International Journal of Multimedia Data Engineering and Management
Volume 7 • Issue 3 • July-September 2016

SSIM-Based Distortion Estimation for Optimized Video Transmission over Inherently Noisy Channels

Arun Sankisa, Northwestern University, Evanston, IL, USA

Katerina Pandremmenou, University of Ioannina, Ioannina, Greece

Peshala V. Pahalawatta, AT&T, Inc., El Segundo, CA, USA

Lisimachos P. Kondi, University of Ioannina, Ioannina, Greece

Aggelos K. Katsaggelos, Northwestern University, Evanston, IL, USA

ABSTRACT

The authors present two methods for examining video quality using the Structural Similarity (SSIM) index: Iterative Distortion Estimate (IDE) and Cumulative Distortion using SSIM (CDSSIM). In the first method, three types of slices are iteratively reconstructed frame-by-frame for three different combinations of packet loss and the resulting distortions are combined using their probabilities to give the total expected distortion. In the second method, a cumulative measure of the overall distortion is computed by summing the inter-frame propagation impact to all frames affected by a slice loss. Furthermore, the authors develop a No-Reference (NR) sparse regression framework for predicting the CDSSIM metric to circumvent the real-time computational complexity in streaming video applications. The two methods are evaluated in resource allocation and packet prioritization schemes and experimental results show improved performance and better end-user quality. The accuracy of the predicted CDSSIM values is studied using standard performance measures and a Quartile-Based Prioritization (QBP) scheme.

KEYWORDS

Cross-Layer Optimization, Cumulative Distortion, Iterative Distortion Estimate, Packet Scheduling and Prioritization, Structural Similarity, Video Quality

1. INTRODUCTION

Smart phones and mobile devices that contain sophisticated video processing elements have become an integral part of our daily lives and have created a dramatic rise in demand for multimedia services over wireless networks. This demand underscores the need for efficient algorithms that provide optimal end-user quality, while taking into account capacity constraints, like storage and bandwidth. Limited resources on transmission systems inherently prone to dropping packets provide strong motivation for video processing and network entities to implement efficient encoding, resource allocation, packet prioritization and scheduling techniques. The end-user experience and overall perceived quality can be influenced by many factors but most notably by compression and transmission impairments. Towards this end, research in video codecs has moved at a fast pace through standards like H.264/

DOI: 10.4018/IJMDEM.2016070103

Moving Picture Experts Group (MPEG) 4/Advanced Video Coding (AVC), Scalable Video Coding (SVC) and H.265/High Efficiency Video Coding (HEVC) (Wiegand, Sullivan, Bjøntegaard, & Luthra, 2003; Schwarz, Marpe, & Wiegand, 2007; Sullivan, Ohm, Han, & Wiegand, 2012). Similarly, wireless communication has also made rapid strides with 3G Universal Mobile Telecommunications System (UMTS), High Speed Downlink/Uplink Packet Access (HSDPA/HSUPA), WiMax, 4G/Long Term Evolution (LTE), and plans to introduce 5G before the end of this decade (HSDPA, 2006; DC-HSPA, 2010; LTE; METIS, 2013). These advances help address the growing demand for video streaming services but also emphasize the need for innovative algorithms that offer complete end-to-end solutions. Research in cross-layer optimization, rate-distortion modeling, packet scheduling and resource allocation in multi-user environments (e.g. Maani, Pahalawatta, Berry, Pappas, & Katsaggelos 2008; Li, Li, Chiang, & Calderbank, 2009; Luo, Ci, & Wu, 2011; Ismail, Zhuang, & Elhedhli, 2013; Sankisa, Katsaggelos, & Pahalawatta, 2015) has shown that transmission methods that are content-aware provide noticeable performance improvements than content-agnostic techniques.

The three components for defining a cross-layer, content-aware system are the encoding mechanism, the transmission network and the quality assessment technique. Video coding creates compression artifacts that directly translate to a perceived degradation in overall quality. During the encoding process, 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 the coding modes usually depends on the frame in which a block resides and a natural outcome of differentiated coding is the formation of data entities with unequal importance, a key incentive 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. 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, although it is usually accompanied by propagation of errors between frames.

Quality assessment on the source side is essential in designing a system that prioritizes packets for transmission. Full-Reference (FR) models that measure quality through distortion use the Mean-Squared Error (MSE), Cumulative MSE (CMSE) and Peak-Signal-to-Noise Ratio (PSNR) metrics. These metrics provide objective mathematical models that easily extend to computational analysis and application of optimization techniques. However, these methods are limited in the way they model the quality that is perceived by the Human Visual System (HVS). Extensive research has been done in enhancing MSE/PSNR-based schemes to accommodate perceptual quality but they are inadequate in that distortion is not assessed as the degradation of structural information detected by HVS. Human perception is naturally adapted to extract luminance, contrast and structure in an image and is the basis for the Structural SIMilarity (SSIM) metric proposed by (Wang, Bovik, Sheikh, & Simoncelli, 2004). 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 (Gao, Kwong, Zhou, & Yuan, 2016).

1.1. Related Work and Motivation

Developing methods that estimate perceptive quality of encoded videos has been the focus of a lot of research but these methods suffer from certain drawbacks that the proposed Iterative Distortion Estimate (IDE) and Cumulative Distortion using SSIM (CDSSIM) methods overcome. While (Moorthy

& Bovik, 2006) provides FR video quality assessment by using SSIM with motion compensation, it does not consider source encoding modes and distortion at the decoder due to packet loss, which are key components in streaming video transmission over wireless networks. Models presented in (Seshadrinathan & Bovik, 2007; Seshadrinathan & Bovik, 2009) provide methods for motion-based video quality assessment in the frequency domain using a complex wavelet version of the SSIM metric. A FR video quality model called MOtion-based Video Integrity Evaluation (MOVIE) index is proposed in (Seshadrinathan & Bovik, 2010; Seshadrinathan, Soundarajan, Bovik, & Cormack, 2010). The MOVIE model presents specific indexes that capture spatial and temporal distortions as a function of squared difference between Gabor coefficients. In (Wang, Rehman, Wang, Ma, & Gao, 2012), a Reduced-Reference (RR) statistical SSIM estimate and a rate model are proposed using information and coefficients in the Discrete Cosine Transform (DCT) domain, while (Zhao, Wang, & Kwong, 2013) presents a framework for coding mode selection based on user-defined complexity factors on partitions larger than MBs.

Perceptive quality measures like Video Quality Metric (VQM), Just Noticeable Difference (JND), Perceptual Distortion Metric (PDM) and Digital Video Quality (DVQ) (Pinson & Wolf 2004; Yu & Wu, 2000; Winkler, 1999; Watson, 1998), including the MOVIE and the Video Intrinsic Integrity and Distortion Evaluation Oracle (VIIDEO) model (Mittal, Saad, & Bovik, 2016), although effective, are computationally complex for real-time resource allocation and low bit rate encoding scenarios. The reference in (Lin, Kanumuri, Zhi, Poole, Cosman, & Reibman, 2010) 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 (Paluri, Kambhatla, Kumar, Bailey, Cosman, & Matyjas, 2012; Paluri, Kambhatla, Bailey, Cosman, Matyjas, & Kumar, 2014). In addition, an extension of these works is presented in (Paluri, Kambhatla, Medley, Matyjas, & Kumar, 2015), where a joint packet fragmentation and error protection scheme is proposed, for transmitting H.264/AVC compressed video over Rayleigh fading channels. However, these three papers, along with the work presented in (Pandremmenou, Tziortziotis, Paluri, Zhang, Blekas, Kondi, & Kumar, 2015), use RR features that depend on access to information in the original video for predicting CMSE. 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. On the other hand, (Wang, Zhang, & Agrafiotis, 2015) introduces a very low complexity metric, which is incorporated into SSIM, for making video quality estimations. Similarly, in (Aabed & AlRegib, 2015) the authors propose a video quality monitoring metric, using optical flow features. While both of these metrics correlate well with the Differential Mean Opinion Score (DMOS), they both rely on RR features, which still require an ancillary channel and access to the reference at some point.

