Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/6/20230850

An integrated functionality framework for robust video streaming heterogeneous networks

D. Kirubha[1], Sathishkumar V. E.[2], Baiju B. V.[3], Maheshwari V.[4] and Prasanna M.[4]

[1]Department of Computer Science and Engineering, Rajarajeswari College of Engineering and Technology, Bangalore, Karnataka, India.
[2]Department of Industrial Engineering, Hanyang University, Seoul, Republic of Korea
[3]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India.
[4]School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India.

srisathishkumarve@gmail.com

Abstract. Video streaming is one of the drastically growing applications and more researchers are focused on their researchers in same. All over the world, the cost of video streaming process is increased up to 37$ billion. Even though video streaming process meets various problems like time, Bit Rate Error (BER) and buffers usage with cost. To provide solution to the above issues, this paper proposed an integrated framework to provide a video streaming method by increased Quality of Service (QoS). The proposed framework integrates (Integrated Framework – IF) a popular standard named H.264/AVC video coding and efficient QoE prediction on similar frames in the video. Using the framework, a QoS based video streaming method is proposed, whereas the packet loss is reduced significantly including extra BER value for the channel coding. From the simulation results it is illustrated that high flexibility and efficacy of the proposed framework is more effective in terms of preventing from frequent loss of frames.

Keywords: video streaming, video compression, quality of service, quality of experience, key frame detection

1. Introduction

One of the most important services is video streaming among mobile users through wireless channel. It is important because of the mass capacity of the channels. The main contest of this is to offer a reliable one and battle the packet loss throughout the channel [1]. Video streaming is affected by various factors like congestion in network, outage in connection and channel fading, etc. In real-time streaming applications, the video packets are protected by using the Forward error correction (FEC). It can achieve a high end-to-end quality while transmitting a video sequence [2]. The videos are protected based on priorities or importance of that video. The priority of the video packets is determined based on the coding methodologies used. In single layer coding, a variety of methods is used for partitioning and classifying the key frames. The key frames are I, P and B frames [3]. In scalable video coding, various scalable layers are having their various priorities. Live streaming and video-on-demand are expanding at a fast pace. 63% stream on-request media and online video more than weekly. Over-the-best TV review will develop from 3.4% of TV survey hours (2013) to 20.4% (2017) in North America. Worldwide OTT video market is assessed to develop to $37.2 billion by 2017 [4]. Nonetheless, live video spilling keeps on anguish from high buffering proportions, high join times, high join disappointments and low normal piece rates. The financial effect of these client encounter measurements is tremendous [5]. A current white paper from Akamai states that 10-second deferral in video spilling caused by startup time triggers over a 45% decrease in viewership (and income). Late reviews have demonstrated that customary CDNs represent over 20% of these join disappointment and bit rate debasement issues. Conventional CDNs were implied for serving static records and page sections which were commonly pulled by the site guests everywhere throughout the world [6]. Be that as it may, a live video stream needs to go from a live occasion area to its watchers all around the globe continuously. A few M&E organizations want to stream live video encourage of their substance starting with one landmass then onto the next to their business accomplices (B2B) or end clients (B2C) [7]. Given the imperatives on ongoing conveyance and nature of administration required, they cannot utilize any current substance conveyance systems for this reason and frequently depend on committed rented lines for such exchanges [8].

2. Related works

Video on Demand (VoD) suggests to a video that is facilitated on at least one server in the cloud and is rushed to a watcher when it expressly asks for that video. Live Internet video, then again, alludes to situations wherein a live occasion is communicated to at least one watcher over the world over the Internet. This is ending up noticeably progressively prominent for games occasions and for client produced content shared over online networking [9]. Live video is commonly caught by a camera (which could be on a cell phone), changed over into at least one coveted organization at foreordained piece rates, and after that either sent specifically to the watchers or first exchanged to at least one cause servers on the cloud. Once a live video sustain is facilitated on an inception server, it is spilled practically like Video on Demand. Commonly live Internet video has a spilling deferral of around 30 seconds because of the time is taken in sending the bolster initially to a cloud server, transcoding and afterward pushing it [10]. RTSP (Real Time Streaming Protocol) and HTTP are the two most famous conventions for spilling live video and VoD. HTTP is broadly upheld over all gadgets and firewalls and henceforth is generally the most favored method for gushing video (e.g., HTTP Live Streaming (HLS)). Both live and VoD utilize ABR (Adaptive Bit Rate) spilling (or HTTP dynamic downloads if there should be an occurrence of HTTP). A video document is first broken into little lumps or portions through a procedure called ABR piecing. Each piece is a little span (regularly under 10 seconds) and is encoded at various piece rates. The customer video player asks for the underlying lump in light of the watched data transfer capacity to the server (there are different approaches to identify the accessible transmission capacity between the customer and the video server). Customary CDNs confront many difficulties in live-streaming and on request video conveyance: CDNs are as of now in charge of a critical part of video quality issues for both live and VoD (counting 20% of join disappointments and 22% of bitrate corruption). Moreover, live video is moving to viral client made streams [11]. Traditional CDNs were implied for serving static documents and website page parts which were commonly pulled by the site guests everywhere throughout the world. A standout amongst the most vital strategies followed in video gushing is QoE. While QoE forecast is effortlessly roused, it remains a troublesome assignment. Demonstrating the impression of video twists is a mind-boggling issue that is exacerbated by an assortment of time Ward behavioral components, for example, regency, which essentially influences the apparent QoE. An assortment of review and persistent time QoE forecast models have been progressed [12].

