



International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Analysis of Different Types of Video

Manish P Bhatt, Dr. Priyanka Sharma, Prof. Purvi Kansara

CSE Dept. Institute of Technology, Nirma University,
Amedabad, India

Abstract—In recent years, multimedia application like video conference, video on demand, video streaming are used over internet very widely. Multimedia application are real time application with variable bit rate and drastic change in traffic so researchers and students consider multimedia application for their research work. In this paper, we study different type of videos their analyse and their behaviour when transmitted over network. This paper can be helpful to researchers or students for further research work.

Keywords—PSNR; MOS; DiffServ; WRED; Delay; Jitter

I. INTRODUCTION

Multimedia file is the integration of text, graphics, audio, video, etc. If you are working on multimedia application, you need to be aware of multimedia container formats, codecs and multimedia frames. Codecs are used to compress and decompress video signal. Container format is to store information on disks. Multimedia file extension, .mov, .ogg, .wmv, .flv, .mp4, .mpeg, .avi etc., indicate the container format. As shown in figure-1, videos have a lot of information including the video stream, audio stream, metadata, subtitles, chapter-information, synchronization information, that can be encapsulated together in a container format. Before encapsulating information, all the various information are compress into a compact file using codec. MPEG-4, MPEG-2, H.264, H.263, VP6, Sorenson Spark, etc., are video codec and AAC, MP3, Vorbis, etc., are audio codec. For example, you may encode a video stream using MPEG-4 and an audio stream using the AAC compression scheme. These may then be wrapped up within a .mov file container.

A digital frame of a black and white 640×480 pixel standard-definition(SD) television image that uses 8-bits to represent each pixel's grayscale consumes 2.45 Mbits of memory. If we updated this image with a frame rate of 30 frames per second (fps), the resulting video bandwidth requirement would be roughly 73.5 Mbps. This requirement increases significantly when we add more bits to encode the image and add color and audio channels. Access network bandwidth constraints make streaming such high-bandwidth streams into homes impractical, so we use compression to reduce the video stream's bit-rate.

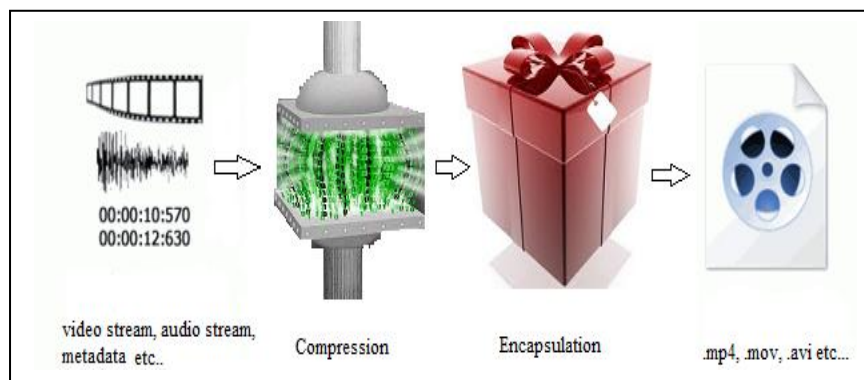


Fig. 1. Multimedia file compression.

This paper is organized as follows. Section II describes Literature Surbey. Sections III describes QoS assessment metric. Section IV describes Analysis of Different videos. Finally section VII concludes the paper.

II. LITERATURE SURVEY

Multimedia file should be compress at sender side and decompress at receiver end to transfer it over network or internet. Encoders are used to compress a multimedia files. In this paper, we used MPEG-4 encoders to compress multimedia files. MPEG-4 Standard are used for multimedia and Web compression. MPEG-4 is object-based compression technique, meaning individual objects within a scene are tracked separately and compressed together to create an MPEG4 file. It also allows developers to control objects independently in a scene, and therefore introduce interactivity. MPEG-4 encoders convert video signal into a series of frames. Generally, only limited change occurs between one frame and the next, so an encoder can compress the video signal significantly by transmitting only the

differences. MPEG-4 encoders convert Multimedia file into three types of frame. Intra, or I-frames, carry a complete video picture and are coded without reference to other frames. I-frames compress using spatial compression. A received I-frame provides the reference point for decoding a received other frames. Predictive-coded or P-frames to be coded from a preceding I-frame or P-frame using temporal compression. P-frames can provide increased compression compared to I-frames, with a P-frame typically 20 to 70 percent the size of an associated I-frame.