The main contribution of this paper is the introduction of two metrics to evaluate perceived distortion using SSIM from individual packet loss during the transmission of encoded video over noisy, error-prone channels. The SSIM metric was originally defined for still images but through both these methods we modify its usage to evaluate the distortion in individual GOBs and video frames and extend it bi-directionally among dependent frames to obtain the total distortion. In each case, the total distortion depends on the source-side encoding mechanism and the resulting inter-frame error propagation due to packet loss and error concealment. The first metric, Iterative Distortion Estimate, utilizes the single-scale SSIM in the pixel/spatial domain to compute the overall source-to-receiver distortion. The estimate is obtained by iteratively reconstructing three types of slices for three different combinations of packet loss, where each type would be considered a "random" variable attached with a specific probability of occurrence. The distortions of these reconstructed slices when compared with the original ones are then stochastically combined to provide the overall expected distortion. The IDE-based distortion is incorporated into the content-aware utility function proposed in (Maani, Pahalawatta, Berry, Pappas, & Katsaggelos, 2008) to perform packet scheduling and resource allocation for the transmission of video packets. For the second metric, Cumulative Distortion using SSIM, we develop a mechanism that also uses SSIM to measure the overall degradation in perceived quality due to a packet loss. The cumulative distortion of each slice is obtained by removing the slice from the frame and, for every frame impacted by the loss of this slice due to inter-frame error propagation, comparing the error-concealed reconstructed frame with the compressed original. This operation is performed for every slice in every frame so that each slice can be individually assessed for its importance. However, this process is computationally very intensive. To avoid the per-slice computational overhead in real-time applications, we provide a No-Reference (NR) linear regression model using the Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani, 1994; Tibshirani, 1997) to predict the measured distortion using key features that are specifically related to slice loss. We evaluate the efficiency of the results using a simple prioritization method, called Quartile-Based Prioritization (QBP) (Pandremmenou, Tziortziotis, Paluri, Zhang, Blekas, Kondi, & Kumar, 2015) 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 start with the necessary background that lays the foundation for the rest of the paper and present the proposed iterative and cumulative distortion estimation models. Experimental results of incorporating the IDE values into a gradient-based utility function for resource allocation and packet scheduling are discussed in Section 3. Additionally, results from comparing the measured with predicted values and the prioritization efficiency of the proposed CDSSIM model are also presented in Section 3. We conclude the paper in Section 4.

2. PROPOSED DISTORTION METRICS

The structural similarity index (Wang, Bovik, Sheikh, & Simoncelli, 2004) measures the similarity of two discrete images using three quality measures – luminance, contrast and structure. SSIM is a widely used metric in image and video processing, especially in assessing decoding and reconstruction quality of images and videos (Zhao, Zeng, Rehman, & Wang, 2013; Mai, Yang, Kuang, & Po, 2006; Yang, Wang, & Po, 2007) that experience distortion from compression artifacts, lost slices or other error concealment. The equations comparing luminance, contrast and structure for two discrete image signals, assuming two dimensional signal arrays **x** and **y**, are given by:

$$l(\mathbf{x},\mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{1}$$

$$c(\mathbf{x},\mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\mu_x^2 + \mu_y^2 + C_2} \tag{2}$$

$$s(\mathbf{x},\mathbf{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{3}$$

where μ_x and μ_y are the means and σ_x^2 and σ_y^2 are the variances of signals **x** and **y**, respectively, and σ_{xy} is the cross covariance between them. Three constants C_1, C_2 and C_3 are introduced to provide stability to the calculations and are selected based on the criteria specified in (Wang, Bovik, Sheikh, & Simoncelli, 2004).

The overall structural similarity measure, SSIM is given by:

$$SSIM(\mathbf{x},\mathbf{y}) = l(\mathbf{x},\mathbf{y})\ c(\mathbf{x},\mathbf{y})\ s(\mathbf{x},\mathbf{y}) \tag{4}$$

The overall SSIM index is the average of the individual similarity indexes of 8x8-pixel sliding window that slides one pixel at a time. A greater similarity between compared images results in SSIM closer to 1 and if the two images are identical, they take the maximum value of 1.

In this work, a single row of MBs is assumed to form a slice/GOB and we use the terms "GOB" and "slice" interchangeably throughout the presentation without any loss or change in meaning. SSIM is used at a sub-image/frame-level, i.e., it is applied on transmission units, GOBs/slices, for IDE and at a frame-level for the CDSSIM computations. Using this metric, the corresponding distortion is defined as:

$$DSSIM = 1 - SSIM \tag{5}$$

The encoding of a video sequence is performed for each frame at a MB-level and after all MBs are fully coded, each slice is packetized for transmission. Each constituent MB is reconstructed from the previous slice and/or previously decoded frame using error concealment if a slice is lost during transmission. However, this reconstruction varies from the encoded version of the image in predictively coded, motion-compensated video sequences. This error propagates to all dependent frames and can combine to create a more complex inter-frame relationship if further packet loss is experienced in subsequent frames. Decisions about packet scheduling, protection, priority and resource allocation prior to transmission can only be made with proper distortion analysis performed at the source side. The sender uses probability of packet loss to estimate the expected source-to-receiver distortion after applying appropriate error concealment. Therefore, to accurately compute distortion, both encoder and decoder need to know and use the same error concealment algorithm.

The process for computing IDE, called the GOB-based Iterative Distortion Estimation using SSIM (GIDE-S), and the CDSSIM metric are presented in the following subsections.

2.1. GOB-Based Iterative Distortion Estimate using SSIM (GIDE-S)

We use the following notation to describe the iterative distortion estimation process. Slice i from frame n of the original video, $S_o(i,n)$, is compressed/encoded at the source and the resulting slice is denoted by $S_e(i,n)$. The decoder reconstruction after error concealment is represented as $S_d(i,n)$. We use s to denote a pixel when referred to in the context of a slice and m for a pixel in the context of a macroblock. The original slice i in frame n comprising v pixels is represented as $S_o(i,n)=\{s_o(v,i,n)\}_v$ i.e., the set of all pixels in the slice from 1 to v; the encoded version is $S_e(i,n)=\{s_e(v,i,n)\}_v$ and the reconstructed decoder version (at the sender) is $S_d(i,n)=\{s_d(v,i,n)\}_v$. An original MB j with k pixels in GOB i of frame n is $M_o(j,i,n){:}\{m_o(k)\}_k$ i.e., the set of all pixels from 1 to k; the corresponding encoded MB $M_e(j,i,n){:}\{m_e(k)\}_k$ and the decoded MB after error concealment is $M_d(j,i,n){:}\{m_d(k)\}_k$. The distortion from SSSIM as given by Equation (5), between the original and decoded GOB i in frame n, is defined as $D_{i,n}$.

We base our iterative distortion calculations on the techniques presented in the ROPE algorithm (Zhang, Regunathan, & Rose, 2000). However, in this paper we construct three different GOBs with pixel values based on whether the current and previous GOBs are lost or received. The three combinations of packet loss are *i)* the packet containing the current GOB is received (R), *ii)* the packet containing the current GOB is lost but packet with the previous GOB is received (LR), and *iii)* the packets containing both the current and previous GOBs are lost (LL). The *expected distortion* of the reconstructed slice i of frame n is then given by combining the three random variables:

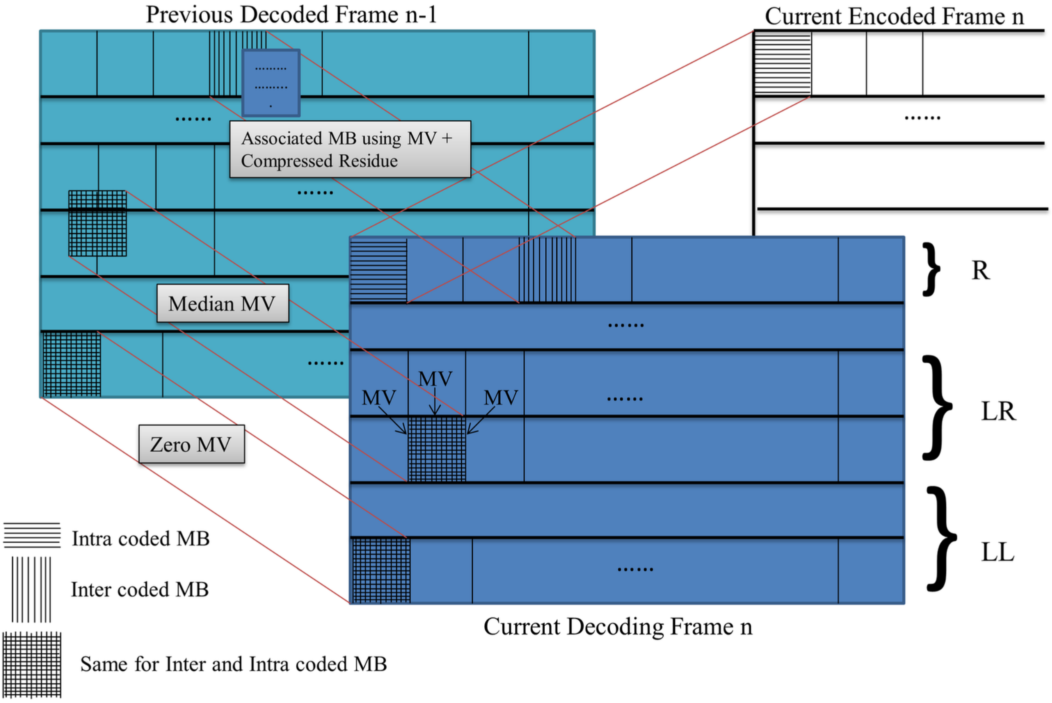
$$E\left\{D_{i,n}\right\} = p_0\left(D_{i,n}^{R}\right) + p_1\left(D_{i,n}^{LR}\right) + p_2\left(D_{i,n}^{LL}\right) \tag{6}$$

where D^X_i are GOB-level distortions for each combination of packet loss $X = \left\{R, LR, LL\right\}$ and p_0, p_1 and p_2 are the probabilities assigned to each packet-loss combination.

Figure 1 illustrates the process where three GOBs are constructed based on the packet loss and coding modes for each macroblock. We use basic error concealment where, if a GOB is lost in transmission it is concealed using the median Motion Vector (MV) from the previous GOB. Since error concealment is done on a per-MB basis, each MB in the current (lost) GOB is concealed using the median motion vectors of its three nearest neighbors i.e. top-left, top and top-right MBs from the previous GOB. If the previous packet is also lost in transmission, the median motion vectors are set to zero and the current (lost) GOB is replaced with one from the same location in the previous decoded frame. It should be noted that all three variations are applied for every GOB. We compute SSIM, and therefore, the corresponding distortion given by Equation (5) for each case, and apply the probability of occurrence to obtain the overall expected distortion. To accomplish this, rather than calculate the first and second moments of the "random" decoder pixel value, we construct the corresponding slices one pixel (and MB) at a time. Since all pixels in a GOB are lost if a packet is lost, pixel loss probability is the same as the probability of packet loss, ε. The selection of the values for ε is described in Section 3.1. Depending on whether an MB is inter- or intra-coded, each pixel in the MB, and collectively the entire GOB, is reconstructed.

Two components are used for calculating the distortion in the current frame: a) the previous decoded frame n-1, and b) the SSIM-based distortion for each packet in the queue. The previous decoded frame is the estimated decoder frame after applying error concealment due to randomly

Figure 1. GOB building process using Intra/Inter coding modes and Motion Vectors (MV). Packet loss combinations are denoted by: R, slice is correctly received; LR, current slice is lost but previous received; and LL current and previous slices are lost.



dropped slices (it is a function of packet loss probability) that simulate the "actual" packet losses during transmission. The detailed process for the GOB reconstruction is described below.

2.1.1. Packet Containing the Current GOB is Correctly Received

When the slice is correctly received by the decoder, each pixel in the slice is reconstructed based on its coding mode. For an intra-coded MB *j* of *k* pixels in the current slice *i* of frame *n*, each decoded pixel value in every MB has the same value as the corresponding encoded MB, i.e.:

$$M_d(,j,i,n):\left\{m_d(k)\right\}_k = M_e(j,i,n):\left\{m_e(k)\right\}_k \tag{7}$$

For an inter (predictive)-coded MB *j* with *k* pixels in the current slice *i* of frame *n*, each pixel is reconstructed using the motion vector to an associated pixel in slice *p* in the previous decoded frame *n-1* and the compressed residue. Therefore:

$$M_d(j,i,n):\left\{m_d(k)\right\}_k = M_d(j,i,n):\left\{m_d(k)\right\}_k + \left\{\hat{e}(k,j,i,n)\right\}_k \tag{8}$$

where *ê(k,j,i,n)* is the compressed residue of each pixel occupying *1* to *k* location in the MB *j* of the current slice *i* in frame *n*. From Equations (7) and (8), we have:

$$S_d(i,n) = \left\{s_e(j,i,n)\right\}_j \tag{9}$$

$$D_{i,n}^{R} = gDSSIM\left(S_o(i,n), S_d(i,n)\right) \tag{10}$$

where *gDSSIM* is the GOB-level distortion using SSIM obtained from Equation (5). The probability of occurrence of this event, as given in Equation (6) is $p_0 = (1\text{-}\varepsilon)$.

2.1.2. Packet Containing Current GOB is Lost but Previous GOB is Correctly Received

The reconstruction for both intra and inter-coded MBs is the same - each pixel in an MB of the current slice is error concealed and reconstructed using the median motion vectors of the three closest MBs (top left, top, top right of the same frame). For instance, a pixel in the current slice *i* of frame *n* is associated with a pixel in slice *q* of the previous decoded frame *n-1*. Therefore:

$$s_d(.,i,n) = s_d(.,q,n-1) \tag{11}$$

Although not every pixel in slice *i* is associated with the same slice *q* or is at the same location in the previous (decoded) frame, for notational simplicity we denote the whole GOB as $S_d(q,n\text{-}1)$:

$$S_d(i,n):\left\{s_d(m,i,n)\right\}_m = S_d(q,n-1):\left\{s_d(m,q,n-1)\right\}_m \tag{12}$$

$$D_{i,n}^{LR} = gDSSIM\left(S_o(i,n), S_d(i,n)\right) \tag{13}$$

Again, *gDSSIM* is the GOB-level distortion using SSIM as it is given by Equation (5). The probability of occurrence of this event, as given in Equation (6) is $p_1 = \varepsilon\ (1\text{-}\varepsilon)$.

2.1.3. Packets Containing Current and Previous Gobs are Lost

The motion vector is set to zero to reconstruct each MB of the slice when both previous and current slices are lost in transmission. Therefore:

$$S_d(i,n):\left\{s_d(m,i,n)\right\}_m = S_d(i,n-1):\left\{s_d(m,i,n-1)\right\}_m \tag{14}$$

$$D_{i,n}^{LL} = gDSSIM\left(S_o(i,n),S_d(i,n-1)\right) \tag{15}$$

As above, *gDSSIM* is the GOB-level distortion using SSIM obtained from Equation (5). The probability of occurrence of this event, as given in Equation (6) is $p_2 = \varepsilon^2$.

The overall expected Iterative Distortion Estimate of a slice *i* in frame *n* based on GOB-SSIM is obtained from Equations (10), (13) and (15):

$$IDE\left\{S(i,n)\right\} = (1-\varepsilon)\left(D_{i,n}^{R}\right) + \varepsilon(1-\varepsilon)\left(D_{i,n}^{LR}\right) + \varepsilon^2\left(D_{i,n}^{LL}\right) \tag{16}$$

The loss of a slice that results in a higher IDE in Equation (16) is given higher priority over one that has a lower value. This distortion estimate is then incorporated into a predefined gradient-based utility function for packet ordering and scheduling as it is discussed in Section 3.

