# Video Quality Loss Model on Communication Networks: An Approach Based on Frame Loss

E. M. C. Matos, T. A. Costa, M. C. Alcântara Neto, B. S. L. Castro, F. S. Farias, J. P. L. Araujo and F. J. B. Barros

*Abstract*—This letter proposes a mathematical model that estimates video quality loss in error-prone networks through Peak Signal-to-Noise (PSNR) ratio metric. Also, it presents experimental results by correlating the frame losses, the video resolution, and the received video visual complexity to obtain video quality loss models. The results show that frame loss satisfies the model to predict the PSNR loss for three resolutions.

*Index Terms*—Video quality loss, frame loss, PSNR loss, modeling.

## I. INTRODUCTION

AS widely observed in the literature, packet loss in communication networks directly impacts end-users information [1]–[7]. The flow of multimedia information has been gaining prominence globally, emphasizing the video streams representing a high consumption rate by global customers [8]. Quality of Service (QoS) and Quality of Experience (QoE) metrics are being evaluated to aid video streaming services [9]. In this field, research has been investigating the impact of QoS and QoE on error-prone networks as a means of finding monitoring solutions that can improve the provided services [10], [11].

Therefore, experiments are being conducted to design video quality prediction models taking into account different communication systems. These models generally evaluate video quality in terms of errors during the video reception (packet loss). However, it has been suggested that packet loss is not the primary metric for estimating video streaming systems behavior [3]. For instance, in [2], a predictive video quality loss model was proposed for Digital Television (DTV) transmission. However, experimental results showed variability in video quality for the same packet loss values.

The way of assessing received video by users is mainly carried out subjectively or objectively [9], [12]. The former requires human beings to undergo visual tests and measure the video through the Mean Opinion Score (MOS). The latter employs mathematical models to measure video quality, such as the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). Although relevant,

subjective metrics require many resources, such as training for the participants, video testing, and naked eye analysis. Also, they tend to result in better evaluations if they are influenced by human beings [9]. In this context, some wide-ranging studies are discussing the choice of video quality metric tools for evaluating video quality [9], [10]. For instance, the authors in [4] concluded that PSNR is a reliable metric for measuring end-user satisfaction by demonstrating during tests a close correlation between objective PSNR and subjective MOS metrics. However, as discussed, subjective video quality metrics require additional effort to implement. In comparison, objective metrics are analytical and provide the same degree of reliability as subjective metrics. That way, the PSNR metric will be adopted for the study, removing subjectivity and maintaining reliability.

In papers [3], [9], [13], [14], authors have discussed the correlation between QoS and QoE, checking how robust the video compression standards are in error-prone communication systems. While in [15], authors investigate how to reduce the bandwidth consumption required by the streaming service. However, the impact exerted by frame loss (e.g., *I*, *P*, *B*, and overall frame losses) in user-received video quality is barely researched [16], [17].

Papers presented in the literature show that packet loss rate can be used to determine the perceived degradation of the received video quality and that the *I* frames are the most important to reconstruct the *P* and *B* frames [1]–[7]. However, to the best of our knowledge no study analyzed video quality loss in error-prone networks and correlated the packet losses with frame losses.

This paper investigates video transmission over an error-prone network to develop a mathematical model to predict PSNR losses based on the relationship between PSNR and frame losses of different video contents. The investigation is carried out by considering three resolution videos, i.e., 1280x720, 1920x1080, and 3840x2160 *pixels* for four different videos. The main contributions of the research stand out:

- Definition of linearity between packet loss and frame loss.
- Development of a mathematical model to estimate PSNR loss through overall frames loss in error-prone networks.
- Model evaluation for different resolutions and video contents.

In Section II, the theoretical basis for the proposed model development will be presented. Then, in Section III, results will be presented and discussed, and the model will finally be validated. Finally, in Section IV, the study's conclusions are drawn.

## II. Theory

### A. Peak Signal-to-Noise Ratio

The objective metric, Peak Signal-to-Noise Ratio (PSNR), quantifies users' video quality compared to the original video. It is measured on a logarithmic scale (in $dB$) and depends on another method denominated Mean Squared Error (MSE), which is also calculated for received (estimated) and original video qualities. The PSNR and MSE are defined by Equations 1 and 2, respectively:

$$PSNR = 10\log_{10}\left(\frac{(2^n - 1)^2}{MSE}\right) \qquad (1)$$

where $(2^n - 1)^2$ represents the square of higher possible signal value within the image, and $n$ is the number of bits needed to represent one image of the pixels.

$$MSE = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}[I_{(i,j)} - K_{(i,j)}]^2 \qquad (2)$$

where MSE is the Mean Squared Error, $I_{(i,j)}$ is the matrix that composes a transmitted frame, $K_{(i,j)}$ is the matrix that composes a received frame, $M$ and $N$ are the frame dimensions.

