SCALABLE VIDEO CODING IN NETWORK APPLICATIONS

Marta Mrak, Mislav Grgic, Sonja Grgic

University of Zagreb, Faculty of Electrical Engineering and Computing
Unska 3 / XII, HR-10000 Zagreb, Croatia
marta.mrak@fer.hr

Abstract: *Streaming video over packet networks is considered a joint source/channel coding problem, where the necessity of real-time delivery, bandwidth constraints, and error resilience determine the design parameters for the video transmission system. The objective of video coding for network applications is changed to optimising the video quality over a given bit rate range instead of at a given bit rate. Meeting bandwidth requirements and maintaining acceptable image quality simultaneously is a challenge. This paper provides an overview on the scalable video coding methods.*

Key words: *Real-Time Video Delivery, Video Compression, Scalability, Internet*

1. INTRODUCTION

Delivery of video in the presence of bandwidth constraints is one of the most important video processing problems. Network channel capacity varies over wide range depending on the type of connections and the traffic at any given time, so it is difficult to predict the traffic on a network when video is to be delivered. 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]. Most current compression techniques require that parameters, such as data rate, be set at the time of encoding. Compression techniques that allow the change of the compression parameters at the time of decoding are very valuable due to the flexibility they provide. These techniques are said to be "scalable." Rate scalability, the capability of decoding a compressed image or video sequence at different data rates, is the one of the most important modes for video streaming over packet networks such as the Internet.

Rate scalable compression that allows the decoded data rate to be dynamically changed, is appealing for many applications, such as video streaming and multicasting on the Internet, video conferencing, video libraries and databases, and wireless communication. In these applications, the bandwidth available cannot be guaranteed due to variations in network load.

2. QoS ISSUES AND CONGESTION CONTROL

Since the Internet was designed for computer data communication, satisfying the necessary requirements for the effective delivery of multimedia streams poses significant challenge. 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].

Bandwidth: To achieve acceptable quality, a streaming application video typically has minimum bandwidth requirement (e.g. 28 Kbps). Congestion control for video streaming 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). Video packet that arrives beyond a time constraint is useless and can be considered lost. Since the network 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. Specifically, the packet loss ratio is required to be kept below a threshold (e.g. 1%) to achieve acceptable visual quality. The packet loss ratio could be very high during network congestion, causing severe degradation of video quality. 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.

When a video sequence is transmitted over a heterogeneous network, network congestion may occur, decreasing the quality observed by the user. 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, [3].

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. 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.

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.

Fig. 1 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.

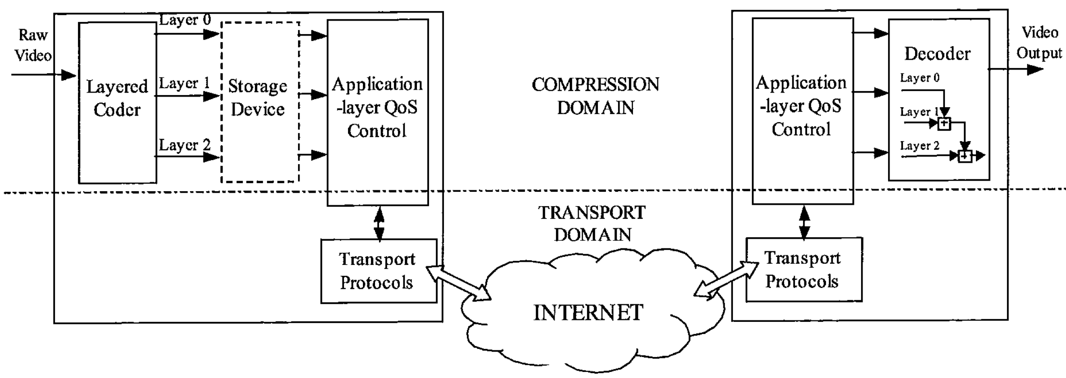

Fig. 1. Architecture for streaming video

3. VIDEO CODING

3.1. Video coding standards

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. The objective of a video compression algorithm is to exploit both the spatial and temporal redundancy of a video sequence such that fewer bits can be used to represent the video sequence at an acceptable visual distortion. 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 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 scalable video coding

The objective of video coding for Internet is to optimise the video quality over a given bit rate range. In rate scalability compression systems, receiver can request a particular data rate, either chosen form a limited set of rates (layered rate scalability) or from a continuous set of data rates (continuous rate scalability). 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. 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.

The distortion-rate curve in Fig. 2 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 continuous curve, which is parallel with distortion-rate curve with a single bitstream.

