Hindawi Publishing Corporation
EURASIP Journal on Applied Signal Processing
Volume 2006, Article ID 42083, Pages 1–21
DOI 10.1155/ASP/2006/42083

# A Framework for Advanced Video Traces: Evaluating Visual Quality for Video Transmission Over Lossy Networks

**Osama A. Lotfallah,[1] Martin Reisslein,[2] and Sethuraman Panchanathan[1]**

*[1] Department of Computer Science and Engineering, Arizona State University, Tempe, AZ 85287, USA*
*[2] Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287-5706, USA*

Received 11 March 2005; Revised 1 August 2005; Accepted 4 October 2005

Conventional video traces (which characterize the video encoding frame sizes in bits and frame quality in PSNR) are limited to evaluating loss-free video transmission. To evaluate robust video transmission schemes for lossy network transport, generally experiments with actual video are required. To circumvent the need for experiments with actual videos, we propose in this paper an advanced video trace framework. The two main components of this framework are (i) advanced video traces which combine the conventional video traces with a parsimonious set of visual content descriptors, and (ii) quality prediction schemes that based on the visual content descriptors provide an accurate prediction of the quality of the reconstructed video after lossy network transport. We conduct extensive evaluations using a perceptual video quality metric as well as the PSNR in which we compare the visual quality predicted based on the advanced video traces with the visual quality determined from experiments with actual video. We find that the advanced video trace methodology accurately predicts the quality of the reconstructed video after frame losses.

## 1. INTRODUCTION

The increasing popularity of video streaming over wireless networks and the Internet require the development and evaluation of video transport protocols that are robust to losses during the network transport. In general, the video can be represented in three different forms in these development and evaluation efforts using (1) the actual video bit stream, (2) a video trace, and (3) a mathematical model of the video. The video bit stream allows for transmission experiments from which the visual quality of the video that is reconstructed at the decoder after lossy network transport can be evaluated. On the downside, experiments with actual video require access to and experience in using video codecs. In addition, the copyright limits the exchange of long video test sequences, which are required to achieve statistically sound evaluations, among networking researchers. Video models attempt to capture the video traffic characteristics in a parsimonious mathematical model and are still an ongoing research area; see for instance [1, 2].

Conventional video traces characterize the video encoding, that is, they contain the size (in bits) of each encoded video frame and the corresponding visual quality (measured in PSNR) as well as some auxiliary information, such as frame type (I, P, or B) and timing information for the frame play-out. These video traces are available from public video trace libraries [3, 4] and are widely used among networking researchers to test novel transport protocols for video, for example, network resource management mechanisms [5, 6], as they allow for simulating the operation of networking and communications protocols without requiring actual videos. Instead of transmitting the actual bits representing the encoded video, only the *number* of bits is fed into the simulations.

One major limitation of the existing video traces (and also the existing video traffic models) is that for evaluation of lossy network transport they can only provide the bit or frame loss probabilities, that is, the long run fraction of video encoding bits or video frames that miss their decoding deadline at the receiver. These loss probabilities provide only very limited insight into the visual quality of the reconstructed video at the decoder, mainly because the predictive coding schemes, employed by the video coding standards, propagate the impact of loss in a given frame to subsequent frames. The propagation of loss to subsequent frames results generally in nonlinear relationships between bit or frame losses and the reconstructed qualities. As a consequence, experiments to date with actual video are necessary to accurately examine the video quality after lossy network transport.

The purpose of this paper is to develop an advanced video trace framework that overcomes the outlined limitation of the existing video traces and allows for accurate prediction of the visual quality of the reconstructed video after lossy network transport without experiments with actual video. The main underlying motivation for our work is that visual content plays an important role in estimating the quality of the reconstructed video after suffering losses during network transport. Roughly speaking, video sequences with little or no motion activity between successive frames experience relatively minor quality degradation due to losses since the losses can generally be effectively concealed. On the other hand, video sequences with high motion activity between successive frames suffer relatively more severe quality degradations since loss concealment is generally less effective for these high-activity videos. In addition, the propagation of losses to subsequent frames depends on the visual content variations between the frames. To capture these effects, we identify a parsimonious set of visual content descriptors that can be added to the existing video traces to form advanced video traces. We develop quality predictors that based on the advanced video traces predict the quality of the reconstructed video after lossy network transport.

The paper is organized as follows. In the following subsection, we review related work. Section 2 presents an outline of the proposed advanced video trace framework and a summary of a specific advanced video trace and quality prediction scheme for frame level quality prediction. Section 3 discusses the mathematical foundations of the proposed advanced video traces and quality predictors for decoders that conceal losses by copying. We conduct formal analysis and simulation experiments to identify content descriptors that correlate well with the quality of the reconstructed video. Based on this analysis, we specify advanced video traces and quality predictors for three levels of quality prediction, namely frame, group-of-pictures (GoP), and shot. In Section 4, we provide the mathematical foundations for decoders that conceal losses by freezing and specify video traces and quality predictors for GoP and shot levels quality prediction. In Section 5, the performance of the quality predictors is evaluated with a perceptual video quality metric [7], while in Section 6, the two best performing quality predictors are evaluated using the conventional PSNR metric. Concluding remarks are presented in Section 6.

### 1.1. Related work

Existing quality prediction schemes are typically based on the rate-loss-distortion model [8], where the reconstructed quality is estimated after applying an error concealment technique. Lost macroblocks are concealed by copying from the previous frame [9]. A statistical analysis of the channel distortion on intra- and inter-macroblocks is conducted and the difference between the original frame and the concealed frame is approximated as a linear relationship of the difference between the original frames. This rate-loss-distortion model does not account for commonly used B-frame macroblocks. Additionally, the training of such a model can be prohibitively expensive if this model is used for long video traces. In [10], the reconstructed quality due to packet (or frame) losses is predicted by analyzing the macroblock modes of the received bitstream. The quality prediction can be further improved by extracting lower-level features from the received bitstream such as the motion vectors. However, this quality prediction scheme depends on the availability of the received bitstream, which is exactly what we try to overcome in this paper, so that networking researchers without access to or experience in working with actual video streams can meaningfully examine lossy video transmission mechanisms. The visibility of packet losses in MPEG-2 video sequences is investigated in [11], where the test video sequences are affected by multiple channel loss scenarios and human subjects are used to determine the visibility of the losses.

The visibility of channel losses is correlated with the visual content of the missing packets. Correctly received packets are used to estimate the visual content of the missing packets. However, the visual impact of (i.e., the quality degradation due to) visible packet loss is not investigated. The impact of the burst length on the reconstructed quality is modeled and analyzed in [12]. The propagation of loss to subsequent frames is affected by the correlation between the consecutive frames. The total distortion is calculated by modeling the loss propagation as a geometric attenuation factor and modeling the intra-refreshment as a linear attenuation factor. This model is mainly focused on the loss burst length and does not account for I-frame losses or B-frame losses. In [13], a quality metric is proposed assuming that channel losses result in a degraded frame rate at the decoder. Subjective evaluations are used to predict this quality metric. A nonlinear curve fitting is applied to the results of these subjective evaluations. However, this quality metric is suitable only for low bit rate coding and cannot account for channel losses that result in an additional spatial quality degradation of the reconstructed video (i.e., not only temporal degradation).

We also note that in [14], video traces have been used for studying rate adaptation schemes that consider the quality of the rate-regulated videos. The quality of the regulated videos is assigned a discrete perceptual value, according to the amount of the rate regulation. The quality assignment is based on empirical thresholds that do not analyze the effect of a frame loss on subsequent frames. The propagation of loss to subsequent frames, however, results in nonlinear relationships between losses and the reconstructed qualities, which we examine in this work. In [15], multiple video coding and networking factors were introduced to simplify the determination of this nonlinear relationship from a network and user perspective.

## 2. OVERVIEW OF ADVANCED VIDEO TRACES

In this section, we give an overview of the proposed advanced video trace framework and a specific quality prediction method within the framework. The presented method exploits motion information descriptors for predicting the reconstructed video quality after losses during network transport.

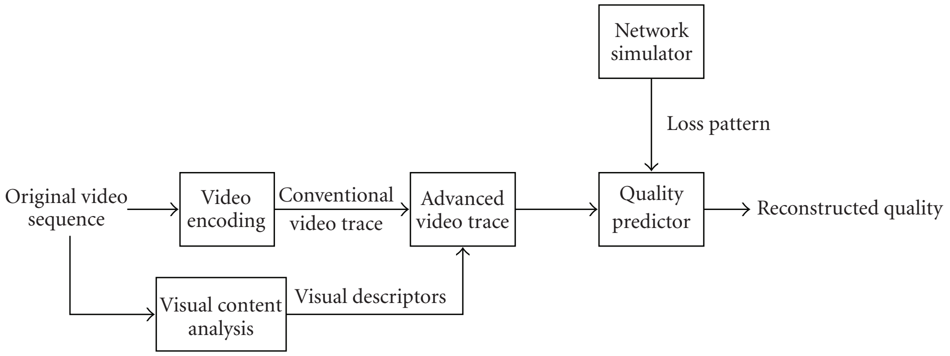


Figure 1: Proposed advanced video trace framework. The conventional video trace characterizing the video encoding (frame size and frame quality of encoded frames) is combined with visual descriptors to form an advanced video trace. Based on the advanced video trace, the proposed quality prediction schemes give accurate predictions of the decoded video quality after lossy network transport without requiring experiments with actual video.

### 2.1. Advanced video trace framework

The two main components of the proposed framework, which is illustrated in Figure 1, are (i) the advanced video trace and (ii) the quality predictor. The advanced trace is formed by combining the conventional video trace which characterizes the video encoding (through frame size in bits and frame quality in PSNR) with visual content descriptors that are obtained from the original video sequence. The two main challenges are (i) to extract a parsimonious set of visual content descriptors that allow for accurate quality prediction, that is, have a high correlation with the reconstructed visual quality after losses, and (ii) to develop simple and efficient quality prediction schemes which based on the advanced video trace give accurate quality predictions. In order to facilitate quality predictions at various levels and degrees of precision, the visual content descriptors are organized into a hierarchy, namely, frame level descriptors, GoP level descriptors, and shot level descriptors. Correspondingly there are quality predictors for each level of the hierarchy.

### 2.2. Overview of motion information based quality prediction method

In this subsection, we give a summary of the proposed quality prediction method based on the motion information. We present the specific components of this method within the framework illustrated in Figure 1. The rationale and the analysis leading to the presented method are given in Section 3.

#### 2.2.1. Basic terminology and definitions

Before we present the method, we introduce the required basic terminology and definitions, which are also summarized in Table 1. We let $F(t,i)$ denote the value of the luminance component at pixel location $i$, $i = 1,\ldots,N$ (assuming that all frame pixels are represented as a single array consisting of $N$ elements), of video frame $t$. Throughout, we let $K$ denote the number of P-frames between successive I-frames and let $L$ denote the difference in the frame index between successive P-frames (and between I-frame and first P-frame in the GoP as well as between the last P-frame in the GoP and the next I-frame); note that correspondingly there are $L-1$ B-frames between successive P-frames. We let $D(t,i) = |F(t,i) - F(t-1,i)|$ denote the absolute difference between frame $t$ and the preceding frame $t-1$ at location $i$. Following [16], we define the motion information $M(t)$ of frame $t$ as

$$M(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(D(t,i) - \overline{D(t)}\right)^2}, \qquad (1)$$

where $\overline{D(t)} = (1/N)\sum_{i=1}^{N} D(t,i)$ is the average absolute difference between frames $t$ and $t-1$. We define the aggregated motion information between reference frames, that is, between I- and P-frames, as

$$\mu(t) = \sum_{j=0}^{L-1} M(t-j). \qquad (2)$$

For a B-frame, we let $v_f(t,i)$ be an indicator variable, which is set to one if pixel $i$ is encoded using forward motion estimation, is set to 0.5 if interpolative motion estimation is used, and is set to zero otherwise. Similarly, we set $v_b(t,i)$ to one if backward motion estimation is used, set $v_b(t,i)$ to 0.5 if interpolative motion estimation is used, and set $v_b(t,i)$ to zero otherwise. We let $V_f(t) = (1/N)\sum_{i=1}^{N} v_f(t,i)$ denote the ratio of forward-motion-estimated pixels to the total number of pixels in frame $t$, and analogously denote by $V_b(t) = (1/N)\sum_{i=1}^{N} v_b(t,i)$ the ratio of backward-motion-estimated pixels to the total number of pixels.

