Enhancing Play-out Performance for Internet Video-conferencing

S. C. Hui, S. Foo and S.W. Yip
School of Applied Science, Nanyang Technological University
Nanyang Avenue, Singapore 639798

Abstract

The high transmission delay and packet loss conditions of the Internet environment have caused jerky video play-out and poor quality for real-time video communication. This research proposes a video play-out agent that employs the techniques of dynamic video frames reconstruction and virtual video play-out to enhance the play-out performance of Internet video-conferencing. The video play-out agent consists of three main components, namely, buffering mechanism, dynamic video frames reconstruction mechanism and play-out controller. The buffering mechanism stores incoming video frames in buffers. The dynamic video frames reconstruction mechanism is used for reconstructing intermediate frames based on two successive received frames to reduce the jerky effects. The play-out controller uses virtual video play-out mechanism to play back previously received video frames to give users the perception of uninterrupted video transmission.

Keywords: Internet video-conferencing; video play-out; dynamic video reconstruction.

1. INTRODUCTION

The Internet is a relatively harsh environment for real-time multimedia services. The high transmission delay and packet loss of the Internet environment which are characteristics of a packet switched network without resource reservation mechanisms, have made real-time video communications difficult. Under varying network load conditions, the video packets will suffer varying degrees of delay. The variance in delay produces jerky video play-out that is undesirable for real-time
services. In addition, the unreliability of the network can give rise to packet loss that will deteriorate the quality of a video-conferencing session.

To tackle these problems, the Transmission Control Protocol/Internet Protocol (TCP/IP) [1] of the Internet can be extended with new mechanisms and descriptors to deliver real-time video communication. Different mechanisms have been developed for this approach to handle the delay jitters and data packet loss problems. A buffering mechanism [2] can be used to adjust the play-out time of arriving packets at the destination to minimise the impact of delay jitters. Various rate control mechanisms [3-7] have been proposed to eliminate the impact of packet loss. In addition, compression techniques are necessary to reduce the bandwidth requirements for video data transmissions.

However, the use of rate control mechanisms and compression techniques might not be sufficient to guarantee the transmission of video packets when the network is congested. The use of rate control mechanism on a congested network by itself may result in the deterioration of the transmission rate to a point where play-out can be very jerky due to low frame rate. To eliminate the unpleasant jerky play-out effect due to packet loss, we propose a video play-out agent to enhance the play-out quality of video frames based on the techniques of dynamic video frames reconstruction and virtual play-out mechanism. In dynamic video frames reconstruction, intermediate frames are reconstructed to simulate the lost frames based on two successive received frames to smoothen out the transition between the two frames during video play-out. Moreover, in cases when no video frames are transmitted during congested network condition, the virtual play-out mechanism can be used to play out video frames received in the past instead of freezing the play-out. This is based on the assumption that there are little changes in the captured video frames as the objects involved in the scene do not move about very often and rigorously especially for video-telephony communication. This attempts to give users the impression of continuous video transmission.

In this paper, we first describe the system architecture for an Internet video communication system in Section 2. An overview of the video play-out agent and its main components are presented in Section 3. The performance evaluation of the video agent and its main components are discussed in Section 4. Finally, conclusions are given in Section 5.
2. INTERNET VIDEO COMMUNICATION

Fig. 1 shows the system architecture of a video communication system that is part of an Internet video-conferencing system [8-11]. Typically, the video communication system consists of two main modules. The Transmitter module is responsible for video capturing, packetisation and transmission of video data across the Internet to the communicating party in a peer to peer video-conferencing session. The Receiver module is responsible for the video play-out upon the video packets’ reception.

