MULTIPLEXING VIDEO TRAFFIC USING FRAME-SKIPPING AGGREGATION TECHNIQUE

BY

ALAN YEUNG

A THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF MASTER OF PHILOSOPHY

DIVISION OF INFORMATION ENGINEERING

THE CHINESE UNIVERSITY OF HONG KONG

JUNE 1998

Abstract

This thesis concerns a frame-skipping aggregation technique for multiplexing several VBR MPEG video streams for feeding into a CBR communications channel. The data rate of VBR video fluctuates over time according to its scene complexity. When the aggregate rate of the video streams being multiplexed exceeds the CBR channel capacity, traffic congestion occurs and the data may not reach the receivers on time for display. This may result in severe image degradation, especially if the missing frames are I or P frames. In the proposed scheme, the multiplexer transmits frames from the video streams in a round-robin manner while monitoring the receiver-buffer levels at the same time. Under traffic congestion, rather than waiting until the receiver buffers underflow before taking remedy actions, the multiplexer skips the transmission of B frames when the buffers fall below a threshold level. The multiplexer makes use of the bandwidth thus conserved to bring forth the transmission of subsequent I or P frames. In place of the missing B frames, the receivers will redisplay the preceding received frames. A subjective assessment test is conducted to evaluate viewers' perception of the frame-skipped video streams. The test result shows that skipping 5% of frames is barely visible to the viewers. With this technique, more users can be supported compared with regular video transmission systems. Excellent performance can be achieved when the number of video streams being multiplexed is eight or more. Most importantly, this proactive frame-skipping technique allows us to adopt a simple call-admission strategy in which video requests are granted based on their mean rates.

使用圖片略過聚集技術於視象交通之複用

摘要

這篇論文是關於「圖片略過聚集技術」在複用可變位元率的「移動圖片專家組」的視象流在固定位元率的通訊管道上之應用，可變位元率視象的數據率會因應畫面的複雜程度而隨時間起伏不定，當複用視象流的總率超於固定位元率管道之容量時便會產生傳送交通擠塞，以致視象數據不能準時傳至接收端作顯示之用。這樣導致嚴重的視象質素下降，尤其當內編碼圖片或預測編碼圖片遺失的時候。在建議的方案裡，複用器以循環分配的方式傳送多個視象流中的圖片，同時亦監察接收端緩衝的水平，如遇上傳送交通擠塞，複用器會在接收端緩衝下降至低於某特定水平時略過雙向內插編碼圖片的傳送，而不會等到緩衝短缺時才作出補救行動。這樣，複用器便可利用省回的頻寬來傳送往後的內編碼圖片或預測編碼圖片。接收端會重播之前所收到的圖片，以取代失去的雙向內插編碼圖片。為了評核觀賞者對略過了圖片的視象的滿意程度，我們亦進行了一項主觀評估測驗，測驗結果指出略過百份之五的圖片會對觀賞者做成僅可察覺的視象質素下降。比較於一般的視象傳送系統，採用「圖片略過聚集技術」的系統能支援更多使用者。在傳送八個或以上的視象流時，表現更為優秀。最重要的是「圖片略過聚集技術」容許系統採用簡單的要求接納策略——根據視象流的平均數據率來決定是否接納傳送該視象流。

Acknowledgement

Special thanks are given to Prof. Soung C. Liew for his endless support without which this thesis could not have been finished. Of no less importance is the support from the members of the Broadband Communications Laboratory. Thanks are also given to my dear parents and friends for their love and concerns.

Abstract

This thesis concerns a frame-skipping aggregation technique for multiplexing several VBR MPEG video streams for feeding into a CBR communications channel. The data rate of VBR video fluctuates over time according to its scene complexity. When the aggregate rate of the video streams being multiplexed exceeds the CBR channel capacity, traffic congestion occurs and the data may not reach the receivers on time for display. This may result in severe image degradation, especially if the missing frames are I or P frames. In the proposed scheme, the multiplexer transmits frames from the video streams in a round-robin manner while monitoring the receiver-buffer levels at the same time. Under traffic congestion, rather than waiting until the receiver buffers underflow before taking remedy actions, the multiplexer skips the transmission of B frames when the buffers fall below a threshold level. The multiplexer makes use of the bandwidth thus conserved to bring forth the transmission of subsequent I or P frames. In place of the missing B frames, the receivers will redisplay the preceding received frames. A subjective assessment test is conducted to evaluate viewers' perception of the frame-skipped video streams. The test result shows that skipping 5% of frames is barely visible to the viewers. With this technique, more users can be *supported compared with regular video transmission systems. Excellent performance can be achieved when the number of video streams being multiplexed is eight or more. Most importantly, this proactive frame-skipping technique allows us to adopt a simple call-admission strategy in which video requests are granted based on their mean rates.*

Contents

List of Tables

List of Figures

Chapter 1

Introduction

VBR (variable bit-rate) coding is used in advanced video coding schemes, such as MPEG (Moving Picture Experts Group), to achieve constant image quality and high compression ratio at the same time. When VBR coding is used, frame size of video varies according to the scene complexity and motion. For the transport of video, however, CBR (constant bit-rate) transmission channels are often used. CBR channels are preferred to VBR channels primarily because of the ease with which they can be managed from the networking standpoint. An interesting issue is how to adapt VBR-coded video for transmission over CBR communication channels.

When a single VBR video stream is transmitted over a CBR channel, some video information may need to be dropped when the CBR channel cannot support the instantaneous bandwidth requirement of the video. This leads to image degradation at the receiver. For a VBR video stream with highly fluctuating bit rate, frequent image degradation may lead to an overall unacceptable viewing experience. One may of course allocate a CBR channel with very high bandwidth so as to support even the peak rate of the VBR stream; but this is inefficient and costly.

A solution is to multiplex several video streams together before transmitting them over the CBR channel [1, 21]. In this way, the CBR channel is shared among the streams so that a VBR stream with high instantaneous bandwidth requirement can "borrow" bandwidth from other streams. With many streams, it is statistically unlikely that all streams will peak together, and image-quality fluctuations due to the loss of video information can be minimized.

There are many practical settings in which a number of video streams can be sent over a common channel. An example is in satellite communication. For cost-effectiveness, a number of video streams can be transmitted over a common satellite channel to a remote location, at which point the individual video streams are further distributed to their respective destinations. A number of other practical applications can be found in [8].

Generally, it is better to multiplex video traffic outside the communication network rather than inside the network [2, 5, 8]. When the video streams are multiplexed outside the network, a video server can selectively drop video information whose omission results in the least severe image degradation under traffic congestion. If they are multiplexed inside the network, unless the network has some way of knowing which data are more important than others (through, for example, priority schemes which add complexity to the network operation), it is generally difficult to select the least significant data to discard. The theme of the thesis concerns multiplexing of video outside the network prior to its transmission.

Video multiplexing can also be categorized as being lossy and lossless. When the channel bandwidth is insufficient to support the instantaneous aggregate video bandwidth, lossless multiplexing delays the data delivery [5]. Lossy multiplexing, on the other hand, discards data. In the thesis, we propose a lossy multiplexing technique called frame-skipping aggregation to multiplex MPEG video streams[1].

In this scheme, a multiplexer transmits frames from the video streams in a round-robin manner. When the channel bandwidth is insufficient for the transmission of every frame of every stream, the transmission of B frames is skipped to conserve bandwidth[13]. More specifically, the decision of whether to skip B frames is based on the status of the receiver-buffer levels at the receivers. A proactive frame-skipping technique is used in which B frames are skipped when the receiver-buffer levels fall below a threshold. We show in this thesis that this technique can support more users on the communications channel compared with regular video transmission schemes. Excellent performance can be achieved when the number of video streams being multiplexed is eight or more. Most importantly, the technique simplifies network operation by allowing straightforward call-admission strategy [22] in which video requests are granted based on their mean rates.

B frames are skipped rather than I or P frames because B frames are not referenced by other frames during decoding, and thus their absence does not cause more frames to be skipped as in the case of skipping I or P frames. Motion discontinuity, however, may still result from B frame skipping.

In most research on video impairment, quality is measured on a per-frame basis using an analytical expression such as the signal-to-noise ratio [8]. However, to evaluate impairment effects due to missing frames as in our system, analytical methods may not be a suitable choice[12, 13]. More appropriate are subjective assessment methods that directly evaluate viewers' perception of the impaired video [9, 14].

Test results using an assessment method recommended in ITU-R BT.500-7[9, 11] are presented. The test result shows that skipping 5% (B frames) of the total frames is barely visible to the viewers. Given a fixed frame-skipping rate, video with milder motion is perceived to be of higher quality compared with motion-intensive video. In addition, it is found that when the frame-skipping rate is high, skipping extra frames simply to bring about video-sequence smoothness (i.e. uniform gaps between skipped frames) may actually give rise to better perceived quality, and thus quality cannot be mathematically correlated to the percentage of skipped frames in absolute terms.

The remainder of the thesis is organized as follows. To put the discussion of other chapters in context, Chapter II gives an overview of the MPEG coding scheme, focusing on the frame layer. Chapter III describes the framework of our proposed frame-skipping aggregation. Chapter IV discusses how missing B frames should be dealt with at the receiver for decoding purposes assuming MPEG video. Chapter V presents the results of a subjective assessment test on frame-skipped video. Chapter VI focuses on the performance evaluation of the proposed frame-skipping technique and the investigation of how different parameters in the system should be set. Chapter VII is devoted to the consideration of various implementation issues. Chapter VIII concludes this thesis.

Chapter 2

MPEG Overview

For primative video coding schemes, a picture is coded and compressed using only information from the picture itself. Compression techniques of this kind is called *intraframe* techniques. Each picture can be decoded independent of other pictures and the decoding and display processes are simple and robust. However, video sequences coded using only intraframe coding are usually very large in size and the compression ratio is expected to be low.

Advanced video coding schemes take into account of similarities among pictures. For normal context video, adjacent pictures are usually of similar video content (except for scene changes or video with vigorous motion). If we code only the difference of a particular picture with the adjacent pictures, many bits can be saved, and therefore higher compression ratio can be achieved. Compression techniques using information from other pictures in the sequence are referred to as *interframe* techniques.

The MPEG coding scheme [3, 6, 7] is designed for compression of series of pictures taken at constantly spaced time intervals. It takes advantage of both

the *interframe* and the *intraframe* coding techniques. The proposed aggregation technique discussed in the thesis will focus on MPEG-I, although the discussion and the majority of the results are applicable to MPEG-II also.

To accommodate both intraframe and interframe compression techniques, in MPEG-I pictures, three different types of encoded pictures (frames) are needed in an MPEG-I stream: I (intra-coded), P (predicted), and B (bi-directionally predicted) frames[1]

I frames are coded by itself without reference to any other frames. A P frame is coded according to the prediction from the previously displayed I or P frame. A B frame is coded with reference to the prediction both from the I or P frames which preceed and proceed it in display order. In fact, I frames can only be coded using intraframe techniques, while P and B frames can be coded using either intraframe or interframe techniques, or both. To be exact, different regions of a P frame may use prediction from the preceding frame or use no prediction. If no prediction is used, that region of the picture is coded by intraframe technique only. Similarly, different regions of a B frame may use different predictions, and may predict from the preceding I/P picture, proceeding I/P picture, both, or neither.

Among the three types of frames, I frame takes the largest number of encoding bits, then P frames and then B frames in that order. For a typical VBR MPEG stream, the ratio of the sizes of I, P and B frames is about 1:3:6.

These three types of frames form a sequence called a group of pictures (GOP) [3, 6], which specifies the display order of frames. Some possibilities are

[1]Actually, there is one more picture type defined in MPEG-I, the D frame. However, it is rarely used.

IBBPBBPBB, IBPBPBPB, etc. Typical length of a GOP is around 15 frames. The GOP also provides resynchronization points in the bitstream. When decoding, the bitstream can always be played at the start of a GOP (I frame positions). The GOP usually (although not necessarily) repeats itself for the whole video duration.

In an MPEG-I video, the order in which frames are stored and transmitted (transmitted order) is not the same as the order in which frames are presented to a viewer (display order). Frames are stored and transmitted in an order that is most convenient for decoding purposes: a frame that is needed for the decoding of another frame is stored and transmitted before that frame. This is illustrated in Fig. 2.1. Since P_1 is needed for the decoding of B_1 and B_2, it should reach the decoder prior to the two B frames. In general, the transmitted (or stored) order is formed by placing P frames ahead of the B frames which depend on them for decoding. In the header of each frame, a frame count is present to determine the display order and ensure proper picture presentation.

