2—3 Bandwidth-Adaptive Video Transmission Method for Heterogeneous Network Environment

S. Sakazawa[1] Y. Takishima[1] M. Wada[1] K. Amano[2]
KDD[1] Hiroshima City University[2]

sakazawa@img.lab.kdd.co.jp

Abstract

For the purpose of a flexible coded video transmission over a heterogeneous network, we propose a new packetization method for coded video data. The proposed method achieves small degradation of coded picture quality in case of packet discard at the network node and does not require heavy processing load for bitrate control operation. Computer simulation results show that the bitrate reduction from 384 kb/s to 192 kb/s does not cause severe degradation in picture quality.

1. Introduction

The Internet is growing at an enormous speed, and the LAN is expanding in every office. In these situations, video applications attract people's attention and there are many kinds of products for video conferencing and broadcasting[1][2].

As these applications are all point-to-point based, extension to a multipoint network environment is required in order to utilize network resources effectively. However, as the network resources (ex. available bandwidth) may not be uniform throughout all networks, there are various requirements for video streams including bitrates which are difficult to satisfy completely. For example, if one terminal in a "rich" environment generates a high bitrate video stream, it may cause congestion in a "poor" network and the received video quality at the poor network terminal may be severely degraded. On the other hand, if one terminal generates a stream conforming to the "poor" network limitation, all the receivers will receive poor quality video.

To cope with this problem it is necessary to convert the bitrate of the video stream at a network node (ex. gateway or router) which connects one network domain to the other[3]. In other words, bitrate reduction is needed when the stream enters the poor bandwidth network domain.

In this paper, we propose a bitrate reduction method that is applicable for not only point-to-point communication but also multicast communication. The feature of our method is that does not cause severe degradation in picture quality and it is easier to implement than other methods. In the next section, various bitrate reduction and control methods are introduced and discussed. In Section 3, the proposed method is described in detail. The proposed method is evaluated by a computer simulation in the following section and the result is discussed in Section 5. Section 6 gives conclusions.

2. Review of various bitrate control methods

In this section, various bitrate reduction methods are introduced and discussed concerning their advantages and disadvantages.

2.1 Control of encoding bitrate

In this method, the encoder controls the bitrate of the generated video stream according to the network congestion, demand from receivers and so on.

The advantages of this method are:

1) Arbitrary bitrate reduction is possible.
2) Network utilizing efficiency is high.
3) Encoding and decoding algorithms are simple.
4) A rate conversion function is not needed in the network node.

The disadvantages are:

1) Some communication mechanism is required between the encoder and a large number of receivers for bitrate negotiation.
2) All receivers must receive the same bitrate video stream.

This causes congestion in poor networks or unreasonable low quality video in rich networks as described in Section 1.

2.2 Simulcast

In this method, multiple video streams are transmitted simultaneously on the network and the network node selects the appropriate bitrate stream.

The advantages are:

1) Encoding and decoding algorithms are simple.
2) Rate conversion at the network node is achieved just by selecting the stream.

The disadvantages are:

1) Network utilizing efficiency is low.
2) Arbitrary bitrate reduction is difficult.

In order to leave choices to the end receivers in poor networks, each network node must relay a low bitrate stream best suited for its own network domain. Therefore the number of streams is limited.

2.3 Transcoding

In this method, the video stream is decoded and re-encoded at the network node in order to adjust the bitrate to the bandwidth of the network domain.

This method has the following advantages:

1) An arbitrary bitrate can be chosen at the network node.
2) Encoding and decoding algorithms at both ends are simple.
3) Network utilizing efficiency is high.

On the other hand, the following disadvantages are raised:

1) The network node has a heavy load in performing transcoding.
2) Video quality is inferior to direct coding due to transcoding loss.

2.4 Layered coding

In this method, the source video is encoded as a layered stream on the encoder side, in which one stream contains core information and other stream(s) contain additional information to improve the video quality. Bitrate reduction is achieved by dropping the additional stream(s) at a network node.