For a video shot, which is defined as a sequence of frames captured by a single camera in a single continuous action in space and time, we denote the intensity of the motion activity by $\theta$. The motion activity $\theta$ ranges from 1 for a low level of motion to 5 for a high level of motion, and correlates well with the human perception of the level of motion in the video shot [17].

Table 1: Summary of basic notations.

| Variable | Definition |
|---|---|
| $L$ | Distance between successive P-frames, that is, $L$–1 B frames between successive P frames |
| $K$ | Number of P-frames in GoP |
| $R$ | Number of affected P-frames in GoP as a result of a P-frame loss |
| $N$ | Number of pixels in a video frame |
| $F(t, i)$ | Luminance value at pixel location $i$ in original frame $t$ |
| $\widehat{F}(t, i)$ | Luminance value at pixel location $i$ in encoded frame $t$ |
| $\widetilde{F}(t, i)$ | Luminance value at pixel location $i$ in reconstructed frame $t$ (after applying loss concealment) |
| $A(t, i)$ | Forward motion estimation at pixel location $i$ in P-frame $t$ |
| $v_f(t, i)$ | Forward motion estimation at pixel location $i$ in B-frame $t$ |
| $v_b(t, i)$ | Backward motion estimation at pixel location $i$ in B-frame $t$ |
| $e(t, i)$ | Residual error (after motion compensation) accumulated at pixel location $i$ in frame $t$ |
| $\Delta(t)$ | The average absolute difference between encoded luminance values $\widehat{F}(t, i)$ and reconstructed luminance values $\widetilde{F}(t, i)$ averaged over all pixels in frame $t$ |
| $M(t)$ | Amount of motion information between frame $t$ and frame $t - 1$ |
| $\mu(t)$ | Aggregate motion information between P-frame $t$ and its reference frame $t$–$L$ for frame level analysis of decoders that conceal losses by copying from previous reference (in encoding order) frame |
| $\gamma(t)$ | Aggregated motion information between P-frame $t$ and the next I-frame for frame level analysis of decoders that conceal losses by freezing the reference frame until next I-frame |
| $\mu$ | Motion information $\mu(t)$ averaged over the underlying GoP |
| $\gamma$ | Motion information $\gamma(t)$ averaged over the underlying GoP |

#### 2.2.2. Advanced video trace entries

For each video frame $t$, we add three parameter values to the existing video traces.

(1) The motion information $M(t)$ of frame $t$, which is calculated using (1).

(2) The ratio of forward motion estimation $V_f(t)$ in the frame, which is added only for B-frames. We approximate the ratio of backward motion estimation $V_b(t)$, as the compliment of the ratio of forward motion estimation, that is, $V_b(t) \approx 1$–$V_f(t)$, which reduces the number of added parameters.

(3) The motion activity level $\theta$ of the video shot.

#### 2.2.3. Quality prediction from motion information

Depending on (i) the concealment technique employed at the decoder and (ii) the quality prediction level of interest, different prediction methods are used. We focus in this summary on the concealment by "copying" (concealment by "freezing" is covered in Section 4) and the frame level prediction (GoP and shot levels predictions are covered in Subsections 3.4 and 3.5). For the loss concealment by copying and the frame level quality prediction, we further distinguish between the lost frame itself and the frames that reference the lost frame, which we refer to as the *affected* frames. With the loss concealment by copying, the lost frame itself is reconstructed by copying the entire frame from the closest reference frame. For an affected frame that references the lost frame, the motion estimation of the affected frame is applied with respect to the reconstruction of the lost frame, as elaborated in Section 3.

For the lost frame $t$ itself, we estimate the quality degradation $Q(t)$ with a logarithmic or linear function of the motion information if frame $t$ is a B-frame, respectively, of the aggregate motion information $\mu(t)$ if frame $t$ is a P-frame, that is,

$$Q(t) = a_0^B \times M(t) + b_0^B, \qquad Q(t) = a_0^P \times M(t) + b_0^P,$$
$$Q(t) = a_0^B \times \ln(M(t)) + b_0^B, \qquad Q(t) = a_0^P \times \ln(M(t)) + b_0^P. \tag{3}$$

(A refined estimation for lost B-frames considers the aggregated motion information between the lost B-frame and the closest reference frame, see Section 3.) Standard best-fitting curve techniques are used to estimate the functional parameters $a_0^B$, $b_0^B$, $a_0^P$, and $b_0^P$ by extracting training data from the underlying video programs.

If the lost frame $t$ is a P-frame, the quality degradation $Q(t+nL)$ of a P-frame $t+nL$, $n = 1,\dots,K-1$, is predicted as

$$Q(t+nL) = a_n^P \times \mu(t) + b_n^P,$$
$$Q(t+nL) = a_n^P \times \ln(\mu(t)) + b_n^P, \tag{4}$$

using again standard curve fitting techniques.

Finally, for predicting the quality degradation $Q(t+m)$ of a B-frame $t+m$, $m = -(L-1),\dots-1,1,\dots,L-1,L+1,\dots,2L-1,2L+1,\dots,2L+L-1,\dots,(K-1)L+1,\dots,(K-1)L+L-1$, that references a lost P-frame $t$, we distinguish three cases.

*Case 1.* The B-frame precedes the lost P-frame and references the lost P-frame using backward motion extimation. In this case, we define the aggregate motion information of the affected B-frame $t+m$ as

$$\mu(t+m) = \mu(t)V_b(t+m). \tag{5}$$

*Case 2.* The B-frame succeeds the lost P-frame and both the P-frames used for forward and backward motion estimation are affected by the P-frame loss, in which case

$$\mu(t+m) = \mu(t), \tag{6}$$

that is, the aggregate motion information of the affected B-frame is equal to the aggregate motion information of the lost P-frame.

*Case 3.* The B-frame succeeds the lost P-frame and is backward motion predicted with repect to the following I-frame, in which case

$$\mu(t+m) = \mu(t)V_f(t+m). \tag{7}$$

In all three cases, linear or logarithmic standard curve fitting characterized by the funtional parameters $a_m^B$, $b_m^B$ is used to estimate the quality degradation from the aggregate motion information of the affected B-frame.

In summary, for each video in the video trace library, we obtain a set of functional approximations represented by the triplets $(\varphi_n^P, a_n^P, b_n^P)$, $n = 0,1,\dots,K-1$, and $(\varphi_m^B, a_m^B, a_m^B)$, $m = -(L-1),\dots-1,0,1,\dots,L-1,L+1,\dots,2L-1,2L+1,\dots,2L+L-1,\dots,(K-1)L+1,\dots,(K-1)L+L-1$, whereby $\varphi_n^P$, $\varphi_m^B$ = "lin" if the linear functional approximation is used and $\varphi_n^P$, $\varphi_m^B$ = "log" if the logarithmic functional approximation is used.

With this prediction method, which is based on the analysis presented in the following section, we can predict the quality degradation due to frame loss with relatively high accuracy (as demonstrated in Sections 5 and 6) using only the parsimonious set of parameters detailed in Subsection 2.2.1 and the functional approximation triplets detailed above.

## 3. ANALYSIS OF QUALITY DEGRADATION WITH LOSS CONCEALMENT BY COPYING

In this section, we identify for decoders with loss concealment by copying the visual content descriptors that allow for accurate prediction of the quality degradation due to a frame loss in a GoP. (Concealment by freezing is considered in Section 4.) Toward this end, we analyze the propagation of errors due to the loss of a frame to subsequent P-frames and B-frames in the GoP. For simplicity, we focus in this first study on advanced video traces on a single complete frame loss per GoP. Single frame loss per GoP can be used to model wireless communication systems that use interleaving to randomize the fading effects. In addition, single frame loss can be seen with multiple descriptions coding, where video frames are distributed over multiple independent video servers/transmission paths. We leave the development and evaluation of advanced video traces that accommodate partial frame loss or multiple frame losses per GoP to future work.

In this section, we first summarize the basic notations used in our formal analysis in Table 1 and outline the setup of the simulations used to complement the analysis in the following subsection. In Subsection 3.2, we illustrate the impact of frame losses and motivate the ensuing analysis. In the subsequent Subsections 3.3, 3.4, and 3.5, we consider the prediction of the quality degradation due to the frame loss at the frame, GoP, and shot levels, respectively. For each level, we analyze the quality degradation, identify visual content descriptors to be included in the advanced video traces, and develop a quality prediction scheme.

### 3.1. Simulation setup

For the illustrative simulations in this section, we use the first 10 minutes of the *Jurassic Park I* movie. The movie had been segmented in video shots using automatic shot detection techniques, which have been extensively studied and for which simple algorithms are available [18]. This enables us to code the first frame in every shot as an intraframe. The shot detection techniques produced 95 video shots with a range of motion activity levels. For each video shot, 10 human subjects estimated the perceived motion activity level, according to the guidelines presented in [19]. The motion activity level $\theta$ was then computed as the average of the 10 human estimates. The QCIF (176 × 144) video format was used, with a frame rate of 30 fps, and the GoP structure IBBPBBPBBPBB, that is, we set $K = 3$ and $L = 3$. The video shots were coded using an MPEG-4 codec with a quantization scale of 4. (Any other quantization scale could have been used without changing the conclusions from the following illustrative simulations.) For our illustrative simulations, we measure the image quality using a perceptual metric, namely, VQM [7], which has been shown to correlate well with the human visual perception. (In our extensive performance evaluation of the proposed advanced video trace framework both VQM and the PSNR are considered.) The VQM metric computes the magnitude of the visible difference between two video sequences, whereby larger visible degradations result in larger VQM values. The metric is based on the discrete cosine transform, and incorporates aspects of early visual processing, spatial and temporal filtering, contrast masking, and probability summation.