A picture is divided into small coding blocks called macroblocks. However, an MPEG picture is not simply an array of macroblocks. Every picture consists of slices which are contiguous sequences of macroblocks in raster scan order. More detailed description of the MPEG-I coding scheme can be found in [3, 6, 7].

MPEG encoding provides great flexibility in choosing parameters such as picture rate, aspect ratio and GOP which directly control the resulting bit-rate. In MPEG, a standard list of picture rates and aspect ratios can be chosen. These parameters affect the motion continuity and also the video display size. The GOP can also be adjusted to satisfy different target quality and resynchronization performance. Since I frames are most costly to encode, extending the

GOP length means encoding less I frames, which in turn reducing the resulting bit-rate and reducing the image quality of P and B frames. In general, lowering the target bit-rate degrades output video quality.

Whether an MPEG video stream is encoded as VBR or CBR also affects the resulting video quality. VBR MPEG video comprise frames of variable sizes as are encoded so that constant image quality is guaranteed. On the other hand, CBR MPEG video try to achieve an overall constant data transfer rate by truncating some data of larger frames and stuffing redundant bits to smaller frames. Hence, at the same average bit-rate, CBR MPEG video are considered as of lower quality. There is no strict restriction on choosing the target encoding bit-rate. As a typical choice for encoding normal context video, 1.5 Mbps should be a suitable choice.

Apart from the video part, the MPEG-I coding scheme also contains the audio part and the system part. The audio part defines the MPEG audio encoding standard. The video data and the audio data are encoded independently, and the corresponding bitstreams are interleaved to make the final coded output. The system part describes the control information for parsing and playback of the bitstream.

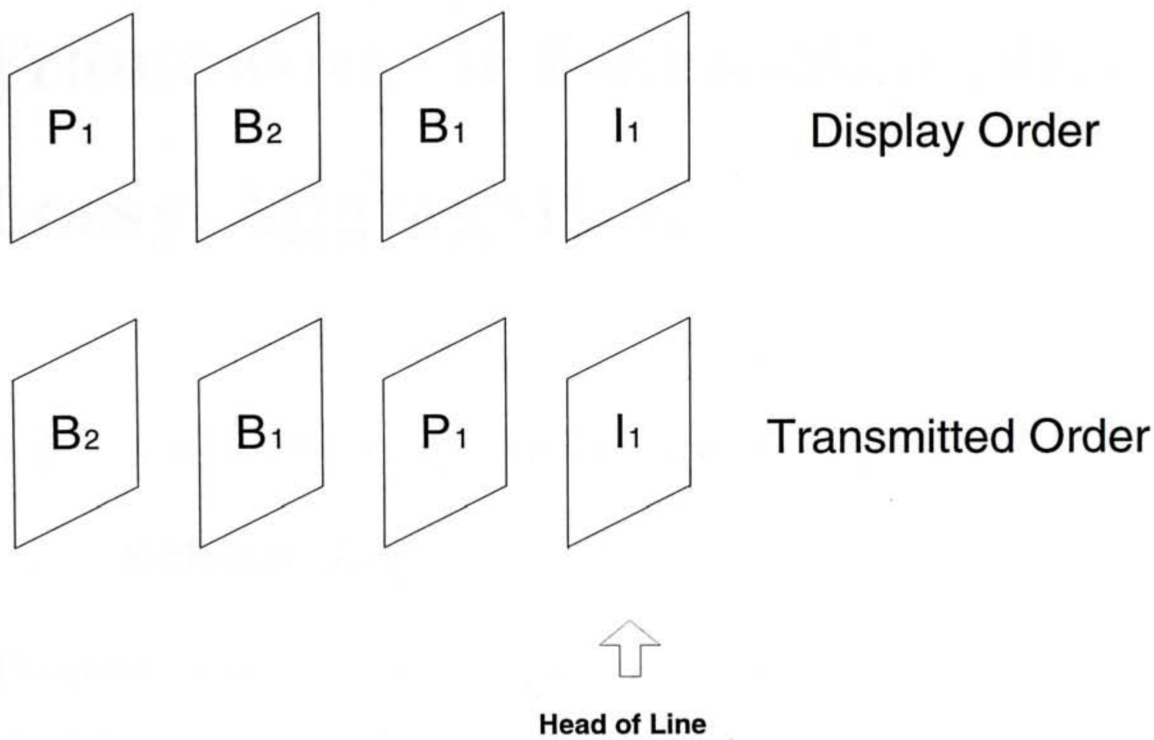

Figure 2.1: Display Order and Transmitted Order of MPEG Frames

Chapter 3

Framework of Frame-Skipping Lossy Aggregation

3.1 Video Frames Delivery using Round-Robin Scheduling

The architecture of the frame-skipping video transmission system is shown in Fig. 3.1. A controller multiplexes the frames from a number of video sources onto a transmission channel. The video sources as well as the controller are outside the communication network. The controller transmits frames from the video sources in a round-robin manner while monitoring the receiver-buffer levels at the same time. Figure 3.2 shows an example of the transmission order with three video streams.

When the receiver-buffer level falls below a certain threshold (details contained in Section 3.2), the controller may skip the transmission of a B frame and move on to serve the next video stream with an I or P frame. By skipping B frames and using the "extra" channel bandwidth to transmit the subsequent frames, the subsequent frames can reach the receivers earlier than otherwise. At the receivers, the frame preceding a missing B frame will be displayed twice, and the net effect is that the buffer levels can be built up quickly.

To elaborate, the round-robin transmission operates as follows. We define a time-slot as the display time for one frame of data, and the data limit as the number of bytes that can be sent out on the transmission link in a time-slot. The data limit divided by a time-slot is equal to the normalized channel bandwidth. In each time-slot, the controller subtracts the frame size of the frame to be sent from the data limit. If the data limit is not exceeded, this frame can be sent successfully. The controller then tries to send the next frame of the next video stream in the chain. If this next frame is a B frame and the system is in the frame-skipping mode, the controller simply discards the B frame and proceeds to the next video source in the chain. This continues until the data limit is exhausted.

Just before the data limit is exhausted, a partial frame may be sent. In the next time slot, the remaining portion of the frame will be sent. Note that a variable number of frames can be sent out in a given time slot depending on the instantaneous aggregate bit rate of the video streams. It should also be noted that with this multiplexing scheme, the receiver buffer occupancies among the receivers differ by at most one frame. In deciding whether B frames should be skipped, the "worst-case" buffer occupancy is examined, as detailed in the next section.

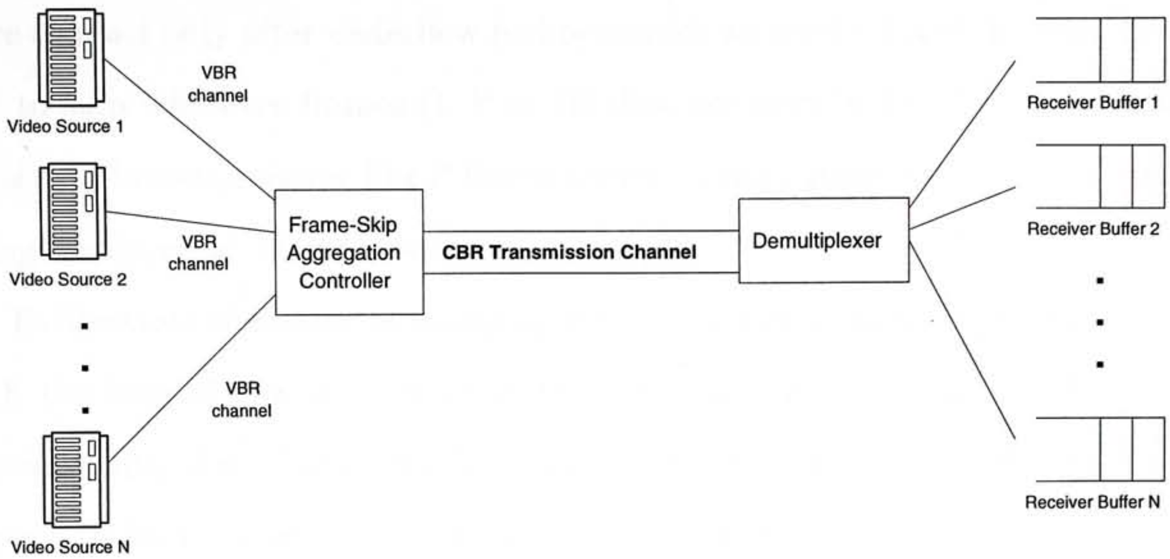

Figure 3.1: The Frame-Skipping Aggregation System

3.2 Underflow Safety Margin on Receiver Buffers

In order to prevent underflow at receiver buffers, the controller monitors the interval between the currently displayed frame number and the latest received frame number. Given a constant display rate, this interval corresponds to the amount of time left before underflow would occur if no more data were to arrive. Therefore, this is some sort of an "urgency" measure for the controller to take actions to prevent buffer underflow.

Let us refer to the interval as the underflow safety margin (USM). When this interval drops below a pre-defined threshold (USMT), the controller starts to skip the transmission of B frames in the next time-slot. When a B frame is skipped, the receiver will repeat the display of the previous frame in place of the B frame. Setting this safety margin enables us to skip B frames at the transmitting side in advance before the onset of underflow, and this in turn minimizes image degradation and ensures more continuous video display. If we were to react only after underflow had occurred, we would simply have no choice but to skip whatever frames (I, P or B) that are next in line for transmission. As is well-known, missing I or P frame results in error-propagation of the images being displayed at the receivers.

To illustrate the frame-skipping aggregation scheme, an example is explained with the help of Fig. 3.2. Assume that the USM drops below the USMT in previous time-slot. The controller starts to skip the B frames in this time-slot. For example, suppose that I_1, P_2 and B_5 frames are of size 50000 , 20000 and 7000 bytes respectively. If 100000 bytes are the data-limit, then I_1 and P_2 can be sent successfully. After this, 30000 bytes are left for the remaining frames. Since B_5 is skipped, the next frame to send is P_1. Assuming P_1 is 20000 bytes, we have 10000 bytes left after its transmission. Then, B_3, B_6, B_1 and B_4 are skipped, and the remaining 10000 bytes are used to send P_4. Note that if we did not adopt the frame-skipping scheme, we could only send the first four frames and about half of B_3. By virtue of the frame-skipping scheme, we see that the USM can be built up quickly. The frame-skipping mechanism will be in effect until the USM goes above the USMT.

3.3 Algorithm in Frame-Skipping Aggregation Controller

The above descriptions of the frame-skipping operations are specified more precisely in the form of a pseudo-algorithm in Table 3.1. In the table, BW denotes the data limit, and the USMT denotes the underflow safety margin threshold. USM, or ReceiverOccupancy, is measured in frames.