2.2. Cumulative Distortion using SSIM and Feature-Based LASSO Regression

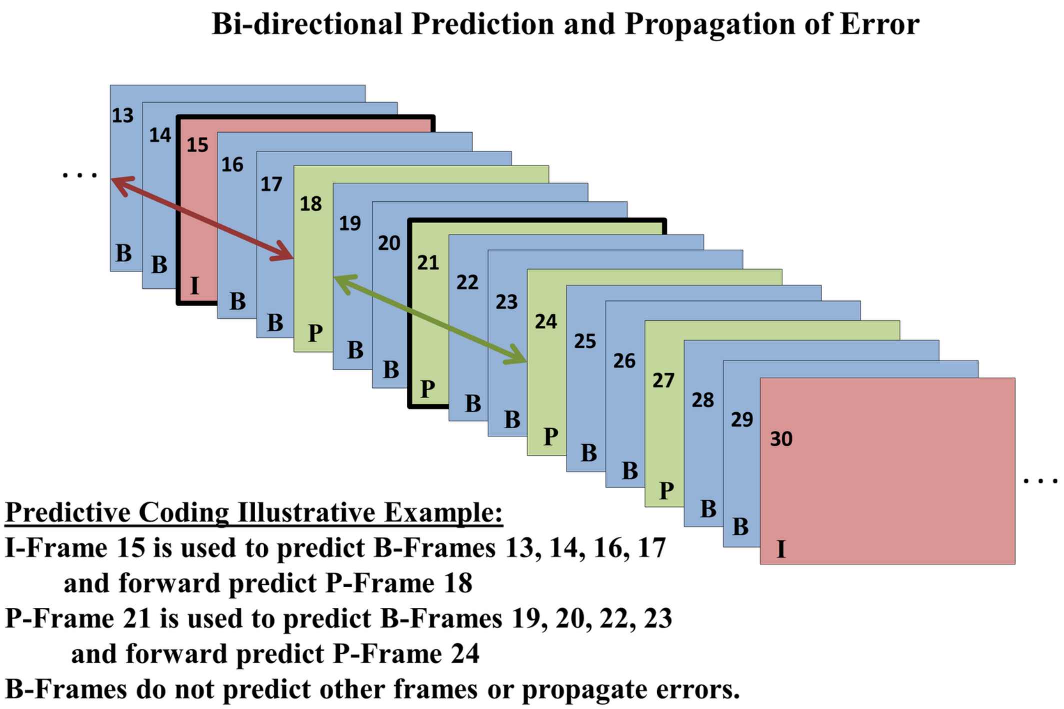
During motion compensation, when content is reconstructed, there exists a strict decoding order that must be followed, based on data dependencies arisen from the spatio-temporal correlations of a video content. Therefore, when an error occurs within a frame, it propagates to all dependent frames, degrading video quality of the thereby reconstructed content. In such cases of induced distortion, the overall impact of error propagation to multiple frames is better captured through a summation of the per-frame distortion instead of a summation of the similarity. In this context, we define the proposed cumulative metric, as the sum of distortion in the current frame and induced distortion in dependent frames as a result of a slice loss (see Equation (5)).

2.2.1. Cumulative Distortion using SSIM (CDSSIM)

In this work, the cumulative distortion metric was calculated for a typical Group-Of-Pictures (GOP) structure of “IBBP”, consisting of 16 frames. The considered GOP structure provides an I-frame with a sufficient frequency to allow the decoder to quickly begin correct decoding. As it is shown in Figure 2, for an “IBBP” GOP structure, 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 within the same GOP, while B-frames are not used as a reference for previous or subsequent frame predictions.

This bi-directional prediction results in the propagation of errors between the frames. As each of the I- and P-frames is used next as a reference to predict subsequent P-frames, the loss of a slice from a frame of such a type does not only affect the frames that depend on them, but the error is propagated to all future frames in the same GOP. For instance, as shown in Figure 2, the loss of a P-slice from frame 21, will not only affect B-frames 19, 20, 22, 23 and P-frame 24, but also all the frames until the end of the GOP, i.e., B-frames 25, 26, 28, 29 and P-frame 27.

Figure 2. GOP structure for video encoding in CDSSIM computation



Bi-directional Prediction and Propagation of Error

Predictive Coding Illustrative Example:
I-Frame 15 is used to predict B-Frames 13, 14, 16, 17 and forward predict P-Frame 18
P-Frame 21 is used to predict B-Frames 19, 20, 22, 23 and forward predict P-Frame 24
B-Frames do not predict other frames or propagate errors.

The cumulative distortion due to a lost slice *n* in frame *i* based on the SSIM index can be expressed differently according to the type of the considered slice loss. Specifically, if the lost slice is from I- or P-frames, *CDSSIM* is given by:

$$I,P:CDSSIM(n,i)=\sum_{k=i-2}^{GOP}\left(fDSSIM_k\right) \tag{17}$$

and when it comes to a slice loss from a B-frame, *CDSSIM* is computed as:

$$B:CDSSIM(n,i)=(fDSSIM_i) \tag{18}$$

where $fDSSIM_j$ is the frame-level distortion using structural similarity from Equation (5) of frame *j*, between the uncorrupted encoded video sequence and the packet-loss-impaired sequence. It is to be noted that the distortion summation is performed to the end of the GOP, and that the same procedure of cumulative distortion calculation is made for every slice loss of each frame in a sequence. It is expected that I-frames incur higher CDSSIM than P- and B-frames, and P-frames higher CDSSIM than B-frames. Finally, each "previous" P-frame incurs a higher CDSSIM as compared to each "subsequent" P-frame in a GOP, as it is used as a reference for the prediction of more frames.

2.2.2. LASSO Regression using Network Features

In the previous subsection we presented the rationale for calculating the CDSSIM metric, taking into account the type of each particular slice loss. Although the indicated method is the most accurate strategy for calculating the cumulative distortion in terms of SSIM, it involves a huge computational complexity that renders its use prohibitive in real-time applications; a complete decoding of the frame that includes the loss and of all the frames that depend on it, is required. For this purpose, we propose a method that is able to estimate CDSSIM values, making use of a number of NR features that are expected to describe perceptual video quality with high accuracy.

In this direction, we apply a simple linear regression model that performs simultaneously feature selection and response variable estimation. More specifically, the use of LASSO is proposed in order to estimate both the regression coefficients and the response variables at the same time. This approach is less complex than other methods because it eliminates the need to first do feature selection using a specific technique and then to apply another technique to perform regression. LASSO was originally proposed in (Tibshirani, 1994; Tibshirani, 1997) as a linear regression analysis tool that helps solve ill-posed multi-variable estimation problems by providing sparse and interpretable solutions.

In more detail, LASSO minimizes 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:

$$\min_{w,w_0}\left\{\frac{1}{2p}\sum_{i=1}^{p}\left(y_i-w_0-w^T x_i\right)^2+\lambda\|w\|_1\right\} \tag{19}$$

where p represents the total number of slices, y_i is the actual CDSSIM value of slice i and x_i is the vector of the values of all examined features (explained in detail later) of slice i. The term w is the vector of regression coefficients, w_0 is the intercept and λ is the regularization parameter.

The l_1 norm of Equation (19) forces some of the regression coefficients to take zero values. For larger λ values, i.e., as the penalty increases, the number of coefficients that take a zero value also increases, and vice versa. Therefore, LASSO is able to shrink a broader set of features to a smaller one, improving the estimation accuracy of the model through the elimination of the "prediction noise".