### B. Linear Regression

The model proposed in this paper seeks to estimate PSNR loss as a function of frame loss. Therefore, the logarithmic function is adopted since the estimated value will be the PSNR given in $dB$, and the coefficients calculated using the linear regression function. Equation 3 presents a logarithmic function that utilizes linear regression, and it is the proposed model.

$$L_{PSNR} = \alpha + \beta\log_{10}(OFL) \qquad (3)$$

$L_{PSNR}$ is the predicted PSNR loss calculated from the measured experiments values, $OFL$ is the overall frame loss measured, $\alpha$ and $\beta$ are coefficients.

In order to obtain Equation 3, is used the Least Squares Method [18], which expressed as:

$$A = \begin{bmatrix} 1 & log_{10}(OFL) \\ 1 & log_{10}(OFL) \end{bmatrix}; \ coef = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}; \ B = \begin{bmatrix} L_1 \\ L_2 \end{bmatrix};$$
$$coef = (A'A)^{-1}A'B \qquad (4)$$

where $\alpha$ and $\beta$ coefficients are obtained for resolution (720p, 1080p, and 2160p). $L_1$ and $L_2$ are the calculated PSRN loss values.

## III. Results

This section presents the video coding configuration and the methodology adopted to obtain the data and its classification according to the visual video complexity. Furthermore, the linearity between packet loss and overall frame loss is found by calculating the correlation to indicate the predictor variable of the proposed model. Finally, the mathematical model of video quality for each resolution and type of video is presented.

### A. Video Transmission Over Error-Prone Network

The impact of frame loss on PSNR loss is demonstrated by employing four videos obtained from Xiph.org[1], designated as: "Crowd Run", "Ducks Take Off", "Park Joy" and "Into Tree" for three spatial resolutions: *720p* (1280x720 pixel), *1080p* (1920x1080 pixel) and *2160p* (3840x2160 pixel).

The videos are encoded with H264/AVC and have a GOP (Group of Picture) size of 25 and a *bitrate* of 32 *Mbps* [13]. Each video has 500 frames, with a duration of 10 seconds. The GOP consists of one *I* frame, twelve *P* frames, and twelve *B* frames. Thus, every video sequence comprises 20 *I* frames, 240 *P* frames, and 240 *B* frames, making 500 frames. Its coding parameters are identical, and it will allow the evaluation of the performance of the proposed model to different video complexities.

In this investigation, 50 simulations are carried out to assess PSNR (in dB), and frame loss values (in %). During simulations, twelve videos are transmitted over an error-prone network. Also, it is used a random packet loss ranging from 0.5% to 10% with a 0.5% step size. The proposed setup applied the probability density function with uniform distribution for generating packet loss in video transmission [19].

For each packet loss variation, the PSNR value for each video is obtained. In order to evaluate PSNR, Table I presents the reference values to compute the PSNR losses, which corresponds to the values PSNR of videos without any transmission loss factors, i.e., original PSNR value. The PSNR losses are computed by comparing the original PSNR value and the PSNR values retrieved from the degraded videos. Equation 5 is responsible for computing the PSNR loss ($L$) for each video sequence.

TABLE I
REFERENCE PSNR VALUES PER VIDEO.

| Video | PSNR ($dB$) | | |
|---|---|---|---|
| | 720p | 1080p | 2160p |
| Crowd Run | 39.21 | 36.67 | 32.72 |
| Ducks Take Off | 36.96 | 31.61 | 28.58 |
| Park Joy | 37.93 | 32.48 | 31.16 |
| Into Tree | 41.67 | 37.61 | 35.16 |