As shown in Fig. 1, the video communication system is implemented as follows. Video data is first captured and digitised through a video capture device such as video camera. Although the transmission is carried out in sequential order, the packets may arrive out of order due to different network paths traversed by the data packets and data transmission. Therefore, time-stamps are inserted into video packets before data transmission. The time-stamp allows the packets arriving at the recipient’s host computer to be ordered in the correct time sequence before being played-out. The time-stamp is stored in the time-stamp field of the header of each video packet. Since each video frame is discrete, a single frame should not be packetised into multiple packets, as the original video frame will not be recoverable once a packet is lost. This is especially true when the Internet is used as the transmission medium. Hence, a video packet should encapsulate one complete video frame before transmission.

Video data is compressed prior to its transmission. The size of video frames can be greatly reduced with the use of compression methods such as discrete cosine transform (DCT) used in JPEG [12] and motion compensation algorithm used in MPEG [13,14]. A compression ratio of 1:15 is usually achievable with JPEG and
MPEG compression. The rate control mechanism [3-7] is used to ensure the quality of service of the video transmission by making dynamic adjustments to the transmission based on the current network condition. It obtains the necessary information regarding the network conditions in an attempt to adjust control parameters (such as sampling rates, choice of compression algorithms, compression ratios, etc.) and provide an optimum level of performance based on existing network conditions. For example, if the network condition is congested, a higher compression ratio with lower quality JPEG compression can be selected to compress video images in order to reduce the bandwidth used.

Received video data has to be decompressed into a format that is used by the video play-out agent. Network condition information such as packet loss rate and transmission delay experienced is transmitted to the calling party by the feedback mechanism for the purpose of dynamic rate control. Finally, the video play-out agent manages the video play-out for the video-conferencing session. The video play-out agent will be discussed in detail in next section. Fig. 2 shows a video-conferencing session between two users.

![Fig. 2. A video-conferencing session between two users.](image-url)
3. VIDEO PLAY-OUT AGENT

Fig. 3 shows the architecture of the video play-out agent. The video play-out agent will not participate in the video communication process. Instead, it will act as a front-end component interfacing with the video communication system for playing out video frames received through the Internet. The video play-out agent consists of three major components: Buffering Mechanism, Dynamic Video Frames Reconstruction Mechanism and Play-out Controller.

Upon reception of video frames, the play-out agent stores these frames in buffers using the buffering mechanism. Based on the video reception rate, the play-out controller will determine the type of play-out to use. Two types of play-out are supported, namely, real-time play-out and virtual video play-out. When the reception rate is high, real-time play-out will be used for playing out video frames immediately once received. Otherwise, virtual video play-out will be used to play out previously received video frames to give the user the perception of uninterrupted video transmission. The play-out controller will then determine the number of intermediate frames required to be reconstructed by the dynamic video frames reconstruction mechanism which reconstructs intermediate frames based purely from two successive frames without requiring any additional information on the relationship between the two frames. Once reconstructed, the original frames together with the reconstructed frames will be used for playback on the user's terminal.
3.1 Buffering Mechanism

The video frames received from the video communication system are stored in buffers by the buffering mechanism. Information pertaining to each individual video frame such as the time-stamp is also stored to facilitate the play-out process. The buffers used for the implementation of the buffering mechanism can reside either in main memory or on disk. Memory buffers allow fast data access required for efficient video play-out. However, each video frame takes up a considerable amount of storage space. Hence, there is a limit to the number of video frames that can be stored in memory buffers. In contrast to memory buffers, the storage capacity of file buffers is only limited to the available disk space. Although access to file buffers is slower as compared to memory buffers, the access time for modern day hard disk drives is sufficient for efficient video play-out. Therefore, file buffers are chosen for implementing the buffering mechanism. In addition, a double buffering mechanism is used to control the use of buffers efficiently as well as allow uninterrupted supply of video frames during virtual play-out.