In the following, we cite the network features associated with transmission and slice loss that we extracted from each bitstream. Feature **TD** represents the Temporal Duration i.e., the number of frames that are affected by a slice loss. Evidently, a higher TD value is expected for the loss of an I-slice as compared to a P and/or B-slice loss. **FrameCenter** is a Boolean feature, which is set to true if a slice lies in the center of a frame i.e., if it is one of the middle slices of a frame. For a Common Intermediate Format (CIF) resolution frame, we consider six slices in the center of a frame, while for a 4CIF resolution frame, we consider 12 slices in the center, as it results by dividing the total number of slices of each resolution by three, and thus having an "upper", a "middle" and a "lower" frame part. **DistToRef** feature refers to the distance that a slice/frame is from the reference frame, measured in frames. For our considered GOP structure, it holds that P-frames are concealed using images three frames ago, while both I-frames and B-frames are concealed using images one frame ago. **FarConceal** is a Boolean feature set to true if DistToRef ≥ 3. **SBM** is the Slice Boundary Mismatch metric described in (Reibman & Poole, 2007), which captures the impact of the impairment on the boundary between correctly received and concealed slices. **MeanResEngy** and **MaxResEngy** are the mean and maximum values, respectively, 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. The value of the residual energy of a slice provides some information about the goodness of the error concealment as well as about the level of motion described by a slice. Particularly, a high value for the residual energy implies that the motion vectors probably do not represent the actual scene motion precisely and that the particular slice captures a high degree of motion. On the contrary, the opposite comes true for a small value of residual energy. **SigMean** and **SigVar** are the mean and variance, respectively, of the Y-component of the signal. **DMVX** and **DMVY** are the average motion vector difference values of a slice in the x and y axes, respectively. Finally, the features **absMVX** and **absMVY** are the average measures of the absolute motion vector values of a slice along the x and y axes, respectively.

It is worth highlighting that the features that are related to a packet loss are evaluated at the slice-level, while the motion-related features, i.e., DMVX, DMVY, absMVX and absMVY, are computed in the context of a macroblock and hence, they are averaged over an entire slice.

3. EXPERIMENTAL RESULTS

In this section, we present experimental results after testing each of the proposed SSIM-based distortion metrics under different scenarios and parameter settings.

3.1. Gradient-Based Utility Function and Packet Scheduling using IDE

We conducted experiments by incorporating the IDE values obtained from Equation (16) into a content-aware, gradient-based utility function, specifically, the two-step implementation for resource allocation described in (Maani, Pahalawatta, Berry, Pappas, & Katsaggelos, 2008). As a first step, the probability of packet loss ε is fixed for each packet in the transmission queue and the optimal values of spreading code n_i and power assignment P_i, for each user, are determined by varying the rate r_i. Then, with n_i and P_i fixed to these optimal values, the probability of packet loss ε becomes a function of the rate, i.e., at a small r_i, when the rate is gradually increased, the per-user expected distortion decreases due to the fact that more bits are transmitted. But at a larger r_i, the probability of packet loss increases as the rate increases (due to higher collisions, errors, dropped packets etc.), causing an increase in the expected distortion. Therefore, the expected distortion is a convex function of the rate assignment, and the optimal r_i that leads to the minimum expected distortion, can be calculated separately for each user as a simple one-dimensional line search.

The proposed GOB-based Iterative Distortion Estimate using SSIM algorithm is executed for five video sequences (users), *carphone, hall monitor, mother and daughter, news and silent*, using 10 different channel realizations. The sequences used were in Quarter Common Intermediate Format (QCIF) (176x144) format at 30 frames per second with 33ms for transmission and playback at the decoder. They were encoded using the H.264 (JVT reference software, JM) using variable bit rate encoding to achieve an average PSNR of 35dB for each decoded frame. All frames except for the first one were encoded as P-frames and 15 random I-MBs (encoded using constrained intra prediction) were inserted into each frame. This was done mainly to limit the propagation of error due to packet loss and to apply the techniques on different MBs. A 2ms transmission timeslot HSDPA wireless channel with a total base station power P = 25W, total number of spreading codes for all users, N = 15, and maximum per-user Signal-to-Interference-plus-Noise Ratio (SINR) constraint of 1.8dB are used. A Nakagami channel with a shaping parameter m=10, models the channel characteristics. Media Access Control (MAC) layer partitioning of the application layer packets mandates that all fragments of the application layer packet be received at the decoder for the decoding process to be complete and correct. For this purpose, a 10ms ACKnowledge/Negative ACKnowledge (ACK/NACK) feedback delay is assumed for each transmission. A NACK during the 10ms feedback delay of the application layer packet would reinsert and reorder the packet into the transmission queue provided that the NACK has arrived within the decoding deadline of the transmitted packet.

The efficiency of the proposed algorithm is evaluated both in terms of the average per-user PSNR and SSIM and the overall per-frame PSNR and SSIM. The per-user PSNR and SSIM are the frame averages of each user from all runs and they represent the average quality of each user; the per-frame PSNR and SSIM are the averages of each corresponding frame of all video sequences from multiple executions of the algorithm and they represent the average quality over all users. Figure 3 shows a comparison of the per-user quality in PSNR (dB) and SSIM, where each bar is the average over 130 QCIF frames for the 10 channel realizations. Figure 4 is the per-frame PSNR and per-frame SSIM of each corresponding frame averaged over all five video sequences and 10 separate realizations. The MSE-based method presented in (Maani, Pahalawatta, Berry, Pappas, & Katsaggelos, 2008) studied four different resource allocation schemes – expected distortion gradient with variable and fixed probability of packet loss and content-agnostic methods such as Queue Length and Max C/I methods. The values shown for the MSE-based method in Figure 3 and Figure 4 represent the best performance as compared to the other four aforementioned schemes. The results in these two figures show that the proposed GIDE-S algorithm performs consistently better than the MSE-based method.

3.2. Evaluation of CDSSIM and a Packet Prioritization Scenario

The specific metric was examined on a set of CIF resolution sequences, including *foreman, hall monitor, mobile* and *paris* and a set of 4CIF resolution sequences, including *crowdrun, harbour, ice,* and *soccer*. These sequences were encoded using the H.264/AVC variable bit rate encoding, while the first 100 frames of each sequence were used to conduct our experiments. At the decoder, Motion Copy Error Concealment (MCEC) was applied to conceal any slice losses in P- and B-frames, and spatial interpolation was used to conceal losses in the IDR-frames. Due to the different resolution of the video sequences, for the CIF case we had a total of 1800 slices per sequence (18 slices per frame) and 1800 x 4 = 7200 slices (and thus, CDSSIM values) in total. Similarly, for the 4CIF case, we obtained a total of 3600 slices per sequence (36 slices per frame) and 3600 x 4 = 14400 slices (and thus, CDSSIM values) in total.

Figure 3. Average per-user quality in (a) PSNR and (b) SSIM

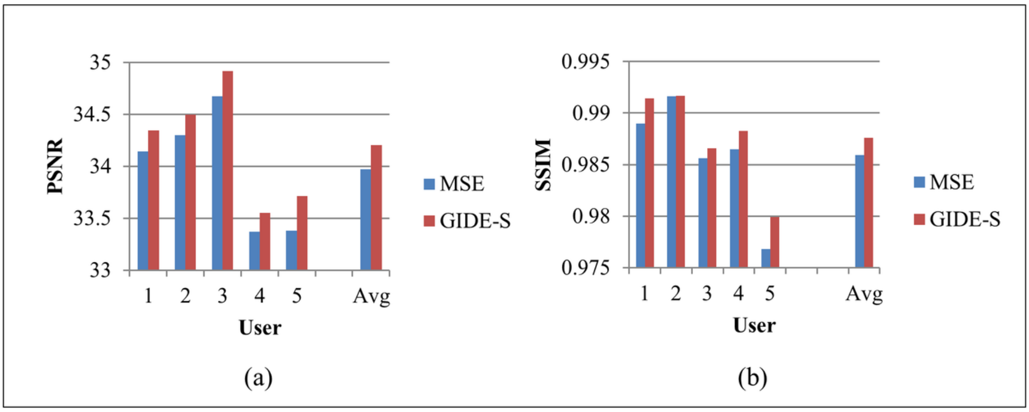
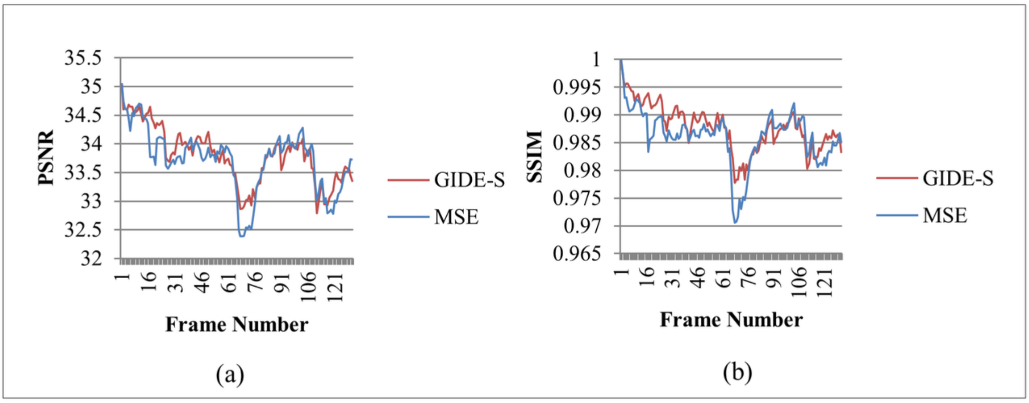