$$L = L_{ref} - L_{sim} \qquad (5)$$

where $L_{ref}$ are the original video PSNR value (Table I), and $L_{sim}$ is the PSNR degraded by packet loss simulation. Next, the PSNR loss ($L$) is calculated for each video sequence. Among all degradation rates, some samples vary widely from the average values, and these are considered in the literature as outliers [10]. Due to that, they are removed during the data selection stage. After removing outliers from the data, the videos' PSNR loss values (as expressed in Equation 5) and their respective resolutions are calculated, as shown in Fig. 1. It should be pointed out "crowd" is a reference to the video "Crowd Run", "duck" to "Ducks Take Off", "park" to "Park Joy", and "tree" to "Into Tree".
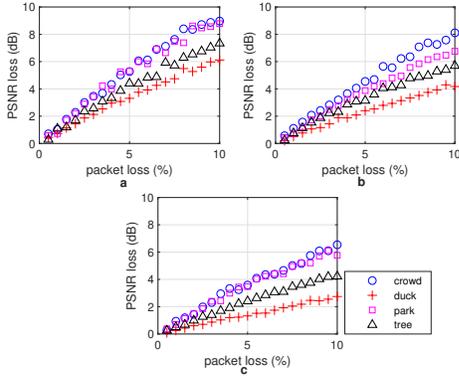
[1]https://media.xiph.org/video/derf/

Fig. 1. Video degradation with no outliers for resolution: *a* the 1280x720 *pixel b* the 1920x1080 *pixel c* the 3840x2160 *pixel*



Fig. 3. Correlation between *I*, *P*, *B*, overall frame loss and packet loss for resolution: *a* the 1280x720 *pixel b* the 1920x1080 *pixel c* the 3840x2160 *pixel*

The results show that the 3840x2160 pixel spatial resolution is the least affected by packet loss (see Fig. 1c) because its curves have a lower slope than the other resolutions.

The metric used to classify the visual video complexity are Spatial Information (SI) and Temporal Information (TI) assessment values, which are metrics recommended by ITU TREC-P.910. The videos' SI and TI values are shown in Fig. 2, where the higher the SI and TI values, the greater the visual complexity of the video [20]. Also, the PSNR loss values become more significant when SI and TI values are higher. Conversely, the videos with the lowest degradation have the lower visual complexity, i.e., "Duck Take Off" and "Into Tree since".
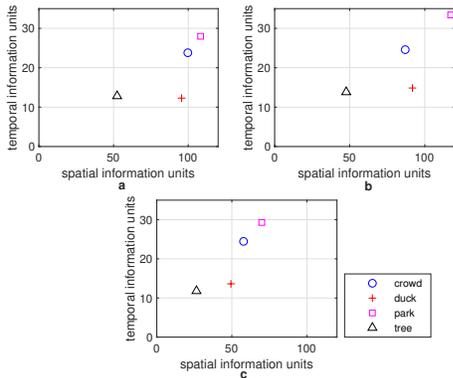


Fig. 2. Spatial Information (SI) and Temporal Information (TI) for resolution: *a* the 1280x720 *pixel b* the 1920x1080 *pixel c* the 3840x2160 *pixel*

### B. Correlation Between Packet Loss and Frame Loss

Fig. 3 presents the frame loss values for I, P, B, and the overall frames calculated (in %) for each video sequence. Overall, the frames comprise the sum of the *I*, *P*, and *B* frames. In Fig. 3, it can be seen that the loss of *I* frames are more widely dispersed than *B* and *P* – this behavior is due to a lower compression factor applied to the *I* frames by the H264/AVC codec. Generally, the frame loss behavior displayed in Fig. 3 shows a packet loss correlation, regardless of the frame type.
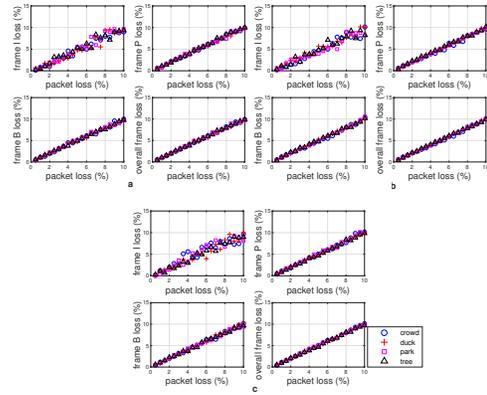
The correlation coefficient is calculated to find the existence of linearity between the packet loss and the frame loss. Table II shows the Pearson correlation coefficient between packet losses and frame losses, calculated to establish the relationship between frame losses and perceived video degradation (PSNR loss), i.e., concerning *I*, *P*, *B*, and overall frames.

