VIDEO CODING SCHEMES FOR TRANSPORTING VIDEO OVER THE INTERNET

MARTA MRAK, NIKOLA ŠPRLJAN, SONJA GRGIĆ
University of Zagreb, Faculty of Electrical Engineering and Computing
Unska 3 / XII, HR-10000 Zagreb, Croatia
marta.mrak@fer.hr

Abstract: Transporting video over the Internet is an important component of many multimedia applications. Video communication systems should provide the user with services that offers video at an acceptable quality. The current networking structure is loosely defined and heterogeneous, i.e. consisting of a range of interconnected networks with different technologies and capabilities. In order to effectively store and transmit video information, which has an inherently high bandwidth, it has been necessary to develop techniques for video data coding and compression. Transmission of video with its bandwidth, delay and loss requirements poses many challenging problems for designing video coding schemes.

Keywords: Internet, real-time video delivery, video compression, MPEG-4

1. INTRODUCTION

Multimedia distribution over the Internet is becoming increasingly popular. Since the Internet was designed for computer data communication, satisfying the necessary requirements for the effective delivery of multimedia streams poses significant challenge. For example, the Internet is characterised by large bandwidth variations due to heterogeneous access-technologies of the receivers (e.g., analog modem, cable modem, xDSL, etc.). In video multicast the heterogeneity of the networks and receivers makes it difficult to achieve bandwidth efficiency and service flexibility. There are many challenging issues that need to be addressed in designing protocols and mechanisms for Internet video transmission, and specially in designing video coding schemes, [1].

Real-time transport of live video or stored video is the predominant part of real-time multimedia. Video streaming refers to real-time transmission of video. There are two modes for transmission of stored video over the Internet, namely, the download mode and the streaming mode. In the download mode, a user downloads the entire video file and then plays it back. However, full file transfer in the download mode usually suffers long and perhaps unacceptable transfer time. In contrast, in the streaming mode, the video content need not be downloaded in full, but is being played out while parts of the content are being received and decoded. Due to its real-time nature, video streaming typically has bandwidth, delay, and loss requirements, as transmission of real-time video. There is no quality of service (QoS) guarantee for video transmission over current Internet, [2].

2. VIDEO TRANSMISSION OVER THE INTERNET

2.1. Unicast and multicast video distribution

Unicast and multicast delivery of video are important building blocks of many Internet multimedia applications. Unicast video distribution

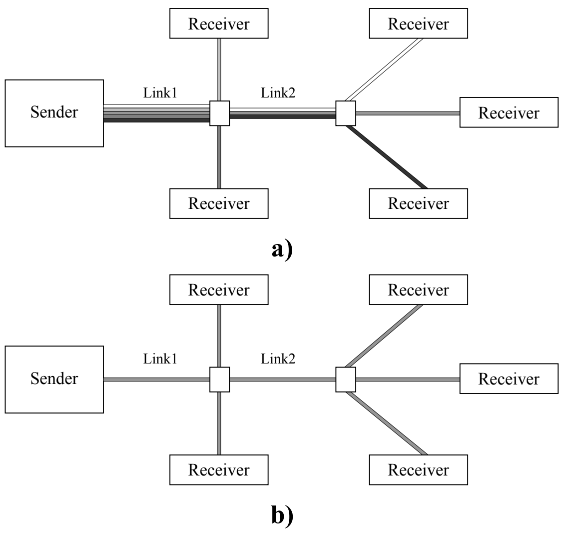

Figure 1. Internet video distribution
a) unicast, b) multicast

uses multiple point-to-point connections, while multicast video distribution uses point-to-multipoint transmission, Figure 1.

For applications such as video conferencing and Internet television, Figure 2, delivery using multicast can achieve high bandwidth efficiency since the receivers can share links. The efficiency of multicast is achieved at the cost of losing the service flexibility of unicast, because in unicast each receiver can individually negotiate service parameters with the source. The heterogeneity of the networks and receivers makes it difficult to multicast Internet video in an efficient and flexible way.

2.2. QoS issues

Real-time transport of video typically has bandwidth, delay and loss requirements, [3].