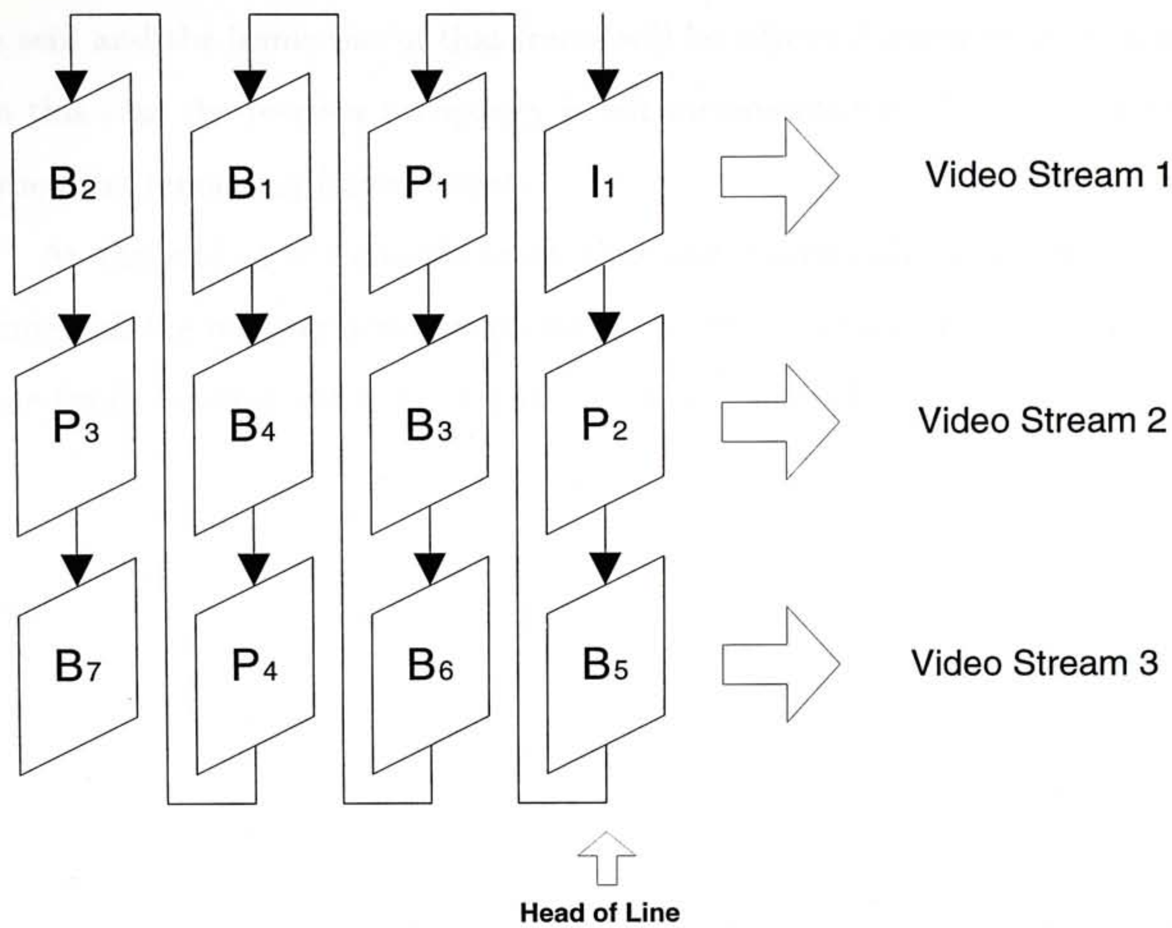

Figure 3.2: The Frame Sending and Frame Skipping sequence

In each time-slot, the controller keeps sending frames to the receivers until the data limit is exhausted. The skipping mode is triggered when the receiver occupancy drops below the USMT. During the skipping mode, any B frame encountered will be skipped, and the corresponding receiver buffer occupancy being monitored will be incremented by one to reflect the fact that the receiver will display the frame preceding the missing B frame twice. After the data limit has been exhausted, the controller checks whether the receiver occupancy is below the USMT. If so, the skipping mode is triggered again. If not, the normal mode is resumed.

It should be noted that just before the data limit is exhausted, a partial frame is sent and the frame size of that frame will be adjusted accordingly. Obviously, in this case the receiver occupancy is not incremented until the next time-slot when the remaining frame is sent.

At the end of a time-slot (each time-slot corresponds to a frame display time), all the receiver occupancies should be decremented by one to reflect that one frame is taken out from the receiver and displayed.

Table 3.1: Frame-Skipping Multiplexing Algorithm Executed in Each Time-slot

```
BW Reset
BWExhausted = False
While (BWExhausted == False) do
 For the next video source j // in a round-robin manner
  If (Mode == Skipping) and (NextFrame == B)
   Skip B frame
   ReceiverOccupancy of j = (ReceiverOccupancy of j) + 1
  Else if BW - FrameSize ≥ 0
   Send frame (or remaining frame)
   BW = BW - FrameSize
   ReceiverOccupancy of j = (ReceiverOccupancy of j) + 1
  Else
   Send partial frame
   FrameSize = FrameSize - BW
   BWExhausted = True
   If ReceiverOccupancy of j < USMT
    Mode = Skipping
   Else
    Mode = Normal
   EndIf
  EndIf
 EndFor
EndWhile

For each receiver
 ReceiverOccupancy = ReceiverOccupancy - 1
 // receivers will have displayed a frame from their buffers
 // by the end of time slot
EndFor
```

Chapter 4

Replacement of Skipped Frames in MPEG Sequence

Frame skipping is a crucial element of the proposed aggregation technique. Since MPEG video coding adopts interframe techniques, the display of the frame sequence is more complicated. At the receiver, post-processing method is thus needed to handle a bitstream with skipped frames. Otherwise, the receiver may not decode the frame-skipped bitstream and display the frame sequence properly.

As far as the consistency with the MPEG standard is concerned, frame stuffing is chosen as the post-processing method used to facilitate frame skipping. In essence, any missing B frames are replaced by stuffed frames prior to being fed to the decoder. Any editing performed on the bitstream is thus transparent to the decoder. The decoder decodes the input bitstream as if it is a normal bitstream. Since the resulting bitstream is a valid MPEG bitstream, display continuity and synchronization of the video with the audio are ensured.

There is a subtlety in the operation of stuffing frames. Consider the frame sequence IBBP (in display order) for example. Figure 4.1 depicts the display of an original sequence, and the expected display sequence when some B frames are skipped.

Imagine that in the uncorrupted sequence, there is an image of a ball rolling from the left to the right of the screen. The position of the ball in each frame is shown.

Case I: If the first B frame is missing, we replace it with an artificially constructed frame. The artificial frame is a B frame in which all the macroblocks are coded as prediction of the previous I or P frame with all motion vectors set to zero. In other words, the decoder will interpret this frame as being the same as the previous I or P frame visually. Details about the construction of this artificial B frame are given in the appendix. The net effect after decoding is that the I frame will be displayed twice. A copy of such an artificial B frame is created once and stored at the receiver to be used throughout the display of the sequence.

Case II: If only the second B frame is missing, instead of the aforementioned artificial B frame, its replacement should be the first B frame. If the artificial frame is used, a wrong displaying sequence of pictures may result as shown in fig. 4.1).

Case III: If both the B frames are missing, we can replace both of them with the artificially constructed frames for proper display sequence.

More generally, to ensure correct visual continuity of a sequence of consecutive B frames, a missing B frame which is preceded by another non-missing B frame should be replaced by that B frame. A missing B frame that is not preceded by any non-missing B frame must be replaced by the aforementioned artificially-constructed B frame.

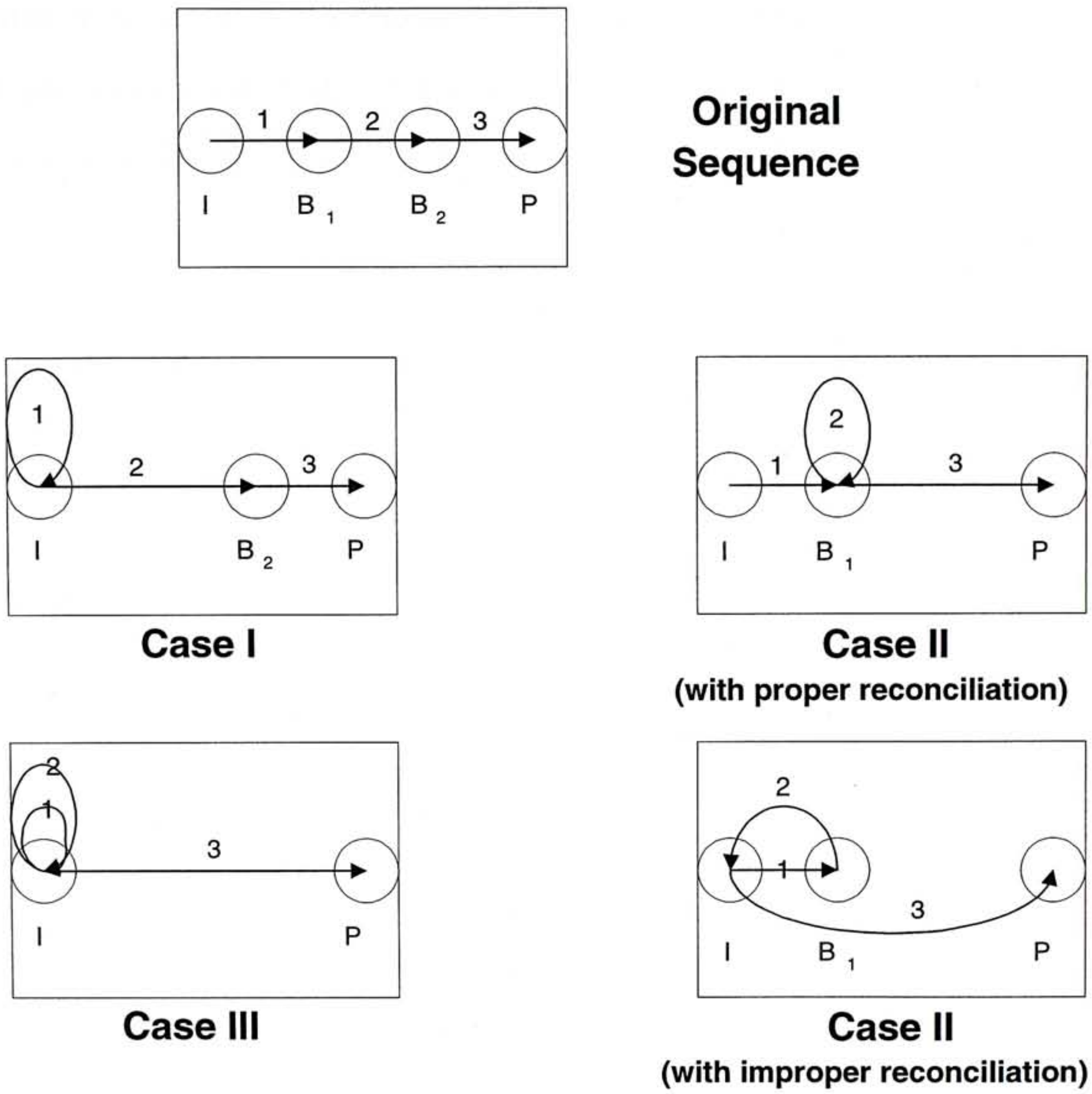

Figure 4.1: Illustration of Frame-Skipping Cases (the small number denotes sequence of motion)

Fig. 4.2 depicts the whole processing flow taking place at the receiver. First of all, the bitstream is taken out from the receiver buffer. Before the bitstream is sent to the decoder, it is stuffed with stuffing frames (artificial frames or previous B frames). Note that the frame sequence shown in the figure corresponds to case

I of the missing frame scenerio discussed above. The subscripts beside the frame types (I, P, or B) denote the temporal display order of the frames. Since the frame B_2 is found missing, an artificial B frame should be stuffed to the bitstream for proper decoding. After stuffing all necessary frames, the bitstream is sent to the decoder for decoding and is further sent to the display.

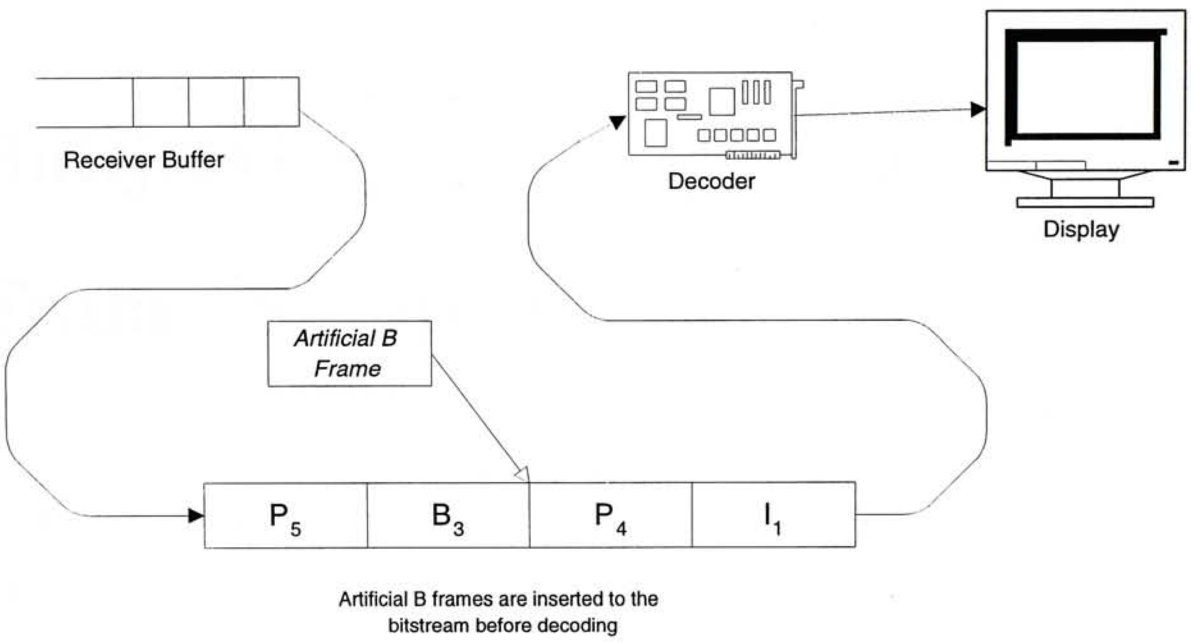

Figure 4.2: Frame Stuffing at the Receiver (Subscripts beside Frame Types Denote Temporal Display Order)

Chapter 5

Subjective Assessment Test on Frame-Skipped Video

For evaluation of the impairment effect thus caused by skipping frames, a subjective assessment test is conducted. Unlike other methods of reducing video bitrate where objective assessment [12] of the video quality can be applied, whether a frame-skipped video is of a certain quality or not is rather subjective. Seeing that recent evaluation of MPEG-4 video coding scheme [9] and many other video distribution methods [14, 17, 18, 19] rely much on subjective assessment tests, subjective assessment tests play an even more important role in evaluation of multimedia coding schemes and applications.