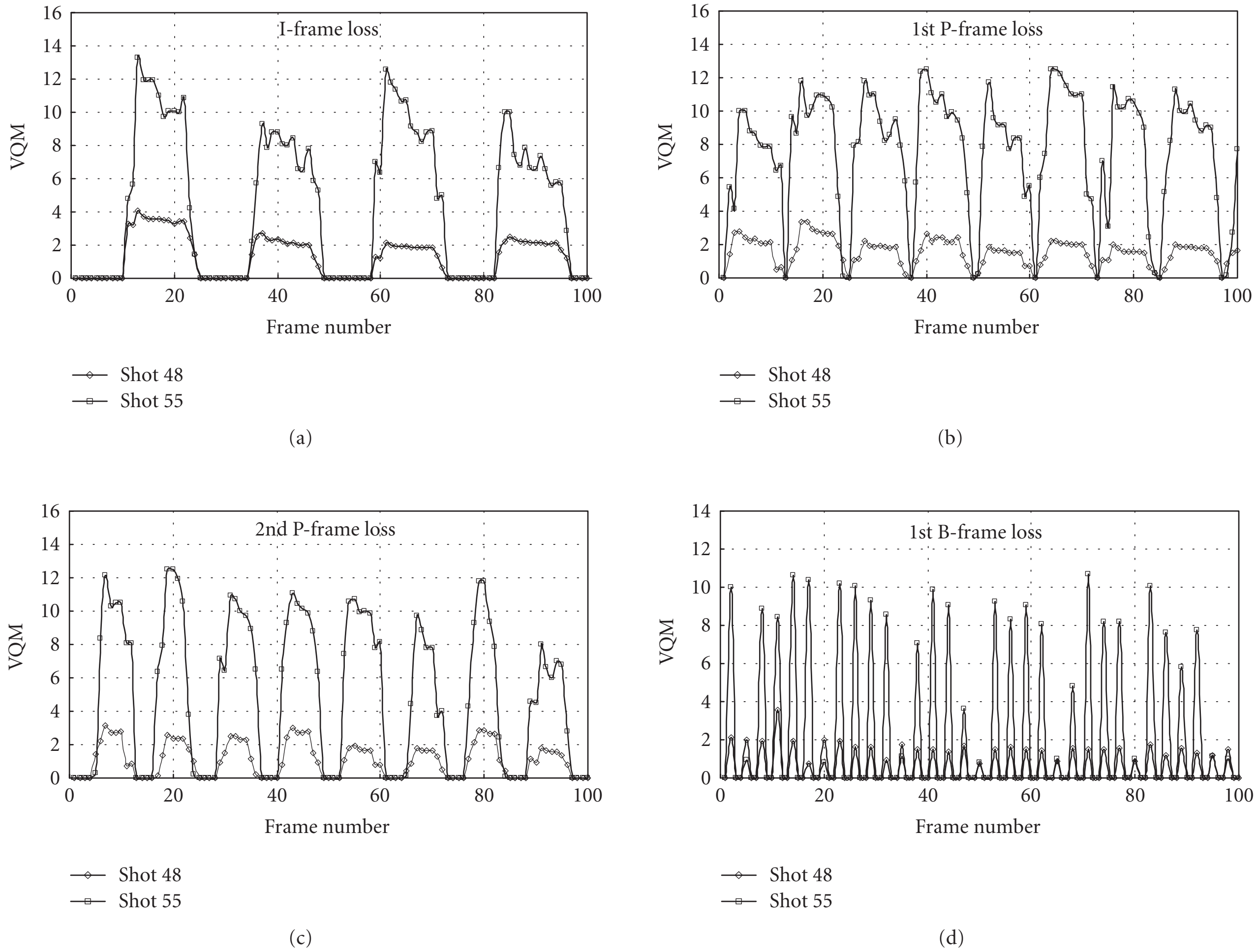


FIGURE 2: Quality degradation due to a frame loss in the underlying GoP for low motion activity level (shot 48) and moderately high motion activity level (shot 55) video.

### 3.2. Impact of frame loss

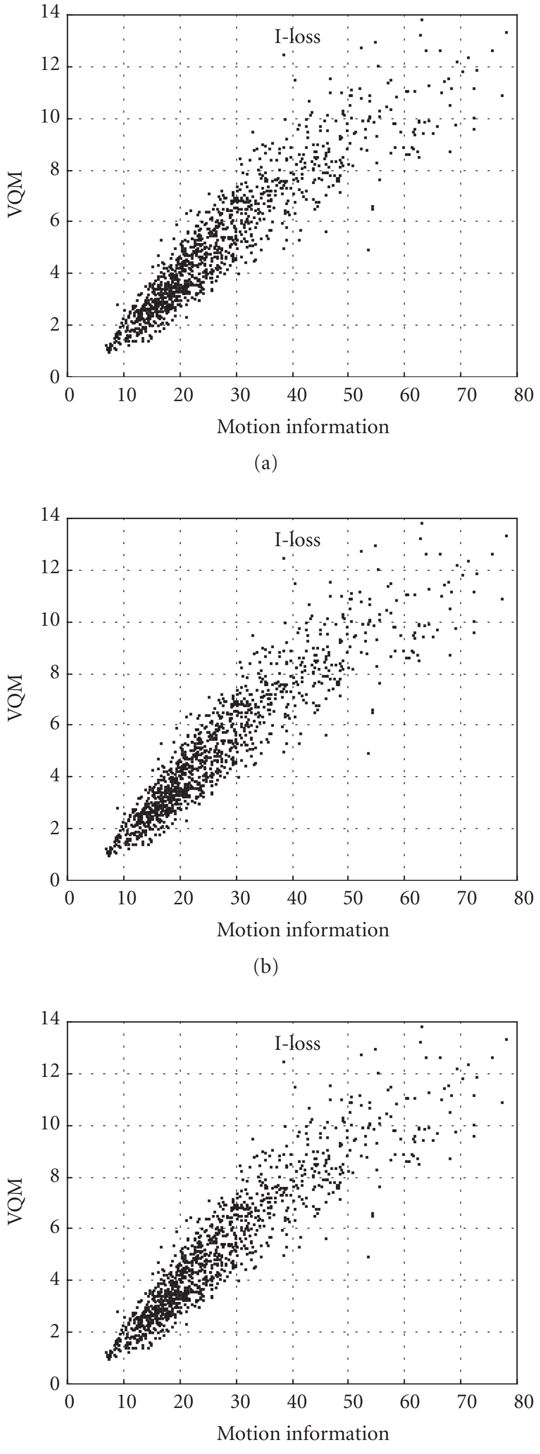
To illustrate the effect of a single frame loss in a GoP, which we focus on in this first study on advanced video traces, Figure 2 shows the quality degradation due to various frame loss scenarios, namely, I-frame loss, 1st P-frame loss in the underlying GoP, 2nd P-frame loss in the underlying GoP, and 1st B-frame loss between reference frames. Frame losses were concealed by copying from the previous (in decoding order) reference frame. We show the quality degradation for shot 48, which has a low motion activity level of 1, and for shot 55 which has moderately high motion activity level of 3. As expected, the results demonstrate that I-frame and P-frame losses propagate to all subsequent frames (until the next loss-free I-frame), while B-frame losses do not propagate. Note that Figure 2(b) shows the VQM values for the reconstructed video frames when the 1st P-frame in the GoP is lost, whereas Figure 2(c) shows the VQM values for the reconstructed frames when the 2nd P frame in the GoP is lost. As we observe, the VQM values due to losing the 2nd P-frame can generally be higher or lower than the VQM values due to losing the 1st P-frame. The visual content and the efficiency of the concealment scheme play a key role in determining the VQM values. Importantly, we also observe that a frame loss results in smaller quality degradations for low motion activity level video.

As illustrated in Figure 2, the quality degradation due to channel losses is highly correlated with the visual content of the affected frames. The challenge is to identify a representation of the visual content that captures both the spatial and the temporal variations between consecutive frames, in order to allow for accurate prediction of the quality degradation. The motion information descriptor $M(t)$ of [16], as given in (1), is a promising basis for such a representation and is therefore used as the starting point for our considerations.

### 3.3. Quality degradation at frame level

#### 3.3.1. Quality degradation of lost frame

We initially focus on the impact of a lost frame $t$ on the reconstructed quality of frame $t$ itself; the impact on frames



(a)

(b)

(c)

Figure 3: The relationship between the aggregate motion information of the lost frame $t$ and the quality degradation $Q(t)$ of the reconstructed frame.

that are coded with reference to the lost frame is considered in the following subsections. We conducted simulations of channel losses affecting I-frames (I-loss), P-frames (P-loss), and B-frames (B-loss). For both a lost I-frame $t$ and a lost P-frame $t$, we examine the correlation between the aggregate motion information $\mu(t)$ from the preceding reference frame $t$–$L$ to the lost frame $t$, as given by (2), and the quality degradation $Q(t)$ of the reconstructed frame (which is frame $t$–$L$ for concealment by copying).

Table 2: The correlation between motion information and quality degradation for lost frame.

| Frame type | Pearson correlation | Spearman correlation |
|---|---|---|
| I | 0.903 | 0.941 |
| P | 0.910 | 0.938 |
| B | 0.958 | 0.968 |

For a lost B-frame $t+m$, $m = 1, \ldots, L-1$, whereby frame $t$ is the preceding reference frame, we examine the correlation between the aggregate motion information from the closest reference frame to the lost frame and the quality degradation of the lost frame $t + m$. In particular, if $m \leq (L-1)/2$ we consider the aggregate motion information $\sum_{j=1}^{m} M(t+j)$, and if $m > (L-1)/2$ we consider $\sum_{j=m+1}^{L} M(t+j)$. (This aggregate motion information is slightly refined over the basic approximation given in (3). The basic approximation always conceals a lost B-frame by copying from the preceding frame, which may also be a B-frame. The preceding B-frame, however, may have been immediately flushed out of the decoder memory and may hence not be available for reference. The refined aggregate motion information approach presented here does not require reference to the preceding B-frame.)

Figure 3 shows the quality degradation $Q(t)$ (measured using VQM) as a function of the aggregate motion information for the different frame types. The results demonstrate that the correlation between the aggregate motion information and the quality degradation is high, which suggests that the aggregate motion information descriptor is effective in predicting the quality degradation of the lost frame.

For further validation, the correlation between the proposed aggregate motion information descriptors and the quality degradation $Q(t)$ (measured using VQM) was calculated using the Pearson correlation as well as the nonparametric Spearman correlation [20, 21]. Table 2 gives the correlation coefficients between the aggregate motion information and the corresponding quality degradation (i.e., the correlation between $x$-axis and $y$-axis of Figure 3). The highest correlation coefficients are achieved for the B-frames since in the considered GoP with $L - 1 = 2$ B-frames between successive P-frames, a lost B-frame can be concealed by copying from the neighboring reference frame, whereas a P- or I-frame loss requires copying from a reference frame that is three frames away.

Overall, the correlation coefficients indicate that the motion information descriptor is a relatively good estimator of the quality degradation of the underlying lost frame, and hence, the quality degradation of the lost frame itself is predicted with high accuracy by the functional approximation given in (3). Intuitively, note that in the case of little or no motion, the concealment scheme by copying is close to perfect, that is, there is only very minor quality degradation.

The motion information $M(t)$ reflects this situation by being close to zero; and the functional approximation of the quality degradation also gives a value close to zero. In the case of camera panning, the close-to-constant motion information $M(t)$ reflects the fact that a frame loss results in approximately the same quality degradation at any point in time in the panning sequence.

### 3.3.2. Analysis of loss propagation to subsequent frames for concealment by copying

Reference frame (I-frame or P-frame) losses affect not only the quality of the reconstructed lost frame but also the quality of reconstructed subsequent frames, even if these subsequent frames are correctly received. We analyze this loss propagation to subsequent frames in this and the following subsection. Since I-frame losses very severely degrade the reconstructed video qualities, video transmission schemes typically prioritize I-frames to ensure the lossless transmission of this frame type. We will therefore focus on analyzing the impact of a P-frame loss in a GoP on the quality of the subsequent frames in the GoP.

In this subsection, we present a mathematical analysis of the impact of a single P-frame loss in a GoP. We consider initially a decoder that conceals a frame loss by copying from the previous reference frame (frame freezing is considered in Section 4). The basic operation of the concealment by copying from the previous reference frame in the context of the frame loss propagation to subsequent frames is as follows. Suppose the I-frame at the beginning of the GoP is correctly received and the first P-frame in the GoP is lost. Then the second P-frame is decoded with respect to the I-frame (instead of being decoded with respect to the first P-frame). More specifically, the motion compensation information carried in the second P-frame (which is the residual error between the second and first P-frames) is "added" on to the I-frame. This results in an error since the residual error between the first P-frame and the I-frame is not available for the decoding. This decoding error further propagates to the subsequent P-frames as well as B-frames in the GoP.

