A NOVEL CUMULATIVE DISTORTION METRIC AND A NO-REFERENCE SPARSE PREDICTION MODEL FOR PACKET PRIORITIZATION IN ENCODED VIDEO TRANSMISSION

Arun Sankisa[1], Katerina Pandremmenou[2], Lisimachos. P. Kondi[2], Aggelos K. Katsaggelos[1]

[1] EECS Department, Northwestern University, Evanston, IL, USA
[2] Department of Computer Science and Engineering, University of Ioannina, Ioannina, Greece

ABSTRACT

In this paper we propose a new quality metric to estimate the impact of packet loss on the perceptual quality of encoded video sequences transmitted over error-prone networks. The proposed metric, henceforth referred to as Cumulative Distortion using Structural Similarity (CDSSIM), quantifies the overall structural distortion resulting from bidirectional error propagation in predictively coded, motion compensated videos. Furthermore, we present a No-Reference (NR) sparse regression model to predict the proposed CDSSIM metric using pre-defined features associated with slice loss. The Least Absolute Shrinkage and Selection Operator (LASSO) method is applied for two resolution formats with features extracted solely from the encoded bit-stream. Standardized statistical performance measures show that the model can predict the cumulative distortion to a high degree of accuracy. We further evaluate the results using a Quartile-Based Prioritization (QBP) scheme and demonstrate that the predicted data provides an effective way to prioritize packets for video streaming applications.

Index Terms— Cumulative distortion, Structural Similarity, Video quality, LASSO, Packet prioritization

1. INTRODUCTION

Videos have become a ubiquitous part of our daily lives, as have devices and applications that capture, process and transmit them. Today, even the most basic mobile terminals contain sophisticated video processing elements creating a dramatic increase in demand for streaming services. This demand steadily pushes the boundaries of multimedia research and underscores the need for efficient algorithms that provide optimal end-user quality while taking into account capacity constraints like storage and bandwidth. The overall user experience and quality is influenced by many factors but notably by compression and transmission impairments. Research in video codecs has moved at a fast pace through standards like H.263 [1], MPEG4-Part2 [2] to H.264/MPEG4-AVC [3], H.264 SVC [4] and H.265/HEVC [5]. Similarly, wireless communication has also made rapid strides with 3G UMTS, High Speed Packet Access (HS-DPA/UPA), WiMax, 4G/LTE, and plans to introduce 5G before the end of this decade [6]-[9]. While these advances help address the growing demand for video streaming services, there is also a need for innovative techniques that offer complete end-to-end solutions.

Compression techniques are inherently lossy, creating artifacts that directly relate to degradation in quality. During the encoding process, video sequences are broken into frames and different coding modes are applied on their constituent units, macroblocks (MBs) and Group of Blocks (GOBs). The decision about coding modes usually depends on the frame in which a block resides and a natural outcome of differentiated coding is the creation of data units with unequal importance, a key motivation for defining a packet prioritization scheme. Additionally, temporal, motion-compensated prediction commonly used by encoders leads to inter-frame dependence and error propagation that needs to be taken into account when designing such a scheme.

With universal high-speed radio access, video streaming over wireless networks is a widely consumed service. When an encoded sequence is ready for transmission, usually over a resource-constrained, loss-prone channel, it is broken and packaged into units that each contains a portion of a video frame (for instance, a GOB). All packets belonging to a frame need to be correctly received for error-free reconstruction at the decoder. But if some packets are lost, data can be recovered by applying the appropriate error-concealment technique and this is usually accompanied by propagation of errors between frames. Recent research (e.g. [10]-[15]) has shown that transmission methods that utilize cross-layer, content-aware resource allocation, packet ordering and prioritization schemes achieve improved end-user experience than content-agnostic ones.