This chapter concerns the evaluation of the degradation of MPEG video due to missing B frames. A subjective assessment test is adopted. The experimental settings and results are discussed below.

5.1 Test Settings and Material

The video clips used in the test are of MPEG-I format, captured by ourselves using a real-time capture card, from a Laser Disc player video source. All the clips used in the test are captured from the Laser Disc version of the film "Die Hard 3". We chose to capture and encode our own video clips rather than using standard video clips in the public domain primarily because we want to have more control over the experimental settings, such as the GOP pattern, aspect ratio, scene types, and comparable image quality. In our experiment, each video clip has a mean bit-rate equal to 4Mbps, and the aspect ratio is 352×240 pixels. The length of each clip is 10 seconds or 301 frames, and the group of picture pattern is IBBPBBPBBPBBPBB.

In the test, two different sets of video clips are used. The properties of each are tabulated in Table 5.1. Video clips in set A are generally not motion-intensive, with few camera pans, while those in set B contain vigorous motion and lots of camera pans. Impaired video clips are made by skipping B frames in the originally captured video clips in a random manner. Video clips with frame-skipping percentage ranging from 5% to 66%, defined as the number of frames skipped divided by the total number of frames, are obtained.

The video clips are decoded using a personal computer with Pentium Pro 200 MHz CPU with the help of a decoder card. All the clips are stored digitally in the harddisk(IDE). The actual viewing diagonal length of the monitor used is of 15.6 inch with high color depth (16-bit). The viewing distance is about 1 meter, and the maximum viewing angle relative to the normal is limited to 30 degree. The test is conducted in a room with low illumination to assist viewing[11]. A total of 20 subjects are involved in the test, and all of them are non-experts of video, with normal color vision and normal or corrected-to-normal visual acuity.

Table 5.1: Properties of Video Clips used in the Subjective Assessment Test

	Set A			*Set B*		
Sequence no.	1	2	3	1	2	3
No. of scene changes	2	3	1	2	1	1
Steady scene portion (sec.)	8	3	4	0	0	0
Panning portion (sec.)	0	4	4	9	9	8

5.2 Choice of Test Methods

There are many different assessment methods described in the ITU-R recommendation BT.500-7[11]. Among them, there are two methods which are typically used to assess impairments.

Double-Stimulus Impairment Scale Method (the EBU method) : The assessor is first presented with an umimpaired reference, then with the same picture impaired. The assessor is then asked to vote on the second, keeping in mind the first. In sessions, the assessor is presented with a series of pictures or sequences in random order and with random impairments covering all required combinations. At the end of the series of sessions, the mean score for each test condition and test picture is calculated. This method uses the 5-grade impairment scale, in which the assessors are asked to assess the comparative picture quality of the second presentation to the first presentation in each testing pair, by choosing suitable description of it from 5 categories, namely, imperceptible, perceptible but not annoying, slightly annoying, annoying, and very annoying.

It is usually found that the stability of the results is greater for small impairments than for large impairments. Although the method sometimes has been used with limited ranges of impairments, it is more properly used with a full range of impairments.

Double-Stimulus Continuous Quality-Scale Method : The assessor is asked to view a pair of pictures, each from the same source, but one of them via process under examination, and the other directly from the source. The assessor is asked to assess the quality of both. In sessions, the assessor is presented with a series of picture pairs (internally random) in random order, and with random impairments covering all required combinations. At the end of the sessions, the mean scores for each test condition and test picture are calculated. The method uses a continuous quality-scale, in which the assessors are asked to assess the overall picture quality of each presentation by inserting a mark on a vertical scale.

In our test, we use the double-stimulus *impairment* scale method to evaluate the quality of frame-skipped video. Compared with the double-stimulus *continuous* quality-scale method where the observers do not know which video of a test video pair is impaired, our experimentation indicates that the double-stimulus *impairment* scale method provides a more consistent result. This could be attributed to the fact that the observer can concentrate more on the impaired video, thus increasing the observability of the impairments.

5.3 Test Procedures

A test session lasts for about half an hour. Each session comprises several video presentations divided into three stages, namely, the training stage, the stabilizing stage and the main testing stage (Fig. 5.1). During the training stage, the assessors are given explanation on the type of impairments to be expected, the grading scale, and the flow of the test session. Training video sequences other than those used in the main testing stage are presented to give the assessors an idea of the range and the type of the impairments to be assessed. After answering a number of questions by the assessors, the test begins.

In each video test pair, the assessors are first presented with an unimpaired reference video clip, and then they are presented with the same video clip with skipped frames. The assessors are asked to assess the frame-skipped video clip according to the five-grade impairment scale (imperceptible (5), perceptible but not annoying (4), slightly annoying (3), annoying (2), very annoying (1)). During the presentation of the first two video testing pairs, the test is said to be in the stabilizing stage. In this stage, the assessment results are discarded and not processed without the awareness of the assessors. The main goal of the stabilizing stage is to calibrate the assessors' judgement by including some critically impaired video in the presentation. The assessors can then have a clearer measure to evaluate the quality of the forthcoming impaired video. The main testing stage is divided into two parts, corresponding to set A and set B video. In between, there is no training stage and stabilizing stage.

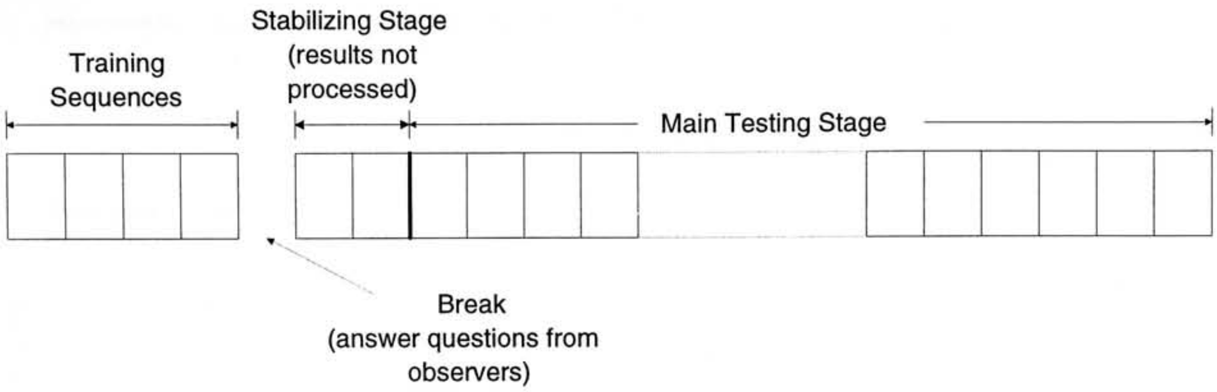

Figure 5.1: Presentation Structure of a Test Session

5.4 Test Results

Figure 5.2 depicts the mean impairment level of set A and set B video against the frame-skipping percentage ranging from 5% to 66%. The result shows that the test subjects consider video sequences in set A to be less impaired than that in set B for the whole range of frame-skipping percentage. This can be accounted for by two observations: 1) video sequences in set A contain mild motion and static scenes, in which skipped frames are hard to detect; 2) the image quality of individual frames is better for set A (Since the encoding rate is the same for both sets, but set B needs more bits for the motion vectors). In other words, the test subjects rate set A higher because set A appears to have more unskipped frames.

We also observe that the 66% frame-skipped video sequences (both sets A and B) are considered less impaired than the 50% frame-skipped video sequences, a seemingly counter-intuitive result until we give it more thought. Consider this. All the B frames in the 66% frame-skipped video sequences are skipped, the resulting video have a steady motion discontinuity throughout (i.e. uniform gaps between skipped frames); whereas the 50% frame-skipped video sequences

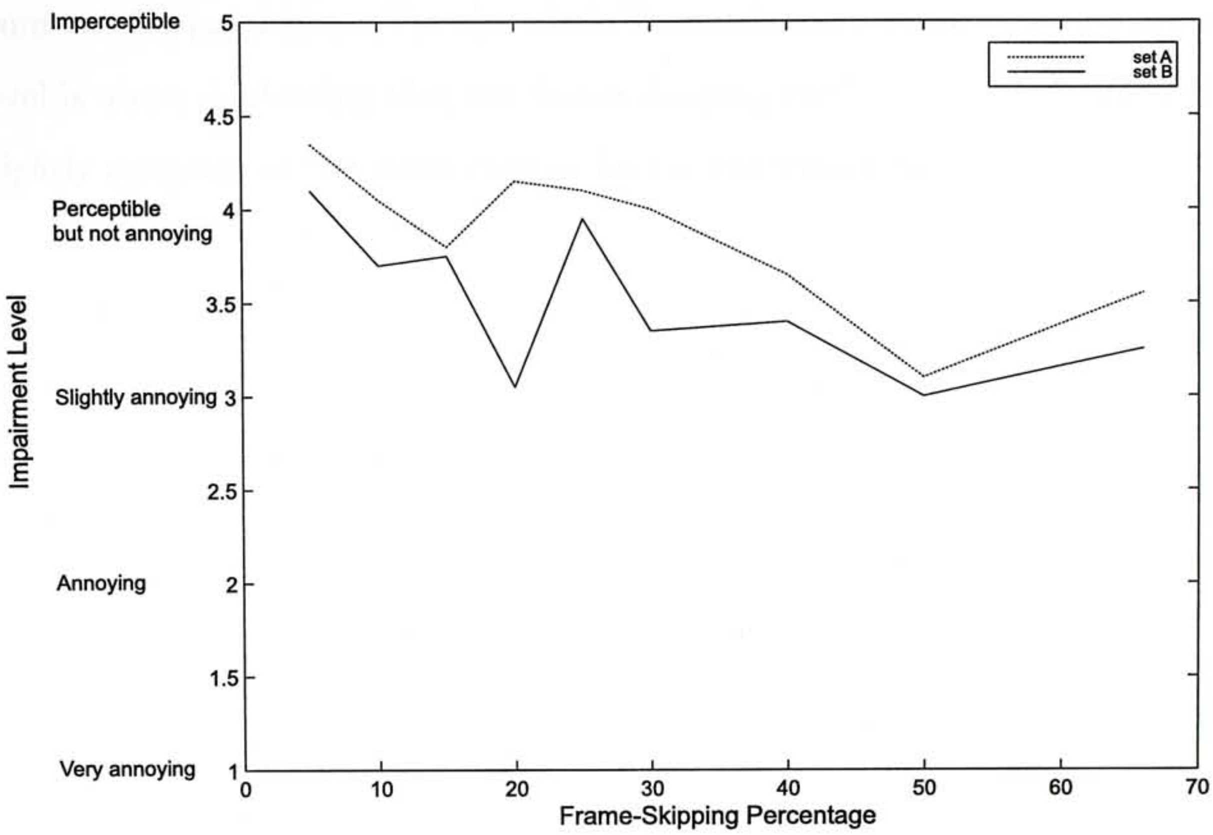

Figure 5.2: Subjective Assessment Test Results on Frame-skipped MPEG Video

contain motion discontinuity mixed with smooth motion. Observers are more annoyed by the unsteady video smoothness even though fewer frames are actually skipped. *This shows that video smoothness is an impairment factor to consider in engineering a video distribution system.*

From our results, the frame-skipping percentage for imperceptible impairment should be less than 5% (impairment level of 4.5 or above), and for perceptible but not annoying impairment, the frame-skipping percentage is shown to be around 10%. As the frame-skipping percentage increases, the impairment level shows a general trend of decrease with some fluctuations. More fluctuations are shown on the set B results. These may be caused by sequence-dependent viewing preference, and the motion content of the video which distracts the observers from careful judgement. For the whole frame-skipping range, the impairment level is above 3, showing that our frame-skipping method is only considered as slightly annoying in the worst case under the experimental conditions set.