To formalize these concepts, we introduce the following notation. We let $t$ denote the position in time of the lost P-frame and recall that there are $L-1$ B-frames between two reference frames and $K$ P-frames in a GoP. We index the I-frame and the P-frames in the GoP with respect to the position of the lost P-frame by $t+nL$, and let $R$, $R \le K-1$, denote the number of subsequent P-frames affected by the loss of P-frame $t$. In the above example, where the first P-frame in the GoP is lost, as also illustrated in Figure 4, the I-frame is indexed by $t-L$, the second P-frame by $t+L$, and $R=2$ P-frames are affected by the loss of the first P-frame. We denote the luminance values in the original frame as $F(t,i)$, in the loss-free frame after decoding as $\hat{F}(t,i)$, and in the reconstructed frame as $\tilde{F}(t,i)$. Our goal is to estimate the average absolute frame difference between $\hat{F}(t,i)$ and $\tilde{F}(t,i)$, which we denote by $\Delta(t)$. We denote $i_0, i_1, i_2, \ldots$ for the trajectory of pixel $i_0$ in the lost P-frame (with index $t+0L$) passing through the subsequent P-frames with indices $t+1L, t+2L, \ldots$.

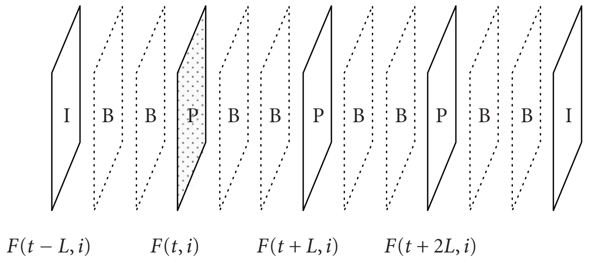


Figure 4: The GoP structure and loss model with a distance of $L = 3$ frames between successive P-frames and loss of the 1st P-frame.

#### 3.3.2.1 Analysis of quality degradation of subsequent P-frames

The pixels of a P-frame are usually motion-estimated from the pixels of the reference frame (which can be a preceding I-frame or P-frame). For example, the pixel at position $i_n$ in P-frame $t+nL$ is estimated from the pixel at position $i_{n-1}$ in the reference frame $t+(n-1)L$, using the motion vectors of frame $t+nL$. Perfect motion estimation is only guaranteed for still image video, hence a residual error (denoted as $e(t,i_n)$) is added to the referred pixel. In addition, some pixels of the current frame may be intra-coded without referring to other pixels. Formally, we can express the encoded pixel value at position $i_n$ of a P-frame at time instance $t+nL$ as

$$\begin{aligned}\hat{F}(t+nL,i_n) &= A(t+nL,i_n)\hat{F}(t+(n-1)L,i_{n-1}) \\ &\quad + e(t+nL,i_n), \quad n = 1,2,\ldots,R,\end{aligned} \tag{8}$$

where $A(t+nL,i_n)$ is a Boolean function of the forward motion vector and is set to 0 if the pixel is intra-coded. This equation can be applied recursively from a subsequent P-frame backwards until reaching the lost frame $t$, with luminance values denoted by $\hat{F}(t,i_0)$. The resulting relationship between the encoded values of the P-frame pixels at time $t+nL$ and the values of the pixels in the lost frame is

$$\begin{aligned}&\hat{F}(t+nL,i_n) \\ &\quad = \hat{F}(t,i_0)\prod_{j=0}^{n-1} A(t+(n-j)L,i_{n-j}) \\ &\qquad + \sum_{k=0}^{n-1} e(t+(n-k)L,i_{n-k})\prod_{j=0}^{k-1} A(t+(n-j)L,i_{n-j}).\end{aligned} \tag{9}$$

This exact analysis is rather complex and would require a verbose content description, which in turn could provide a rather exact estimation of the quality degradation. A verbose content description, however, would result in complex verbose advanced video traces, which would be difficult to employ by networking researchers and practitioners in evaluations of video transport mechanisms. Our objective is to find a parsimonious content description that captures the main content features to allow for an approximate prediction of the quality degradation. We examine therefore the following approximate recursion:

$$\widehat{F}(t+nL,i_n) \approx \widehat{F}(t+(n-1)L,i_{n-1}) + e(t+nL,i_n). \quad (10)$$

The error between the approximated and exact pixel value can be represented as:

$$\zeta(t+nL,i_k) = \begin{cases} F(t+nL,i_k) & \text{if } A(t+nL,i_k) = 0 \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

This approximation error in the frame representation is negligible for P-frames, in which few blocks are intra-coded. Generally, the number of intra-coded blocks monotonically increases as the motion intensity of the video sequence increases. Hence, the approximation error in frame representation monotonically increases as the motion intensity level increases. In the special case of shot boundaries, all the blocks are intra-coded. In order to avoid a high prediction error at shot boundaries, we introduce an I-frame at each shot boundary regardless of the GoP structure.

After applying the approximate recursion, we obtain

$$\widehat{F}(t+nL,i_n) \approx \widehat{F}(t,i_0) + \sum_{j=0}^{n-1} e(t+(n-j)L,i_{n-j}). \quad (12)$$

Recall that the P-frame loss (at time instance $t$) is concealed by copying from the previous reference frame (at time instance $t$–$L$), so that the reconstructed P-frames (at time instances $t+nL$) can be expressed using the approximate recursion as

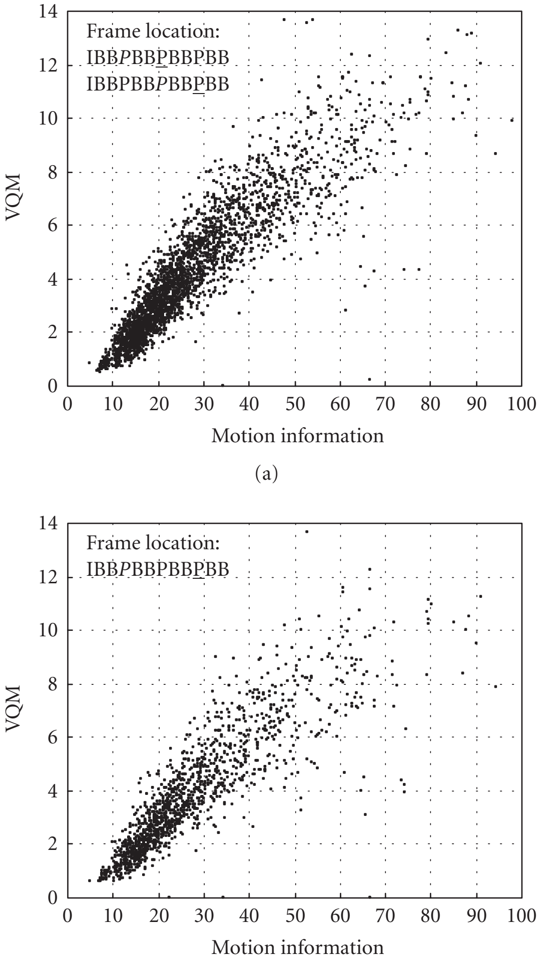
$$\widetilde{F}(t+nL,i_n) \approx \widehat{F}(t-L,i_0) + \sum_{j=0}^{n-1} e(t+(n-j)L,i_{n-j}). \quad (13)$$

Thus, the average absolute differences between the reconstructed P-frames and the loss-free P-frames are given by

$$\begin{aligned} \Delta(t+nL) &= \frac{1}{N}\sum_{i_n=1}^{N} \left| \widehat{F}(t+nL,i_n) - \widetilde{F}(t+nL,i_n) \right| \\ &= \frac{1}{N}\sum_{i_0=1}^{N} \left| \widehat{F}(t,i_0) - \widehat{F}(t-L,i_0) \right|. \end{aligned} \quad (14)$$

The above analysis suggests that there is a high correlation between the aggregate motion information $\mu(t)$, given by (2) of the lost P-frame, and the quality degradation, given by (11), of the reconstructed P-frames. The aggregate motion information $\mu(t)$ is calculated between the lost P-frame and its preceding reference frame, which are exactly the two frames that govern the difference between the reconstructed frames and the loss-free frames according to (11).

Figure 5 illustrates the relationship between the quality degradation of reconstructed P-frames measured in terms of the VQM metric and the aggregate motion information $\mu(t)$ for the video sequences of the *Jurassic Park* movie for a GoP with $L = 3$ and $K = 3$. The quality degradation of the P-frame at time instance $t+3$ and the quality degradation of the P-frame at time instance $t+6$ are considered. The Pearson correlation coefficients for these relationships (between $x$-axis and $y$-axis data in Figure 5) are 0.893 and 0.864, respectively, which supports the suitability of motion information descriptors for estimating the P-frame quality degradation.



(a)

(b)

Figure 5: The relationship between the quality degradations $Q(t+3)$ and $Q(t+6)$ and the aggregate motion information $\mu(t)$ (the lost frame is indicated in italic font, while the considered affected frame is underlined).

### 3.3.2.2 Analysis of quality degradation of subsequent B-frames

For the analysis of the loss propagation to B-frames, we augment the notation introduced in the preceding subsection by letting $t+m$ denote the position in time (index) of the considered B-frame. The pixels of B-frames are usually motion-estimated from two reference frames. For example, the pixel at position $k_m$ in the frame with index $t+m$ may be estimated from a pixel at position $i_{n-1}$ in the previous reference frame with index $t$ and from a pixel at position $i_n$ in the next reference frame with index $t+L$. Forward motion vectors are used to refer to the previous reference frame, while backward motion vectors are used to refer to the next reference frame. Due to the imperfections of the motion estimation, a residual error $e(t,k)$ is needed. The luminance value of the pixel at position $k_m$ of a B-frame at time instance $t+m$ can thus be expressed as

$$\begin{aligned}\widehat{F}(t+m,k_m) &= v_f(t+m,k_m)\widehat{F}(t+(n-1)L,i_{n-1})\\ &\quad + v_b(t+m,k_m)\widehat{F}(t+nL,i_n)+e(t+m,k_m),\end{aligned} \tag{15}$$

where $m = -(L-1), -(L-2), \dots, -1, 1, 2, \dots, (L-1), L+1, \dots, 2L-1, \dots 2L+1, \dots 2L+L-1, \dots (K-1)L+1, \dots, (K-1)L+L-1, n = \lceil (m/L) \rceil$, and $v_f\ (t,k)$ and $v_b(t,k)$ are the indicator variables of forward and backward motion prediction as defined in Subsection 2.2.

There are three different cases to consider.