Assessing quality on the source side is essential in designing a system that prioritizes packets for transmission. Traditionally used metrics such as Mean-Squared Error (MSE) and Cumulative MSE (CMSE) that measure quality as distortion or PSNR that measures comparative signal strength are objective mathematical models used in analytical and computational techniques such as Rate-Distortion Optimization (RDO). But these metrics suffer from the fact that they do not correlate well to perceptual quality [16]. More recently, the Structural SIMilarity index (SSIM) [17], modelled after the operation of the human visual system, has been shown to provide a better evaluation of perceived end-user experience and can be easily incorporated in motion estimation, mode selection and RDO algorithms. Additionally, localized quality information provided by SSIM is particularly attractive for applications that study the impact of sub-frame units (such as MBs, GOBs etc.) commonly used during transmission. Other perceptive quality measures like Video Quality Metric (VQM), Motion-based Video Integrity Evaluation (MOVIE), Just Noticeable Difference (JND), Perceptual Distortion Metric (PDM) and Digital Video Quality (DVQ) [18]-[22], although effective, are computationally complex for real-time resource allocation and low bit-rate encoding scenarios.

Over the past few years, there has been a lot of research in developing methods that estimate perceptive quality of encoded videos but these suffer from certain limitations that the proposed model intends to overcome. The work presented in [23] evaluates the effect of transmission errors on video quality without utilizing metrics such as MSE or SSIM. Instead, it approaches the issue from the visibility of a lost slice on the entire frame using factors such as scene cuts, motion and distance-to-reference. A Generalized Linear Model (GLM) that predicts quality degradation contributed by individual slice loss in H.264/AVC encoded videos is proposed in [24],[25]. However, in these two papers, along with the work presented in [26], Reduced-Reference (RR) features that depend on access to information in the original video for predicting CMSE are employed. All these models, as mentioned previously, are based on quality metrics that do not provide an accurate measure of the perceptual quality as experienced by the end user.

The main contribution of this paper is a content-aware packet prioritization framework for transmission of encoded video over lossy networks. As a first step, we develop a mechanism to measure the overall degradation in perceived quality due to packet loss by defining a new metric based on the widely adopted SSIM index [17]. Although SSIM was originally defined for still images, we modify its usage to evaluate the distortion in individual video frames and extend it bi-directionally between dependent frames to obtain the total distortion. This is accomplished by sequentially removing one slice from each frame and comparing the error-concealed, reconstructed frame with the compressed original. Since this operation is performed for every slice in every frame, it is computationally very intensive. To circumvent the overhead of per-slice computation in real-time applications, we provide a NR linear regression model using LASSO to predict the measured distortion using key features that are specifically related to slice loss. We show that the predicted values are strongly correlated to the actual measurements and can be used effectively in packet prioritization schemes. We evaluate the efficiency of the results using a simple prioritization method called Quartile-Based Prioritization (QBP) [26] and study the performance of distributing packets into four priority groups using both the measured and predicted values.

The organization of the rest of the paper is as follows. In Section 2 we present the proposed CDSSIM metric and describe the mechanism to compute the metric for encoded video. Section 3 describes the features we extract from the encoded bit-stream, the LASSO regression model, and the QBP method. Experimental results from comparing the measured with predicted values and the prioritization efficiency of the proposed model are presented in Section 4. We conclude the paper in Section 5.

2. PROPOSED CUMULATIVE DISTORTION METRIC

The Structural Similarity Index [17] compares two images and quantifies the similarity between them as viewed through three quality measures – luminance, contrast, and structure. This metric has been widely used in multimedia signal processing, especially in assessing decoding and reconstruction quality of images and videos [27]-[29] that experience distortion from compression artifacts, lost slices or other error concealment.