3. Proposed system

The channel used for video streaming or in the network communication may be a static or dynamic one. It may support circuit-switching or packet-switching, constant or variable bit rate transmission and it may support best support of video streaming or QoS based video streaming. According to the properties of video communication channel and the application used are strongly influence the design of the system. Therefore, in this paper QoS based video streaming is considered in application point of view. In this paper it is aimed to develop a framework which integrates the functionalities like predicting similar frames called as Key Frames, to investigate the QOE measures like memory, VQA, playback status time, bitrate drop and total video duration. The novelty of this paper is increasing the video streaming quality

DOI: 10.54254/2755-2721/6/20230850

by key frame detection and transmits the video by analyzing the QoE. It is proved already that key frame-based video streaming and QoE based video streaming obtained better results. Hence by integrating both key frame detection and QoE analyzation the IF approach will be a promising framework in quality video streaming framework. The entire functionality of the IF approach is depicted in Figure 1. Any sender in the heterogeneous network can transmit video file to any receiver in the network. A complete frame (or full frame) of an image in a video is called as key frame. The other frames containing dissimilar information (changing information) is called as delta or subsequent frames in a video. Key frames are more, and it appears a greater number of times within a specific stream.

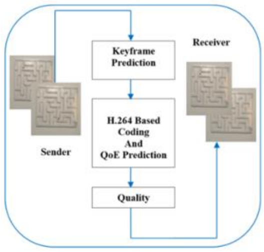

Figure 1. Integrated framework model.

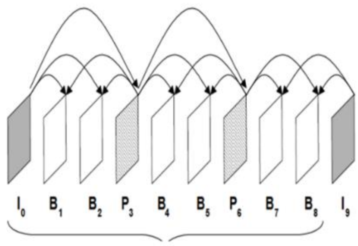

Figure 2. Predicting similar dependent frames.

3.1. Keyframe/similar frame detection

This reduces the time and buffer size effectively. In the receiver side the keyframe is repeatedly streamed according to the number of times the keyframe is available in the video. It is a function avoids redundant frame transmission. The number of keyframes available in the video is set inside the video-encoder. It is assigned by the variable called as “keyframe Interval” and it says that how often the keyframe is created in the video. More number of keyframe, i.e., keyframe interval provides more compression which increases the quality of video streaming in terms of eliminating the noticeable redundancy. For example, the number of keyframe is 80 means, the interval is set as 2 seconds and the frame rate is 40 frames per second. The keyframe availability in a video is shown in Figure 2. The, I, B and P frames are the keyframes available in an interval.

3.2. H.264 based coding

In this paper H.264 method follows inter prediction model where it follows previous encoded frames and form the model from the shifting samples of the reference frame. It takes the block-based motion compensation includes a range of block sizes and fine sub-pixel motion vectors.

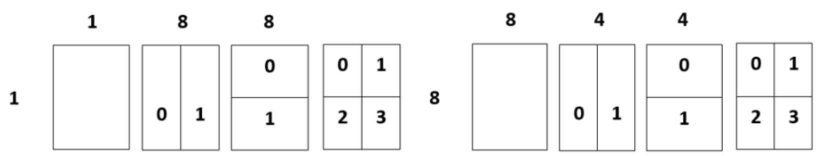

Figure 3. Macro partitions (a). 16 x 16, 8 x 16, 16 x 8, 8 x 8; (b) 8 x 8, 4 x 8, 8 x 4, 4 x 4.

The block size is from 16 x 16 to 4 x 4. The block splitting method is shown in the following Figure 3 and it is represented as macro blocks. This partitioning method increases the large number of possible combinations macroblocks. It is also called as tree-structured motion compensation. A separate motion vector is needed to each partition or to the sub-partition.