3.2 Dynamic Video Frames Reconstruction Mechanism

Dynamic video frames reconstruction is a lightweight mechanism for reconstructing intermediate frames to smoothen out jerky video play-out due to low video transmission rate and packet loss during the video transmission process. Different algorithms can be used to generate intermediate frames based purely on two successive frames. In this research, the interpolated transparency algorithm [15] is adopted due to its simplicity in implementation and its capability in providing satisfactory performance to produce the necessary effects. Conventionally, transparency algorithms have been used for producing fading effects in computer graphics. This fading effect produces a smooth transition between two different scenes.

3.2.1 Interpolated Transparency algorithm

In the interpolated transparency algorithm, pixel values for intermediate frames are determined by linear interpolation of pixel values between two frames. The computation of the intermediate frames’ pixel values between frame 1 and frame 2 is based on the following equation:

\[ I_b = (1 - k_t) I_{b1} + k_t I_{b2} \]

The transmission coefficient \( k_t \), which ranges between 0 and 1, indicates the transparency of frame 1. A value of 0 for \( k_t \) will result in the reconstructed frame being exactly the same as frame 1. Hence, \( k_t \) determines the degree of influence frames 1 and 2 have on the intermediate frames to be reconstructed. \( I_{b1} \), \( I_{b2} \) and \( I_b \) are the pixel values for the reconstructed frame, frame 1 and frame 2 respectively. As pixel values are usually stored as RGB (Red-Green-Blue) triplets, \( I_b \) will contain the combination of individual color component for each pixel of the reconstructed frames. Thus, after computing each individual color component for each pixel using the above equation, the result will be combined to form the whole pixel's value. During reconstruction, \( k_t \) will progressively increase from 0 to 1, resulting in the reconstruction of video frames that produce the fading effect from frame 1 to frame 2.
3.2.2 Reconstruction mechanism

Fig. 4 illustrates the dynamic video frames reconstruction mechanism. Frame buffer 2 is used to store the current video frame to be displayed and frame buffer 1 is used to store the previously displayed video frame. The video frame (i.e. F2) about to be played out serves as the input to the video frames reconstruction mechanism and stored in the frame buffer 2. Control information required for the reconstruction such as the number of frames to be reconstructed will be computed by the play-out controller and passed to the video frames reconstruction mechanism. Intermediate frames (F121, F122, …F12n) are then reconstructed based on the two video frames (F1 and F2) stored in the video frame buffers 1 and 2. Upon completion of the reconstruction, the input video frame (F2) in the frame buffer 2 will be transferred to frame buffer 1 for preparation for the next reconstruction. This can simply be done by re-labeling frame buffer 2 as frame buffer 1.

3.3 Play-Out Controller

The play-out controller determines the play-out scheme to be used during video play-out. In addition, it calculates the number of frames required to be reconstructed based on the time-stamp associated with each video frame.
3.3.1 Play-out schemes

Two play-out schemes are supported, namely, real-time play-out and virtual play-out. For real-time play-out, video frames will be played out immediately once they are received. Real-time play-out is only feasible when the transmission rate is high and the transmission latency is within acceptable range. In this research, the real-time play-out scheme will be used whenever the play-out controller detects a reception rate of more than 1 frame per second (fps). In conjunction with the dynamic video frames reconstruction mechanism, acceptable video play-out is possible from such low transmission rate.

In virtual play-out, previously received video frames are used for play-out. This allows for uninterrupted video play-out even when video transmission has stopped due to network congestion or other reasons. Using virtual play-out, the user may perceive a real-time play-out even when no video frames are actually received from the video communication system. The number of video frames available for playing out will be based on the number of video frames previously stored in the buffer and the buffer size. This scheme will be used when real-time play-out is infeasible.

Fig. 5. Virtual play-out modes.
Video frames stored in the buffer can be played back in either the forward or reverse direction. As shown in Fig. 5, two play-out modes are possible during virtual play-out, namely, unidirectional and bi-directional play-out modes. In the unidirectional play-out mode, video frames in the buffer are played back from the first frame till the last frame with the process repeating itself (Fig. 5(b)). In the bi-directional play-out mode, video frames are played back in the reverse direction after the last frame has been played back (Fig. 5(c)). The bi-directional play-out mode is preferred and used as it gives a better visual effect and smoother transition of video play-out from the last video frame to the next frame in the play-out sequence.