A detailed description of the metric and its usage is available in [17] and we will skip the details for conciseness. Using this metric, the corresponding distortion is defined as,

$$DSSIM = 1 - SSIM. \tag{1}$$

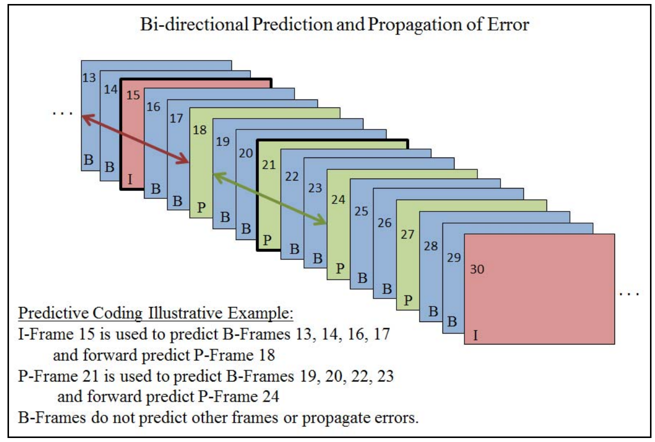


Figure 1. GOP Structure for Video Encoding

In motion-compensated, predictively coded video sequences, the overall impact of error propagation to multiple frames is better captured through a summation of per-frame distortion than a summation of the similarity. We therefore define the proposed cumulative metric as the sum of distortion in the current frame and induced distortion in dependent frames as a result of slice loss.

The Group of Pictures (GOP) structure used in our work is IBBP of size 16. I-frames are used to predict two B-frames in the previous GOP along with the next two B-frames and the first P-frame in the current GOP. A P-frame is used to predict the previous two B-frames, subsequent two B-frames and the next P-frame. B-frames are not used to predict any other frames. This bi-directional prediction results in the propagation of errors between these frames. An example of the prediction and error propagation process using frame numbers is illustrated in Figure 1.

In this work, we assume that a row of MBs forms a slice that is packetized as a transmission unit to be sent over a network. The cumulative distortion due to lost packet n in frame i using SSIM is then given by the following equations. Eq. (2) applies to I- and P-frames and Eq. (3) applies to B-frames.

$$CDSSIM_{n,i}^{I,P} = \sum_{k=i-2}^{LF} \left(1 - SSIM_k\right) \tag{2}$$

$$CDSSIM_{n,i}^{B} = (1 - SSIM_i) \tag{3}$$

where $SSIM_j$ is the Structural Similarity of frame j between the uncorrupted encoded video sequence and the packet-loss-impaired sequence and LF is the frame number of the "last frame" in the GOP. It is clear that I-frames incur higher CDSSIM than P- and B-frames, and P-frames have higher CDSSIM than B-frames. It should be noted that the CDSSIM computation given above is made for every slice of each frame in a sequence.

3. FEATURE EXTRACTION, REGRESSION FRAMEWORK AND PACKET PRIORITIZATION

We continue the presentation with a discussion on the extracted features accompanied by a brief description of NR sparse linear regression using LASSO and the scheme we used to test packet prioritization for video transmission.

3.1 Feature Extraction

The following network features associated with transmission and slice loss have been used in our work.

- **TD** represents the Temporal Duration i.e. number of frames affected by a slice loss.
- **FrameCenter** is a Boolean set to true if a slice lies in the center of a frame i.e. if it is one of six slices in the middle of a CIF frame or one of 12 slices in the middle of a 4CIF frame.
- **DistToRef** refers to the distance the current slice/frame is from the reference frame from which it is concealed.
- **FarConceal** is a Boolean set to true if DistToRef is greater than or equal to 3.
- **SBM** is the Slice Boundary Mismatch [30].
- **MeanResEngy, MaxResEngy** are mean and maximum values of residual energy. Residual energy is the sum of the squares of the motion-compensated transform coefficients taken over all the MBs in a slice. High residual energy implies that the slice captures high degree of motion.
- **SigMean, SigVar** are the signal mean and variance, respectively, of the Y-component (luminance) of the slice.
- **DMVX, DMVY** are the average motion vector difference of a slice in the x and y axes.
- **absMVX, absMVY** are the average motion vector values of a slice along the x and y axes.

While features related to packet loss are evaluated at the slice level, motion related features, DMVX, DMVY, absMVX and absMVY, are computed in the context of a macroblock and hence, they are averaged over an entire slice.