Chapter 6

Performance Study

This chapter presents the results of experiments for investigating the performance of the frame-skipping technique. The video data used is taken from a trace of the film *Star War* [4, 10], and it lasts about two hours and is MPEG-I coded with 24 frames per second and GOP pattern IBBPBBPBBPBB. The trace contains only the number of bits in each frame and does not contain the actual movie contents. The mean GOP sizes of the video stream are shown in Fig. 6.1 and the mean frame size is 15614 bytes.

For the purpose of our multiplexing experiments, it would be preferable to have data from many different video streams. Consistent data sets of two-hour in length, however, are not available in the public domain. We therefore resort to the construction of many "artificial" video streams from the single Star War video. To create an artificial stream, we start from a random I-frame position in the whole sequence and extract data for a 15-min section. After that it is concatenated with another 15-min section starting from another random I-frame position. This continues until we have a two-hour stream. This construction

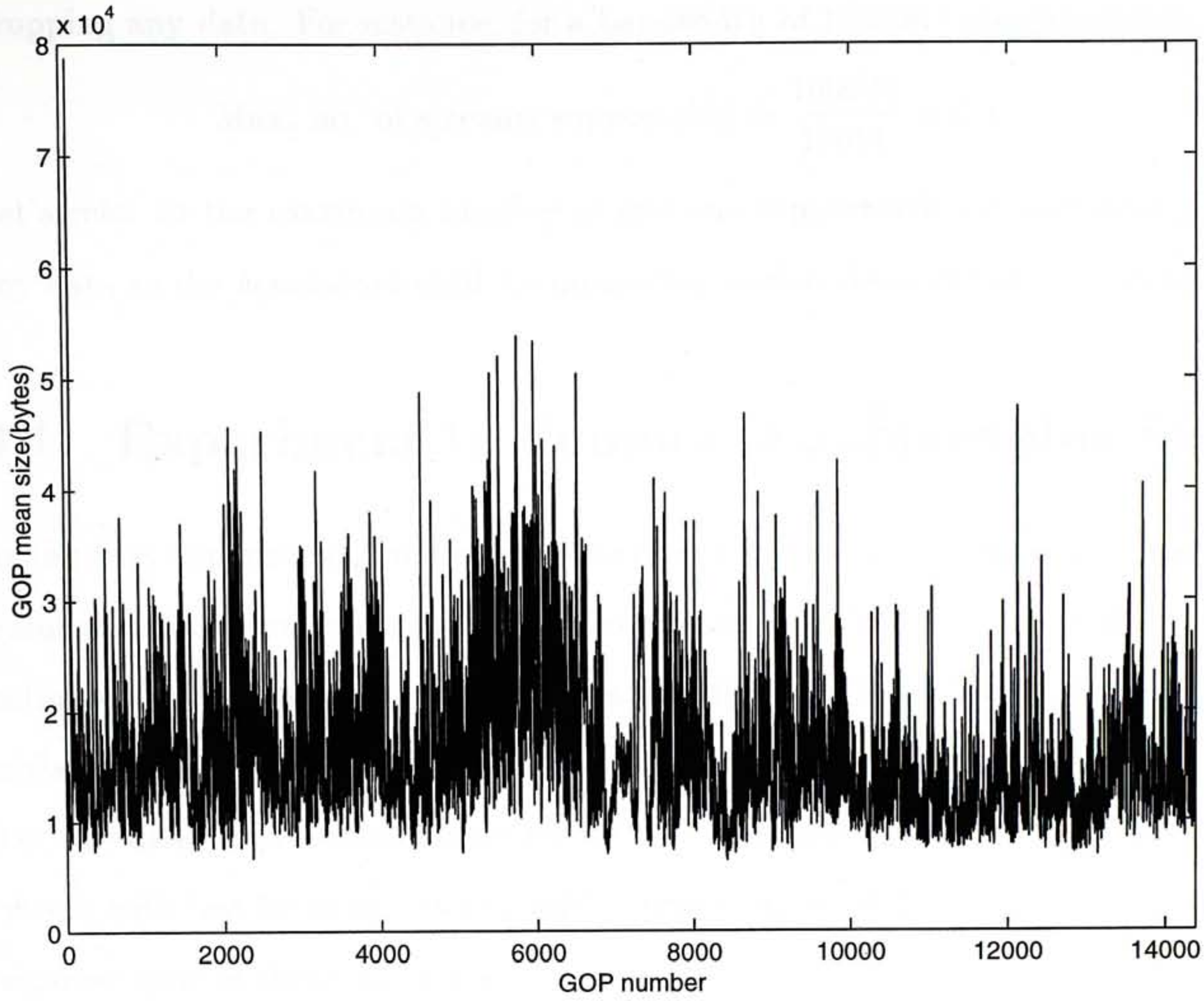

Figure 6.1: Mean GOP Sizes of the Star War Trace

method yields several video streams of different bit-variation profiles.

A performance metric used in our experiments is the number of supportable video streams given a CBR channel with a fixed bandwidth. A number of video streams is said to be supportable if 1) the receiver buffer does not underflow and 2) the frame-skipping rate is not more than some pre-defined percentage. The bandwidth of the channel is represented as the number of bytes that can be sent in one time-slot (1/24 sec.). The bandwidth divided by the mean frame size is thus the maximum number of streams supported by the channel without dropping any data. For instance, for a bandwidth of 100000 bytes/time-slot,

$$\text{Max. no. of streams supportable} = \frac{100000}{15614} = 6.4$$

Let's refer to the maximum number of streams supportable without dropping any data as the *benchmark* used for measuring performance in the experiments.

6.1 Experiment 1: Number of Supportable Streams

In the first experiment, for the frame-skipping systems, we explore a range of frame-skipping percentage in each video stream: 1%, 5%, 10% and 66%. The initial buffer occupancy is set to 8 frames, and the USMT is set to 4 frames. The initial buffer occupancy is proportional to the start-up delay for a video stream (i.e., the delay experienced before the video is presented), and for an interactive system with fast-forward, rewind, and jump functions, it is proportional to the response time of these functions.

The results of the first experiment are shown in Fig. 6.2. It is observed that the number of streams that can be supported by a frame-skipping system of any frame-skipping percentage is greater than or equal to that of a system with no frame skipping. We also observe that the higher the frame-skipping percentage, the more streams can be supported. In particular, if we aim to have no frames skipped at all, the price is under-utilization of bandwidth: the number of supportable streams can be 2 to 4 streams fewer than the benchmark. At low bandwidths (e.g., when the benchmark is less than 8), this difference is quite significant; at high bandwidths, it is not. Therefore, when the number of streams being multiplexed is small, say less than 16, one should perhaps target for a non-zero frame-skipping percentage for bandwidth efficiency. When the number of streams being multiplexed is large, say more than 32, then one can target for zero frame-skipping percentage without sacrificing bandwidth efficiency significantly.

When no frames are skipped, the threshold USMT is never reached and the frame-skipping mechanism is not triggered at all in the experiment. Therefore, the no-frames-skipped data is also representative of a regular system without the frame-skipping mechanism. Note, however, this does not mean we should really go ahead and build such a regular multiplexing system without installing the frame-skipping mechanism, even if we were happy with its experimental performance shown above. This is because in a real system, the set of video being multiplexed will be different and changes over time, and there is no guarantee that underflow will not occur given the same CBR bandwidth as in the experiment. Some precautionary action to prevent underflow, such as skipping frames as proposed here, is still advisable even if we were targeting for a system that rarely ran out of bandwidth.

Referring to the results obtained in the previous section, in which it was shown that a maximum of 5% frames can be skipped with barely perceptible impairment, Figure 6.2 shows that by adopting this percentage we can support a number of streams greater than or equal to the benchmark. If we choose to skip a maximum of 66%, where essentially all B frames are skipped, we can support 37% more video streams at the same bandwidth. With regard to the subjective quality, the impairment score is above 3, corresponding to a fair quality. For a fair-quality video distribution system, skipping all B frames should be a possible solution to deal with channel congestion.

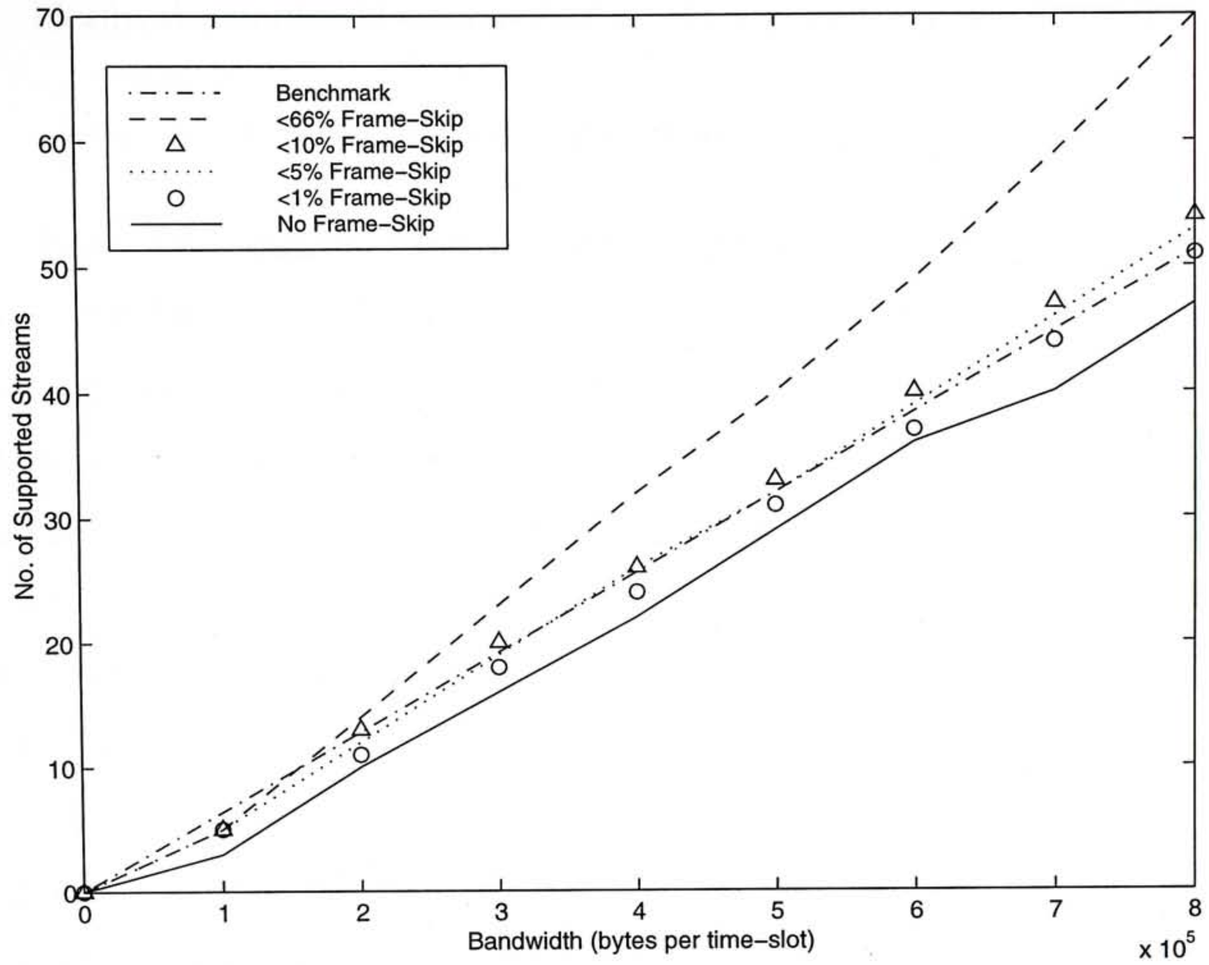

Figure 6.2: Number of Streams Supportable for Different Systems for Different Bandwidth (USMT = 4 frames time)

6.2 Experiment 2: Frame-Skipping Rate When Multiplexing on a Leased T3 Link

The second set of experiment investigates the case when a leased T3 channel with fixed bandwidth is used to multiplex video data. It investigates the relationship of three system parameters in the frame-skipping aggregation system, namely, number of streams supported, USMT of the receiver buffer, and percentage of frames skipped. The video streams used in this experiment are obtained in the same way as those in the last experiment.