Bandwidth: To achieve acceptable quality, a streaming application video typically has minimum bandwidth requirement (e.g. 28 Kbps). However, the current Internet does not provide bandwidth reservation to meet such a requirement. Furthermore, since traditional routers typically do not actively participate in congestion control, excessive traffic can cause congestion collapse, which can further degrade the throughput of real-time video. It is desirable for video streaming applications to employ congestion control to avoid congestion, which happens when the network is heavily loaded. For video streaming, congestion control takes the form of rate control, that is, adapting the sending rate to the available bandwidth in the network. Compared with non-scalable video, scalable video is more adaptable to the varying available bandwidth in the network.

Delay: In contrast to data transmission, which is usually not subject to strict delay constraints, real-time video requires bounded end-to-end delay (e.g. 1 second). That is, every video packet must arrive at the destination in time to be decoded and displayed. This is because real-time video must be played out continuously. If the video packet does not arrive in time, the playout process will pause, which is annoying to human visual system. The video packet that arrives beyond a time constraint is useless and can be considered lost. Although real-time video requires timely delivery, the current Internet does not offer such a delay guarantee. In particular, the congestion in the Internet could incur excessive delay, which exceeds the delay requirement of real-time video. Since the Internet introduces time-varying delay, to provide continuous playout, a buffer at the receiver is usually introduced before decoding.

Loss: Loss of packets can potentially make the presentation displeasing to human eyes, or, in some cases, make the presentation impossible. Thus, video applications typically impose some packet loss requirements. Specifically, the packet loss ratio is required to be kept below a threshold (e.g. 1%) to achieve acceptable visual quality. Although real-time video has a loss requirement, the current Internet does not provide any loss guarantee. In particular, the packet loss ratio could be very high during network congestion, causing severe degradation of video quality. Thus, it is desirable that a video stream be robust to packet loss. Multiple description coding is such a compression technique to deal with packet loss.