Figure 4. Average per-frame quality for all users combined using (a) PSNR and (b) SSIM



Instead of randomly splitting the dataset into training and test sets, we assumed an a priori partitioning into the two sets, such that the performance of the model is evaluated on specific sequences at each time. In this context, three sequences were used for training and one for testing, for each resolution separately. A percentage of 75% of the whole data was used to train our model, while it was tested on the remaining 25% of the dataset, represented by a single sequence. We tried all different dataset partitions, leaving always one sequence for testing. As the performance of each different sub-partition of each corresponding resolution did not show a significant variation, we present results when the *foreman, hall monitor* and *mobile* sequences are used for training and *paris* for testing, for the CIF case, and correspondingly, for the 4CIF case, using *crowdrun, ice,* and *soccer* as our training set and *harbour* for testing. Particular emphasis is given to the tradeoff between the efficiency of the proposed model in terms of prediction accuracy and on the complexity it involves, in terms of the number of the employed features used for making the CDSSIM estimations. The smaller the number of the employed features the lower the computational complexity of the model, usually at the cost of a reduced prediction capability on behalf of the latter.

Table 1 presents the features that were used as input to LASSO and the corresponding sparse coefficients it generates along with the intercept term and the λ value. Two sets of results, Set-1 and Set-2 for CIF and 4CIF sequences, highlight the use of different λ values for each format that result in different feature coefficients. For the CIF sequences, Set-1 results were obtained using lower λ values as compared to Set-2, while for 4CIF, Set-1 had a higher λ value than Set-2. In each case, a higher λ value results in a sparser set of features. It is interesting to notice that each sparser set of a specific resolution includes a subset of the features of the larger set. Moreover, the values of the regression coefficients offer an intuition about the significance of each of the selected features, i.e., a higher regression coefficient value implies a higher importance for the specific feature and vice versa. The data indicates that a fairly sparse representation is able to predict CDSSIM to a high level of accuracy, as it is evident from Table 2, which presents 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) (OQM Evaluation, 2002; Hinkle, Wiersma, & Jurs, 2003; VQEG, 2009).

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. For example, comparing the results for Set-1 and Set-2 for CIF and 4CIF sequences, it can

Table 1. Linear regression coefficients

Features	CIF (Set-1)	CIF (Set-2)	4CIF (Set-1)	4CIF (Set-2)
TD	0.0314	0.0308	0.0132	0.0143
FrameCenter	0	0	0	0
DistToRef	0	0	0	0
FarConceal	0	0	0	0
SBMMeanResEngy	0	0	0	0
MaxResEngy	0.0412	0.0409	0.0121	0.0136
SigMean	0.0025	0.0024	0	0.0002
SigVar	0.0008	0	0	0
DMVX	0	0	0	0
DMVY	0	0	0	0
absMVX	0	0	0	0
absMVY	0	0	0	0.0010
	0	0	0.0003	0.0012
Intercept	0.0534	0.0528	0.0183	0.0174
λ	0.0085	0.0093	0.0041	0.0019

Table 2. Standardized performance metrics

	CIF (Set-1)	CIF (Set-2)	4CIF (Set-1)	4CIF (Set-2)
PCC	0.9141	0.9142	0.9466	0.9497
SROCC	0.8220	0.8085	0.8601	0.8578
RMSE	0.0399	0.0392	0.0077	0.0077
MAE	0.0287	0.0533	0.0058	0.0185

be seen that with a larger λ, we select a sparser set of features but it does not necessarily result in worse performance. This is mainly because some features with non-zero coefficients may actually be considered unrelated, without a big influence on the prediction accuracy. More specifically, by observing the results of Table 2, we note that both Set-1 and Set-2 of each video sequence resolution offer comparable performance in terms of all examined metrics. It can be argued that the inclusion of one or two additional features, while it can have a positive impact on the accuracy of the prediction process it can also adversely impact it by just adding noise and possibly harming the overall quality of the estimation.

Having calculated both the measured and estimated CDSSIM values, as they result from the loss of each possible slice of a video sequence, we followed a slice prioritization approach, by grouping the slices into four categories. Specifically, we applied a Quartile-Based Prioritization scheme (Pandremmenou, Tziortziotis, Paluri, Zhang, Blekas, Kondi, & Kumar, 2015) on both the actual and estimated CDSSIM values. Therefore, we computed the three points (quartiles) that divide the data into four equal sized groups. In more detail, the values in each corresponding vector were sorted in an ascending order and the median values (corresponding to Q_2 quartiles) of each of them were computed. The same procedure was also followed with the lower and upper halves of each dataset, giving the Q_1 and Q_3 quartiles, respectively. Figure 5 illustrates this procedure.

As shown in Figure 5, the slices belonging to the class with the highest CDSSIM values are assigned the highest priority (1st priority), the slices with CDSSIM values between the Q_3 and Q_2 quartiles are assigned the 2nd priority, slices with CDSSIM between the Q_2 and Q_1 quartiles the 3rd priority and last, the rest of the slices are assigned the lowest priority (4th priority).

In an effort to evaluate the efficiency of our prediction model in terms of the CDSSIM estimations it was able to provide, we checked for correspondences between the actual and estimated CDSSIM values in each of the priority classes. A slice was considered "misclassified" if it was placed in a priority group that was different in the predicted set as compared to the measured one. Using this method, the percentage of slice misclassifications was considered as an indicator of the efficiency of the proposed model. The lower the misclassification percentages the better the performance of the prediction mechanism and the greater the potential for its use as a prioritization scheme.

An evaluation of the prediction accuracy using QBP for comparing the predicted CDSSIM with the computed values for Set-1 results, as representative sets for CIF and 4CIF sequences, is shown in Table 3. The low percentages of misclassifications for each category demonstrate that the NR

Figure 5. Quartile-based prioritization procedure

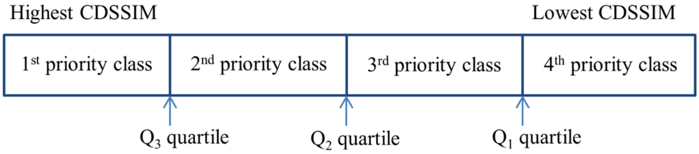


Table 3. QBP misclassification percentages

Misclassification	CIF (Set-1)	4CIF (Set-1)
1st→ 2nd	2.9%	3.1%
1st→ 3rd	1.5%	0.0%
1st→ 4th	1.1%	0.0%
2nd→ 1st	5.4%	3.1%
2nd→ 3rd	4.4%	7.4%
2nd→ 4th	2.3%	1.3%
3rd→ 1st	0.1%	0.0%
3rd→ 2nd	8.5%	5.7%
3rd→ 4th	1.2%	8.4%
4th→ 1st	0.0%	0.0%
4th→ 2nd	0.8%	2.3%
4th→ 3rd	3.9%	6.8%

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. The efficiency of the regression model is further underlined by the small numbers of misclassifications that extend beyond one priority group (e.g., $1^{st} \leftrightarrow 3^{rd}, 1^{st} \leftrightarrow 4^{th}, \; or \; 4^{th} \leftrightarrow 2^{nd}$).