TABLE II
CORRELATION COEFFICIENTS BETWEEN PACKET LOSS AND FRAME LOSS
PER RESOLUTION.

| Resolution | Name video | Frame I | Frame P | Frame B | Overall Frame |
|---|---|---|---|---|---|
| 720p | Crow | 0.9859 | 0.9982 | 0.9961 | 0.9990 |
| | Ducks | 0.9695 | 0.9979 | 0.9977 | 0.9985 |
| | Park | 0.9821 | 0.9945 | 0.9961 | 0.9965 |
| | Tree | 0.9774 | 0.9982 | 0.9990 | 0.9993 |
| 1080p | Crow | 0.9775 | 0.9950 | 0.9970 | 0.9977 |
| | Ducks | 0.9861 | 0.9983 | 0.9991 | 0.9969 |
| | Park | 0.9697 | 0.9996 | 0.9986 | 0.9971 |
| | Tree | 0.9710 | 0.9978 | 0.9978 | 0.9975 |
| 2160p | Crow | 0.9466 | 0.9971 | 0.9977 | 0.9992 |
| | Ducks | 0.9699 | 0.9988 | 0.9969 | 0.9992 |
| | Park | 0.9637 | 0.9962 | 0.9971 | 0.9988 |
| | Tree | 0.9885 | 0.9986 | 0.9975 | 0.9992 |

Considering Table II and performing data analysis of the Pearson correlation coefficient, it is confirmed that *I*, *P*, *B*, and overall frame loss hold very proximate degrees of influence (above 0.9) for all three resolutions utilized in the study. This means that *I*, *P*, *B*, and the overall frame losses have a strong linear relationship with packet losses. Hence, the relationship between frame loss and packet loss is proven, which justifies the adoption of frame loss as the model-independent variable. Therefore, in numerical terms, the overall frame loss keeps the most significant correlation and is chosen to design the video quality loss model.

### C. Modeling PSNR loss and Overall Frame Loss

Following an analysis carried out previously, a linear regression based on the Least Squares Method (Equation 4) technique is employed to predict PSNR loss (Equation 3) in terms of the video sequence in terms of percentage of Overall Frame Loss (OFL). For that, the PSNR metric with a

logarithmic base function is adopted to evaluate video quality, as seen in [13]. The $OFL$ values will be used in matrix $A$, and the $L$ (Equation 5) values will be used in Matrix $B$ to calculate the $\alpha$ and $\beta$ coefficients ($coef$) of Equation 4. The results are displayed in Fig. 4, and the equations used for the calculations are shown in Table III.
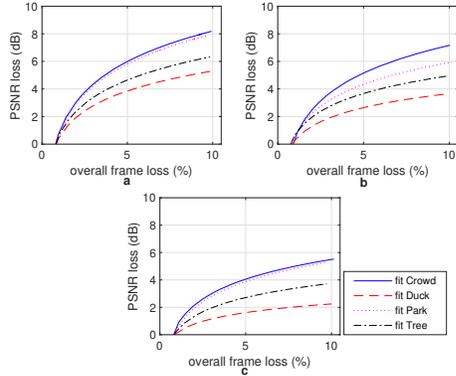


Fig. 4. Fitness values of video sequences for resolution $a$ the 720p $b$ the 1080p $c$ the 2160p

TABLE III
PSNR LOSS MODELS PER RESOLUTION.

|  |  |  |
|---|---|---|
| **720p** | Crowd Run | $0.84 + 7.37\log_{10}(OFL)$ |
|  | Park Joy | $0.79 + 7.18\log_{10}(OFL)$ |
|  | Ducks Take Off | $0.49 + 4.82\log_{10}(OFL)$ |
|  | Into Tree | $0.67 + 5.69\log_{10}(OFL)$ |
| **1080p** | Crowd Run | $0.49 + 6.67\log_{10}(OFL)$ |
|  | Park Joy | $0.66 + 5.26\log_{10}(OFL)$ |
|  | Ducks Take Off | $0.28 + 3.38\log_{10}(OFL)$ |
|  | Into Tree | $0.47 + 4.27\log_{10}(OFL)$ |
| **2160p** | Crowd Run | $0.71 + 4.78\log_{10}(OFL)$ |
|  | Park Joy | $0.48 + 4.86\log_{10}(OFL)$ |
|  | Ducks Take Off | $0.15 + 2.1\log_{10}(OFL)$ |
|  | Into Tree | $0.31 + 3.44\log_{10}(OFL)$ |

Note that in Table III, as the resolution increases, the slope coefficient $\beta$ decreases due to higher resolution robustness. Therefore, higher resolution videos lose less video quality as the overall frame loss increases.