Ideally, the number of streams that can be supported by the T3 channel is

$$\text{Ideal no. of streams supported (benchmark)} = \frac{45 \times 10^6}{8 \times 15614 \times 24} \approx 15$$

Figure 6.3 shows the frame-skipping percentage under different USMT settings and different numbers of supported streams. We note that the results are not very sensitive to values of USMT. We do not include results for less than 12 streams because very few frames are skipped in these cases during the two-hour movies. The USMT is seldom reached, if at all. We therefore conclude that the bandwidth is quite abundant for the sending of nearly all frames when less than 80% (12 streams divided by 15 supportable streams in the ideal case) of the ideal capacity is used.

As the number of streams increases, the USMT becomes reachable and the frame-skip percentage increases. When 15 streams are supported by the system (benchmark for this bandwidth), only about 5% of frames are skipped. This frame-loss rate, as suggested by the subjective assessment results in previous section, corresponds to a generally imperceptible degradation in video quality.

As the number of streams increases, the frame-skipping percentage goes up in a rather linear manner with a slope of about 10 percent per extra stream. At first glance, it may seem odd that the system can work beyond the benchmark. But thanks to the frame-skipping mechanism of our system, the amount of data being sent is actually less than what we have to send when the system is overloaded.

According to our experiment results (not shown in Fig. 6.3), the system reaches its operation limit when there are 20 or more streams because all B frames plus some I or P frames must then be skipped. We do not recommend skipping non-B frames in general.

An observation we made during the experiment is worth pointing out. Underflow does occur for small USMT. When very large frames are to be sent, frame skipping may not increase the receiver buffer occupancy sufficiently within one frame time-slot. If the USMT is large enough, in the next few rounds of frame transmission, frame skipping can increase the receiver buffer gradually to the USMT. If the USMT is small, there would not be enough room for such adjustment before underflow occurs. If underflow occurs, the receivers may choose to freeze the display by continuously stuffing artificial B frames (presentation of audio will also be frozen) until the buffer is built up again to a safe level.

6.3 Experiment 3: Bandwidth Usage

In the third set of experiment, we look into the efficiency of bandwidth usage. Figure 6.4 shows the bandwidth needed per stream versus the number of streams for various frame-skipping percentages fixing USMT at 4. In the experiment, to obtain a data point, we fixed the number of streams and then increased the bandwidth until that number of stream can be supported without underflow. The corresponding bandwidth per stream was then obtained. As a reference, the per-stream mean rate of the video streams is also plotted in Fig. 6.4.

When there is only one stream, the bandwidth needed is much higher than the mean bandwidth of the stream. This is expected for VBR streams, such as that used here. When the number of streams increases, thanks to multiplexing gain, the bandwidth needed per stream decreases rather dramatically in the beginning and then tapers off.

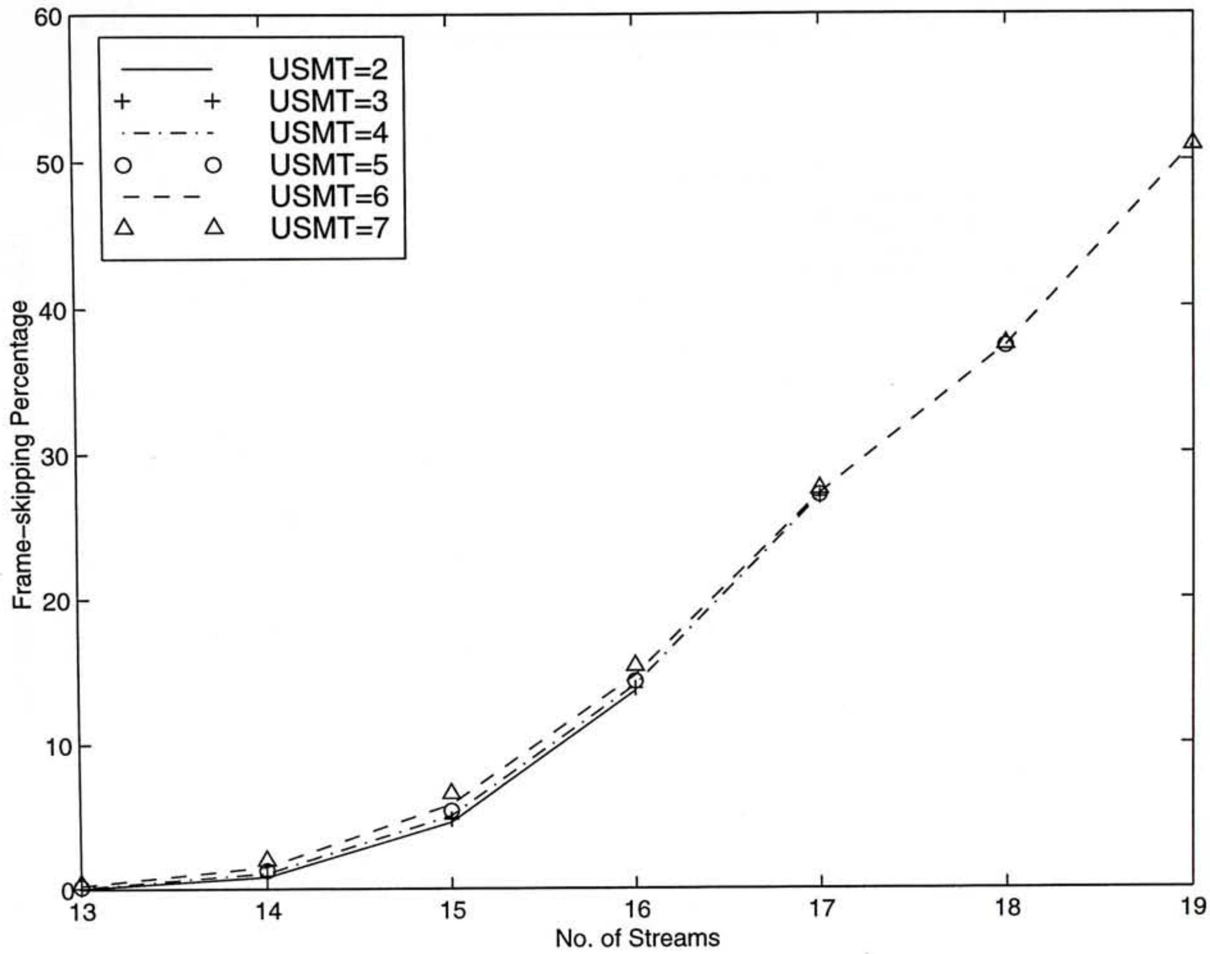

Figure 6.3: Frame-Skipping Percentage under Different USMT and Number of Streams using a T3 Link as the CBR Channel

When the number of streams being multiplexed is small, say four or fewer streams, their aggregate bandwidth may still vary quite dynamically over time. Even with frame skipping, the bandwidth needed cannot approach the mean rate of the streams.

When the number of streams increases beyond eight, the bandwidth needed per stream gradually approaches the mean rate. For 10% frame-skipping rate, the bandwidth needed can even be less than the mean rate. In this situation, bandwidth allocation becomes simple: we simply support a stream according to its mean rate.

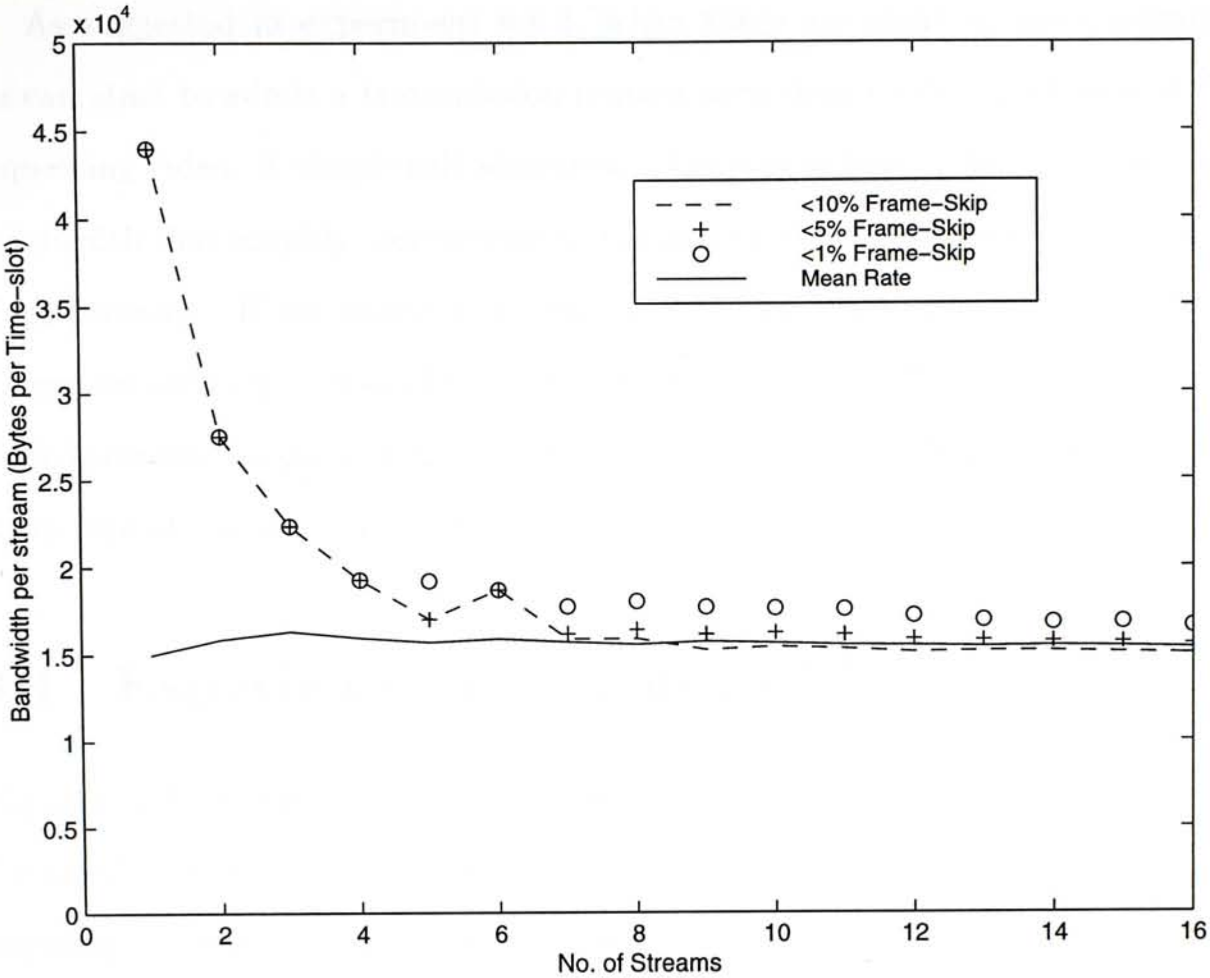

Figure 6.4: Bandwidth needed per Stream using Frame-Skipping Aggregation

We have considered in the second set of experiment the case when a fixed-bandwidth link is leased to transport video. For video transmission incorporating advanced network services such as ATM (asynchronous transfer mode) [20] which supports real-time bandwidth allocation [15, 16], the bandwidth can be allocated in a dynamic manner according to needs. The fourth set of experiment investigates the performance of our system in which an extra bandwidth corresponding to the mean rate of a new video is allocated when the video is admitted. The aggregate bandwidth may therefore change in a dynamic fashion according to the set of video being multiplexed.

As suggested in experiment set 3, when there are eight or more streams, we can start to admit a transmission request according to the mean rate of the requesting video. A simple call admission scheme is to have a start-up minimum bandwidth that roughly corresponds to the aggregate mean rates of eight typical video streams. If the number of video is eight or fewer, this bandwidth will always be used regardless of the exact number of video. When the number of video increases beyond eight, we allocate additional bandwidth according to the mean rate of the new video stream.

6.4 Experiment 4: Optimal USMT

Figure 6.5 shows the frame-skipping percentage and the optimal USMT versus the number of streams. An USMT is said to be optimal when the lowest frame-skipping percentage can be achieved without underflow. Experimentally, we simply decrease the USMT slowly until underflow occurs and take the previous USMT value as the optimal USMT. The video streams used are generated as in the previous experiments.

From Fig. 6.5, we find that the average frame-skipping percentage can be kept to about 5%, which corresponds to a comparable quality with the original video streams according to our subjective test. It is also observed that the frame-skipping percentage and the optimal USMT decreases gradually as the number of streams increases. This shows that if we admit an extra stream based on its mean rate, we are actually improving the performance of all streams. We therefore conclude that bandwidth allocation based on mean rate is a rather viable approach with our proposed frame-skipping aggregation mechanism.