3.2 Regression Framework

Least Absolute Shrinkage and Selection Operator (LASSO), originally proposed in [31],[32], is a linear regression analysis tool that helps solve ill-posed multi-variable estimation problems and provides sparse, interpretable solutions. It is used to estimate both regression coefficients and response variables and is especially useful in cases where the objective is to shrink a broader set of features to a smaller set to improve estimation accuracy and eliminate "prediction noise".

The central idea behind LASSO is in minimizing the square of the input-output residuals with an additional constraint imposed through the sum of the absolute values of the regression coefficients, given by,

$$\min_{\beta,\beta_0} \left\{ \frac{1}{2p} \sum_{i=1}^{p} \left(y_i - \beta_0 - \beta^T x_i \right)^2 + \lambda \|\beta\|_1 \right\} \quad (4)$$

where p is the total number of observations, i.e. the total number of slices we have examined for our experiments, y_i is the measured CDSSIM of slice i and x_i is a vector that contains the values of all examined features for slice i. The term β is the vector of the regression coefficients, β_0 is the intercept term, and the regularization coefficient is the term λ.

LASSO performs covariate selection while also shrinking the number of coefficients through the l_1 norm regularization term that forces some of the coefficients to take zero values. As the value of regularization coefficient λ increases, i.e., as the penalty increases, the number of coefficients that take a zero value also increases.

3.3 Packet Prioritization

We propose to use a Quartile-Based Prioritization scheme [26] to evaluate the prediction efficiency of our method and its reliability for prioritizing packets for video transmission. Towards this goal, we group all the CDSSIM values from the test sequences into four different priority categories. The groups are formed by sorting the values from the lowest to highest CDSSIM, taking the median of the sorted set and then the median of the resulting halves giving the 25^{th}, 50^{th}, and 75^{th} percentile breakdown of the full range of CDSSIM values.

Next we assign the highest priority (Priority 1) to the group of slices with CDSSIM values that fall above the 75^{th} quartile, Priority 2 to the set of slices with CDSSIM in the 50^{th}-75^{th} quartile, Priority 3 is for slices in 25^{th}-50^{th} quartile and the lowest Priority 4 is assigned to slices that fall inside the 25^{th} quartile. This exercise is performed on both the computed and predicted values of CDSSIM.

After the various groups are formed, we compare the sets of slices that fall into each priority category for both the measured and predicted CDSSIM values. A slice is considered "misclassified" if it is placed in a priority group that is different in the predicted set as compared to the measured one. The percentage of slices that are misclassified using the predicted model indicates the level of divergence from the measured-value-based prioritization. This QBP-based comparison has been shown to provide an accurate evaluation of the prediction mechanism and its effectiveness as a foundation for defining a prioritization scheme.

4. EXPERIMENTAL RESULTS

For our experiments we used four CIF sequences *(foreman, hall, mobile* and *paris*) and four 4CIF sequences *(crowdrun, harbour, ice,* and *soccer*). These sequences were encoded using H.264/AVC variable bit-rate encoding. Of the four CIF sequences, *foreman, hall* and *mobile* were used for training and *paris* was used as the test sequence. Correspondingly, for 4CIF, we used *crowdrun, ice,* and *soccer* as our training set and *harbour* for testing. The first 100 frames of each sequence were used to conduct the experiments, with a CIF frame containing 18 slices and 4CIF containing 36 slices. As discussed earlier, one slice from each frame is dropped sequentially and the corresponding CDSSIM is computed. This results in a total of 1800 x 4 = 7200 CDSSIM values for the CIF sequences, of which 5400 (75%) were used for training and 1800 (25%) were used for comparison with the LASSO predictions. Similarly, for the 4CIF case, we obtain a total of 3600 x 4 = 14400 CDSSIM measurements, where 10800 (75%) were used for training and 3600 (25%) for test.