3.3.2 Play-out sequence

![Diagram of play-out sequence](image)

(a) Real-time play-out during high video transmission rate.

(b) Real-time video when transmission is disrupted.

(c) Virtual video play-out when video transmission stopped or real-time play-out de-activated.

Fig. 6. Play-out sequence.
As the play-out controller supports the use of two types of play-out schemes according to network congestion conditions, the resulting play-out sequence may switch between real-time and virtual depending on the video reception rate. Fig. 6 illustrates some of the possible play-out sequence as a result of the integration of real-time and virtual play-out schemes.

When there is an uninterrupted transmission of video frames, real-time play-out is in effect (Fig. 6(a)). Video frames will be played out immediately upon reception. In addition, the video frames will also be stored in the buffer to be ready for use in virtual play-out. When no video frames are received (Fig. 6(b)), the play-out controller will switch from real-time play-out to virtual play-out. In this case, video frames previously stored in the buffer will be played out one at a time. Intermediate frames will also be constructed and played out with the original frames. After playing out each video frame, the system will check for the availability of any newly received video frames. If new video frames are received, real-time play-out will be used. If no new video frames are received, then virtual play-out will continue (Fig. 6(c)).

3.3.3 Determination of the number of intermediate frames

Upon reception of video frames, the play-out controller will determine the number of intermediate frames to be reconstructed by the dynamic video frames reconstruction mechanism. The time interval between successive received video frames may not be the same due to packet loss during the transmission process. The number of intermediate frames to be reconstructed will directly affect the quality of the play-out.

To determine the number of frames to be reconstructed, it is assumed that this number is directly related to the difference in time-stamps between two consecutive video frames. In addition, the reconstructed video stream should be played out at a rate approximately equal to the video capture rate. The number of intermediate frames for the \(n^{th}\) video frame in the buffer, \(N_{\text{inter},n}\), to be reconstructed is computed based on the following equations:

\[
N_{\text{compute},n} = \frac{(TS_{n+1} - TS_n) }{ TR } + N_{\text{carry},n-1} \\
N_{\text{inter},n} = \lceil N_{\text{compute},n} \rceil \\
N_{\text{carry},n} = N_{\text{compute},n} - N_{\text{inter},n}
\]

where \(TS\) is the time-stamp of the video frame and \(TR\) is the amount of time required for reconstructing a single frame. Since no fractional portion of a video frame will be displayed, the fractional portion, \(N_{\text{carry},n}\), will be carried over to the next frame and used in its determination of the required intermediate frames. This procedure provides a conservative estimation that will maximise but not exceeding the amount of time available, reflected from the time-stamps between two successive frames, for reconstruction. The video play-out agent will buffer the intermediate frames prior to its play-out. The buffering mechanism enables the video play-out agent to introduce the time interval required between the playing-out of two successive frames.
4. PERFORMANCE ANALYSIS

As the video play-out agent employs the techniques of dynamic video reconstruction and virtual play-out to provide uninterrupted video play-out services, experiments are conducted to measure the performance of the dynamic video reconstruction mechanism and virtual video play-out. The measurements will gauge whether dynamic video reconstruction can sustain a high video play-out rate so as to minimise the jerky effect and whether the virtual video play-out can maintain uninterrupted play-out services.

The performance analysis of the virtual video play-out will be purely subjective as its effectiveness depends on its ability in giving the user the perception of a continuous real-time video play-out when used. The performance of the dynamic reconstruction mechanism will be affected mainly by two factors: the current network performance and the available computing resources.

- **Network performance.** The quality of the network directly affects the effectiveness of the dynamic video reconstruction mechanism since large transmission delay and high packet loss rate may render the mechanism helpless as no useful reconstruction and play-out can be performed under such circumstances. In such cases, video simulation with virtual video play-out will be used to simulate an ongoing video communication session.