*Case 1.* The pixels of the considered B-frame are referencing the error-free frame by forward motion vectors and the lost P-frame with backward motion vectors. Using the approximation of P-frame pixels (12), the B-frame pixels can be represented as

$$\begin{aligned}\widehat{F}(t+m,k_m) &= v_f(t+m,k_m)\widehat{F}(t-L,i_{-1})\\ &\quad + v_b(t+m,k_m)\widehat{F}(t,i_0)+e(t+m,k_m).\end{aligned} \tag{16}$$

The lost P-frame at time instance $t$ is concealed by copying from the previous reference frame at time instance $t$–$L$. The reconstructed B-frames can thus be expressed as

$$\begin{aligned}\widetilde{F}(t+m,k_m) &= v_f(t+m,k_m)\widehat{F}(t-L,i_{-1})\\ &\quad + v_b(t+m,k_m)\widehat{F}(t-L,i_0)+e(t+m,k_m).\end{aligned} \tag{17}$$

Hence, the average absolute difference between the reconstructed B-frame and the loss-free B-frame is given by

$$\Delta(t+m) = \frac{1}{N}\sum_{k_m=1}^{N} v_b(t+m,k_m)\left|\widehat{F}(t,i_0)-\widehat{F}(t-L,i_0)\right|. \tag{18}$$

*Case 2.* The pixels of the considered B-frame are motion-estimated from reference frames, both of which are affected by the P-frame loss. Using the approximation of the P-frame pixels (12), the B-frame pixels can be represented as

$$\begin{aligned}&\widehat{F}(t+m,k_m)\\ &\quad = v_f(t+m,k_m)\left[\widehat{F}(t,i_0)+\sum_{j=0}^{n-2}e(t+(n-j)L,i_{n-j})\right]\\ &\qquad + v_b(t+m,k_m)\left[\widehat{F}(t,i_0)+\sum_{j=0}^{n-1}e(t+(n-j)L,i_{n-j})\right]\\ &\qquad + e(t+m,k_m).\end{aligned} \tag{19}$$

The vector $(i_{n-1}, i_{n-2}, \dots, i_0)$ represents the trajectory of pixel $k_m$ using backward motion estimation until reaching the lost P-frame, while the vector $(i_{n-2}, i_{n-3}, \dots, i_0)$ represents the trajectory of pixel $k_m$ using forward motion estimation until reaching the lost P-frame. P-frame losses are concealed by copying from the previous reference frame, so that the reconstructed B-frame can be expressed as

$$\begin{aligned}&\widetilde{F}(t+m,k_m)\\ &\quad = v_f(t+m,k_m)\left[\widehat{F}(t-L,i_0)+\sum_{j=0}^{n-2}e(t+(n-j)L,i_{n-j})\right]\\ &\qquad + v_b(t+m,k_m)\left[\widehat{F}(t-L,i_0)+\sum_{j=0}^{n-1}e(t+(n-j)L,i_{n-j})\right]\\ &\qquad + e(t+m,k_m).\end{aligned} \tag{20}$$

Thus, the average absolute difference between the reconstructed B-frame and the loss-free B-frame is given by

$$\begin{aligned}\Delta(t+m) &= \frac{1}{N}\sum_{k_m=1}^{N}\left(v_b(t+m,k_m)+v_f(t+m,k_m)\right)\\ &\qquad \times\left|\widehat{F}(t,i_0)-\widehat{F}(t-L,i_0)\right|.\end{aligned} \tag{21}$$

*Case 3.* The pixels of the considered B-frame are referencing the error-free frame (i.e., I-frame of next GoP) by backward motion vectors and to the lost P-frame using forward motion vectors. Using the approximation of the P-frame pixels (12), the B-frame pixels can be represented as

$$\begin{aligned}\widehat{F}(t+m,k_m) &= v_f(t+m,k_m)\widehat{F}(t+RL,i_R)\\ &\quad + v_b(t+m,k_m)\widehat{F}(t+(R+1)L,i_{R+1})\\ &\quad + e(t+m,k_m),\\ &\widehat{F}(t+m,k_m)\\ &\quad = v_f(t+m,k_m)\left[\widehat{F}(t,i_0)+\sum_{j=0}^{R-1}e(t+(R-j)L,i_{R-j})\right]\\ &\qquad + v_b(t+m,k_m)\widehat{F}(t+(R+1)L,i_{R+1})+e(t+m,k_m),\end{aligned} \tag{22}$$

where $R$ is the number of affected (subsequent) P-frames that are affected by the P-frame loss at time instance $t$ and $\widehat{F}(t+(R+1)L,i)$ is the I-frame of the next GoP.

The reconstructed B-frames can be expressed as

$$\begin{aligned}&\widetilde{F}(t+m,k_m)\\ &\quad = v_f(t+m,k_m)\left[\widehat{F}(t-L,i_0)+\sum_{j=0}^{R-1}e(t+(R-j)L,i_{R-j})\right]\\ &\qquad + v_b(t+m,k_m)\widehat{F}(t+(R+1)L,i_{R+1})+e(t+m,k_m).\end{aligned} \tag{23}$$

Thus, the average absolute difference between the reconstructed B-frame and the loss-free B-frame is given by

$$\Delta(t+m) = \frac{1}{N}\sum_{k_m=1}^{N} v_f(t+m,k_m)\,\big|\hat{F}(t,i_0) - \hat{F}(t-L,i_0)\big|. \tag{24}$$

The preceding analysis suggests that the following aggregate motion information descriptors achieve a high correlation with the quality degradation of the B-frames.

$$\begin{aligned}
Case1:\ \mu(t+m) &= \left(\sum_{j=0}^{L-1} M(t-j)\right)\frac{1}{N}\sum_{k_m=1}^{N} v_b(t+m,k_m).\\
Case2:\ \mu(t+m) &= \left(\sum_{j=0}^{L-1} M(t-j)\right)\frac{1}{N}\\
&\quad\times\sum_{k_m=1}^{N}\big(v_b(t+m,k_m)+v_f(t+m,k_m)\big).\\
Case3:\ \mu(t+m) &= \left(\sum_{j=0}^{L-1} M(t-j)\right)\frac{1}{N}\sum_{k_m=1}^{N} v_f(t+m,k_m).
\end{aligned} \tag{25}$$

The first summation term in these equations represents the aggregate motion information $\mu(t)$ between the lost P-frame and its preceding reference frame (see (2)). The second summation term represents the ratio of the backward motion estimation $V_b(t+m)$, the ratio of non-intra-coding (which we approximate as one in the proposed prediction method), and the ratio of forward motion estimation $V_f(t+m)$ in the B-frame, respectively, as summarized in (5)–(7).

Figure 6 shows the correlation between the aggregate motion information $\mu(t+m)$ and the quality degradation of B-frames for the loss scenario presented in Figure 4. The Pearson correlation coefficients for these relationships (shown in Figure 6) are 0.899, 0.925, 0.905, and 0.895, respectively, which indicates the ability of the motion information descriptors to estimate the reconstructed qualities of the affected B-frames.

### 3.4. Quality degradation at GoP level

The frame level predictor requires a predictor for each frame in the GoP. This fine-grained level of quality prediction may be overly detailed for practical evaluations and be complex for some video communication schemes. Another quality predictor can be applied at the GoP level, whereby the quality degradation is estimated for the entire GoP. When a frame loss occurs in a GoP, a summarization of the motion information across all affected frames of the GoP is computed. This can be accomplished by using (2), (5), (6), and (7), and averaging over all $((R+2)L-1)$ frames that suffer a quality degradation due to a P-frame loss at time instance $t$:

$$\mu = \frac{1}{(R+2)L-1}\sum_{n=-(L-1)}^{RL-1}\mu(t+n). \tag{26}$$

To see this, recall that $R$ P-frames are affected by the loss due to error propagation from the lost P-frame, for a total of $R+1$ P-frames with quality degradations. Also, recall that $(L-1)$ B-frames are coded between P-frames for a total of $(R+2)(L-1)$ affected B-frames.

Figure 7 shows the average quality degradation (measured using the VQM metric) for the GoP, where the $x$-axis represents the summarization of the motion information $\mu$. Three illustrative simulations were conducted, corresponding to 1st P-frame loss, 2nd P-frame loss, and 3rd P-frame loss. Similarly to the functional approximations of Subsection 2.2.2, the quality degradation of the GoP can be approximated by a linear or logarithmic function of the averaged aggregate motion information $\mu$. The functional approximations can be represented by the triplets ($\varphi_r^{\mathrm{GoP}}$, $a_r^{\mathrm{GoP}}$, $b_r^{\mathrm{GoP}}$), $r = 1,\ldots,K$.

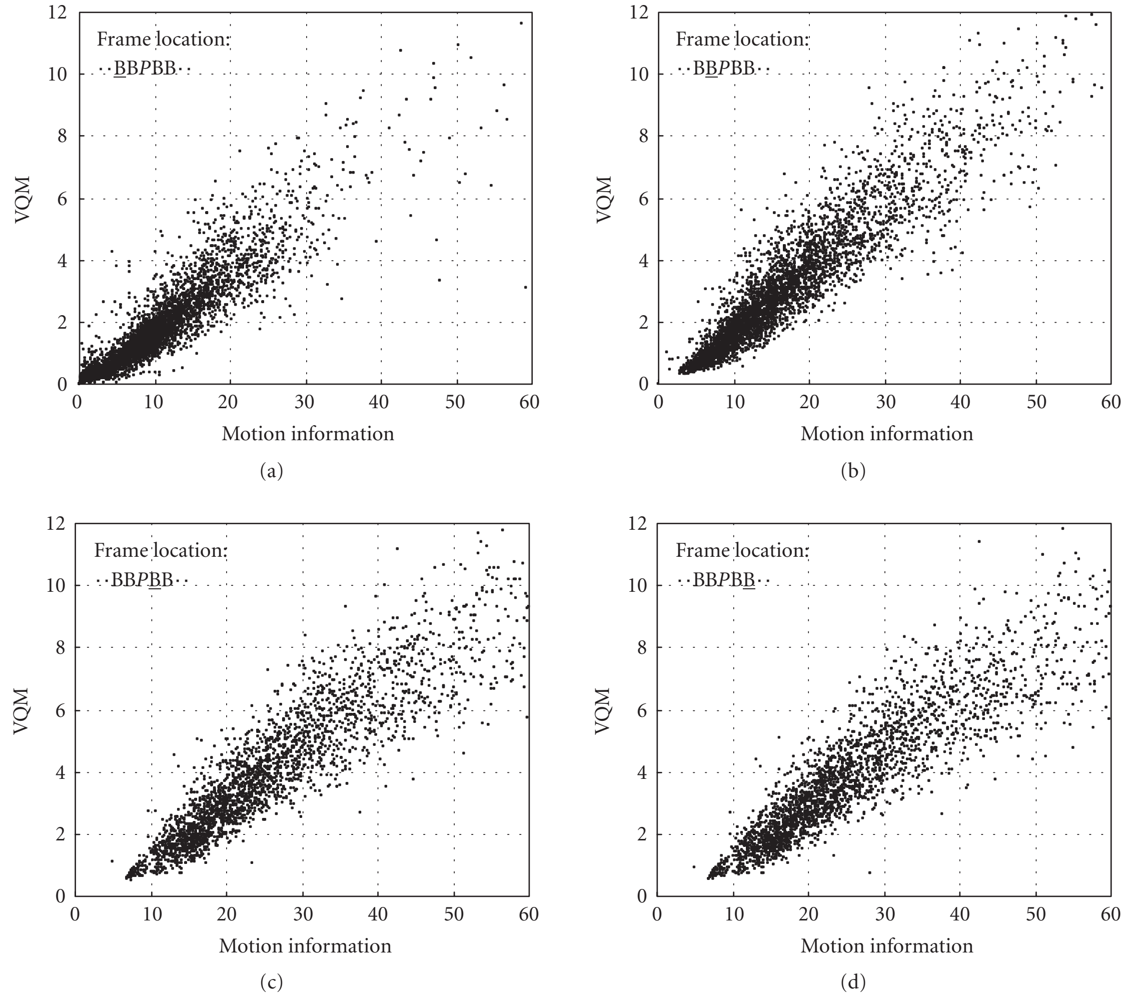
### 3.5. Quality degradation at shot level

The next coarser level in the logical granularity of a video sequence after the GoP level is the shot level, which can provide networking researchers with a rough approximation of the reconstructed quality. For the shot level analysis, we employ the motion activity level $\theta$, which correlates well with the human perception of the motion intensity in the shot.