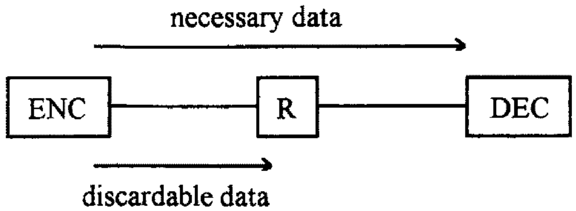


Fig. 1 Proposed method

This method has the following advantages:

1) The bitrate reduction process is simple at the network node.
2) Network utilizing efficiency is high.

On the other hand, the following disadvantages arise:

1) Encoding and decoding algorithms are complicated.
2) The choice of reduced bitrates is limited.

Practically, layered coding may lose coding efficiency when many layers are employed, therefore the number of layers is limited to some extent.

3. Proposed method

In this section, we first consider the requirement for multicast over a heterogeneous network environment. Then, we propose a new video multicast architecture and its controlling algorithms which meet the above requirement.

3.1 Requirements

A multicast protocol is very generic, and it can be used in both small networks and huge networks such as MBONE. It can be foreseen that the range of available bandwidth for each network domain is large, the number of network nodes is huge, and the number of receivers is enormous.

With the above considerations, we defined the requirements as follows.

1) Arbitrary bitrate reduction is mandatory.
2) The rate reduction process performed by the network node should be simple.

The function of network nodes should be limited to the transport layer.

3) The decoding process must be simple and be executable only by software.
4) Coding efficiency should be high.
5) Video quality should be maintained to some extent even in the case of poor bandwidth, compared with direct coding.

3.2 Proposed architecture

3.2.1 Encoder

An MC+DCT video encoder such as H.261[4] is used. In this algorithm, a video signal is encoded in units of pixel blocks. The block video data is first predicted using the previous video frame, and the prediction error is DCT-transformed and quantized. The quantized data is zigzag-scanned and encoded with a Huffman code.

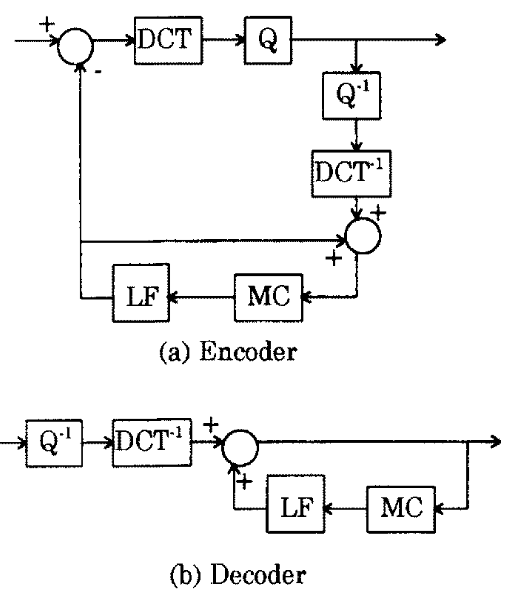
Here, DCT coefficients corresponding to high frequency have less importance in subjective picture quality compared with the lower frequency coefficients. This means severe degradation does not occur in subjective picture quality when the higher frequency coefficients are discarded. On the other hand, if the prediction parameters such as motion vector or the lower frequency coefficients are damaged, it may severely deteriorate the decoded video quality. For this reason we employ a data partitioning method that separates the zigzag-scanned data into lower and higher frequency parts. Thus, important data and less important data are separately packetized based on the above data partitioning. It should be noted that the minimum bitrate, which is achieved by discarding all less important packets, is decided at this stage.

3.2.2 The network node

A function of the network node here is discarding the less important packets to reduce the bitrate, while the important packets must be protected. Although there may be various and complicated discarding methods, we employ a very simple method in which a discarded packet is selected regularly: one out of two less important packets or one out of three and so on.

3.2.3 Decoder

The decoder reconstructs a video signal from the received data. When only an important packet is received, the decoder adds an EOF code at the end of zigzag-scanned data and decodes it by the usual decoding process. When both important and less important packets are received, the decoder concatenates them to make the complete zigzag-scanned data and decodes them.