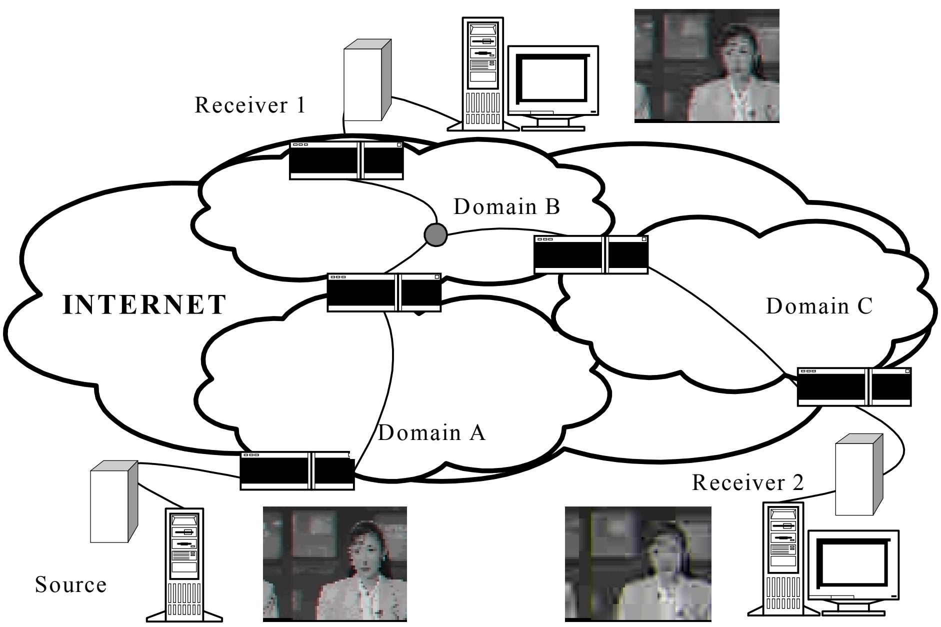

Figure 2. Internet television uses multicast

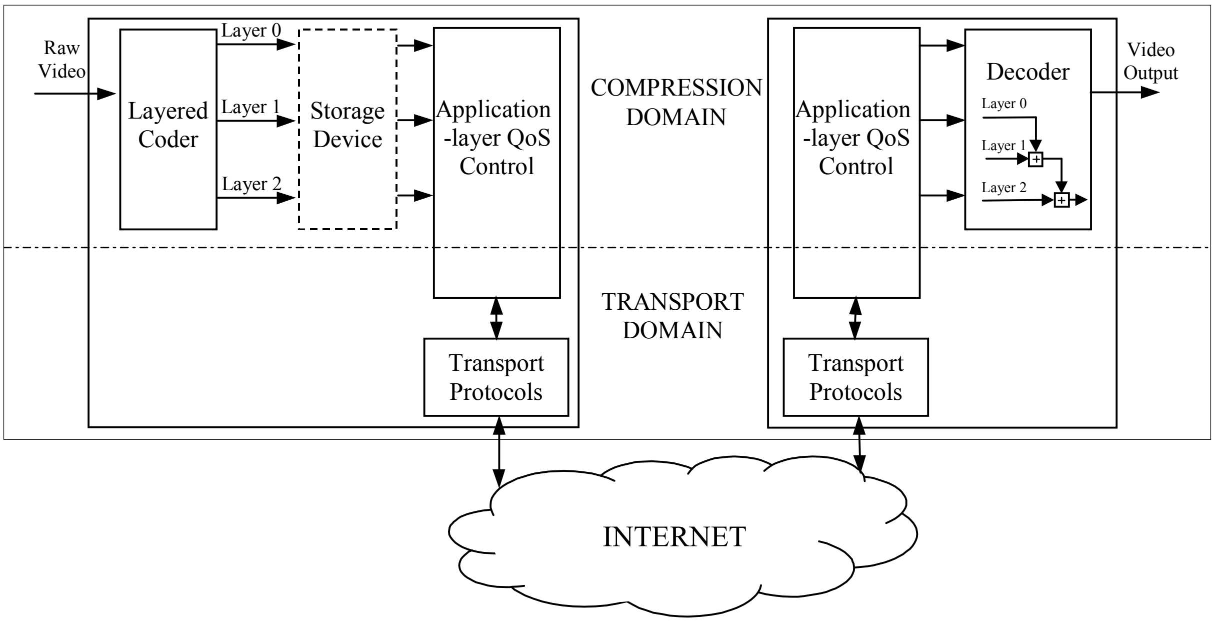

Figure 3. Architecture for streaming video

2.3. Congestion control

Bursty loss and excessive delay have devastating effect on video presentation quality and they are usually caused by network congestion. The purpose of congestion control is to prevent packet loss. However, packet loss is unavoidable in the Internet and may have significant impact on perceptual quality. Thus, other mechanisms must be in place to maximize video presentation quality in presence of packet loss. Existing rate control schemes can be classified into three categories, namely, source-based, receiver-based and hybrid rate control.

Source-based Rate Control: Under the source-based rate control, the sender is responsible for adapting the transmission rate of the video stream. The source-based rate control can minimize the amount of packet loss by matching the rate of the video stream to the available network bandwidth. In contrast, without rate control, the traffic exceeding the available bandwidth could be discarded in the network. Typically, feedback is employed by source-based rate control mechanisms to convey the changing status of the Internet. Based upon the feedback information about the network, the sender could regulate the rate of the video stream. The source-based rate control can be applied to both unicast and multicast.

Receiver-based Rate Control: Under the receiver-based rate control, the receivers regulate the receiving rate of video streams by adding/ dropping channels. In contrast to the sender-based rate control, the sender does not participate in rate control here. Typically, the receiver-based rate control is applied to layered multicast video rather than unicast video. This is primarily because the source-based rate control works reasonably well for unicast video and the receiver-based rate control is targeted at solving heterogeneity problem in the multicast case.

Hybrid Rate Control: Under the hybrid rate control, the receivers regulate the receiving rate of video streams, e.g. by adding/dropping channels while the sender also adjusts the transmission rate of each channel based on feedback information from the receivers. The hybrid rate control is targeted at multicast video and is applicable to both layered video and non-layered video.

Figure 3 shows an architecture for video delivery over the Internet. Raw video can be saved in storage devices after compression. Upon the client's request, the application-layer QoS control module adapts the video bitstreams according to the network status and QoS requirements. After the adaptation, the transport protocols packet the compressed bitstreams and send the video packets to the Internet. Packets may be dropped or experience excessive delay inside the Internet due to congestion. To improve the quality of video/ audio transmission, continuous media distribution services (e.g., caching) are deployed in the Internet. Packets that are successfully delivered to the receiver first pass through the transport layers and then are processed by the application layer before being decoded at the video decoder. Under this architecture, which uses layered based coding, a QoS monitor is maintained at the receiver side to control network congestion status based on the behaviour of the arriving packets, e.g., packet loss and delay. Such information is used in the feedback control protocol, which sends information back to the video source. Based on such feedback information, the rate control module, application-layer QoS control, estimates the available network bandwidth and regulates the video output rate of the video stream according to the estimated network bandwidth.