Finally, bi-directionally predictive-coded, or B-frames, use the previous and next I-frame or P-frame as their reference points for motion compensation. B-frames provide further compression, typically 5 to 40 percent the size of an associated I-frame.

In MPEG encoding, frames are arranged into groups of pictures (GoPs) that include an I-frame and all subsequent frames leading up to the next I-frame. GoP size is defined by number of frames in GoP. typically GoP size is 15 or 12.

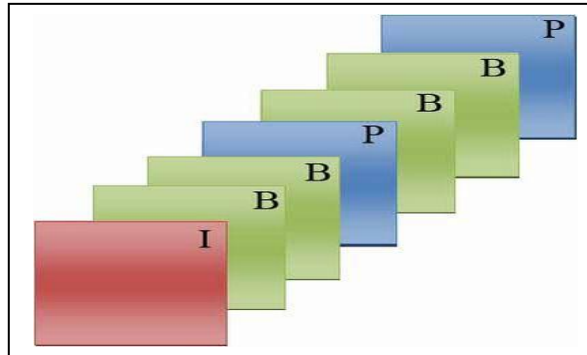


Fig. 2. Group Of Picture (GoP).

To transport MPEG-encoded video over IP networks, MPEG frame information is encapsulated within MPEG Transport Stream (TS) packets, which are in turn transported in IP packets. Figure shows MPEG-TS packet. A typical IP packet for transporting MPEG video contains seven 188-byte MPEG-TS packets. An MPEG frame can span multiple IP packets, and a single packet can contain information from two consecutive frames, so the loss of a single packet can result in a loss of information from two frames. MPEG frame information is encapsulated within MPEG Transport Service (TS) packets.

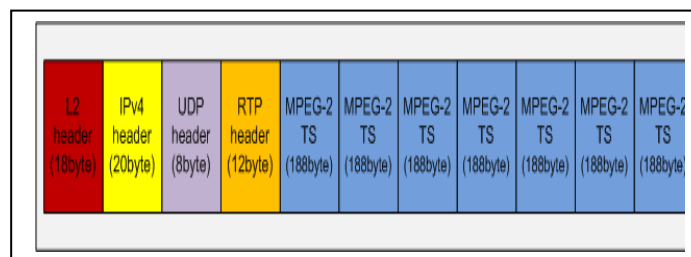


Fig. 3. MPEG-TS PACKET

III. QoS ASSESSMENT METRIC : PSNR AND MOS

There are basically two approaches to measure digital video quality, namely subjective quality measures and objective quality measures. Subjective quality metrics always grasp the crucial factor, the impression of the user watching the video while they are extremely costly. The human quality impression usually is given on a scale from 5 (best) to 1 (worst) as in Table 1. This scale is called Mean Opinion Score (MOS). The expensive and complex subjective tests can often not be afforded. Therefore, objective metrics have been developed to emulate the quality impression of the human visual system (HVS).

The most widespread method is the calculation of peak signal to noise ratio (PSNR) image by image. It is a derivative of the well known signal to noise ratio (SNR), which compares the signal energy to the error energy. The PSNR compares the maximum possible signal energy to the noise energy, which has shown to result in a higher correlation with the subjective quality perception than the conventional SN

TABLE 1

ITU-R quality and impairment scale	Scale	Quality	Impairment
	5	Excellent	Imperceptible
	4	good	Perceptible but not annoying
	3	fair	Slightly annoying
	2	poor	Annoying
	1	bad	Very annoying

Let p_n^{di} denote d_i^{th} pixel of the n^{th} reconstructed frame at the decoder, and p_n^{ei} denote e_i^{th} pixel of the n^{th} original coded frame at the encoder. The total frame error at frame n is defined as

$$e_n = \sum_{i=1}^M (p_n^{di} - p_n^{ei}) \tag{1}$$

Where M is the number of pixels in each frame. The Mean Square Error (MSE) associated with frame error e_n is given by

$$d_n = E[(p_n^{di} - p_n^{ei})^2] = \frac{1}{M} \sum_{i=1}^M [(p_n^{di} - p_n^{ei})^2] \tag{2}$$