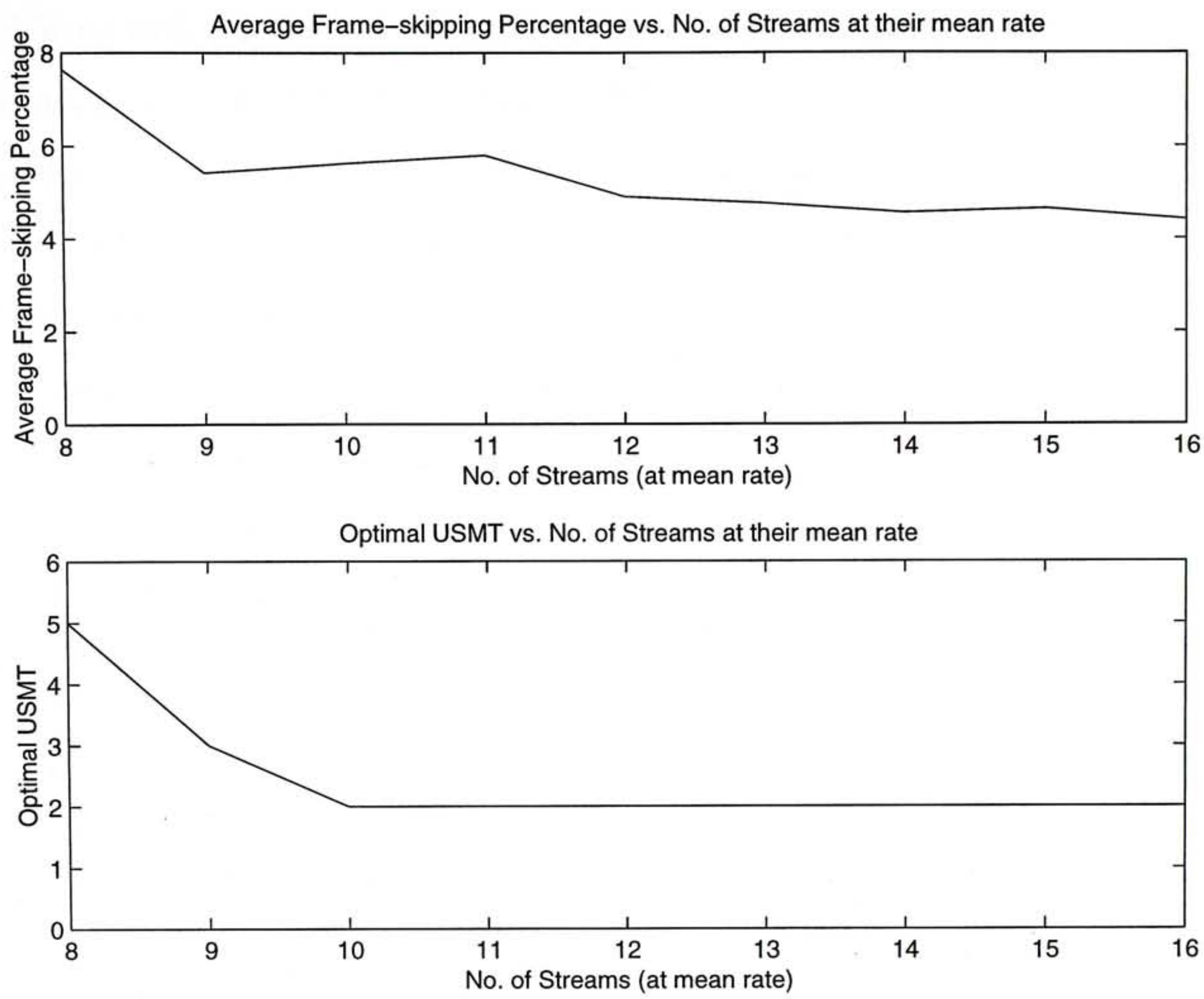

Figure 6.5: Frame-Skipping Percentage and Optimal USMT under Mean Rate Call Admission

The above experiments give us some indications as to how to set the essential frame-skipping aggregation operating parameters. These parameters should be properly set so that 1) more video streams can be supported, 2) underflow probability is kept low, and 3) acceptable video quality is preserved.

In the subjective test results, we have found that 5% frame-skipping provides video quality comparable to the original quality; in the above experiments, we have also observed that at all the bandwidth used, the benchmark can be achieved with 5% frame skipping. Therefore, we should aim to keep the frame-skipping percentage to 5% or below. When multiplexing video streams in a fixed bandwidth channel, we should not admit over 20% more streams than the benchmark to preserve acceptable video quality. When multiplexing video streams in a channel with dynamic bandwidth allocation, we can start out with a constant shared bandwidth when the number of streams is less than eight. We can admit any extra stream beyond the first eight streams by requesting an additional bandwidth according to the mean rate of the stream.

The USMT value should be chosen after taking into consideration the likelihood for underflow and the frame-skipping percentage. As the USMT is set smaller, the probability of buffer underflow will be larger; as the USMT is set larger, more frames will be skipped resulting in video quality degradation. A good balance should thus be striked in choosing the USMT value. From our experimentation, in all cases, a USMT of about 3 or 4 is already appropriate.

Chapter 7

Implementation Considerations

Let us now consider the details related to the implementation of a stored MPEG video distribution system based on the frame-skipping technique. The controller plays a key role in the whole system. The scheduling of the transmission of frames depends on the knowledge of the frame size. Real-time frame-size calculation may burden the controller severely, especially when the number of video streams is large. To solve this problem, we may construct a frame-size table file in an offline manner for each video stream before actual transmission. The MPEG frame size can be specified using 24 bits (i.e., 3 bytes). With this, frame size up to 16777216 bytes can be denoted. For a video stream that is 2 hours long, the file size for the table will be about 648000 bytes.

Receiver buffer status is also a crucial element. In the preceding study, we have assumed that the controller has perfect knowledge of the receiver buffer level. This is easy to achieve if the clocks of all receivers and the clock of the transmitter are mutually synchronized and run at the same rate. With synchronized timing, all the receivers display at the same rate and the receiver-buffer levels of different receivers do not drift apart.

There are many situations in which the synchronization of the receiver clocks cannot be assumed. When clocks are not synchronized, the receivers need to send feedback messages to the controller to inform it of their buffer levels. There are two important issues to be considered: 1) the feedback delay may vary among the receivers; 2) the receivers may display at varying rates and the buffer levels may vary over time. The first problem is easy to tackle in a high-performance communication network in which the delay can be guaranteed to be below 33 milliseconds (a frame period for 30 fps video). In this case, the system can be operated in a time-slotted manner in which the feedback messages during the current frame period are collected and used only during the next frame period in the scheduling of the transmission of the frames. That is, the delay variations are immaterial if all the delays are below 33 milliseconds.

To solve the second problem, the controller may occasionally need to transmit an extra frame for a stream if its buffer level is falling behind others' buffer level. That is, when the controller scans through the video stream in the aforementioned round-robin manner, two successive frames of a lagging video stream may be sent out in just one round. By the same token, if a stream is ahead of other streams, the transmission of a frame of the stream can be deferred to the next round.

At the receiver, a front-end should be present to get data from the receiver buffer, do necessary frame stuffing and send video data to the decoder for decoding and display. If datagram-type protocols such as UDP are used, where the data sequence is not guaranteed, special message indicating the positions of missing frames should be generated at the controller and sent to the receiver. If video frames are transmitted using connection-oriented protocols such as TCP/IP which conserve data sequence, missing B frames can be detected easily by examining the temporal information of the received frames.

The missing frame detection mechanism is further explained with the help of Fig. 7.1. The left-most frame in a frame sequence is the latest received frame, and the subscripts beside the frame types denote the display order.

Case I: After I_1 and P_4 have been received, B_2 and B_3 are expected to come soon. On receiving P_7, which means B_2 and B_3 should have come, B_2 (not received yet) is considered missing.

Case II: After I_1 and P_4 have been received, B_2 and B_3 are expected to come soon. B_2 and B_3 are considered missing when P_7 is received.

Case III: After P_4 and P_7 have been received, B_5 and B_6 are expected to come soon. B_6 is considered missing when P_10 is received.

In general, on receiving any P frame, an expected list of B frames (identified by the display order number) can be generated by comparing the display order number of that P frame and the display order number of the last received I or P frame. Any B frame arriving afterwards is removed from the expected list. On receiving the next P frame, those B frames still on the list are considered missing.

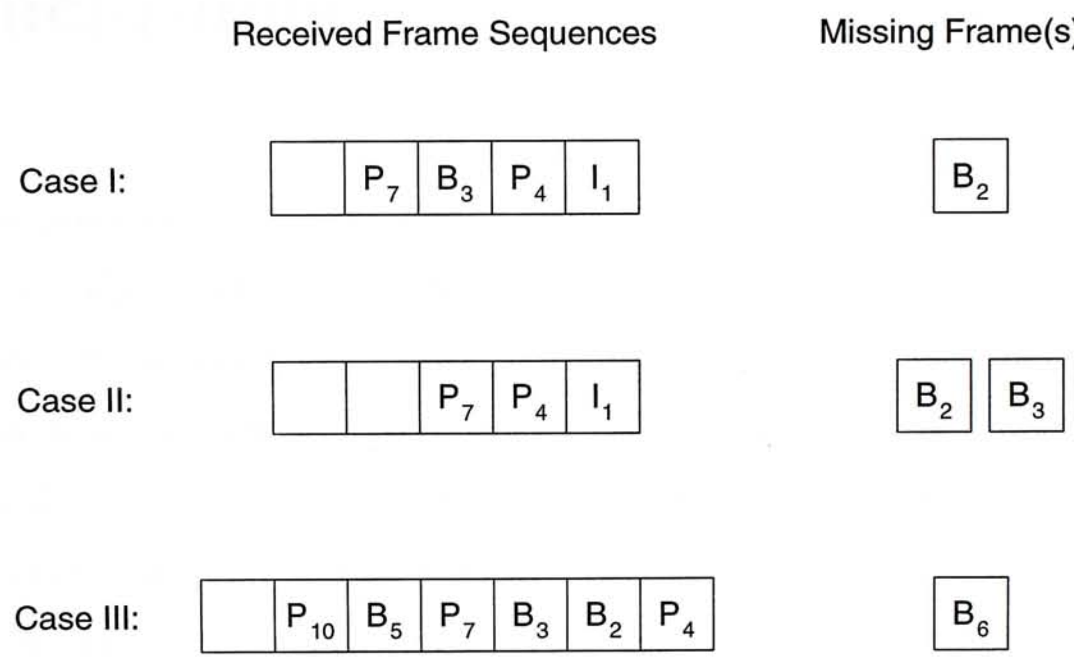

Figure 7.1: Missing Frame Cases

Chapter 8

Conclusions

We have proposed a new multiplexing method for video traffic that makes use of a frame-skipping aggregation technique. This is a lossy multiplexing scheme which discards B frames in MPEG video streams when the channel capacity is insufficient to carry all data. In this scheme, the transmission of B frames is skipped when the multiplexer finds that underflow at the receivers is imminent. The channel bandwidth is used to transmit the subsequent frames instead, so that they can reach the receivers earlier than otherwise. At the receivers, the frame preceding a missing frame will be displayed twice, and the net effect is that the buffer levels can be built up quickly.

We have investigated the performance of this technique relative to that of the regular video transmission system without the frame-skipping mechanism. Our result shows that the proposed technique can support more video streams with only a very small number of frames (less than 5% for quality comparable to originals) being skipped.

We have conducted subjective tests to gauge the response of human observers to skipped frames. Our work indicates that MPEG video containing vigorous motion is more vulnerable to frame skipping than less motion-intensive video, as expected. In addition, the observers found that up to 5% of frames can be skipped with barely perceptible image impairment, and 15% frames can be skipped with acceptable quality. In fact, under critical traffic-congestion situations, it appears that skipping all B frames (66% frames being skipped in our samples) is a possible solution, as the corresponding subjective impairment score is not too low.

The proposed frame-skipping scheme has an important implication for call admission. Suppose we adopt a simple call admission strategy in which a new connection request will be accepted so long as the aggregate *mean rate* of all the video streams does not exceed the CBR channel bandwidth. This call admission strategy, although simple, does not guarantee that the aggregate bit rate of the video streams would not occasionally exceed the channel bandwidth.

Two consequences with a regular multiplexing system are that 1) the receivers would experience underflow and I and P frames may not arrive in time for decoding, resulting in severe image degradation due to the well-known error-propagation effect of MPEG-type video; 2) since the transmitter keeps on transmitting without skipping frames that are no more useful to the receivers (because their display time has passed), the channel bandwidth is wasted, and it may take a long time for the display process at the receiver to catch up. The simple call-admission scheme is therefore not viable when a regular multiplexing system is used.