3. VIDEO CODING

3.1. Video coding standards

The most important video codec standards for streaming video are H.261, H.263, MPEG-1, MPEG-2 and MPEG-4. Compared to video codecs for CD-ROM or TV broadcast, codecs designed for the Internet require greater scalability, lower computational complexity, greater resiliency to network losses, and lower encode/decode latency for video conferencing. New algorithms specifically targeted at Internet video are being developed. Most recent efforts on video compression for streaming video have been focused on scalable video coding, which is included in MPEG-4 standard in many ways. The primary objectives of on-going research on scalable video coding are to achieve high compression efficiency at affordable cost/complexity. A promising direction on scalable video coding is to integrate several video coding techniques to deal with QoS fluctuations in the networks. Scalable video coding is capable of coping with bandwidth variations, [4].

MPEG-4 is an ISO/IEC standard developed by MPEG (Moving Picture Experts Group) and adopted in 1998. The mandate for MPEG-4 was to standardize algorithms for audio-visual coding in multimedia applications so as to support content-based interactivity, high compression, and/or universal accessibility and portability of audio and video contents. The visual part of the standard provides profiles for object-based coding of natural, synthetic, and hybrid visual contents, in which bit rates targeted are within 5-64 kbps for Internet and mobile applications, 2 Mbps for TV/film applications and 19 Mbps for HDTV applications, [5].

3.2. Layered video coding

A non-scalable video encoder generates one compressed bitstream. In contrast, a scalable video encoder compresses a raw video sequence into multiple layers. One of the compressed layers is the base layer, which can be independently decoded and provide coarse visual quality. Other compressed layers are enhancement layers, which can only be decoded together with the base layer and can provide better visual quality. The complete bitstream (i.e., combination of all the layers) provides the highest quality.

Specifically, compared with decoding the complete bitstream (Figure 4.a), decoding the base layer or multiple layers produces pictures with degraded quality (Figure 4.b), or a smaller image size (Figure 4.c), or a lower frame rate (Figure 4.d). The scalabilities of quality, image sizes, or frame rates, are called SNR (signal-to-noise ratio), spatial, or temporal scalability, respectively. These three scalabilities are basic scalable mechanisms and these scalabilities have been included in MPEG-2 and MPEG-4. There can be combinations of the

a)

b)

c)

d)

