technology provides the capability to insert compressed digital commercials into digital channels at the headend and removes the need to match the bitrate of the locally compressed commercial with that of the remotely transmitted program, or creating and storing different bitrate versions of the commercials in the ad server.

Targeted Insertion Systems using MPEG-4

While insertion of MPEG-2-coded advertisements into a channel with stat-muxed video requires bitrate transcoding of the advertisement stream, MPEG-4 coding schemes not only allow use of coding tools that are more attractive for advertisements (such as natural video merged with synthetic video with lots of scene changes), they allow for lowering of bit rates to values that are far lower than MPEG-2 stat-mux stream rates. Using a combination of synthetic and natural hybrid tools, advertisements could be authored at rates lower than 500 Kbps with quality that matches that of the MPEG-2-coded video. MPEG-4 provides tools, such as simple, advanced simple, and FGS profiles (visual) capability, and animated 2d mesh, basic animated texture, scalable texture and simple face for coding advertisements on television today.

As advertisements are typically not coded in real-time, MPEG-4's combination of natural video (including sprites) and synthetic video tools provide a very efficient coding scheme for these applications, even without the use of MPEG-4 system tools, such as BIFS. With the use of MPEG-4 coding, multiple advertisement streams can now be inserted into a stat-mux channel instead of a bitrate-transcoded single advertisement stream using MPEG-2. Currently, standards have been completed in both WG11 for carriage of MPEG-4 streams in networks that carry MPEG-2 transport [1], and Advanced Television Systems Committee (ATSC) that provide for implementation of "targeted advertisements" at the consumer premises equipment by use of appropriate signaling in broadcast streams [11].

Amendment 7 to MPEG-2 systems [1] specifications specifies the insertion of MPEG-4

audio and visual elementary streams into an MPEG-2 transport multiplex with synchronization using the MPEG-2 STC (System Time Clock). In addition, the amendment specifies the use of a MPEG-4 system layer, such as Synchronization Layer (that duplicates some of the PES layer functions). Applications such as AICI (Advanced Interactive Content Initiative) use this amendment for insertion and synchronization of MPEG-4 content with MPEG-2 content. In addition, they also have used BIFS streams to generate advanced compositions on the display screen with user interaction enabled even in a broadcast environment.

The ATSC specification defines signaling based on its system information part of the standard (called Program and System Information Protocol PSIP) for a function called "Directed Channel Change DCC." This is a "virtual" re-tuning of the viewer channel to another part of the transport multiplex at specified timing based on events, as well as user preference settings in the receiver. The switch from the channel being watched to a "directed" channel can occur for criterion such as program identifier, demographic category, postal codes, content subject category, authorization level (premium subscribers) and content advisory values. This information is sent as an MPEG private_section at regular intervals and at the activation time (which could be the cue time of ad-insertion), the receiver can switch to the audio, video and data PID's based on the directed channel change table and switch criterion. Switching back to the original viewed program occurs at the end of the ad-insertion event.

There are two methods of "targeted" ad-insertion that can occur using the above two standards. In the first method, the multiple MPEG-4 based ad-streams are inserted into the stat-mux channel during the ad-insertion period and the DCC directs the receiver to appropriate ad-channel based on user preferences. In the second method, several MPEG-4-based ad-streams can be generated in a multi-program transport stream that can be shared between several stat-mux channels and the DCC for each of the stat-mux channel can direct the user to one of the ad-streams in the large multiplex. The second method allows for larger targeting of streams and for more user preference categories.

Conclusion

This paper discusses some of the advanced features of MPEG-4 standards, which can be implemented to realize advanced A/V content delivery system and targeted commercial insertion in an all-digital environment. Streaming media applications are one of such delivery systems which utilizes network capability effectively and delivers viewer-preferred content on one-to-one basis. Besides, streaming media, which delivers content using IP transport provides unlimited choice to consumers in terms of number of contents available anywhere in the world. This paper also discusses the enabling of low bit rate "effective" advertisement authoring by the emerging MPEG-4 natural and synthetic video coding tools as well as the use of two additional standards that enable "targeted" ad-insertion.

Acknowledgment

References

[1] ISO/IEC 13818-1 (Systems), ISO/IEC 13818-2 (Video) and ISO/IEC 13818-1 Amendment 7, December 2000, ISO/IEC 14696-1 (Systems), ISO/IEC 14696-2 (Video).

[2] Hyder M. Radha, Mihaela van der Schaar, and Yingwei Chen, " The MPEG-4 Fine-Grained Scalable Video Coding Method for Multimedia Streaming over IP" IEEE trans. On Multimedia, Vol. 3, No.1, March 2001.

[3] ISO/IEC 14696-2 MPEG-4 Video Amendment 4: Streaming Video profile, MPEG output document N3518.

[4] Mpeg-4 Overview & Profiles, MPEG home page (www.cselt.it/mpeg/)

[5] Mukta l. Kar, and Sam Narasimhan," Targeted advertisement insertion using MPEG-4 coding and SCTE standards for cue-messaging (DVS 253) and API (DVS 380)"

[6] Mukta L, Kar, Bill Kostka, Majid Chelehmal, and Munsi Haque, " Streaming Over HFC-MPEG-2 or IP or Both?" NCTA paper, 2000.

[7] Julien Signes, " Binary Format for Scene (BIFS): Combining MPEG-4 media to build rich multimedia services", France Telecom R&D Document, CA, USA 1(650) 875-1516.

[8] Digital Program Insertion For Local Advertising, Mukta Kar, Majid Chelehmal, and Richard S. Prodan, NCTA paper, 1998.

[9] Local Commercial Insertion in the Digital Headend, by Mukta Kar, Sam Narasimhan and Richard S. Prodan, NCTA Technical Papers, 2000.

[10] Rate-remultiplexing: An Optimum Bandwidth Utilization Technology, Richard S. Prodan, Mukta Kar and Majid Chelehmal, NCTA Technical Paper, 1999.

[11] Amendment 1A to ATSC standard A/65A, "Program and System Information Protocol for Terrestrial and Broadcast and Cable", May 31, 2000.

[12] Digital Program Insertion Cueing Message for Cable, SCTE Standard DVS 253, December, 1999. Digital Program Insertion Splicing API, SCTE Standard DVS 380 (committee draft).