

行政院國家科學委員會補助專題研究計畫期末報告

(計畫名稱)

視訊傳輸的互動迴授型式及錯誤防止

計畫類別： \*整合型計畫

計畫編號：NSC 90-2213-E-009-141

執行期間： 90年 08月01日至 91年 07月31日

計畫主持人：張文鐘

執行單位：交通大學電信系

中華民國 91年 8月

# Interactive feedback message and Error resilience in Video Transmission

## 視訊傳輸的互動迴授型式及錯誤防止

計畫編號：NSC 90-2213-E-009-141

執行期限：90年08月01日至91年07月31日

主持人：張文鐘 交通大學電信系

E-mail: wtchang@cc.nctu.edu.tw

### 一,中文摘要

關鍵字：(速率控制，自動重傳，緩衝器控制) 選擇性的自動重傳是一個回授信號用來作為指示錯誤的封包。這種方法被用於目前的一些數據傳輸。當此種方法被用來傳即時視訊時，由於資料源源不絕，緩衝器管理變成一個問題，本研究探討此一問題並提出一些解決即時視訊因為緩衝器溢滿而引起跳格的問題。

**ABSTRACT** (keywords: rate control, ARQ, buffer control): Selective ARQ is a kind of feedback message to indicate transmission error and request re-transmission. This kind of method is commonly used in data communication. When this scheme is used for real time video, buffer fullness arises and buffer throughput reduces. From the buffer condition, rate control for video coding is discussed to take this real time buffer management into consideration such that frame skip in real time video can be reduced.

### 二,緣由與目的

For real time video transmission over the Internet, issues such as video streaming, flow control, network traffic monitoring, call setup, capacity exchange have been discussed under the umbrella of the standard H.323. The framework for video streaming includes adaptive source encoding, elementary stream packetization, error control for the loss packet, feedback message based on RTP/RTCP, TCP and ARQ in MAC layer, packetization with RTP/UDP/IP format for transport. is discussed. Video

streaming implies that the video is being played out while the system is still receiving and decoding the future video stream. In video transmission, the compression layer generates the compressed elementary stream(ES). The ES is first packetized to provide timing, synchronization and random access point. All the packetized ES streams are then multiplexed into a transport stream. The packetized transport stream is then passed to the transport layer for Ethernet packetization. The ES layer packetization and MUX operation will add headers (start code) to the ES packet before the including of the RTP/UDP/IP format.

The purpose of the adaptive source encoding algorithm is to adjust the output bit rate to match the network bandwidth and buffer fullness. An upper layer packetization algorithm for the encoded bit stream at the sync (session) layer before passing to the RTP layer is for the purpose of efficiency and robustness. The packet size is limited by the size of the message transfer unit (576 bytes for Ethernet). If an GOB can not fit into a packet, the GOB is broken and as many as MB are packed in a packet. There will be no fragmentation of an MB into two packets. After this session layer packetization, RTP payload format for the multiplexed elementary stream is used for end-to-end real time transport. However, the RTP does not guarantee a reliable delivery. Therefore, a companion control protocol RTCP is used to convey the traffic information between the two ends. There are two categories in the RTCP control protocols. They are the Sender

Report and the Receiver Report. These two constitute the mechanism for feedback control based on RTCP. H.323 makes use of the logical channel signaling procedure of H.225 for call setup and H.245 for flow control. H.245 specifies the syntax and semantics of terminal information messages as well as procedures to use them for in-band communication. The components of H.323 include terminals, gateway, gatekeeper, multipoint controller, multipoint processor. An H.323 terminal includes the video/audio codec, telematic applications, H.225 system control function, H.245 communication control, H.225 RAS and Call control, and LAN interface. The H.225 layer formats streams into messages for transmission and retrieve streams from received message. It performs logical framing, sequencing numbering, error detection, and error correction. This is the layer where the RTP and RTCP are defined. The LAN interface provides reliable service (TCP) for H.245 control channel, data channel, call signaling channel and unreliable (UDP) service for Audio, Video, and RAS channel.