4. CONCLUSION

In this paper, we presented two new perceptive quality metrics, an iterative estimate and a cumulative index, based on SSIM, to evaluate the overall distortion due to dropped packets during video transmission over loss-prone networks. For both these methods, the decoder distortion was evaluated on the source side using a motion-compensated predictive coding mechanism employed in video compression. The iterative expected distortion was plugged into a utility function in order to perform optimal resource allocation and packet ordering in a multi-user wireless transmission environment. Experiments showed that the proposed IDE metric provided, on average, a better estimate of the expected distortion when compared to existing MSE-based approaches and it is reflected in the overall performance and end-to-end video quality. We also developed a cumulative distortion index and an NR sparse prediction model to circumvent the complexity of its computation in real-time streaming applications. Standard performance measures showed that the predicted results were highly correlated with the computed CDSSIM values and the prediction was achieved with only a subset of features extracted from the encoded bitstream. A Quartile-based Prioritization scheme demonstrated that the distortion prediction provides a reliable basis for prioritizing packets for video transmission. We utilized QCIF, CIF and 4CIF video sequences encoded with the H.264/AVC standard, but future work can include extending these algorithms to videos encoded with H.265/HEVC coding standards and transmitted over LTE/4G or 5G high-speed wireless networks.

REFERENCES

Aabed, A., & AlRegib, G. (2015, September). Reduced-reference perceptual quality assessment for video streaming. *Proceedings of theIEEE Int. Conf. on Image Processing*, Quebec City, Canada (pp. 2394-2398).

High Speed Downlink Packet Access; overall description (2006). 3GPP Std. TS 25.308 v7.0.0, 2006.

Dual Carrier HSPA specification. (2010). 3GPP Std TS 25.101 v9.4.0, 2010.

Gao, W., Kwong, S., Zhou, Y., & Yuan, H. (2016, March). SSIM-based game theory approach for rate-distortion optimized intra frame CTU-level bit allocation. IEEE Trans on Multimedia, PP(99). doi:10.1109/TMM.2016.2535254

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral sciences*. Boston: Houghton Mifflin.

Ismail, M., Zhuang, W., & Elhedhli, S. (2013, July). Energy and content aware multi-homing video transmission in heterogeneous networks. *IEEE Transactions on Wireless Communications*, *12*(7), 3600–3610. doi:10.1109/TWC.2013.062713.130302

Li, Y., Li, Z., Chiang, M., & Calderbank, A. R. (2009, October). Content-aware distortion-fair video streaming in congested networks. *IEEE Transactions on Multimedia*, *11*(6), 1182–1193. doi:10.1109/TMM.2009.2026102

Lin, T.-L., Kanumuri, S., Zhi, Y., Poole, D., Cosman, P., & Reibman, A. (2010, March). A versatile model for packet loss visibility and its application to packet prioritization. *IEEE Transactions on Image Processing*, *19*(3), 722–735. doi:10.1109/TIP.2009.2038834 PMID:20028623

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

Luo, H., Ci, S., & Wu, D. (2011, August). A cross-layer design for the performance improvement of real-time video transmission of secondary users over cognitive radio networks. *IEEE Transactions on Circuits and Systems for Video Technology*, *21*(8), 1040–1048. doi:10.1109/TCSVT.2011.2129810

Maani, E., Pahalawatta, P. V., Berry, R., Pappas, T. N., & Katsaggelos, A. K. (2008, September). Resource allocation for downlink multiuser video transmission over wireless lossy networks. *IEEE Transactions on Image Processing*, *17*(9), 1663–1671. doi:10.1109/TIP.2008.2001402 PMID:18713672

Mai, Z.-Y., Yang, C.-L., Kuang, K.-Z., & Po, L.-M. (2006, May). A novel motion estimation method based on structural similarity for H.264 inter prediction. Proceedings of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP) (pp. 913-916).

METIS. (2013, July). 2020 Project – Mobile and Wireless Communication Enablers for the 2020. *The Information Society*.

Mittal, A., Saad, M. A., & Bovik, A. C. (2016, January). A completely blind video integrity oracle. *IEEE Transactions on Image Processing*, *25*(1), 289–300. doi:10.1109/TIP.2015.2502725 PMID:26599970

Moorthy, A. K., & Bovik, A. C. (2006). *A motion compensated approach to video quality assessment* (pp. 872–875). Asilomar Conf. on Signals, Systems and Computers.

Objective Quality Model Evaluation Criteria. (2002). *Full reference television phase II subjective test plans*.

Paluri, S., Kambhatla, K., Bailey, B., Cosman, P., Matyjas, J., & Kumar, S. (2014, November). A low complexity model for predicting slice loss distortion for prioritizing H.264/AVC video. *Multimedia Tools and Applications*, 75(2), 961-985.

Paluri, S., Kambhatla, K., Kumar, S., Bailey, B., Cosman, P., & Matyjas, J. (2012, September). Predicting slice loss distortion in H.264/AVC video for low complexity data prioritization. Proceedings of the IEEE Int. Conf. on Image Processing (pp. 689-692). doi:10.1109/ICIP.2012.6466953

Paluri, S., Kambhatla, K., Medley, M., Matyjas, J., & Kumar, S. (2015, December). Priority-aware joint packet fragmentation and error protection scheme for H.264 video over wireless channels. Proceedings of the IEEE Int. Symposium on Multimedia, Miami, FL, USA (pp. 95-100). doi:10.1109/ISM.2015.47

Pandremmenou, K., Tziortziotis, N., Paluri, S., Zhang, W., Blekas, K., Kondi, L. P., & Kumar, S. (2015, August). Quality optimization of H.264/AVC video transmission over noisy environments using a sparse regresion framework. Conf. on Visual Information Processing and. Communications VI, Proc. SPIE IST & Electronic Imaging, San Francisco *(Vol. 9410)*.

Pinson, M. H., & Wolf, S. (2004, September). A new standardized method for objectively measuring video quality. *IEEE Transactions on Broadcasting*, *50*(3), 312–322. doi:10.1109/TBC.2004.834028

Reibman, A. R., & Poole, D. (2007, September). Characterizing packet loss impairments in compressed video. Proceedings of the IEEE Int. Conf. on Image Processing. doi:10.1109/ICIP.2007.4379769

Sankisa, A., Katsaggelos, A. K., & Pahalawatta, P. V. (2015, December). Distortion estimation using structural similarity for video transmission over wireless networks. Paper presented at the IEEE Int. Symposium on Multimedia, Miami, FL, USA. doi:10.1109/ISM.2015.88

Schwarz, H., Marpe, D., & Wiegand, T. (2007, September). Overview of the scalable video coding extension of the H.264/AVC standard. *IEEE Transactions on Circuits and Systems for Video Technology*, *17*(9), 1103–1120. doi:10.1109/TCSVT.2007.905532

Seshadrinathan, K., & Bovik, A. C. (2007, April). A structural similarity metric for video based on motion models. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Honolulu, HI, USA *(Vol. 1*, pp. 869-872). doi:10.1109/ICASSP.2007.366046

Seshadrinathan, K., & Bovik, A. C. (2009, February). Motion-based perceptual quality assessment of video. *Paper presented at the SPIE Proc. Human Vision and Electronic Imaging* (Vol. 7240, pp. 1-12). doi:10.1117/12.811817

Seshadrinathan, K., & Bovik, A. C. (2010). Motion tuned spatio temporal quality assessment of natural videos. *IEEE Transactions on Image Processing*, *19*(2), 335–350. doi:10.1109/TIP.2009.2034992 PMID:19846374

Seshadrinathan, K., Soundarajan, R., Bovik, A. C., & Cormack, L. K. (2010, June). Study of subjective and objective quality assessment of video. *IEEE Transactions on Image Processing*, *19*(6), 1427–1441. doi:10.1109/TIP.2010.2042111 PMID:20129861

Sullivan, G. J., Ohm, J. R., Han, W. J., & Wiegand, T. (2012, December). Overview of the high efficiency video coding (HEVC) standard. *IEEE Transactions on Circuits and Systems for Video Technology*, *22*(12), 1649–1668. doi:10.1109/TCSVT.2012.2221191

Tibshirani, R. (1994). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society. Series B. Methodological*, *58*, 267–288.