- **Computing resources.** Sufficient processing power is required to process the video packets and reconstruct intermediate frames whenever necessary. The interpolated transparency algorithm is used so that real-time reconstruction is feasible without compromising the performance of other processes such as video packet processing, transmission and reception. The amount of processing power available will also affect the responsiveness of the system while handling incoming and outgoing packets.

To conduct performance analysis on the video play-out agent, it is necessary to test the video play-out agent under various Internet conditions. As there are large variations for Internet behavior at different routes and at different times of the day, it is difficult for any single simulation model to correctly model all the characteristics of the Internet [16]. An analysis that is performed only on a few Internet routes will not be sufficient in capturing all possible scenarios necessary to verify the actual performance of the video play-out agent. Under such constraints, a synthetic and controlled environment that can generate video packets at different transmission rates and packet loss conditions can be used.

The environment used for performance analysis comprises two Pentium 200 MHz PCs connected over a LAN (local area network). A video sequence consisting of 225 frames captured at a resolution of 160x120 using Indeo compression technique [17] is used as the video stream for testing. The use of a pre-captured video sequence enables various transmission rates and packet loss rates to be simulated.

Two experiments are carried out to determine the performance of the video play-out agent’s dynamic video reconstruction mechanism. The first experiment is used to determine the play-out rates for various transmission rates. The second experiment tests the video play-out agent with varying packet loss conditions. In the two experiments, virtual video play-out will not be activated since it does not participate
in real-time play-out. In addition, different surveys are conducted to analyse the effectiveness of the reconstruction mechanism in producing a smoothing visual effect and the virtual video play-out in simulating real-time play-out.

4.1 Play-out rates for various transmission rates

The first experiment provides an insight into the relationship between the play-out and transmission rates. During the experiment, the pre-recorded video stream was transmitted at various transmission rates from 1 fps to 15 fps in steps of 0.5 fps. Since video play-out at 15 fps produces satisfactory performance for a video-conferencing session, it is not necessary to simulate transmission rates above 15 fps. Packet loss is not simulated in this experiment. The transmission rate is computed over the receiver side as follows:

\[ \text{Transmission Rate} = \frac{1000}{T_{sc} - T_{sp}} \]  

(4)

where \( T_{sc} \) is the time-stamp (in ms) of the current video packet and \( T_{sp} \) is the time-stamp (in ms) of the previous video packet. The number of intermediate frames to be reconstructed by the reconstruction mechanism is computed based on equations (1), (2) and (3). The play-out rate is then determined for each of the transmission rate using the following equation:

\[ \text{Play-out Rate} = \frac{N_t}{T_t} \]  

(5)

where \( T_t \) is the time interval used to measure the play-out rate and \( N_t \) is the total number of intermediate frames and the original video frame used for the playing-out during the interval \( T_t \).

The number of intermediate frames reconstructed and the resulting play-out rates (with \( T_t = 1 \) sec) for various transmission rates are shown in Fig. 7 and Fig. 8 respectively. The resulting play-out rates remain relatively constant at around 23 fps for transmission rates between 1 fps to 15 fps. Hence, the video play-out agent is able to maintain a constant and high play-out rate for transmission rates between 1 fps to 15 fps.
4.2 Play-out rates under packet loss conditions

In the second experiment, the video play-out agent is tested under varying packet loss conditions. Delay variation and out of order delivery will not be simulated in this experiment since the problem arises from them can be solved by the buffering mechanism and virtual video play-out. During the experiment, video packets are randomly discarded by the transmitting system. As a result of these packet losses, the receiving system will perceive varying transmission rates, computed from the time-stamps of two adjacent packets. Fig. 9 shows the perceived transmission rates under the packet loss conditions and Fig. 10 gives the corresponding play-out rates. As seen from Fig. 9 and Fig. 10, a relatively constant play-out rate can be achieved under the simulated packet loss conditions. Given a more lenient environment on the Internet, the video play-out agent will be able to sustain the same performance as demonstrated in this experiment. However, when the transmission delay is too high for effective reconstruction, the virtual video play-out, as discussed earlier, will be used instead.
4.3 Subjective measurement

This section briefly describes the surveys carried out on the visual performance of the reconstruction mechanism and the effectiveness of the virtual video play-out mechanism.