Since the codec employs motion compensation and inter prediction to encode consecutive frames, distortion propagates to subsequent frames. Thus, the total distortion for a single lost frame at n is

$$D_n = \sum_{i \geq n} d_i \tag{3}$$

The PSNR of the video signal of frame n is given by

$$PSNR_{dB}[n] = 20 \log_{10} (V_{peak}/RMSE) = 20 \log_{10} (V_{peak}/\sqrt{d_n}) \tag{4}$$

Where V_{peak} is the maximum possible pixel value of the frame and RMSE is the root mean square error between received and original frames.

we consider luminance component Y of image and we represent pixel in row and column then equation 4 can be written as below

$$PSNR_{dB}[n] = 20 \log_{10} \left[\frac{V_{peak}}{\sqrt{N_{col} N_{row} \sum_{i=0}^{N_{col}} \sum_{j=0}^{N_{row}} [Y_S(n,i,j) - Y_D(n,i,j)]^2}} \right] \tag{5}$$

Where $V_{peak} = 2^k - 1$ and $k = \text{number of bits per pixel (luminance component)}$. Denominator is nothing but a mean square error (MSE). PSNR measures the error between a reconstructed image and the original one. Prior to transmission, it is possible to compute a reference PSNR value sequence on the reconstruction of the encoded video as compared to the original raw video. After transmission, the PSNR is computed at the receiver for the reconstructed video of the possibly corrupted video sequence received. Conversion of PSNR to MOS shown in table 2

TABLE 2

Possible PSNR to MOS conversion	PSNR (dB)	MOS
	> 37	5 (Excellent)
	31-37	4 (good)
	25-31	3 (fair)
	20-25	2 (poor)
	< 20	1 (bad)

IV. ANALYSIS OF DIFFERENT VIDEOS

To study behavior of video quality, we have taken four different videos in terms of number of objects in frame and changes per frame. i) Claire news reader (only facial movement and static background) has total 494 frames which include 56 I-frame, 328 B-frame and 110 p-frame. ii) Deaf news reader (continuous facial movement, hand movement and static background) has Total 300 frames which include 34 I-frame, 199 B-frame and 67 P-frame. iii) Coastguard (moving boat and steamer in opposite direction with high motion speed, changing background having many objects) has total 300 frames which include 34 I-frame, 199 B-frame and 67 P-frame. iv) Foreman (site man explaining site have many object and continuous movement throughout frames) has total 400 frames which include 45 I-frame, 266 B-frame and 89 P-frame.



Fig. 4. Claire news reader: only facial movement and static background) has total 494 frames which contain 56 I-frame, 328 B-frame and 110 p-frame



Fig. 5. Deaf news reader : continuous facial movement, hand movement and static background video sequence has Total 300 frames which contain 34 I-frame, 199 B-frame and 67 P-frame.



Fig. 6. Coastguard : moving boat and steamer in opposite direction with high motion speed, changing background having many objects has total 300 frames which contain 34 I-frame, 199 B-frame and 67 P-frame.



Fig. 7. Foreman: site man explaining site have many objects and continuous movement throughout frames has total 400 frames which contain 45 I-frame, 266 B-frame and 89 P-frame.

Analysis of Video Frame size : I-, P-, B- frames' size varies according to video types. I-, P-, B- frames' size varies based on number of object in frame and speed of motion in video. I-frame Size : I-frame size is always greater than P-, B-frames' size. I-frame size varies from video to video based on number of object in video frame. We can understand this by analysis I-frame size for above videos. Figure-8 shows I-frame size for videos.