Q : Quantizer Q^{-1}: inverse Quantizer
LF : Loop Filter DCT^{-1}: inverse DCT

Fig. 2 Block diagram of H.261

3.3 Control algorithm of data partitioning[5][6]

Classification of zigzag-scanned data has a definite impact on the decoded video quality when a packet discard occurs. In this subsection, we propose two different algorithms. It must be noted that the difference between the data partitioning algorithms does not have any effect on the network node and the decoder, because they treat data only as a packet.

3.3.1 Method 1

In this method, the boundary of important and less important coefficients in the zigzag-scan is fixed over all blocks as depicted in Figure 3. In addition, a kind of rate control mechanism is needed at the encoder to keep the minimum bitrate to the target value. Such a mechanism will employ a virtual buffer for the important packets, and when the buffer is full the boundary is changed to the lower frequency and vice versa.

3.3.2 Method 2

The basic idea for this method is to minimize the degradation of picture quality caused by the packet discard. In order to estimate the degradation, the sum of each DCT coefficient importance index is calculated. Here, we use the visually-weighted DCT coefficient energy[7] as a visual importance index. The weighting function is Equation (1).

$$w = \frac{2.46(0.1 + 0.25f)\exp(-0.25f)}{\sqrt{2.02\log_{10} S}} \tag{1}$$

Here, S is the AC power of the block, and f is the spatial frequency. Therefore the weighting factor changes according to the frequency and local luminance activity, which reflects human visual sensitivity.

The minimization of degradation is achieved by the following procedure for a set of DCT blocks such as a macroblock.

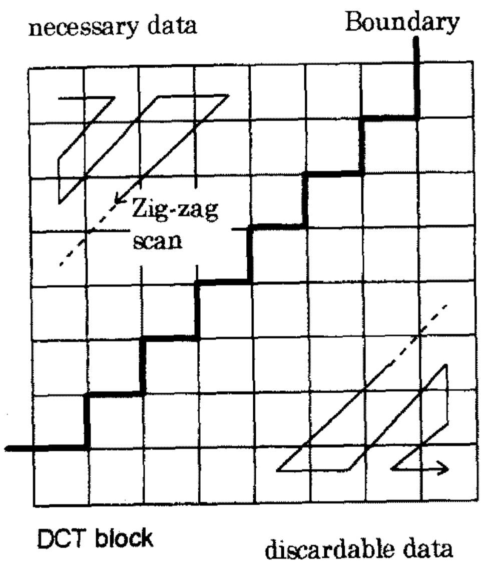
1) calculate the importance index of each DCT coefficient
2) arrange them in a zigzag-scan order for each DCT block
3) select the least index among the last coefficients of the arranged set and move the

corresponding DCT coefficient to the less important packet
4) reduce the number of bits for that DCT coefficient
5) go back to step 3) until the total reduced bits reaches the target bits

The advantage of this method is that the data partitioning can be completely achieved to meet the target minimum bitrate.

4. Computer simulation

In this paper H.261 is employed as a coding algorithm for the computer simulation. The coded frame rate is about 11 frames/sec. Intra-refresh macroblock method is used and its refresh period is



(a) method 1

zigzag-scanned (run, level) event
=> (bit length, importance index) sequence

The least index is selected among the last event of the arranged set of blocks and put into the discardable packet

Block #1 : (bit, index), ..., (16, 1) → discardable packet
Block #2 : ...., (20, 15)
....
Block #N : ...., (10, 15)

(b) method 2

Fig. 3 Control algorithm of data partitioning

132 frames, that is, 3 macroblocks per frame are encoded in an intra-frame mode for a CIF picture. The test sequences are CLAIRE and SALESMAN in a CIF format of 150 frames. When the proposed methods are applied, intra-refresh macroblocks are excluded from the packet discard.

The simulation is carried out in order to examine the decoded video quality in case of packet discard. Figure 4 shows the coding characteristics for H.261 in which the proposed methods reduce the bitrate of the video stream originally encoded at 384 kbps to the minimum bitrate of 128 kbps. Figure 5 shows the decoded picture of the 100th frame in the case of a target bitrate is 192 kbps.