Tibshirani, R. (1997). The LASSO method for variable selection in the Cox model. *Statistics in Medicine*, *16*(4), 385–395. doi:10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3 PMID:9044528

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." (2009, June).

Wang, M., Zhang, F., & Agrafiotis, D. (2015, September). A very low complexity reduced reference video quality metric based on spatio-temporal information selection. Proceedings of the IEEE Int. Conf. on Image Processing, Quebec City, Canada (pp. 571-575). doi:10.1109/ICIP.2015.7350863

Wang, S., Rehman, A., Wang, Z., Ma, S., & Gao, W. (2012, April). SSIM-motivated rate-distortion optimization for video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, *22*(4), 516–529. doi:10.1109/TCSVT.2011.2168269

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004, April). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612. doi:10.1109/TIP.2003.819861 PMID:15376593

Watson, A. B. (1998). Toward a perceptual video quality metric. *SPIE Proc. Human Vision and Electronic Imaging III* (pp. 139-147).

Wiegand, T., Sullivan, G. J., Bjøntegaard, G., & Luthra, A. (2003, July). Overview of the H.264/AVC video coding standard. *IEEE Transactions on Circuits and Systems for Video Technology*, *13*(7), 560–576. doi:10.1109/TCSVT.2003.815165

Winkler, S. (1999). A perceptual distortion metric for digital color video. *SPIE Proc. of Human Vision and Electronic Imaging*, San Jose, CA (Vol 3644, pp. 175-184).

Yang, C.-L., Wang, H.-X., & Po, L.-M. (2007 Nov). Improved inter prediction based on structural similarity in H.264. Proceedings of the Int. Conf. on Signal Processing and Communications, Dubai, UAE (pp. 340-343). doi:10.1109/ICSPC.2007.4728325

Yu, Z., & Wu, H. (2000). Human visual system based object digital video quality metric. Proceedings of the Int. Conf. on Signal Processing (pp. 1088-1095).

Zhang, R., Regunathan, S. L., & Rose, K. (2000, June). Video coding with optimal inter/intra-mode switching for packet loss resilience. *IEEE Journal on Selected Areas in Communications*, *18*(6), 966–976. doi:10.1109/49.848250

Zhao, T., Wang, Z., & Kwong, S. (2013, December). Flexible mode selection and complexity allocation in high frequency video coding. *IEEE Journal of Selected Topics in Signal Processing*, *7*(6), 1135–1144. doi:10.1109/JSTSP.2013.2271421

Zhao, T., Zeng, K., Rehman, A., & Wang, Z. (2013). *On the use of SSIM in HEVC* (pp. 1107–1111). Pacific Grove, CA: IEEE Asilomar Conf. on Signals, Systems and Computers. doi:10.1109/ACSSC.2013.6810465

Arun Sankisa received the Bachelor of Technology degree from Jawaharlal Nehru Technological University (JNTU), India majoring in Electrical and Electronics Engineering. He currently works at Nokia (formerly Alcatel-Lucent and Lucent Technologies, Bell Laboratories) in their wireless Long Term Evolution (LTE) division. He is a Ph.D. candidate in the Electrical Engineering and Computer Science department at Northwestern University, Evanston, IL, USA. He holds multiple patents in ad-hoc location-based wireless communication and applications. His research interests lie at the intersection of image/video processing and wireless telecommunications.

Katerina Pandremmenou received the BSc degree in computer science in 2008 from the Computer Science Department, University of Crete, Heraklion, Greece. In 2011, she received the MSc degree in technologies-applications and in 2015 the PhD degree in video processing and communications from the Computer Science and Engineering Department, University of Ioannina, Ioannina, Greece. Her current research interests include video quality assessment, video quality metrics and resource allocation over wireless networks.

Peshala Pahalawatta received his Bachelor of Science degree at Lafayette College, Easton, PA majoring in Electrical Engineering, and Masters and PhD in Electrical and Computer Engineering at Northwestern University Evanston, IL. He specializes in video compression and transmission. Currently, he is Senior Video Technologist at AT&T. He has multiple patents in video signal processing. His interests are in video quality evaluation, 3D video compression, and High Dynamic Range and wide color gamut video.

Lisimachos P. Kondi received the Diploma in electrical engineering from the Aristotle University of Thessaloniki, Greece, in 1994 and the MS and PhD degrees in electrical and computer engineering from Northwestern University, Evanston, IL, USA, in 1996 and 1999, respectively. He is currently an Associate Professor in the Department of Computer Science and Engineering, University of Ioannina, Greece. He was previously with the faculty of the University at Buffalo, The State University of New York and has held summer appointments at the Naval Research Laboratory (Washington, DC) and the Air Force Research Laboratory (Rome, NY). His research interests are in the general areas of signal and image processing and communications, including image and video compression and transmission over wireless channels and the Internet, sparse representations and compressive sensing, super-resolution of video sequences, and shape coding. Kondi is a co-author of a book entitled “4G Wireless Video Communications” (Wiley, 2009). He has been an Associate Editor of the IEEE Signal Processing Letters (2008-2012), the EURASIP Journal on Advances in Signal Processing (2005-present), and the International Journal of Distributed Sensor Networks (2013-2015). He was Technical Program Committee (TPC) Chair of the International Conference on Digital Signal Processing (DSP), Santorini, Greece, 2013, and will serve as Technical Program Committee Chair of the IEEE International Conference on Image Processing (ICIP), Athens, Greece, 2018.

Aggelos K. Katsaggelos received the Diploma degree in electrical and mechanical engineering from the Aristotelian University of Thessaloniki, Thessaloniki, Greece, in 1979, and the MS and PhD degrees in electrical engineering from the Georgia Institute of Technology, Atlanta, in 1981 and 1985, respectively. In 1985, he joined the Department of Electrical Engineering and Computer Science at Northwestern University, Evanston, IL, where he is currently a Professor holder of the Joseph Cummings Chair. Before that he was the holder of the Ameritech Chair of Information Technology and teh AT&T Chair and was the co-founder and Director of the Motorola Center for Seamless Communications. He is also a member of the Academic Affiliate Staff, NorthShore University Health System, an affiliated faculty at the Department of Linguistics, and he has an appointment at the Argonne National Laboratory. He has published extensively (5 books, 220 journal papers, 600 conference papers, 40 book chapters, 20 patents). He is the editor of Digital Image Restoration (Springer-Verlag, 1991), co-author of Rate-Distortion Based Video Compression (Kluwer, 1997), co-editor of Recovery Techniques for Image and Video Compression and Transmission (Kluwer, 1998), and co-author of Super-Resolution for Images and Video (Claypool, 2007), Joint Source-Channel Video Transmission (Claypool, 2007), and Machine Learning Refined (Cambridge University Press, 2016). Katsaggelos has served the IEEE and other Professional Societies in many capacities; he was, for example, Editor-in-Chief of the IEEE Signal Processing Magazine (1997–2002), a member of the Board of Governors of the IEEE Signal Processing Society (1999–2001), and a member of the Publication Board of the IEEE PROCEEDINGS (2003-2007). He is the recipient of the IEEE Third Millennium Medal (2000), the IEEE Signal Processing Society Meritorious Service Award (2001), the IEEE Signal Processing Society Technical Achievement Award (2010), an IEEE Signal Processing Society Best Paper Award (2001), an IEEE International Conference on Multimedia and Expo Paper Award (2006), an IEEE International Conference on Image Processing Paper Award (2007), an ISPA Best Paper Award (2009) and a EUSIPCO Paper Award (2013). He was a Distinguished Lecturer of the IEEE Signal Processing Society (2007–2008) and he is a Fellow of IEEE (1998) and SPIE (2009).