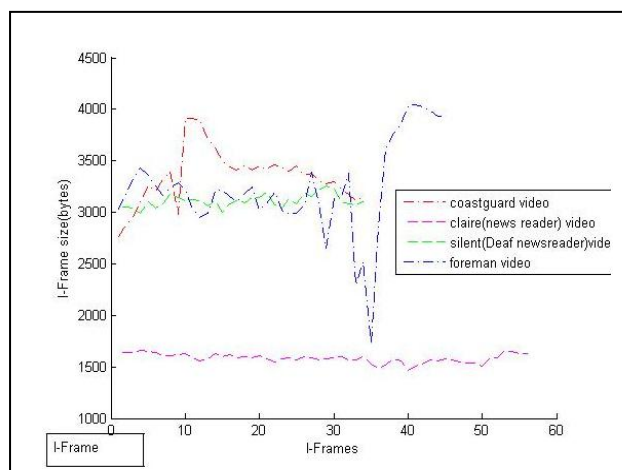


Fig. 8. I-frame size for different videos.

Claire video has very few changes (only facial movement) per frame and very slow in motion, has I-frame size is nearly 1600bytes. Deaf news reader video has continuous facial and hand movement, has I-frame size is nearly 3100bytes (approx. Double than Claire video). Coastguard video has moving object and background with high motion speed, has I-

frame size varies between 2800byte to 4000bytes. There are drastic changes in I-frame size for coastguard video, when objects are passing in opposite direction with high motion speed. Foreman video explain new cite with zooming effect has I-frame size from 3000byte to more than 4000bytes. P-frame Size: P-frame is coded by referencing I-frame. Figure-9 shows P-frame size for above videos. Claire video has P-frame size around 200bytes. Deaf news reader video has P-frame size from 200bytes to 600bytes. Coastguard video has object overlapping and high motion, has P-frame size 800bytes to 2000bytes. We can see drastic change in P-frame size for coastguard video. Foreman video has more number of objects per frame and zooming effect, has P-frame size from 600bytes to more than 1600bytes. B-frame Size: B-frame size is smaller compared to other frames. B-frame size is varies as shown in Figure-10. Claire video has B-frame size between 100bytes to 200bytes. Deaf news reader video has B-frame size between 200bytes to 400bytes. Coastguard video has B-frame size start with 200bytes to 1200bytes. Foreman video has B-frame size up to 900bytes. Coastguard and Foreman video have drastic changes in B-frame size.

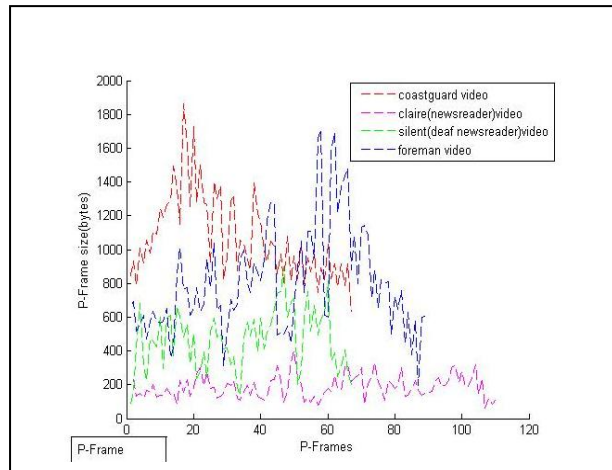


Fig. 9. P-frame size for different videos.

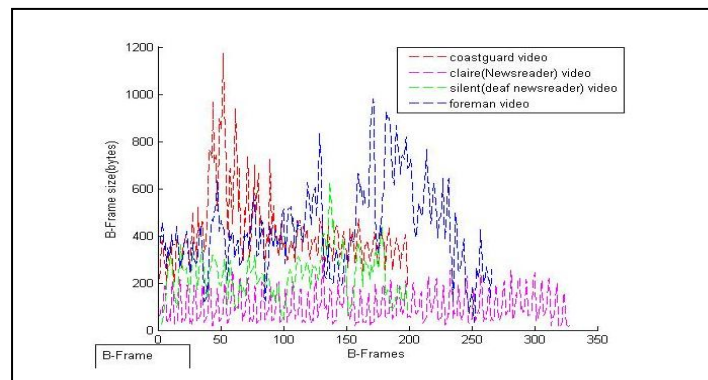


Fig. 10. B-frame size for different videos.

Effect of Frame Loss: We take coastguard video and foreman video to understand importance of frames. If I-frame is lost, PSNR values are continuously low from frame number 40 to 60 in Coastguard video as shown in Figure-11. In contrast, if B-frames are lost, PSNR values vary in odd/even manner. Figure-12 shows user perceive quality when I-frame is lost, that means complete one GoP. When B-frame is lost, PSNR values are in zigzag manners, Figure-13 shows user perceive quality when B-frame is lost.