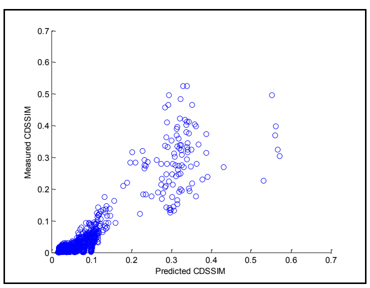
The features that were used as input into LASSO and the corresponding sparse coefficients it generates along with the intercept term and the λ value are shown in Table 1. The data indicate that a fairly sparse representation, four features for the CIF case and three features for 4CIF, are able to predict CDSSIM to a high level of accuracy. This is evident from Table 2, which presents some standard statistical performance measures, i.e. the Pearson

Correlation Coefficient (PCC), Spearman Rank Ordering Correlation Coefficient (SROCC), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) [33]-[35] that highlight the effectiveness of the employed prediction model. These results are representative of our experiments using different combinations of training and test sequences and reflect an optimal compromise between prediction accuracy and number of selected features. In view of compactness of this presentation, we present results that provide a larger reduction in computational complexity through the use of fewer features.

The scatter plots for CIF and 4CIF test sequences in Fig. 2 show high degree of correlation between the predicted and measured values indicating that the prediction model provides a reliable alternative to the overhead of per-slice CDSSIM computations.

Features	CIF	4CIF
TD	0.0314	0.0132
FrameCenter	0	0
DistToRef	0	0
FarConceal	0	0
SBM	0	0
MeanResEngy	0.0412	0.0121
MaxResEngy	0.0025	0
SigMean	-0.0008	0
SigVar	0	0
DMVX	0	0
DMVY	0	0
absMVX	0	0
absMVY	0	0.0003
Intercept	0.0534	0.0183
λ	0.0085	0.0041

Table 1. Linear Regression Coefficients



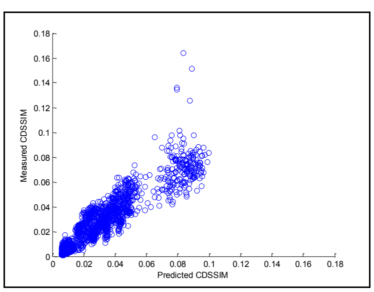


Figure 2. Scatter Plots of Predicted vs. Measured CDSSIM for CIF (top) and 4CIF (bottom) Test Sequences

	CIF	4CIF
PCC	0.9141	0.9466
SROCC	0.8220	0.8601
RMSE	0.0399	0.0077
MAE	0.0287	0.0058

Table 2. Standardized Performance Metrics

Misclassification	CIF	4CIF
$1^{st} \rightarrow 2^{nd}$	2.9%	3.1%
$1^{st} \rightarrow 3^{rd}$	1.5%	0.0%
$1^{st} \rightarrow 4^{th}$	1.1%	0.0%
$2^{nd} \rightarrow 1^{st}$	5.4%	3.1%
$2^{nd} \rightarrow 3^{rd}$	4.4%	7.4%
$2^{nd} \rightarrow 4^{th}$	2.3%	1.3%
$3^{rd} \rightarrow 1^{st}$	0.1%	0.0%
$3^{rd} \rightarrow 2^{nd}$	8.5%	5.7%
$3^{rd} \rightarrow 4^{th}$	1.2%	8.4%
$4^{th} \rightarrow 1^{st}$	0.0%	0.0%
$4^{th} \rightarrow 2^{nd}$	0.8%	2.3%
$4^{th} \rightarrow 3^{rd}$	3.9%	6.8%

Table 3. QBP Misclassification Percentages

Finally, Table 3 evaluates the Quartile-Based Prioritization scheme using the predicted values when compared with the actual ones. The corresponding misclassification percentages for each category demonstrate that the NR sparse prediction model provides a reliable framework for packet prioritization. It should be noted that packets belonging to the highest priority group, i.e., the most important packets needing the highest protection, have very low misclassification percentages. Additionally, misclassifications that go beyond one priority group (e.g., $1^{st} \leftrightarrow 3^{rd}/4^{th}$ or $4^{th} \leftrightarrow 2^{nd}$) are small, underlining the efficacy of the regression model.