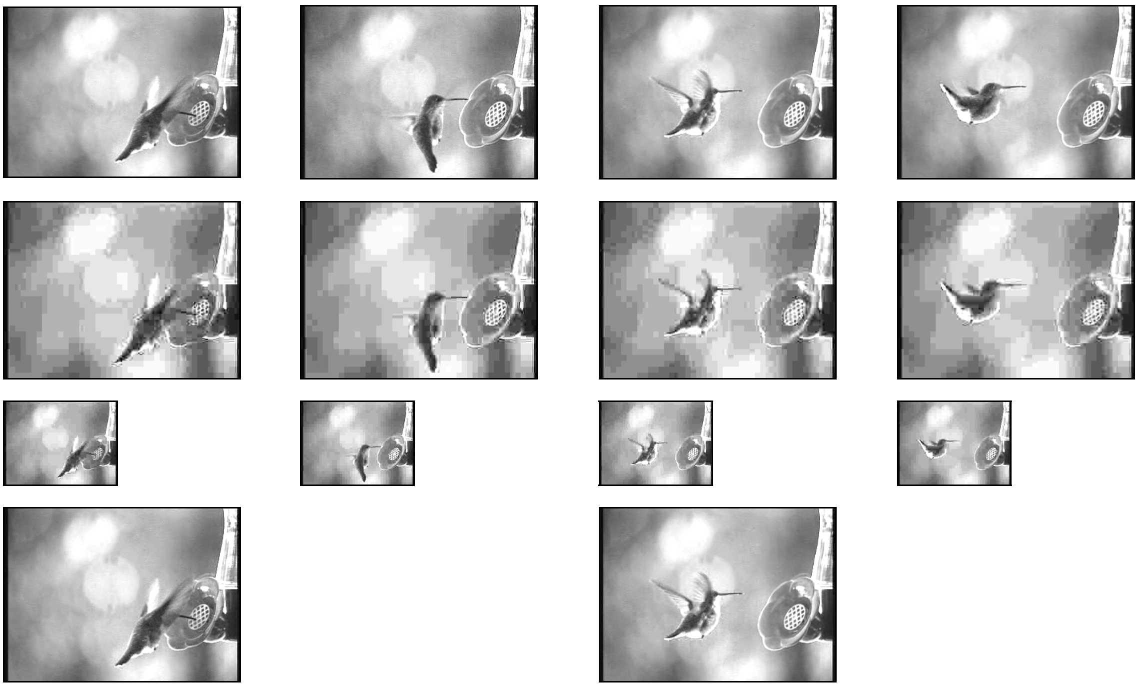

Figure 4. Scalable video; video frames a) reconstructed from the complete bitstream, b) with degraded quality, c) with a smaller image size, d) with a lower frame rate

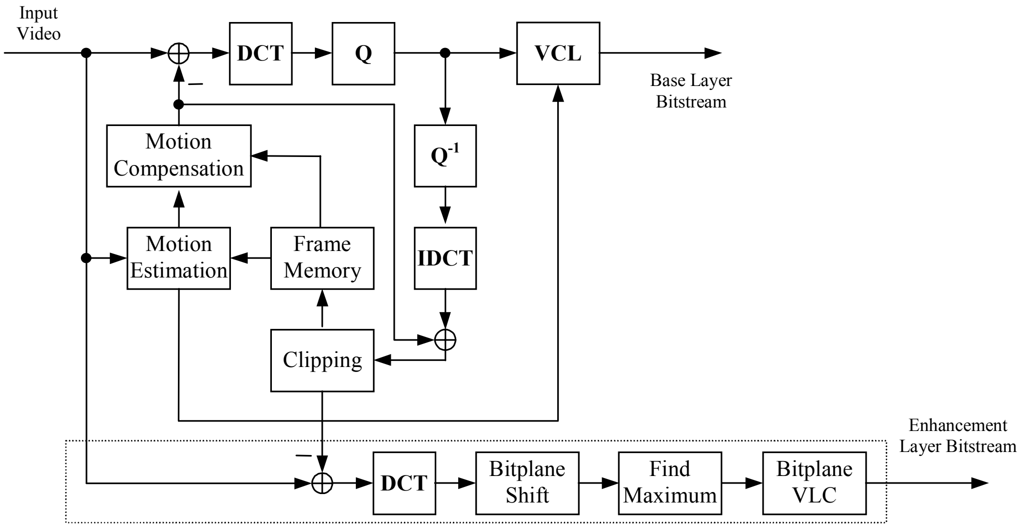

Figure 5. FGS encoder structure

basic mechanisms, such as spatio-temporal scalability. A common characteristic of the layered scalable coding techniques is that the enhancement layer is either entirely transmitted/received/decoded or it does not provide any enhancement at all, [6].

The architecture for video delivery over the Internet from Figure 3 uses layered video coding. At the sender side, a raw video sequence is compressed into multiple layers: a base layer (i.e., Layer 0) and one or more enhancement layers (e.g., Layers 1 and 2 in Figure 3). The base layer can be independently decoded and it provides basic video quality; the enhancement layers can only be decoded together with the base layer and they further refine the quality of the base layer.

3.3. Fine granularity scalability

In response to the growing need for a video coding method for streaming video on Internet applications, the MPEG video group has developed Amendment 4 of MPEG-4, which includes a set of tools for fine granularity scalability (FGS) and its combination with temporal scalability (FGST).

The basis idea of FGS is also to code a video sequence into a base layer and an enhancement layer. The base layer uses nonscalable coding to reach the lower bound of the bit-rate range and the difference between the original picture and the reconstructed picture is coded using bit-plane coding of the DCT coefficients into the enhancement layer. Figure 5 shows the FGS encoder structure. The bitstream of the FGS enhancement layer may be truncated in into any number of bits per picture after encoding is completed. The decoder should be able to reconstruct an enhancement video from the base layer and the truncated enhancement-layer bitstreams. The enhancement-layer video quality is proportional to the number of bits decoded by the decoder for each picture.