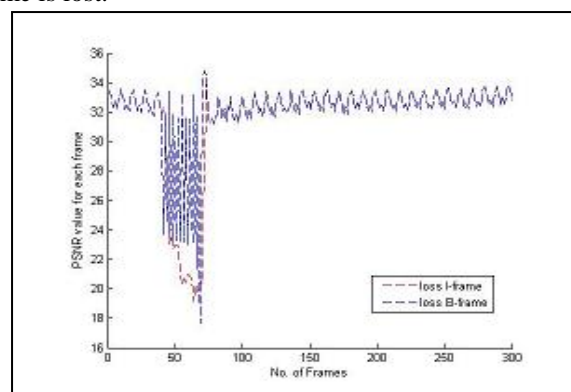


Fig. 11. PSNR value with I-frame loss and B-frame loss.

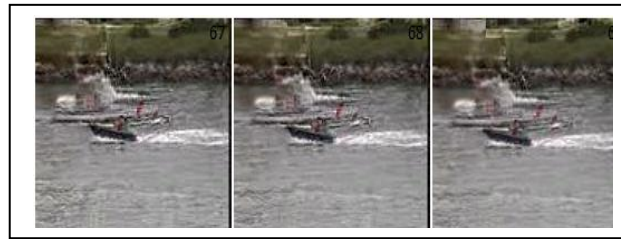


Fig. 12. Frame quality with loss of I-frame.



Fig. 13. Frame quality with loss of B-frame

V. CONCLUSION

In this paper, we study behaviour for different types of video sequences in terms of number of object per frame and changes per frame. We conclude that, if packet loss occur for video sequences, having less number of object in frame and slow in motion, make negligible effect on user perceive quality while packet loss occur for video sequences, having more number of object in frame and more movement, user perceive poor quality of video. When I-frame is lost, consecutive frames are corrupted as I-frame is reference frame for a GoP. We also conclude that user perceive bad quality of video when consecutive frame are corrupted. if a single or alternate B-Frames or P-Frames are corrupted in any GoP then it doesn't make any difference in user perceive quality as frame rate is generally 25/30 frames per second. For high speed motion videos like football game or coastguard sequence, B-Frame also carries information as I-frame, so loss of B-frame makes effect on user perceive quality for such videos

REFERENCES

- [1] J. Klaue, B. Rathke, and A. Wolisz, "EvalVid A framework for video transmission and quality evaluation," in Proceedings of the International Conference on Modelling Techniques and Tools for Computer Performance Evaluation, 2003, pp. 255-272.
- [2] Chih-Heng Ke, Ce-Kuen Shieh, Wen-Shyang Hwang, and Artur Ziviani, "An evaluation framework for more realistic simulations of MPEG video transmission", *Journal of Information Science and Engineering*, vol. 24, pp. 425-440, 2008.
- [3] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An architecture for differentiated services," RFC 2475, IETF, Dec. 1998.
- [4] J. Shin, J. Kim, and C.-C. J. Kuo, "Quality-of-service mapping mechanism for packet video in differentiated services network," *IEEE Trans. Multimedia*, vol. 3, no. 2, pp. 219-231, Jun. 2001.
- [5] C.-H. Ke, C.-K. Shieh, W.-S. Hwang, and A. Ziviani, "Two markers system for improved MPEG video delivery in a DiffServ network," *IEEE Commun. Lett.*, vol. 9, no. 4, pp. 381-383, Apr. 2005.
- [6] L. Chen, G. Liu, and F. Zhao, "An improved marking mechanism for real-time video over DiffServ networks," in Proc. PCM, Dec. 2007, pp. 510-519.
- [7] Lishui Chen, and Guizhong Liu, "A Delivery System for Streaming Video Over DiffServ Networks," *IEEE trans. on circuits and system for video technology*, vol. 20, no. 9, september 2010.
- [8] Manish P.Bhatt, Priyanka Sharma, and Purvi Kansara, "Streaming MPEG-4 Compressed Video Over DiffServ Network", *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 4, 2014.