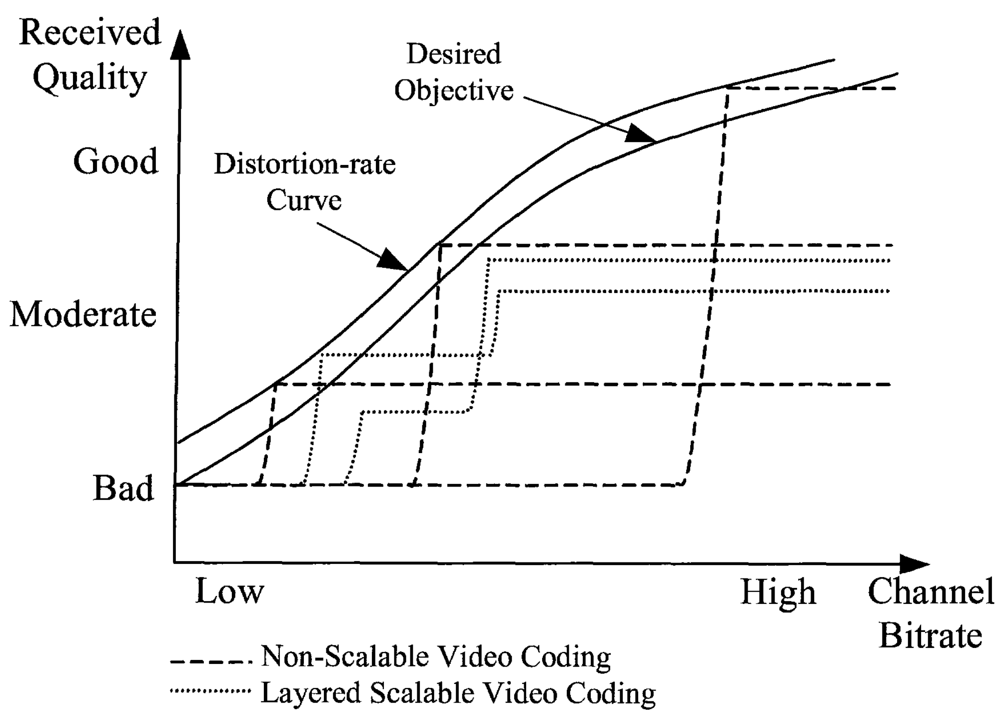

Fig. 2. Video coding performance

The architecture for video delivery over the Internet from Fig. 1 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 Fig. 1). 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.

Specifically, compared with decoding the complete bitstream (Figure 3 a), decoding the base layer or multiple layers produces pictures with degraded quality (Figure 3 b), or a smaller image size (Figure 3 c), or a lower frame rate (Figure 3 d).

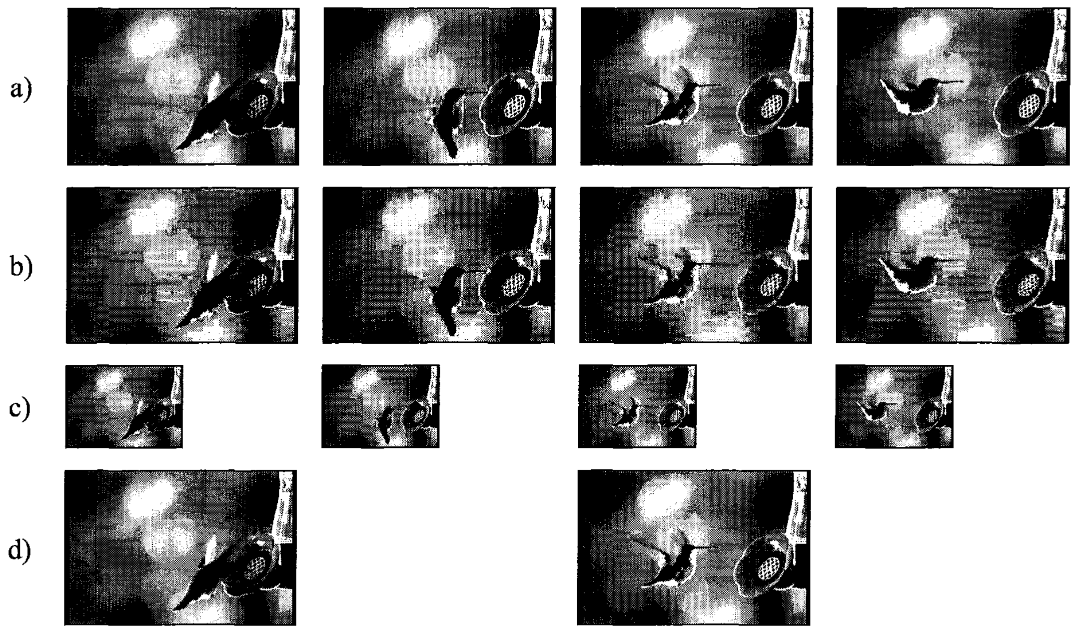

Fig. 3. Scalable video; video frames a) reconstructed from the complete bitstream, b) with degraded quality, c) with a smaller image size, d) with a lower frame rate