Table 3 shows the average quality degradation (per affected frame in the entire video shot) using the VQM metric for various shot activity levels, for 3 different types of P-frame losses (1st P-frame loss, 2nd P-frame loss, or 3rd P-frame loss). Frame losses in shots with high motion activity levels result in more severe quality degradation, compared to the relatively mild degradation of shots with low motion activity levels. Table 3 also illustrates that the average quality degradation of a shot depends on the position of the lost frame. For example, the average quality degradation when losing the 2nd P-frame is 3.84, while the average quality degradation when losing the 3rd P-frame is 3.45. Therefore, when a video shot experiences a P-frame loss, the quality degradation can be determined (using Table 3) based on the location of the P-frame loss as well as the motion activity level of the video shot. For each video in the video trace library, a table that follows the template of Table 3 can be used to approximate the quality degradation in the video shot.

## 4. ANALYSIS OF QUALITY DEGRADATION WITH LOSS CONCEALMENT BY FREEZING

In this section, we consider a decoder that conceals lost frames by freezing the last correctly received frame, until a correct I-frame is received. If a P-frame at time instance $t$ is lost, the reference frame from time instance $t$–$L$ is displayed at all time instances $t+n$, where $n = -(L-1), -(L-2),\ldots,0,1,2,\ldots$. In other words, all received frames at time instances $t+n$ are not decoded but replaced with the reference frame at time instance $t$–$L$. This technique of loss concealment, while simple, results typically in quite significant temporal quality degradation, in contrast to the relatively moderate temporal and spatial quality degradation of the loss concealment by copying considered in the previous



(a)

(b)

(c)

(d)

Figure 6: The relationship between the quality degradations $Q(t-2)$, $Q(t-1)$, $Q(t+1)$, and $Q(t+2)$, and the aggregate motion information $\mu(t-2)$, $\mu(t-1)$, $\mu(t+1)$, and $\mu(t+2)$, respectively (the lost frame is indicated in italic font, while the considered affected frame is underlined).

section. For the GoP structure in Figure 4, for instance, if the 2nd P-frame is lost during transmission, 8 frames will be frozen. Human viewers perceive such quality degradation as jerkiness in the normal flow of the motion. We use a perceptual metric, namely, VQM, to estimate this motion jerkiness in our illustrative experiments since a perceptual metric is better suited than the conventional PSNR metric for measuring this quality degradation. In the following, we present the method for calculating the composite motion information for the frozen frames.

Assuming that the P-frame at time instance $t$ is lost during the video transmission and that there are $R$ affected P-frames $t+L, \dots, t+RL$ in the GoP before the next I-frame, the reference frame at time instance $t$–$L$ is frozen for a total of $RL+2L-1$ frames. The difference between the error-free frames and the frozen frames can be calculated as

$$\Delta(t+n) = \frac{1}{N}\sum_{i=1}^{N} \left| \widehat{F}(t+n,i) - \widehat{F}(t-L,i) \right| \qquad (27)$$

for $n = -(L-1), -(L-2), \dots, 0, 1, 2, \dots, RL+L-1$.

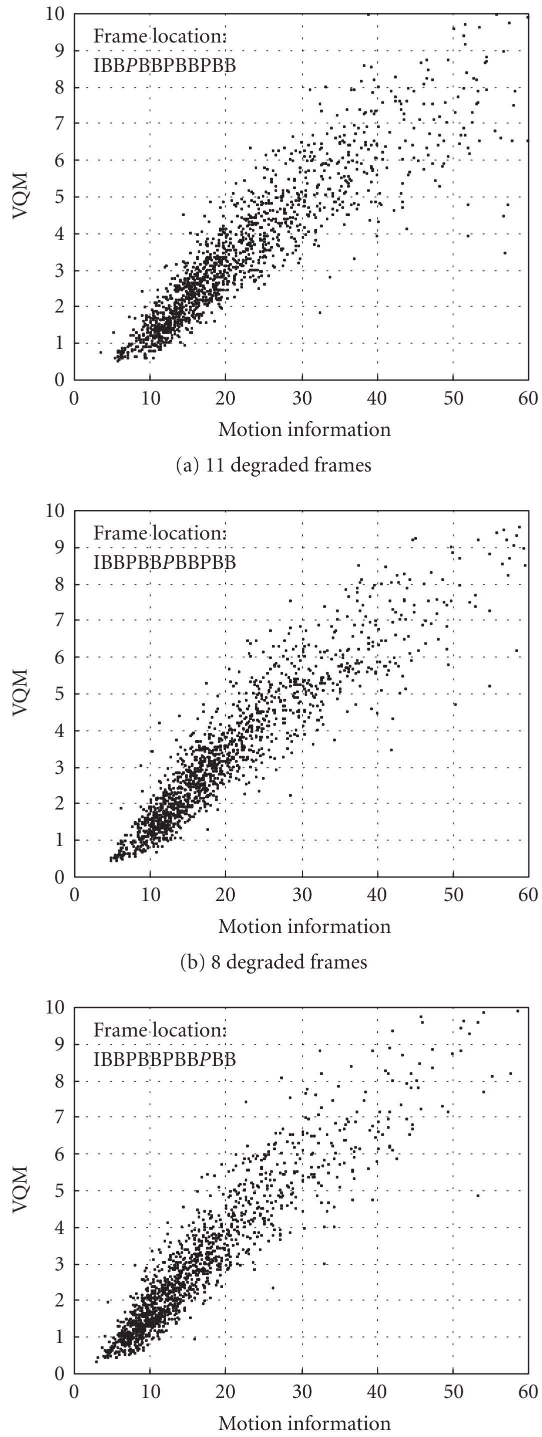
This equation demonstrates that the quality degradation for this type of decoder can be estimated from the motion information between the error-free frame $t+n$ and the frozen frame $t-L$. This effect is captured with the aggregate motion information descriptor

$$\gamma(t+n) = \sum_{k=-(L-1)}^{n} M(t+k). \qquad (28)$$

The degree of temporal quality degradation depends on the length of the sequence of frozen frames as well as the amount of lost motion information. Therefore, estimating the quality degradation for each individual frozen frame is not useful. Instead, we consider a GoP level predictor and a shot level predictor.

### 4.1. Quality degradation at GoP level

The GoP level predictor estimates the quality degradation based on the $\gamma(t+n)$ motion information averaged over all the frozen frames, namely, based on the average aggregate



(a) 11 degraded frames

(b) 8 degraded frames

(c) 5 degraded frames

Figure 7: The relationship between the average quality degradation in the GoP and the average aggregate motion information $\mu$ using concealment by copying (the lost frame is indicated in italic font).

motion information

$$\gamma = \frac{1}{RL + 2L - 1} \sum_{i=-(L-1)}^{RL+L-1} \gamma(t+n). \tag{29}$$

Table 3: The average quality degradation (per affected frame) for each motion activity level for shots from *Jurassic Park* with concealment by copying.

| Activity level | Video shots # | 1st P-frame loss | 2nd P-frame loss | 3rd P-frame loss |
|---|---|---|---|---|
| 1 | 12 | 1.670 | 1.558 | 1.354 |
| 2 | 24 | 2.967 | 2.813 | 2.443 |
| 3 | 45 | 4.459 | 4.425 | 3.989 |
| 4 | 12 | 5.359 | 5.461 | 5.199 |
| 5 | 2 | 7.264 | 7.451 | 5.968 |
| All shots | 95 | 3.896 | 3.844 | 3.455 |

The quality degradation can be approximated as a linear or logarithmic function of $\gamma$.

Figure 8 shows the relationship between the average quality degradation of the underlying GoP, and the average aggregate motion information descriptor for different P-frame loss scenarios. The Pearson correlation coefficients for these relationships are 0.929 for freezing the 2nd P-frame, and 0.938 for freezing the 3rd P frame. According to the GoP structure shown in Figure 4, the 1st P-frame loss results in the freezing of 11 frames of the GoP, and therefore reduces the frame rate from 30 fps to 2.5 fps. This is very annoying to human perception and it is not considered in our study.

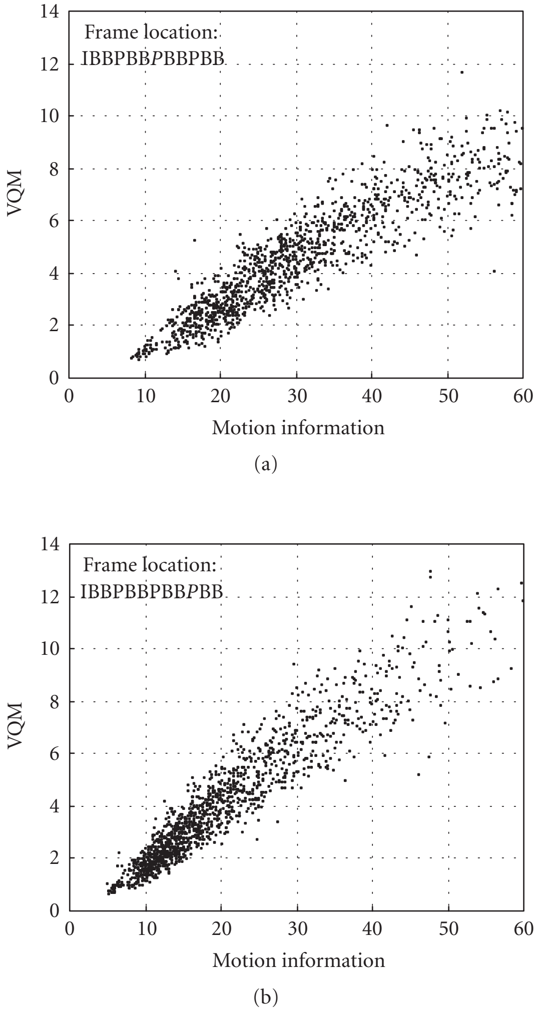
### 4.2. *Quality degradation at shot level*

Table 4 shows the average quality degradation (per affected frame) for video shots of various motion activity levels. We consider the quality degradation due to losing the 2nd P-frame, and the quality degradation due to losing the 3rd P-frame.

Freezing lost frames for shots of high motion activity levels results in more severe quality degradation, compared to shots of low motion activity levels. In addition, the average quality degradation is affected by the position of the lost frame. Comparing with Table 3, we observe that the quality degradation due to losing the 2nd P-frame is 3.84 for decoders that conceal frame losses by copying, while the quality degradation due to losing the 2nd P-frame is 5.45 for decoders that conceal frame losses by freezing. For this quality predictor, when a video shot experiences a P-frame loss, the quality degradation is determined (using Table 4) based on the location of the P-frame loss as well as the motion activity level of the video shot.

## 5. EVALUATION OF QUALITY PREDICTION USING VQM METRIC

In this and the following section, we conduct an extensive performance evaluation of the various quality predictors, derived in Sections 3 and 4. The video quality is measured with the VQM metric in this section and with PSNR in the following section. The accuracy of the quality predictor (which is implemented using the advanced video traces) is



(a)

(b)

Figure 8: The relationship between the average quality degradation $Q(t)$ in the GoP and the average aggregate motion information $\gamma$ using concealment by frame freezing (the lost frame is indicated in italic font).

compared with the actual quality degradation, determined from experiments with the actual video bit streams. The video test sequences used in the evaluation in this section are extracted from the *Jurassic Park I* movie as detailed in Subsection 3.1. In Subsection 5.1 we consider error concealment by copying from the previous reference frame (as analyzed in Section 3) and in Subsection 5.2 we consider error concealment by frame freezing (as analyzed in Section 4).

### 5.1. Evaluation of quality prediction for loss concealment by copying

P-frame losses are the most common type of frame losses that have a significant impact on the reconstructed quality. We have therefore conducted three different evaluations, corresponding to 1st P-frame loss, 2nd P-frame loss, and 3rd P-frame loss.