Figure 4 shows that the proposed method 2 gives 2 dB higher SNR than the proposed method 1. From subjective evaluation of Figure 5, method 2 also gives better quality than method 1.

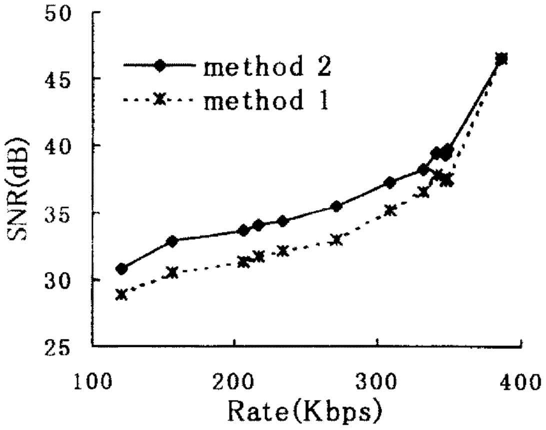
5. Discussion

The MC+DCT coding system has a premise that synchronization of the frame memory is maintained between the encoder and the decoder. Such a synchronization is destroyed in case of data loss. Once synchronization is destroyed, it continues until the frame memories at both ends are refreshed.

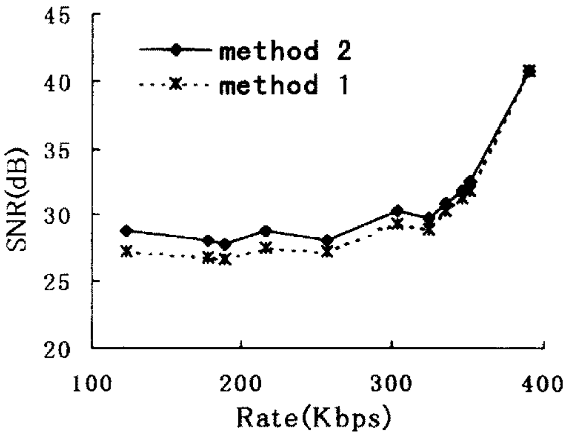
Although the proposed methods use the packet discard approach and may cause this mismatch of the frame memory, the decoded video is not destroyed. A compensation mechanism is necessary for improving picture quality in our system, and the candidates are macroblock refresh and LPF inside the MC+DCT loop.

First, we analyze the mismatch effect. In order to evaluate the mismatch effect, the SNR of the 150th frame is calculated for a bitrate reduction of from 384 kbps to 192 kbps in the case 1) H.261 direct coding, 2) proposed methods (packets for the last frame are not discarded), and 3) H.261 direct coding (packet discard is applied only to the last frame). 2) shows the effect of the mismatch, while 3) shows the instantaneous effect of the packet discard without the mismatch. Table 1 shows that the SNR of 2) is much worse than 3), therefore it can be concluded that frame memory mismatch is a dominant factor in video degradation in our system.

Next, we consider the loop filter and refresh protection. Macroblock refresh was originally employed in H.261 for the purpose of reduction of IDCT mismatch. As it refreshes the frame memory by intra-frame coding, it is also effective for

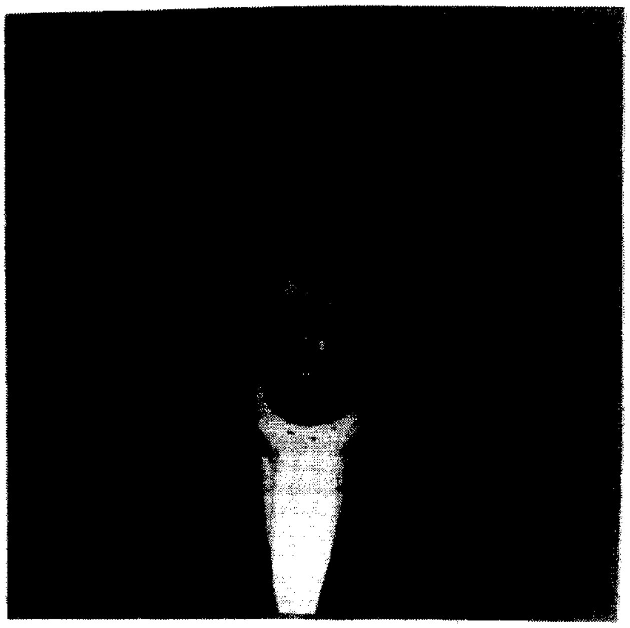


(a) CLAIRE



(b) SALESMAN

Fig. 4 Coding characteristics



(a) Decoded picture at 384 kbps in case of the whole packets are received.