With our proposed frame-skipping technique, both the above problems are solved by starting to skip B frames just before the onset of receiver-buffer underflow. When there is a sudden surge in the aggregate rate of the video streams, the worst that can happen is that more B frames are skipped and the detrimental effect is not as fatal. Thus, a nice consequence of incorporating a frame-skipping mechanism such as ours is that a simple call admission strategy that admits requests based on the mean rate can be adopted.

Appendix A

The Construction of Stuffed Artificial B Frame

The artificial B frame consists of a picture header, a slice header and two macroblocks. The picture header is necessary for keeping the continuity of display. It contains temporal information which defines the display order of that picture. This is needed since the coding order and the display order are different in an MPEG bitstream. Macroblocks are the basic coding units of a picture. In B frames, macroblocks can be coded by itself, or by prediction from the preceding I/P picture, or the succeeding I/P picture, or both. Macroblocks may also be skipped when there are no changes with respect to the reference pictures. A slice is a contiguous sequence of macroblocks in the raster scan order in a picture. The slice header defines the slice starting position in the picture. Fig. A.1 shows a possible slice structure in an MPEG-I picture. The squares are macroblocks and contiguous squares of the same color comprise a slice.

The exact coding detail of the macroblocks can be better explained with the

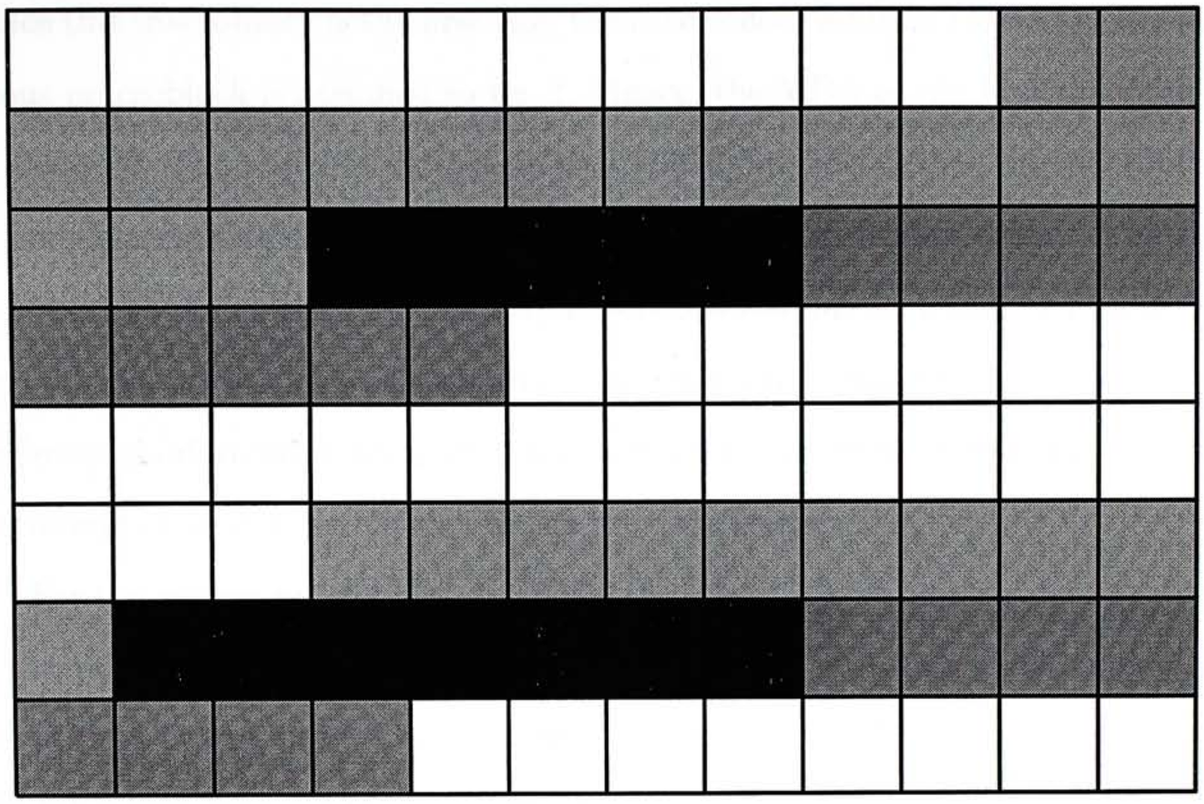

Figure A.1: A Possible Slice Structure in an MPEG-I Picture

help of an example. Consider an artificial B frame of aspect ratio 352×240 pixels. The coding in it is as shown in Table A.1. In the picture header, the picture type is set to B frame and the temporal reference (i.e., display order) is set to that of the skipped frame. In the artificial B frame, there is only one slice, running from the top left corner to the bottom right corner of the display area. At these two corners, there are two macroblocks coded with reference to the previous frame. All the macroblocks in between are skipped.

For the slice header, the Quantizer Scale is set to 00001 which corresponds to the minimum quantization, and the Slice Information bit is set to 0 to indicate that there is no extra information carried by the slice. For the first macroblock, the macroblock address increment (MBAI) is set to 1 which means the address of the macroblock is one more than that of the previous encode d macroblock.

Since this macroblock is the first one, the macroblock address (MBA) of the previous macroblock is assumed to be -1. Hence, the MBA of the first macroblock is always 0. The macroblock type is set to 0010, indicating that the macroblock is coded according to the preceding I or P frame in display order. The motion vectors define the position of that preceding frame to which coding of this macroblock refers. They are set to one which corresponds to zero value. Other differential information between this frame and the previous is also omitted since we intend to repeat the display of the previous frame.

For the second macroblock, the MBAI is set to 329. When the decoder finds a macroblock with MBAI greater than 1, it considers all other macroblocks in between as having been skipped. The skipped macroblocks are assumed to have the same properties as the previous coded macroblock (referencing the previous picture with zero motion vectors). In our example, since the aspect ratio is 352×240 pixels, and the macroblock size is 16×16, there should be 22×15 macroblocks in the whole frame. The macroblock at the bottom right corner should be the 330th macroblock, with MBA of 329. However, the largest code available for coding an MBAI is only 32. Larger MBAI can be coded only by adding macroblock escape codes (MBE). Each occurance of MBE adds 33 to the MBAI. For our example, we need a total of 9 MBEs.

The macroblock type and the motion vectors of the second macroblock are the same as those of the first macroblock. The artificial B frame is ended by stuffing 23 '0' bits for coding alignment. The frame size is only 32 bytes.

For implementation, we can make some artificial B frames of standard aspect ratios at the receivers so that they can be ready for use for different video streams. Should the receiver encounter video streams of arbitrary aspect ratio, a compatible artificial B frame can always be generated by running a simple program. On skipping a frame, the temporal information of the skipped frame is needed to be sent to the receiver, so that the receiver can set the temporal reference of the stuffed frame to the right value.

Table A.1: Coding in Artificial B Frame

Description	Code (binary)	Value/Meaning
Picture Header		
Picture Start Code	00000100 (hex)	
Temporal Reference	Variable	Same as skipped frame
Picture Type	011	B frame
VBV delay	arbitrary	Not used in VBR MPEG
Extra Information	0	
Slice Header		
Slice Start Code	00000101 (hex)	1st Slice
Quantizer Scale	00001	
Slice Information	0	
First Macroblock		
Macroblock Address Increment	1	1
Macroblock Type	0010	Prediction from last frame
Motion Vector (horizontal)	1	0
Motion Vector (vertital)	1	0
Second Macroblock		
Macroblock Escape	00000001000	+33
Macroblock Escape	00000001000	+33
Macroblock Escape	00000001000	+33
Macroblock Escape	00000001000	+33
Macroblock Escape	00000001000	+33
Macroblock Escape	00000001000	+33
Macroblock Escape	00000001000	+33
Macroblock Escape	00000001000	+33
Macroblock Escape	00000001000	+33
Macroblock Address Increment	00000011001	+32
Macroblock Type	0010	Prediction from last frame
Motion Vector (horizontal)	1	0
Motion Vector (vertital)	1	0
Stuffing Bits	0000000	
Stuffing Bits	00000000	
Stuffing Bits	00000000	

Bibliography

[1] Alan Yeung and Soung C. Liew, "Multiplexing Video Traffic using Frame-Skipping Aggregation Technique", *Proceedings IEEE International Conference on Image Processing, 1997*, pp.334-337

[2] Chi-yin Tse and Soung C. Liew, "An Integrated Video Compression and Multiplexing Scheme for Broadband Networks", *Proceedings IEEE Infocom 95*, pp.439-446

[3] Simon S. Lam, Simon Chow, and David K. Y. Yau, "An Algorithm for Lossless Smoothing of MPEG Video"

[4] M. W. Garrett, W. Willinger, "Analysis, Modeling and Generation of Self-Similar VBR Video Traffic", *Proc. ACM SigComm, London*, Sept. 1994

[5] Soung C. Liew and Hanford H. Chan, "Lossless Aggregation: A Scheme for Transmitting Multiple Stored VBR Video Streams over a Shared Communications Channel without Loss of Image Quality", *IEEE Journal of Selected Areas on Communications*, Aug. 1997

[6] Didier Le Gall, "MPEG: A Video Compression Standard for Multimedia Applications", *Communications of the ACM*, April 1991

[7] Joan L. Mitchell, William B. Pennebaker, Chad E. Fogg, and Didier Le Gall, "MPEG Video Compression Standard", *Chapman & Hall*, April 1991

[8] Soung C. Liew and Chi-yin Tse, "Video Aggregation: Adapting Video Traffic for Transport over Broadband Networks by Integrating Data Compression and Statistical Multiplexing", *IEEE Journal on Selected Areas in Communications*, Aug. 1996

[9] Fernando and Thierry Alpert, "MPEG-4 Video Subjective Test Procedures and Results", *IEEE Transactions on Circuits and Systems for Video Technology*, Feb. 1997

[10] MPEG1 *Star War* trace available via anonymous ftp site - *ftp://thumper.bellcore.com/pub/vbr.video.trace*

[11] The ITU Radiocommunication Assembly, "Methodology for the Subjective Assessment of the Quality of Television Pictures", *Recommendation ITU-R BT.500-7*

[12] Stephen Wolf, Margaret H. Pinson, Stephen D. Voran, and Arthur A. Webster, "Objective Quality Assessment of Digitally Transmitted Video", *IEEE Pacific Rim Conference on Communications, Computers and Signal Proceeding*, May. 1991

[13] Akio Ichikawa, Katsunori Yamaoka, Toshiyuki Yoshida, Yoshinori Sakai, "Multimedia Synchronization System for MPEG Video Based on Quality of Pictures", *IEEE Proceedings of Multimedia*, 1996

[14] V. Seferidis, M. Ghanbari and D. E. Pearson, "Forgiveness Effect in Subjective Assessment of Packet Video", *IEE Electronics Letters*, 8th October 1992, Vol. 28, No. 21

[15] ATM Forum Technical Committee, "ATM Forum Traffic Management Specification v.4", April 1996.

[16] Pramod Pancha, Magda El Zarki, "Bandwidth Allocation Schemes for Variable Bit Rate MPEG Sources in ATM Networks", *Trans. of Circ. and Sys. in Video Technology*, May 1992.

[17] C. Grewin, T. Ryden, "Subjective Assessments of Low Bit-rate Audio Codecs", *IEEE ASSP Workshop on Applications of Signal Processing, 1991*

[18] M. B. Amin, "Subjective Assessments of the Quality of HDTV Pictures impaired by Noise", *IEE International Broadcasting Convention, 1990*

[19] B. L. Jones, J. A. Turner, "Subjective Assessment of Cable Impairments on Television Picture Quality", *IEEE Trans. on Consumer Electronics*, Vol. 38, No. 4, Nov. 1992.

[20] Marwan Krunz, Herman Hughes, "A Performance Study of Loss Priorities for MPEG Video Traffic", *IEEE International Conference on Communications, 1995*

[21] D. T. Hoang, J. S. Vitter, "Multiplexing VBR Video Sequences onto a CBR Channel with Lexicographic Optimization", *Proceedings IEEE International Conference on Image Processing, 1997*, pp.369-372

[22] Hon-Wai Chu, D. H. K. Tsang, Tao Yang, "Call Admission Control of Teleconference VBR Video Traffic in ATM Networks", *Proceedings IEEE International Conference on Communications, 1995*