As previously mentioned about the complexity of the videos through SI and TI metrics (see Fig. 2) and the equations formulated for the designed model (see Table III), we classify the videos "crowd " and "park " as more movement. In comparison, "duck " and "tree " refer to less movement. This classification helps define the model limits for each resolution on different videos type since there is no overlap of the coefficient $\alpha$ and $\beta$ for videos with more and less movement. The coefficient limits of Equation 3 are shown in Table IV and allow the videos to be modeled with similar visual complexity by depending on each resolution.

In order to verify the model performance, the Root Mean Square Error (RMSE) is computed for each video sequence by considering the predicted data and the measured data. Finally, the results are listed in Table V, confirming the approach's accuracy. Furthermore, the RMSE values obtained for all videos and resolutions were lower than a similar study found in [13]. Therefore, it is proven that adopting the video visual

TABLE IV
LIMITS FOR COEFFICIENTS $\alpha$ AND $\beta$ PER RESOLUTION AND VIDEO COMPLEXITY.

|  | More movement | Less movement |
|---|---|---|
| **720p** | $0.79 \leq \alpha \leq 0.84$ $7.18 \leq \beta \leq 7.37$ | $0.49 \leq \alpha \leq 0.67$ $4.82 \leq \beta \leq 5.69$ |
| **1080p** | $0.49 \leq \alpha \leq 0.66$ $4.27 \leq \beta \leq 6.67$ | $0.28 \leq \alpha \leq 0.47$ $3.38 \leq \beta \leq 4.27$ |
| **2160p** | $0.48 \leq \alpha \leq 0.71$ $4.78 \leq \beta \leq 4.86$ | $0.15 \leq \alpha \leq 0.31$ $2.11 \leq \beta \leq 3.44$ |

complexity classification system using SI and IT increases the accuracy of video quality prediction modeling.

TABLE V
RMSE VALUES (IN $dB$) FOR ALL VIDEO SEQUENCES.

|  | 720p | 1080p | 2160p |
|---|---|---|---|
| **Crowd Run** | 0.7978 | 0.6633 | 0.6005 |
| **Park Joy** | 0.7675 | 0.6099 | 0.5063 |
| **Ducks Take Off** | 0.5676 | 0.4039 | 0.2833 |
| **Into Tree** | 0.6811 | 0.5278 | 0.3916 |

The designed model performance has proven more accurate for videos with less movement, e.g., "Ducks Take Off" and "Into Tree", which have lower RMSE values at each resolution. The resolutions at *2160p* obtained the lowest values of RMSE values and proved to be the most accurate. However, in general, all the RMSE values obtained are satisfactory and below $0.8$ $dB$. As it is a simple mathematical model, the additional computational effort is not required for its implementation.

## IV. CONCLUSION

A mathematical model to predict PSNR loss is developed for transmission videos in error-prone networks through the existing linearity between packet loss and frame loss. The correlation between packet losses and frame losses has been proven, demonstrating that the model can be implemented for the application layer without monitoring any other layer. It is also evident that the video visual complexity (more movement and less movement) and the video spatial resolution cause variations in the PSNR loss. Thus, to increase the model's accuracy, a pre-classification on video visual complexity is realized, using the SI and IT metrics, and evaluating the PSNR loss by resolution.

It is concluded that the mathematical model proposed to predict PSNR loss presents good results. Furthermore, the 2160p resolution is the most robust for video quality loss for fixed bitrate. Furthermore, videos with less visual complexity are also the ones that show less PSNR loss. Finally, from this study is possible to conclude that GOP and bitrate variations must be considered to improve the future studies proposed model.

## REFERENCES

[1] F. D. Simone, M. Naccari, M. Tagliasacchi, F. Dufaux, S. Tubaro, and T. Ebrahimi, "Subjective quality assessment of h.264/AVC video streaming with packet losses," *EURASIP Journal on Image and Video Processing*, vol. 2011, pp. 1–12, 2011. [Online]. Available: https://doi.org/10.1155/2011/190431

[2] J. Joskowicz and R. Sotelo, "A model for video quality assessment considering packet loss for broadcast digital television coded in h.264," *International Journal of Digital Multimedia Broadcasting*, vol. 2014, pp. 1–11, 2014. [Online]. Available: https://doi.org/10.1155/2014/242531

[3] J. Frnda, M. Voznak, and L. Sevcik, "Impact of packet loss and delay variation on the quality of real-time video streaming," *Telecommunication Systems*, vol. 62, no. 2, pp. 265–275, apr 2015. [Online]. Available: https://doi.org/10.1007/s11235-015-0037-2