Bitplane coding of the DCT coefficients is used as the basic coding technique for FGS (Figure 5). The FGS enhancement layer encoder takes the original frame and reconstructed frame as input and produces an FGS enhancement bitstream. The difference between the original and reconstructed frames is transformed by the DCT to generate the DCT coefficients. After obtaining all the DCT coefficients of a frame, bitplane shift operation can be performed. Then the maximum absolute value of the DCT coefficients is found and the maximum number of bitplanes for the frame is determined. The 64 absolute values of each DCT block are zigzag ordered into an array. A block bitplane is formed as an array of 64 bits, taken one from each absolute value of the DCT coefficients at the same significant position. In order to cover a wide range of bitrate, there is a need to combine FGS with temporal scalability so that not only picture quality can be scalable but also temporal resolution (frame rate) can be scalable, [7, 8].

3.4. Comparison of video coding techniques

The objective of video coding for Internet is to optimise the video quality over a given bit rate range. The bitstream should be partially decodable at any bit rate within the bit rate range to reconstruct a video signal with the optimised quality at that bit rate. The distortion-rate curve in Figure 6 indicates the upper bound in quality for any coding technique at any given bit rate. Layered scalability techniques change the nonscalable single staircase curve to a curve with two stairs. The desired objective is to achieve the continuous curve parallels the distortion-rate curve with a single bitstream and it is done with FGS video-coding technique.

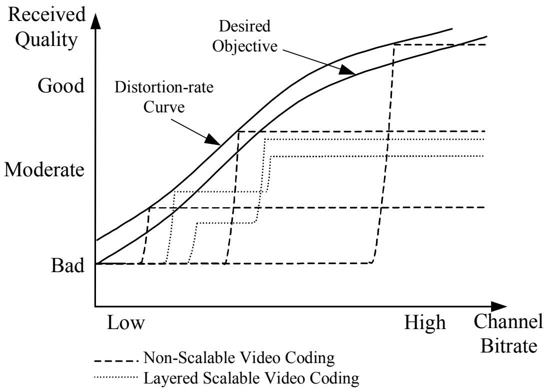

Figure 6. Video coding performance

The major difference between FGS and the layered scalable coding techniques is that, although the FGS coding technique also codes a video sequence into two layers, the enhancement bitstream can be truncated into any number of bits within each frame to provide partial enhancement proportional to the number of bits decoded for each frame. Therefore, FGS provides the continuous scalability curve illustrated in Figure 6.

4. CONCLUSION

Recent advances in computing technology, compression technology, high bandwidth storage devices, and high-speed networks have made it possible to provide real-time multimedia services over the Internet. We have discussed various requirements and control techniques for Internet video, and video compression schemes imposed by Internet applications on video codec. It is necessary for the encoder to be able to discard a part of the video bit stream with graceful degradation to fit the channel bandwidth if the channel to the server is crowded by many client's access. Designers of Internet video transmission service need to choose an appropriate video coding scheme, which meets the target efficiency at an affordable cost/ complexity.

REFERENCES

[1] M. J. Riley, I. E. G. Richardson, *Digital Video Communications*, Artech House Publishers, Boston, 1997.

[2] D. Wo, Y. T. Hou, Y. Zhang, *Transporting Real-Time Video over the Internet: Challenges and Approaches*, Proceedings of IEEE, Vol. 88, No. 12, pp. 1855-1874, December 2000

[3] D. Wo, Y. T. Hou, W. Zhu, Y. Zhang, J. M. Peha, *Streaming Video over the Internet: Approaches and Directions*, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, No. 3, pp. 282-300, March 2001

[4] M. Mrak, *Transporting Video Over the Internet*, Faculty of Electrical Engineering and Computing, University of Zagreb, Technical Report (in Croatian), Zagreb, Jun 2001

[5] MPEG Requirements Group, *MPEG-4 Applications*, Document ISO/IEC JTC1/SC29/ WG11 MPEG 99/N2724, Seoul MPEG meeting, March 1999

[6] *MPEG-4 Video Verification Model version 18.0*, ISO/IEC JTC1/SC29/WG11 N3908, Pisa, January 2001

[7] W. Li, *Overview of Fine Granularity Scalability in MPEG-4 Video Standard*, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, No. 3, pp. 301-317, March 2001

[8] R. Lukac, *Boolean LUM Smoother*, Journal of Electrical Engineering, Vol. 51, No. 9-10, pp. 264-268, 2000