4.3.1 Visual performance

The capability of the reconstruction mechanism in maintaining a high play-out rate does not guarantee good visual performance. In this part of the assessment, the visual performance was measured. Ten test subjects were presented with a live video sequence captured at 1 fps to 15 fps and asked to determine its visual effects before and after the use of the dynamic reconstruction mechanism. For transmission rates between 1 fps to 10 fps, 90% of the subjects agree that the play-out is less jerky with the incorporation of the reconstruction mechanism. However, for transmission rates above 10 fps, about 60% of the subjects agree that the play-out is less jerky with reconstruction mechanism. The remaining 40% of the subjects feel that although the reconstruction mechanism can smoothen out the play-out, it produces blurry visual effects. This is mainly due to the characteristic of the interpolated transparency algorithm.

4.3.2 Virtual video play-out

The effectiveness of the virtual video play-out depends on the extent to which its integration with real-time play-out in providing a continuous video sequence. In this survey, the test subjects were presented with a real-time video sequence simulating a video-conferencing session with virtual video play-out activated. During video communication, the real-time video sequence was occasionally terminated and resumed to either activate or deactivate the virtual video play-out respectively. The test subjects were then asked on their awareness of the termination and resumption of real-time video sequence. The results show that the test subjects were unable to clearly notice the transitions with the occasional switching from real-time play-out to virtual play-out, or vice versa. However, if the real-time video transmission is frequently terminated and resumed, the transitions become obvious.
4.3.3 Discussion

The surveys have demonstrated that the dynamic reconstruction mechanism is effective in minimising jerky effects for transmission rates below 10 fps. The interpolated transparency algorithm, although simple to implement, produces some blurry visual effects. Hence, other alternative algorithms can be investigated for better visual performance.

The virtual video play-out agent produces poor visual effects when the switch from the real-time play-out to virtual play-out or vice versa is frequent. To overcome this problem, the play-out agent has adopted a mechanism to ensure that the play-out time for either real-time play-out or virtual play-out is maintained sufficiently long to avoid too frequent changes. In this mechanism, when real-time transmission has been interrupted, virtual video play-out will be activated. The video play-out agent will not switch back to real-time play-out immediately upon resumption of real-time transmission. Instead, it will monitor the real-time transmission for a period of 15 seconds to ensure that there is no interruption within this period before switching back to real-time play-out. Although this mechanism introduces additional delay in responding to the resumption of real-time transmission, it is more resilient to frequent interruption to real-time transmission due to unreliable network condition and heavy traffic load. Currently, this mechanism has been implemented and virtual video play-out has performed satisfactorily.

5. Conclusion

In this paper, we have described the system architecture of the video play-out agent that consists of three main components: buffering mechanism, dynamic video frames reconstruction mechanism and play-out controller. The buffering mechanism stores incoming video frames in buffers. The play-out controller will determine whether to use real-time play-out or virtual video play-out based on video reception rate. Real-time play-out will be used when the reception rate is high. Otherwise, virtual video play-out will be used to play back previously received video frames to give users the perception of uninterrupted video transmission. The dynamic video reconstruction mechanism is used for reconstructing intermediate frames based on two successive received frames. Finally, the original frames and the reconstructed frames are played out on the user's terminal. The development of the video play-out agent has demonstrated considerable potential in providing low bit-rate uninterrupted video play-out services for Internet video communication applications.

REFERENCES