3.3. QoE prediction

3.3.1. NARX model. Nonlinear Auto Regressive Exogenous model (NARX) is a nonlinear auto regressive model, and it has exogenous inputs in the time series modelling. This model compares the present values of time series to previous values of the same series and the present and previous values of the exogenous series. This exogenous series is the externally determined series that impacts in the

DOI: 10.54254/2755-2721/6/20230850

significance of series. It also has the term "Error" which is used to associate the truth of the knowledge of others which will not predict the present value of the time series. In the recent research NARX based QoE predictive methods shows the best performance when compared to other available methods. The main aim is to design the predictive method for determining the QoE efficiently in video QoE predictions; hence this NARX method is the good choice for this process. In this method, nonlinear collective of subjective QoE parameters are given as inputs which involves the parameters like video quality, rebuffering traces, and the memory of previous event which affects the QoE previously. Here each input is integrated with the auto regressive function non-linearly. This method is mathematically expressed as:

$$y_t = F\left(y_{t-1}, y_{t-2}, y_{t-3}, \dots u_t, u_{t-1}, u_{t-2}, u_{t-3}, \dots\right) + \in_t \tag{1}$$

In this equation, y is the interest variable and u are considered as the externally determined variable. Here this externally defined variable u is to predict y which is previously done for y itself. Term ε is the error variable used in this equation. Function F is the nonlinear function, and it is used for testing the time series non-linearly. Neural network, wavelet network or sigmoid network can be used for this nonlinear functionality. Artificial Neural Network (ANN) is chosen for the function F because of its best performance. This is the nonlinear function having the previous inputs $\{y_{t-1}, y_{t-2}, y_{t-3}, \dots y_{t-d_y}\}$ and the external variables $\{y_t, y_{t-2}, y_{t-3}, \dots y_{t-d_u}\}$.

4. Experimental results and discussion

In this paper the data is taken from internet which is publicly available (http://www.viratdata.org/). The dataset comprises of two different categories as activities involved by human and activities involved by vehicles. It is designed to be natural, realistic and challenging for video surveillance area with respect to background clutter, various scenes, resolution and human activities and more advanced than the existing datasets. It consists of various advanced features comparing with other datasets as: Realism and natural scenes, Diversity, Quantity, Wide range of resolution and frame rates and Ground and Aerial Videos. The obtained results are given in Figure-5. In the Table 1 Median performance metrics of IF model is compared with VQA model for all 50 test sequences. The best results are shown in boldface. To confirm the video quality the experiment is repeated for 10 times and the average PSNR is calculated. That is the average PSNR is calculated manually from the obtained PSNR in each round of operation. Some parameters are used to assess the quality of the video. They are calculated and compared with the existing approach values such as PSNR, SSIM (Bakre et al. 1995), NIQE (A.C.Bovik et al. 2013), VMAF (Z.Li et al. 2016) and MS-SSIM (Z.Wang et al. 2003). The obtained results from IF approach is compared with the existing VQA+R+M approach is shown in Table-1. The figure 5 depicts that the proposed IF approach has applied

H.264/AVC method-based bit streams of Activity video sequences. The obtained PSNR with respect to the bit rate loss is represented in the Figure 5.

DOI: 10.54254/2755-2721/6/20230850

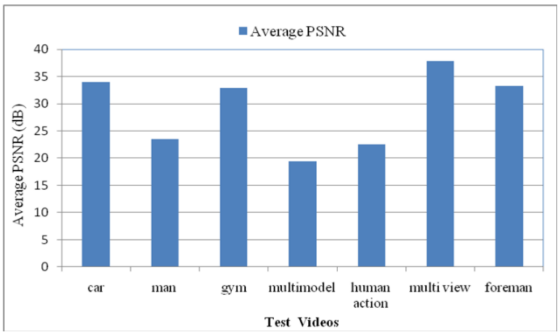

Figure 4. Different videos quality assessment in average PSNR.

Table 1. Median performance metrics.

External Variables	**Existing VQA+R+M [2]**			**Proposed IF**		
Metrics	**RMSE**	**Outage%**	**DTW**	**RMSE**	**Outage%**	**DTW**
PSNR	0.3149	19.2988	25.9719	0.3101	19.0145	24.8430
SSIM	0.2575	14.2081	22.9150	0.2469	14.0109	22.1032
MS-SSIM	0.3326	22.4458	24.3942	0.2985	21.8945	23.7859
NIQE	0.3557	23.4994	31.7299	0.3498	22.8998	31.0321
VMAF	0.3041	16.3525	24.3822	0.4582	29.9994	41.0013

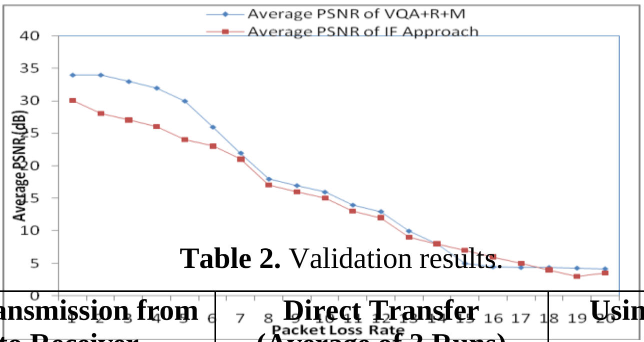

Figure 5. Packet loss versus average PSNR for 50 video sequences.