(b) Method 1 (192 kbps)

(c) Method 2 (192 kbps)

Fig.5 Decoded pictures

Table 1 Comparison of the picture quality

	SNR (dB)
packet discard effect	41.7
mismatch effect	34.8
direct coding	46.8

reduction of mismatch caused by packet discard. Meanwhile, a loop filter is originally employed in H.261 for the purpose of improving coding efficiency by suppressing the high frequency energy of the prediction picture where quantization noise energy may be higher than the signal energy. In our proposed methods, the noise is limited to high frequency coefficients, and the loop filter can suppress that noise too. To evaluate their effects, the proposed methods without the protection of macroblock refresh or without a loop filter are compared in Figure 6. The results of the proposed methods without both are also shown in Figure 6. Figure 6 shows that the effects of each factor are almost the same, and employing these methods improves the video quality considerably.

6. Conclusion

New video packet transmission schemes are proposed. The key techniques employed here are the efficient data partitioning at the encoder and a clever control algorithm, which have the advantages of simple packet discard at the network node and a simple video decoder conforming to the standard. These techniques achieve very flexible traffic control and easy implementation of multicast video. Computer simulation shows that the proposed method gives good video quality even in the case of packet discard.

The reason why the standard decoder can manage the packet discard was also considered. This is because techniques employed in the decoder for other purposes are also effective to reduce the frame memory mismatch between the encoder and the decoder.

Acknowledgment

The authors wish to express their deep appreciation to Dr. Murakami, Dr. Hatori and Mr. Koga of KDD R&D Laboratories for their encouragement and support.

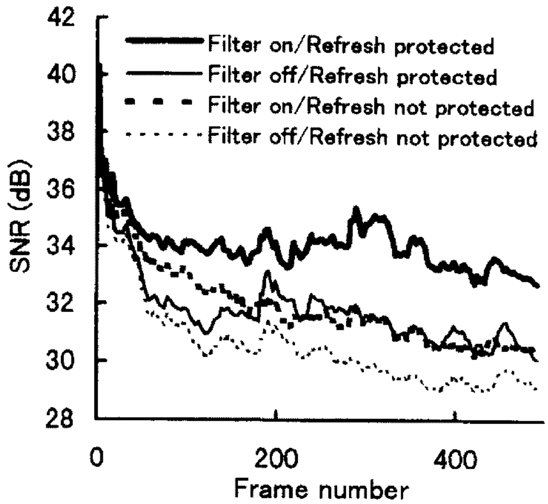


Fig. 6 Quality characteristics

References

[1] ITU-T Draft Recommendation H.323 (1995).
[2] http://www.xingtech.com/
[3] McCanne, Jacobson and Vetterli, "Receiver-driven Layered Multicast", In proceedings ACM SIGCOMM'96, pp.117-130 (1996).
[4] ITU-T Recommendation H.261 (1990).
[5] Sakazawa, Takishima, Wada, "A Study on Rate Conversion for Video Communications Between LANs", 1996 IEICE Spring National Conf., D-214. (in Japanese)
[6] Sakazawa, Takishima, Wada, "A Consideration on Packet Video Coding Method for Bitrate Control on the Network.", 1996 ITEJ National Conf. 24-3. (in Japanese)
[7] Hamada, Matsumoto, "Optimal Quantization Scheme of Orthogonal Transform Coeffcients Considering Noise Masking Effect due to Local Activity of Video Signal", IEICE Trans. B-I Vol.J75-B-I, No.12, pp.791-801 (1992). (in Japanese)