[4] B. Bondzulic, V. Petrovic, B. Pavlovic, and M. Andric, "Performance of peak signal-to-noise ratio quality assessment in video streaming with packet losses," *Electronics Letters*, vol. 52, no. 6, pp. 454–456, mar 2016. [Online]. Available: https://doi.org/10.1049/el.2015.3784

[5] J. Wang and Y. bin Hou, "Packet loss rate mapped to the quality of experience," *Multimedia Tools and Applications*, vol. 77, no. 1, pp. 387–422, dec 2016. [Online]. Available: https://doi.org/10.1007/s11042-016-4254-9

[6] A. Adeyemi-Ejeye, M. Alreshoodi, L. Al-Jobouri, M. Fleury, and J. Woods, "Packet loss visibility across SD, HD, 3d, and UHD video streams," *Journal of Visual Communication and Image Representation*, vol. 45, pp. 95–106, may 2017. [Online]. Available: https://doi.org/10.1016/j.jvcir.2017.02.012

[7] A. O. Adeyemi-Ejeye, M. Alreshoodi, L. Al-Jobouri, and M. Fleury, "Impact of packet loss on 4k UHD video for portable devices," *Multimedia Tools and Applications*, vol. 78, no. 22, pp. 31 733–31 755, jul 2019. [Online]. Available: https://doi.org/10.1007/s11042-019-07996-1

[8] CISCO, "Cisco visual networking index: Forecast and trends, 2017–2022," Cisco Visual Networking Index, Tech. Rep., 2018.

[9] Y. Chen, K. Wu, and Q. Zhang, "From QoS to QoE: A tutorial on video quality assessment," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 1126–1165, 2015. [Online]. Available: https://doi.org/10.1109/comst.2014.2363139

[10] P. Juluri, V. Tamarapalli, and D. Medhi, "Measurement of quality of experience of video-on-demand services: A survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 401–418, 2016. [Online]. Available: https://doi.org/10.1109/comst.2015.2401424

[11] D. Zhang, D. Zhou, and X. Jin, "A content-adaptive video quality assessment method for online media service," *Multimedia Tools and Applications*, vol. 76, no. 5, pp. 7175–7195, mar 2016. [Online]. Available: https://doi.org/10.1007/s11042-016-3359-5

[12] M. Uhrina, J. Hlubik, and M. Vaculik, "Correlation between objective and subjective methods used for video quality evaluation," in *2012 ELEKTRO*. IEEE, may 2012. [Online]. Available: https://doi.org/10.1109/elektro.2012.6225581

[13] J. V. C. Carmona, E. M. C. de Matos, B. S. L. Castro, F. J. B. Barros, M. C. de Alcântara Neto, and E. G. Pelaes, "Video loss prediction model in wireless networks," *PLOS ONE*, vol. 14, no. 3, p. e0212407, mar 2019. [Online]. Available: https://doi.org/10.1371/journal.pone.0212407

[14] J. Frnda, J. Nedoma, J. Vanus, and R. Martinek, "A hybrid QoS-QoE estimation system for IPTV service," *Electronics*, vol. 8, no. 5, p. 585, may 2019. [Online]. Available: https://doi.org/10.3390/electronics8050585

[15] A. Kalampogia and P. Koutsakis, "H.264 and h.265 video bandwidth prediction," *IEEE Transactions on Multimedia*, vol. 20, no. 1, pp. 171–182, jan 2018. [Online]. Available: https://doi.org/10.1109/tmm.2017.2713642

[16] P. Orosz, T. Skopko, and P. Varga, "Towards estimating video QoE based on frame loss statistics of the video streams," in *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*. IEEE, may 2015. [Online]. Available: https://doi.org/10.1109/inm.2015.7140482

[17] T. Zhao, Q. Liu, and C. W. Chen, "QoE in video transmission: A user experience-driven strategy," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 285–302, 2017. [Online]. Available: https://doi.org/10.1109/comst.2016.2619982

[18] J. Wolberg, *Data Analysis Using the Method of Least Squares*. Springer-Verlag GmbH, Feb. 2006.

[19] R. Jain, *The Art of Computer Systems Performance Analysis*. John Wiley & Sons, 1991.

[20] A. Ostaszewska and R. Kłoda, "Quantifying the amount of spatial and temporal information in video test sequences," in *Recent Advances in Mechatronics*. Springer Berlin Heidelberg, pp. 11–15. [Online]. Available: https://doi.org/10.1007/978-3-540-73956-2_3