Table 2. Validation results.

Video file transmission from Sender to Receiver	**Direct Transfer (Average of 3 Runs)**	**Using H.264/AVC Method**
Transfer Time (seconds)	134.87	14.5
Obtained Throughput (Mbps)	4.34	42.43

The packet loss rate is very low, and it proves the highest quality of the video streaming. During the transmission and bit coding the hidden layers, enhanced layer and base layer support the video data not to be lost by optimizing the resource allocation within one another. In this work, since transmission is through key frames the IF approach saves the bit rate and protects the remaining frames in the video. These values are useful to determine the variations in the quality loss of the videos used in this study. The PSNR loss was analyzed to assess the impact of the packet loss on the videos, first in a simulated environment and later in a comparison with the actual losses in the second scenario (validation). The average mean PSNR value of the video is taken to represent the video quality. From the comparison, it has found that the average PSNR obtained using the proposed IF approach is higher than the VQA+R+M approach.

DOI: 10.54254/2755-2721/6/20230850

5. Conclusion

In this framework the entire video coding and streaming fully depends on the H.264/AVC method. In this work the proposed IF approach detects the keyframes and transmits with the help of H.264/AVC approach and with QoE prediction. From the obtained results it is concluded that the proposed IF approach can do video streaming with increased quality of service in terms of PSNR and BER. From the comparison of results and evaluation output it is decided that the proposed IF approach can provide better quality-based video streaming method. In future it can be tried in cloud-based video streaming and the performance can be evaluated.

References

[1] Christos G. Bampis, Zhi Li and Alan C. Bovik, "Continuous Prediction of Streaming Video Qo Eusing Dynamic Networks", IEEE Signal Processing Letters, DOI 10.1109/LSP.2017.2705423.

[2] H. T. Siegelmann, B. G. Horne, and C. L. Giles, "Computational capabilities of recurrent NARX neural networks," IEEE Trans. on Sys. Man, and Cyber., vol. 27, no. 2, pp. 208–215, 1997.

[3] "Ericsson 2013 TV and Media Report," http: // www. ericsson.com / res /docs /2013 /consumer lab /tv-and-mediaconsumerlab-2013.pdf.

[4] J. Jiang, V. Sekar, I. Stoica, and H. Zhang, "Shedding Light on the Structure of Internet Video Quality Problems in the Wild," in ACM CoNext 2013.

[5] "Maximizing Audience Engagement: How online video performance impacts viewer behavior," https://content.akamai.com/PG2114-Audience-Engagement-WP.html.

[6] "ABR Streaming," http://www.tvtechnology.com/cable-satelliteiptv/ 0149/abr-streaming/239523.

[7] C. Bovik, "Automatic prediction of perceptual image and video quality," Proc. IEEE, vol. 101, no. 9, pp. 2008–2024, Sept. 2013.

[8] D. S. Hands and S. Avons, "Recency and duration neglect in subjective assessment of television picture quality," Applied cognitive psychology, vol. 15, no. 6, pp. 639–657, 2001.

[9] S. Tavakoli, S. Egger, M. Seufert, R. Schatz, K. Brunn-strom, and N. Garc´ıa, "Perceptual quality of http adaptive streaming strategies: Cross-experimental analysis of multi-laboratory and crowdsourced subjective studies," IEEE Journal on Selected Areas in Communications, vol. 34, no. 8, pp. 2141–2153, 2016.

[10] R. Soundararajan and A. C. Bovik, "Video quality assessment by reduced reference spatio-temporal entropic differencing," IEEE Transactions on Circuits and Systems for Video Technology, vol. 23, no. 4, pp. 684–694, 2013.

[11] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind prediction of natural video quality," IEEE Transactions on Image Processing, vol. 23, no. 3, pp. 1352–1365, 2014.

[12] Saravanan, S., Kumar, V. V., Sarveshwaran, V., Indirajithu, A., Elangovan, D., & Allayear, S. M. (2022). Computational and Mathematical Methods in Medicine Glioma Brain Tumor Detection and Classification Using Convolutional Neural Network. Computational and Mathematical Methods in Medicine, 2022.