Table 4: The average quality degradation for each shot activity level (freezing).

| Activity level | 2nd P-frame freezing | 3rd P-frame freezing |
|---|---|---|
| 1 | 2.389 | 1.936 |
| 2 | 4.115 | 3.306 |
| 3 | 6.252 | 5.239 |
| 4 | 7.562 | 6.748 |
| 5 | 9.524 | 7.914 |
| All shots | 5.450 | 4.573 |

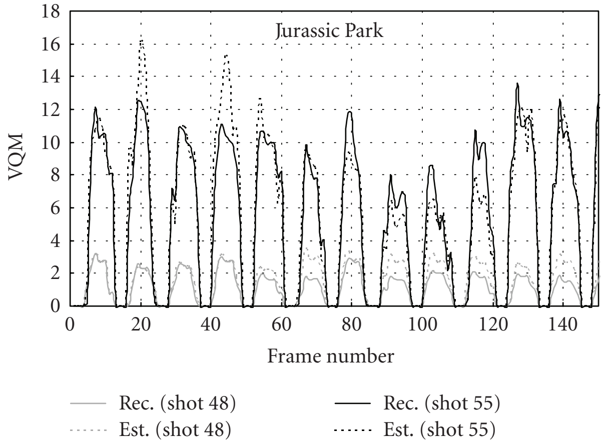


Figure 9: Comparison between actual reconstructed quality and estimated quality per each frame (2nd P-frame is lost in each GoP) for concealment by copying.

#### 5.1.1. Prediction at frame level

Figure 9 shows a comparison between the proposed scheme for frame level quality prediction (est.) (see Subsections 2.2.3 and 3.3) and the actual reconstructed quality (rec.) due to the loss of the 2nd P-frame. We observe from the figure that the proposed frame level prediction scheme provides overall a relatively good approximation of the actual quality degradation. The accuracy of the frame level predictor is examined in further detail in Table 5, which gives the average (over the entire video sequence) of the absolute difference between the actual reconstructed quality and the predicted quality. The frame level predictor can achieve an accuracy of about ±0.65 for predicting the quality degradation of losing the 2nd P-frame, where the average actual quality degradation is about 3.844 (see Table 3) using the VQM metric. We observe that better accuracy is achieved when video shots have a lower motion activity level. For high motion activity videos, the motion information is typically high. As we observe from Figures 5 and 6, the quality degradation values are scattered over a wider range for high motion information values. Hence, approximating the quality degradation of this high motion information by a single value results in larger prediction errors.

Table 5: The absolute difference (in VQM) between actual reconstructed quality and estimated quality using frame level analysis for concealment by copying.

| Activity level | 1st P-frame loss | 2nd P-frame loss | 3rd P-frame loss |
|---|---|---|---|
| 1 | 0.434 | 0.361 | 0.315 |
| 2 | 0.600 | 0.578 | 0.469 |
| 3 | 0.859 | 0.807 | 0.764 |
| 4 | 1.252 | 1.326 | 1.309 |
| 5 | 1.871 | 1.948 | 1.948 |
| All shots | 0.696 | 0.650 | 0.607 |

Table 6: The absolute difference (in VQM) between actual reconstructed quality and estimated quality using GoP level analysis for concealment by copying.

| Activity level | 1st P-frame loss | 2nd P-frame loss | 3rd P-frame loss |
|---|---|---|---|
| 1 | 0.418 | 0.369 | 0.363 |
| 2 | 0.568 | 0.542 | 0.477 |
| 3 | 0.761 | 0.676 | 0.699 |
| 4 | 1.055 | 1.231 | 1.386 |
| 5 | 1.660 | 1.876 | 1.683 |
| All shots | 0.643 | 0.607 | 0.606 |

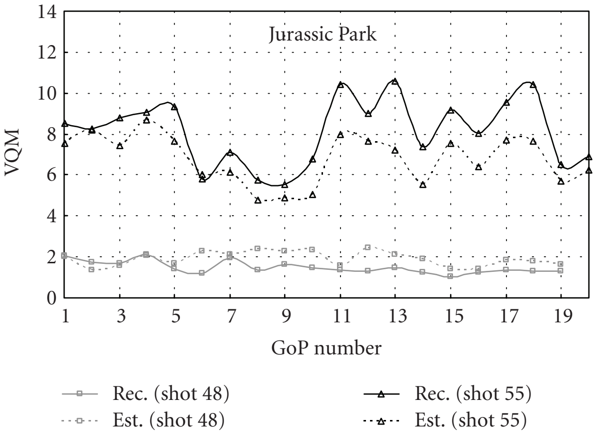


Figure 10: Comparison between actual reconstructed quality and estimated quality per each GoP (2nd P-frame is lost in each GoP) for concealment by copying.

The position of the lost frame has a significant impact on the accuracy of the quality prediction. For example, the accuracy of the quality predictor increases when fewer frames are affected. In particular, when losing the 1st P-frame the accuracy of the quality prediction is around ±0.6963, while it is around ±0.6073 when losing the 3rd P-frame. The activity levels 4 and 5 do not follow this general trend of increasing prediction accuracy with fewer affected frames, which is primarily due to the small number of shots of activity levels 4 and 5 in the test video (see Table 3) and the resulting small statistical validity. The more extensive evaluations in Section 6 confirm the increasing prediction accuracy with decreasing the number of affected frames for all activity levels (see in particular Tables 11, 12, and 13).

#### *5.1.2. Prediction at GoP level*

Figure 10 shows the performance of the GoP level predictor (see Subsection 3.4), compared to the actual quality degradation. The performance over two video shots of motion activity level 1 (shot 48), and of motion activity level 3 (shot 55) is shown. Table 6 shows the average absolute difference between the GoP quality predictor that uses the advanced video traces and the actual quality degradation. Similarly to the frame level predictor, Table 6 shows that better accuracy is achieved when shots are of lower motion activity level. Comparing the results shown in Tables 5 and 6, we observe that more accurate estimates of the quality degradation are provided by GoP level predictors. This is because the frame level predictor estimates the quality degradation for each frame type and for each frame position in the GoP, which results in an accumulated estimation error for the entire GoP. On the other hand, the GoP level predictor estimates the quality degradation for a GoP by a single approximation. In the case of 1st P-frame loss (where 11 frames are affected by the frame loss and hence 11 approximations are used for the frame level predictor), the accuracy of the GoP level predictor is about 0.643, while the accuracy of the frame level predictor is about 0.696. However, in the case of 3rd P-frame loss (where only 5 frames are affected by the frame loss), the reduction of the estimation error with the GoP level predictor is marginal.

#### *5.1.3. Prediction at shot level*

Figure 11(a) shows the performance of the shot level predictor (see Subsection 3.5) compared to the actual quality degradation, when the 2nd P-frame in each GoP is lost during video transmission. Figure 11(b) shows the motion activity level for each video shot. Table 7 shows the accuracy of the shot level predictor. Similarly to frame level and GoP level predictors, improvements in predicting the quality degradation are achieved with shots of lower motion activity level. In general, the accuracy of the shot level predictor is improved when a frame loss is located close to the subsequent correctly received I-frame, because it does not affect many subsequent frames. Comparing the results of Tables 5, 6, and 7, the quality prediction using shot level analysis does not provide any added accuracy compared to the quality prediction using frame level analysis, or the quality prediction using GoP level analysis. The quality prediction using the GoP level analysis is the best, in terms of the accuracy of the quality degradation estimate, and the speed of the calculation.

### *5.2. Evaluation of quality prediction for loss concealment by freezing*

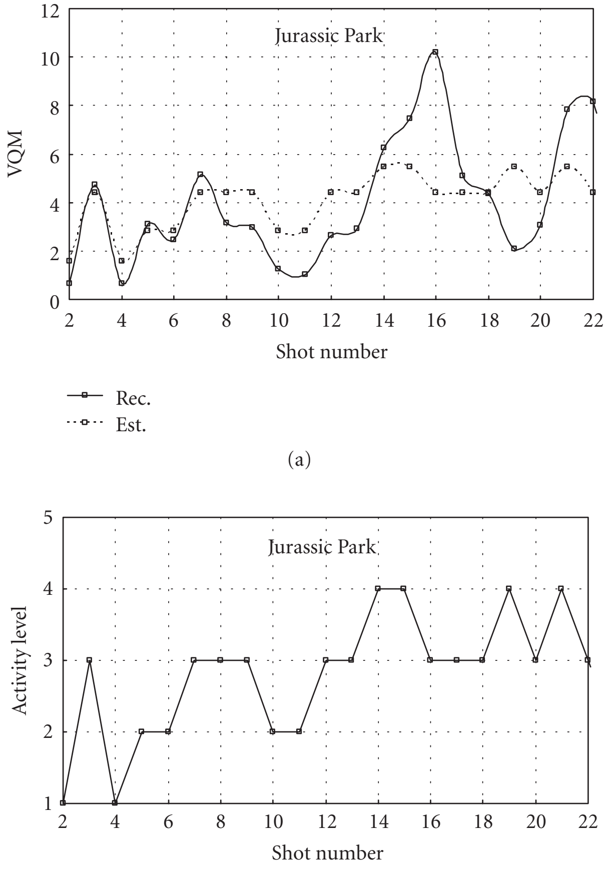
Two different evaluations were conducted, corresponding to 2nd P-frame loss and 3rd P-frame loss.

Table 7: The absolute difference (in VQM) between actual reconstructed quality and estimated quality using shot level analysis for concealment by copying.

| Activity level | 1st P-frame loss | 2nd P-frame loss | 3rd P-frame loss |
|---|---|---|---|
| 1 | 0.719 | 0.619 | 0.586 |
| 2 | 1.216 | 1.187 | 1.092 |
| 3 | 1.647 | 1.597 | 1.529 |
| 4 | 1.976 | 2.015 | 2.356 |
| 5 | 2.638 | 2.070 | 2.204 |
| All shots | 1.482 | 1.431 | 1.417 |

Table 8: The absolute difference (in VQM) between actual reconstructed quality and estimated quality using GoP level analysis for concealment by freezing.

| Activity level | 2nd P-frame freezing | 3rd P-frame freezing |
|---|---|---|
| 1 | 0.542 | 0.537 |
| 2 | 0.640 | 0.541 |
| 3 | 0.847 | 0.772 |
| 4 | 1.278 | 1.324 |
| 5 | 1.624 | 1.639 |
| All shots | 0.740 | 0.698 |



(a)

(b)

Figure 11: (a) Comparison between actual reconstructed quality and estimated quality per each shot (2nd P-frame is lost in each GoP); and (b) motion activity level of the video shots.

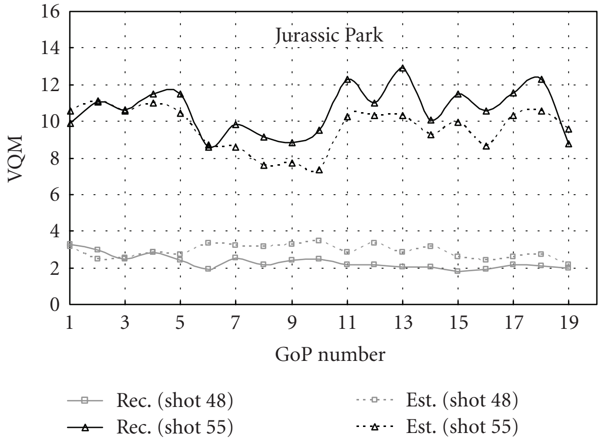


Figure 12: Comparison between actual reconstructed quality and estimated quality per each GoP (2nd P-frame is lost in each GoP) for concealment by freezing.