5. CONCLUSIONS

We presented a new perceptive quality metric based on Structural Similarity to estimate the cumulative distortion due to dropped packets during video transmission over loss-prone networks. The underlying motion-compensated predictive coding mechanism employed in video compression was used to compute the overall impact of lost slices from three different types of encoded frames. We then developed a NR sparse prediction model to circumvent the computational complexity of the estimation process, especially useful in real-time streaming applications. Standard statistical performance measures showed that the predicted results were highly correlated with the actual CDSSIM calculations and this was achieved using only a small set of features extracted from the encoded bit-stream. We finally compared the actual and predicted values by utilizing them as inputs to a Quartile-based Prioritization scheme and demonstrated that the distortion prediction provides a reliable basis for prioritizing packets for video transmission. Our experiments were conducted for CIF and 4CIF video sequences that were encoded using H.264/AVC, but these procedures can be easily extended to other formats, coding standards and prioritization schemes.

6. REFERENCES

[1] ITU-T Rec. H.263, "Video Coding for Low Bit Rate Communication," v1, Nov. 1995; v2, Jan. 1998; v3, Nov. 2000.

[2] ISO/IEC JTC 1, "Coding of Audio-Visual Objects — Part 2: Visual," ISO/IEC 14496-2 (MPEG4 Visual Version 1), Apr. 1999; Amendment 1 (Version 2), Feb. 2000; Amendment 4 (streaming profile), Jan. 2001.

[3] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," IEEE Trans. on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 560–576, Jul. 2003

[4] H. Schwarz, D. Marpe, T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," IEEE Trans. on Circuits and Systems for Video Technology, Vol. 17, No. 9, pp. 1103-1120, Sept. 2007.

[5] G. J. Sullivan, J. R. Ohm, W. J. Han and T. Wiegand, "Overview of the high efficiency video coding (HVEC) standard," IEEE Trans. on Circuits and Systems for Video Technology, Vol. 22, No. 12, pp. 1649-1668, Dec. 2012.

[6] High Speed Downlink Packet Access; overall description. 3GPP Std. TS 25.308 v7.0.0, 2006.

[7] Dual Carrier HSPA specificiation, 3GPP Std TS 25.101 v9.4.0, 2010.

[8] Long Term Evolution (LTE) 3GPP Standards TS 36 Series.

[9] "The METIS 2020 Project – Mobile and Wireless Communication Enablers for the 2020 Information Society". METIS. 6 Jul. 2013.

[10] P. V. Pahalawatta, R. Berry, T. N. Pappas, and A. K. Katsaggelos, "Content-aware resource allocation and packet scheduling for video transmission over wireless networks," IEEE Journal on Selected Areas in Communications, Special Issue on Cross-Layer Optimization for Wireless Multimedia Communications, Vol. 25, No. 4, pp. 749-759, May 2007.

[11] E. Maani, P. V. Pahalawatta, R. Berry, T. N. Pappas, and A. K. Katsaggelos, "Resource allocation for downlink multiuser video transmission over wireless lossy networks," IEEE Trans. Image Processing, Vol. 17, No. 9, pp. 1663-1671, Sept. 2008.

[12] Y. Li, Z. Li, M. Chiang, A. R. Calderbank, "Content-aware distortion-fair video streaming in congested networks," IEEE Trans. Multimedia, Vol. 11, No. 6, pp. 1182-1193, Oct. 2009.

[13] H. Luo, S. Ci, and D. Wu, "A cross-layer design for the performance improvement of real-time video transmission of secondary users over cognitive radio networks," IEEE Trans. Circuits Syst. Video Tech., Vol. 21, No. 8, pp. 1040-1048, Aug. 2011.

[14] M. Ismail, W. Zhuang, and S. Elhedhli, "Energy and content aware multi-homing video transmission in heterogeneous networks," IEEE Trans. on Wireless Comm., Vol. 12, No. 7, pp. 3600-3610, Jul. 2013.