### 三、研究方法及成果

In the above section, an architecture for video delivery in the LAN environment has been briefly described. A major difference of such an environment from others such as ISDN or PSTN is the variation of the transmission bandwidth during the session interval. Therefore complex video processing may have to be designed in order to adapt the output bit rate to the channel bandwidth. The purpose of the RTCP message is to allow the receiver to inform the sender of the network condition through the use of the packet loss ratio. With this information, the sender can adjust its output bit rate to fit the channel condition. With this rate adaptation, the task is to distribute this new rate to the following subsequent frames that are to

be coded. When a frame is allocated a new rate, algorithms are designed to allocate the available bit rates to all the Macroblocks. In frame level rate control, the target number of bits  $B$  will be selected based on the average throughput and the buffer space. The encoder buffer fullness is defined as

$W = \max(W_{prev} + B - R/F, 0)$ ,  $R/F$  is the target bit rate. Due to the encoding variation and the actual transmission rate fluctuation, the buffer will exhibit variation. Therefore, adaptation of the actual encoding rate is used to control the fullness of the buffer such that the buffer will not overflow and underflow. The bit rate is adjusted as  $B = \frac{R}{F} - \Delta$ , for each

frame, with

$$\Delta = \begin{cases} \frac{W}{F} & W > 0.1 \cdot M \\ W - 0.1 \cdot M & otherwise \end{cases}, M \text{ is a}$$

threshold. The current encoding frame would be skipped if the calculated buffer fullness  $W$  were larger than the threshold  $M$ . This threshold is set to  $R/F$  by default which would restrict the longest buffer delay to  $M/R = 1/F$ .

After the frame bit rate is determined, the next step is the allocation of available bits to all the macroblock. The objective is to minimize the frame distortion. The control parameter is the quantization step size  $Q$  for each macroblock. The low bit rate model of average number of bits produced by the  $i$ th macroblock would

$$B_i = A \left( K \frac{f_{L,i}^2}{Q_i^2} + C \right) \text{ where } f_{L,i}^2 \text{ is the}$$

variance of the DCT coefficients of the  $i$ th macroblock,  $Q_i$  is the quantization step size, and  $C$  is the *overhead*,  $A=16^2$  the macroblock size. The parameter  $K$  is the slope that relates the bit rate to the ratio of the variance and the quantizer step size. The value of  $K$  and  $C$  would be adapted to the statistics of each macroblock during the encoding process. The typical

distortion of one macroblock with scalar quantizer is proportional to the square of the step size. When considering N

$$\text{macroblocks, this is } D = \frac{1}{N} \sum_{i=1}^N r_i^2 \frac{Q_i^2}{12},$$

$\hat{a}_i$  is a distortion weight of the  $i$ th macroblock  $\hat{a}_i$  is used to account for the uncertainty of this model as well as to take into account the relative importance of all the macroblocks. The optimal quantization parameter  $Q_i^*$  on the condition that the total number of bits is equal to  $B$ , is derived

$$\text{as } \Rightarrow Q_i^* = \sqrt{\frac{AK}{B - ANCr_i} \frac{t_i}{\sum_{i=1}^N t_i r_i}}. \text{ This}$$

is done by *Define  $\lambda^*$  to minimize*

$$\arg \min_{Q_1, \dots, Q_N, \lambda} \frac{1}{N} \sum_{i=1}^N r_i^2 \frac{Q_i^2}{12} + \lambda \left[ \sum_{i=1}^N B_i - B \right]$$

This equations simultaneously determine all the Q for all the macroblock. However, due to the inconsistency of the model and the actual coding rate, a sequential method is used. That is after encoding one block, the current available bit rate is updated and the next Q is derived. With this sequential method, we also update the K and C for each block. In TMN8, a method based on weighted average of the previous K and C is used for update. In the following, we discuss a different method. The importance of K and C is that the above models about bit rate and distortion is very rough. The estimated distortion and bit rate deviates from the model very much. One of the reasons for this deviation is the pre-determined value of the slope K is not well estimated. As can be seen from the relationship between K and Q, If K is not properly determined, the resultant Q is not the optimum one. It is easy to derive the optimum Q and on the above models. However there is still no ways to derive the optimum K and C. This will need a fundamental study on the property of the DCT transform coefficient and its relation to distortion

as well as the VLC code used to represent these coefficients. With regression method, set  $K$  as the slope of the equation

$$B = A(k \times \frac{t^2}{(2QP)^2} + C). \text{ The way TMN8}$$

updating  $K$  is to derive an weighted average the previous slopes. The averaging weight is a function of the encoding order.