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. Part of MPEG-4 is content scalability, which is the ability to determine the content of a scene at the decoder. There can be combinations of the 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].

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. Fig. 4 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. 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].

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 Fig. 2.

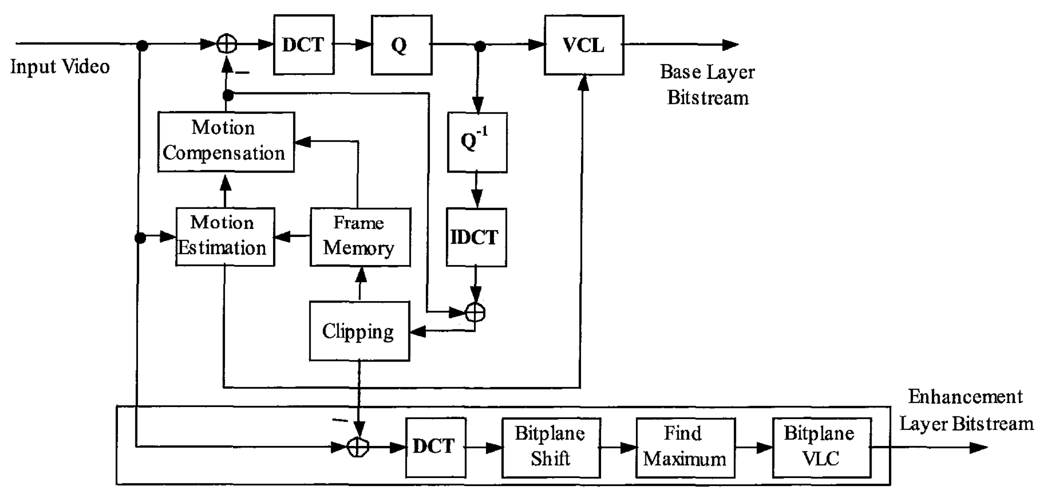

Fig. 4. FGS encoder structure

3.4. Wavelet transform in video coding

More recently, the Discrete Wavelet Transform (DWT) has become a popular tool for signal processing, due to its superior space-frequency decomposition properties. DWT can provide better image quality than DCT, especially at higher compression ratio, and it is an important tool in the constructing of resolution-scalable bit-stream.

Wide variety of novel and sophisticated wavelet image coding schemes, like EZW, SPIHT and EBCOT, have been developed. EZW and SPIHT algorithms have distortion-progressive ordering of the compressed bit-stream, as zero-tree involves downward dependencies between the subbands produced by successive DWT stages. Ones compressed, the bit stream cannot be reordered so as to support decompressors with reduced memory resources, [8], [9].

The basic encoding engine of JPEG2000, new ISO/ITU-T standard for still images coding, is based on DWT and EBCOT algorithm. The basic idea in EBCOT is to code bit-planes of quantised subbands coefficients after octave-band decomposition and dividing of each subbands into block of samples. JPEG 2000 places a strong emphasis on scalability, so the scalability property, in its different forms, pertains to the ordering of information within the bit-stream. Structure of JPEG 2000 coding mechanism can be a basis for further scalable video coding systems, because it provides resolution scalable and SNR scalable bitstream.

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 video compression schemes imposed by network 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 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 and I. E. G. Richardson, *Digital Video Communications*, Artech House Publishers, Boston, 1997

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

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

[4] M. Mrak, *Transporting Video Over the Internet*, Faculty of Electrical Engineering and Computing, University of Zagreb, Technical Report (in Croatian), Zagreb, June 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, March 2001, pp. 301-317

[8] D. Taubman, E. Ordentlich, M. Weinberger and G. Seroussi, *Embedded Block Coding in JPEG2000*, Hewlett-Packard Company Publication, February 2001

[9] R. Lukac, *Vector LUM Smoothers as Impulse Detector for Color Images*, Proceedings of European Conference on Circuit Theory and Design, ECCTD'01, Espoo, Finland, August 28-31, 2001, Vol. III, pp. 137-140