[15] A. Sankisa, A. K. Katsaggelos, and P. V. Pahalawatta, "Distortion estimation using structural similarity for video transmission over wireless networks," IEEE International Symposium on Multimedia, Miami, FL, USA, Dec. 2015.

[16] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," IEEE Signal Processing Magazine, Vol 26, No. 1, pp. 98-117, Jan. 2009.

[17] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. on Image Processing, Vol. 13, No. 4, pp. 600 – 612, Apr. 2004.

[18] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," IEEE Trans. on Broadcasting, Vol 50, No. 3, pp. 312-322, Sept. 2004.

[19] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," IEEE Trans. on Image Processing, Vol 19, No. 2, pp. 335-350, Feb. 2010.

[20] Z. Yu and H. Wu, "Human visual system based object digital video quality metric," Int. Conf. on Signal Processing, pp. 1088-1095, 2000.

[21] S. Winkler, "A perceptual distortion metric for digital color video," Proc. SPIE Human Vis. Electron. Imag. San Jose, Vol 3644, pp. 175-184, 1999.

[22] A. B. Watson, "Toward a perceptual video quality metric," Proc. Human Vis. Electron. Imag. III, pp. 139-147, 1998.

[23] T.-L. Lin, S. Kanumuri, Y. Zhi, D. Poole, P. Cosman, and A. Reibman, "A versatile model for packet loss visibility and its application to packet prioritization," IEEE Trans. on Image Processing 19, pp. 722-735, Mar. 2010.

[24] S. Paluri, K. Kambhatla, S. Kumar, B. Bailey, P. Cosman, and J. Matyjas, "Predicting slice loss distortion in H.264/AVC video for low complexity data prioritization," 19th IEEE ICIP, pp. 689-692, Sept. 2012.

[25] S. Paluri, K. Kambhatla, B. Bailey, P. Cosman, J. Matyjas, and S. Kumar, "A low complexity model for predicting slice loss distortion for prioritizing H.264/AVC video," Multimedia Tools and Applications, Nov. 2014.

[26] K. Pandremmenou, N. Tziortziotis, S. Paluri, W. Zhang, K. Blekas, L. P. Kondi, S. Kumar, "Quality optimization of H.264/AVC video transmission over noisy environments using a sparse regression framework," Conf. Visual Inform. Proc. and Comm. VI, Proc. SPIE IST & Electron. Imag., San Francisco, Vol. 9410, Aug. 2015

[27] T. Zhao, K. Zeng, A. Rehman, and Z. Wang, "On the use of SSIM in HEVC," Proc. 47th IEEE Asilomar Conf. Signals Sys. Comput., Pacific Grove, CA, USA, pp. 1107-1111, 2013.

[28] Z.-Y. Mai, C.-L. Yang, K.-Z. Kuang, and L.-M. Po, "A novel motion estimation method based on structural similarity for H.264 inter prediction," Proc. Int. Conf. Acoust. Speech and Signal Processing (ICASSP), pp. 913-916, May 2006.

[29] C.-L. Yang, H.-X. Wang, and L.-M. Po, "Improved inter prediction based on structural similarity in H.264," Proc. Int. Conf. Signal Process. Comm. ICSPC, Dubai, UAE, pp. 340-343, Nov. 2007.

[30] A. R. Reibman and D. Poole, "Characterizing packet loss impairments in compressed video", IEEE ICIP, Sept. 2007

[31] R. Tibshirani, "Regression shrinkage and selection via the LASSO," Journal of the Royal Statistical Society, Series B 58, 267–288, 1994.

[32] R. Tibshirani, "The LASSO method for variable selection in the Cox model," Statist. Med. 16(4), 385–395, 1997.

[33] "Full reference television phase II subjective test plans," Objective Quality Model Evaluation Criteria, 2002.

[34] D. E. Hinkle, W. Wiersma, and S. G. Jurs, "Applied statistics for the behavioral sciences," Boston, Houghton Mifflin, 2003.

[35] VQEG, "Final report from the Video Quality Experts Group on the validation of reduced-reference and no-reference objective models for standard definition television, Phase I," Jun. 2009.