Our study reveals that K usually bound within a certain range to get the best result for coding. We test another average method based on linear regression. This SLR analysis eliminates the order factor used in TMN8. There will be one data set (K,C) for each macroblock. Substituting  $y_i$  and  $x_i$  for  $\frac{B_i^2}{256}$  and  $\frac{t_i^2}{Q_i^2}$ ,  $K$  and  $C$  can be derived as follows:

$$K = \frac{\sum_{k=1}^i (y_k - \bar{y})(x_k - \bar{x})}{\sum_{k=1}^i (x_k - \bar{x})^2}, \quad C = \bar{y} - K\bar{x},$$

$\bar{y} = \bar{y}(i-1)/i + y/i$ ,  $\bar{x} = \bar{x}(i-1)/i + x/i$ . This estimation begins at the third macroblock. After encoding one frame, the value of  $\bar{y}$  and  $\bar{x}$  will be passed to the next frame like  $K_i$  in TMN 8.

#### 四、結論

The above sections have presented an architecture for the real time video transmission over LAN. When the traffic is heavy, the fluctuation of the transmission rate may be high. In this case, the buffer fullness may exhibit high variation. For this, a two level rate control is implemented. This is the frame bit adaptation and the macroblock bit allocation. The frame bit adaptation is to control the output bit rate and the macroblock bit allocation is to minimize frame distortion. The first one is easier to achieve, while the second one is more complicated. In the following, we use two measures to account for the frame quality, the SNR and the skip block number. The SNR is used in the formulation to derive

the optimum quantizer, so it is a direct indication of the appropriateness of the formula. The second one is used to indicate the variation of K and the subsequent quantizer value for a series of blocks. The skipped block is defined for these block that is skipped. This may be due to two major reasons. The first one is due to the small energy of the block. This is a normal case. The second one is due to a too large step size. This may be used to evaluate the property for the update formula of the parameter K. Table 1 indicate the first level rate control, the produced bit rates in both methods are all very close to the target it rate. We refer to those macroblocks which should have texture bits but the resultant  $B_{l,c}=0$  as the “ Force-skipped macroblocks”. For different image sequence, the threshold for variance such that a block shall have texture bits  $B_{l,c}$  is different. We choose 25 for MissA, 20 for Suzie and 150 for Trevor based on the simulation data. The results in percentage form is shown in Table 4.2. From the data, we can see that with SLR, the number of forced skipped blocks is less than that of the TMN8. This is due to the more smooth property

of the K value in SLR. But the PSNR is about the same, so the different K update methods exhibit different visual effect. This will remain a further research.

**Table 1**

image	target	TMN 8	SLR I
MissA	30	30.02	30.03
Suzie	72	72.07	72.06
Trevor	40	39.25	40.2

**Table 4.3**

**The average PSNR of TMN 8 and SLR**

Test name	TMN 8 rc PSNR dB	SLR I rc PSNR dB
MissA	37.37	37.38
Suzie	34.88	34.9
Trevor	30.35	30.5

**Table 4.2**

**The number of forced-skipped macroblocks in TMN 8 and proposed methods SLR I, II, III, IV**

Test image: Miss American					
	TMN 8	SLR I	SLRII	SLRIII	SLRIV
MB with variance > 25	2360	2329	2326	2360	2360
% of MB with texture bit	77.3	82.8	83.8	80.9	70.2
% of forced-skipped MB (COD=1/COD=0)	22.7(9.1/13.6)	17.2(10/7.2)	16.2(8.6/7.6)	19.1(10/9.1)	29.8(14.7/15.4)
Test image: Suzie					
	TMN 8	SLR I	SLRII	SLRIII	SLRIV
MB with variance > 20	5517	5485	5317	5485	5433
% of MB with texture bit	88.7	90.1	83.8	90.1	87.8
% of forced-skipped MB (COD=1/COD=0)	11.3(4.2/7.1)	9.9(3.9/6)	16.2(7.1/9.1)	9.9(3.9/6)	12.2(4.9/7.3)
Test image: Trevor					
	TMN8	SLR I	SLRII	SLRIII	SLRIV
MB with variance > 150	1430	1276	1271	1271	1272
% of MB with texture bit	80.1	88.7	89.1	89.1	90.8
% of forced-skipped MB (COD=1/COD=0)	19.9(9.6/10.3)	11.3(8.5/2.8)	10.9(8.1/2.8)	10.9(8.1/2.8)	9.2(8.6/0.6)