### 5.2.1. Prediction at GoP level

Figure 12 shows the performance of the GoP level predictor (see Subsection 4.1), compared to the actual quality degradation, when the 2nd P-frame is lost during video transmission. The performance over two video shots of motion activity level 1 (shot 48) and of motion activity level 3 (shot 55) is shown. Table 8 shows the average absolute difference between the GoP quality predictor and the actual quality degradation. In the case of losing the 3rd P-frame, where the average quality degradation for this type of decoder is 4.573 (see Table 4), the accuracy of the GoP quality predictor is about ±0.698 using the VQM metric. When the 2nd P-frame is lost, the accuracy of the GoP level predictor for decoders that conceal losses by copying is 0.6, while the accuracy of GoP level predictor for decoders that conceal losses by freezing is 0.74 (compare Table 6 to Table 8). These results suggest that (1) decoders that conceal losses by copying provide better reconstructed quality (compare the results of Tables 3 and 4), and (2) quality predictions derived from the advanced video traces are better for decoders that conceal losses by copying.

### 5.2.2. Prediction at shot level

Figure 13 shows the performance of the shot level predictor (see Subsection 4.2) compared to the actual quality degradation, when the 2nd P-frame in each GoP is lost during video transmission. Table 9 shows the accuracy of the shot level predictor. We observe that better accuracy is always achieved when shots are of lower motion activity levels. In general, the accuracy of shot level predictor is better when fewer frames

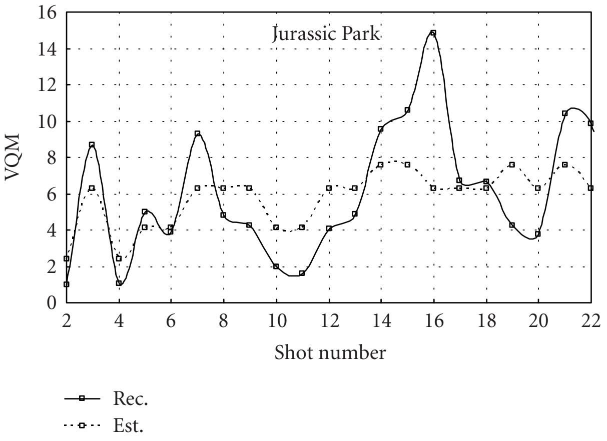


Figure 13: Comparison between actual reconstructed quality and estimated quality per each shot (2nd P-frame is lost in each GoP) for concealment by freezing.

Table 9: The absolute difference (in VQM) between actual reconstructed quality and estimated quality using shot level analysis.

| Activity level | 2nd P-frame freezing | 3rd P-frame freezing |
|---|---|---|
| 1 | 0.934 | 0.727 |
| 2 | 1.552 | 1.329 |
| 3 | 2.08 | 1.891 |
| 4 | 2.440 | 2.355 |
| 5 | 1.927 | 2.256 |
| All shots | 1.842 | 1.666 |

are affected by the channel loss. Comparing the results of Tables 8 and 9, we observe that the accuracy of quality prediction using shot level analysis is significantly lower than the accuracy of quality prediction using GoP level analysis.

## 6. EVALUATION OF QUALITY PREDICTION USING PSNR METRIC

According to the results obtained with the VQM metric in Section 5, the quality prediction for the error concealment by copying and the GoP level quality predictor appear to be the most promising. In this section, we follow up on the exploratory evaluations with the VQM metric by conducting an extensive evaluation of the frame level and GoP level predictors using the PSNR as the quality metric of the reconstructed video. We use the quality predictors analyzed in Subsections 3.3 and 3.4 for decoders that conceal packet losses by copying from the previous reference frame.

For the extensive evaluations reported in this section, we randomly selected 956 video shots of various durations, extracted from 5 different video programs (*Terminator*, *Star Wars*, *Lady and Tramp, Tonight Show,* and *Football with Commercial*). The shots were detected and their motion activity levels were determined using the procedure outlined in Subsection 3.1. Table 10 shows the motion characteristics of the selected video shots. Shots of motion activity level 5 are rare in these video programs and have typically short duration. For television broadcasts and kids programs, shots of motion activity level 2 are common; see results of *Tonight Show* and *Lady and Tramp*. However, for sports events and movie productions, shots of motion activity level 3 are common; see results of *Star Wars*, *Terminator*, and *Football_WC*.

For these 5 video programs, the advanced video traces are composed of (i) the frame size in bits, (ii) the quality of the encoded video (which corresponds to the video quality of loss-free transmission) in PSNR, (iii) the motion information descriptor $M(t)$ between successive frames, which is calculated using (1), (iv) the ratio of forward motion estimation $V_f(t)$, and (v) the motion activity level $\theta$ of the underlying video shot. These video traces are used by the quality predictors to estimate the quality degradation due to frame losses.

### 6.1. Frame level predictor for concealment by copying

The quality predictor presented in Subsection 3.3 is used to estimate the reconstructed qualities when the video transmission suffers a P-frame loss. We have conducted three different evaluations for 1st P-frame loss, 2nd P-frame loss, and 3rd P-frame loss. Tables 11, 12, and 13 show (i) the mean actual quality reduction in dB, that is, the average difference between the PSNR quality of the encoded video and the PSNR quality of the actual reconstructed video, and (ii) the mean absolute prediction error in dB, that is, the average absolute difference between the actual quality reduction in dB and the predicted quality reduction for the frame level quality predictor for each motion activity level, and for the whole video sequence. (We note that for the PSNR metric the quality degradation $Q$ is defined as $Q$ = (encoded quality − actual reconstructed quality)/encoded quality for the analysis in Sections 2–4; for ease of comprehension we report here the quality reduction = encoded quality − actual reconstructed quality.) We observe that the proposed quality predictor gives a relatively good approximation of the actual quality degradation. We observe from Table 13, for instance, that for the *Terminator* movie, where the actual quality reduction is about 9.4 dB when losing the 3rd P-frame, the frame level quality predictor estimates the reconstructed qualities with an accuracy of ±1.4 dB around the actual value.

We observe that the accuracy of this quality predictor is generally monotonically decreasing as the motion activity level increases. Due to the small number of shots of motion activity level 5, the results for activity level 5 have only very limited statistical validity. For some video shots of motion activity level 1, the quality predictor does not effectively estimate the reconstructed qualities. In these video shots, the actual quality reduction (in dB) is larger than the estimated quality reduction. This is mainly because for shots of low motion activity levels, the actual quality reduction measured in PSNR tends to be higher than the actual quality reduction perceived by humans and predicted with our methodology. Indeed, comparing Tables 5 and 11, we observe that when the perceptual quality metric is used, which more closely

Table 10: The characteristics of the video test sequences.

| Video sequence | Number of shots per activity level | | | | | Total number of shots | Duration of shots per activity level (seconds) | | | | | Total duration (minutes) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | | 1 | 2 | 3 | 4 | 5 | |
| Star Wars | 10 | 52 | 89 | 44 | 5 | 200 | 22 | 200 | 427 | 244 | 29 | 15.37 |
| Terminator | 14 | 70 | 98 | 18 | 1 | 201 | 34 | 147 | 491 | 148 | 6.7 | 13.77 |
| Football_WC | 14 | 68 | 90 | 27 | 2 | 201 | 8.4 | 123 | 177 | 53 | 6.3 | 6.14 |
| Tonight Show | 15 | 89 | 42 | 6 | 1 | 153 | 30 | 642 | 186 | 29 | 4.4 | 14.85 |
| Lady and Tramp | 19 | 87 | 73 | 22 | 0 | 201 | 76 | 293 | 531 | 201 | 0 | 18.35 |

Table 11: The mean actual quality reduction and the mean absolute prediction error (in PSNR) between actual reconstructed quality and estimated quality using frame level analysis when the 1st P-frame is lost.

| Video sequence | The mean quality reduction per activity level | | | | | | The mean absolute prediction error per activity level | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | All shots | 1 | 2 | 3 | 4 | 5 | All shots |
| Star Wars | 5.52 | 7.99 | 8.40 | 12.60 | 11.24 | 9.44 | 2.55 | 1.85 | 2.08 | 2.38 | 2.62 | 2.14 |
| Terminator | 4.09 | 9.21 | 11.12 | 12.54 | 11.35 | 10.75 | 1.57 | 1.91 | 2.03 | 2.38 | 3.28 | 2.06 |
| Football_WC | 4.72 | 7.29 | 10.87 | 13.81 | 19.25 | 10.10 | 3.41 | 2.23 | 2.37 | 2.98 | 1.80 | 2.43 |
| Tonight Show | 6.25 | 8.06 | 10.02 | 12.92 | 19.88 | 8.62 | 1.84 | 1.78 | 2.27 | 2.10 | 4.09 | 1.91 |
| Lady and Tramp | 6.31 | 8.41 | 9.10 | 10.15 | — | 8.92 | 1.73 | 1.98 | 2.04 | 2.10 | — | 2.01 |

models the human perception, the prediction accuracy of our methodology for shots of motion activity 1 is higher than for the other shot activity levels. The average accuracy for video programs such as *Tonight Show* is better than that for other video programs because of its statistical distribution of the motion activity levels; see Table 10. Similarly to the results of Section 5, the accuracy of the prediction is improved if the number of affected frames is smaller.

### 6.2. *GoP level predictor for concealment by copying*

Tables 14, 15, and 16 show the quality prediction error of the GoP level quality predictor from Subsection 3.4 for each motion activity level, and for the whole video sequence. Comparing these prediction errors with the average actual quality reductions reported in Tables 11, 12, and 13 demonstrates that the GoP level predictor achieves very good prediction accuracy. Similarly to the observations for the frame level predictor, the accuracy of the GoP quality predictor generally monotonically improves as the motion activity level decreases. For some video shots, the quality predictor cannot effectively estimate the reconstructed qualities for some motion activity levels, since the number of video shots of the motion activity level is underrepresented in the training set which is used to generate the functional approximations of the quality degradation. In addition, the PSNR metric is not suitable for measuring the quality degradation for shots of low motion activity levels, which in turn degrades the accuracy of the GoP level quality predictor. Similarly to the results of Section 5, substantial improvements in the accuracy of estimating the actual quality degradation are achieved if the GoP level predictor is adopted compared to the frame level predictor. Comparing Tables 11 and 14, for instance, a 1 dB improvement in estimating the quality reduction is achieved for the *Star Wars* movie, in the case of 1st P-frame loss.

## 7. CONCLUSION

A framework for advanced video traces has been proposed, which enables the evaluation of video transmission over lossy packet networks, without requiring the actual videos. The advanced video traces include—aside from the frame size (in bits) and PSNR contained in conventional video traces—a parsimonious set of visual content descriptors that can be arranged in a hierarchal manner. In this paper, we focused on motion-related content descriptors. Quality predictors that utilize these content descriptors to estimate the quality degradation have been proposed. Our extensive simulations demonstrate that the GoP level quality predictors typically estimate the actual quality degradation with an accuracy of about ±1 dB. The performance of the proposed quality predictors can be improved by using a perceptual quality metric such as VQM instead of the traditional PSNR. The proposed advanced video trace framework is flexible enough to be used with various packet transmission scenarios, multiple methods of loss concealment, different granularities of the video sequence (frame level, GoP level, shot level), and a different degree of accuracy in estimating the reconstructed qualities. To the best of our knowledge the advanced video traces, proposed in this paper, represent the first comprehensive evaluation scheme that permits communication and networking researchers and engineers without access to actual videos to meaningfully examine the performance of lossy video transport schemes.

There are many exciting avenues for future work on advanced video traces. One direction is to develop advanced

Table 12: The mean actual quality degradation and the mean absolute prediction error (in PSNR) between actual reconstructed quality and estimated quality using frame level analysis when the 2nd P-frame is lost.

| Video sequence | The mean quality reduction per activity level | | | | | | The mean absolute prediction error per activity level | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | All shots | 1 | 2 | 3 | 4 | 5 | All shots |
| Star Wars | 4.62 | 7.60 | 7.96 | 12.04 | 10.34 | 8.96 | 2.13 | 1.66 | 1.83 | 2.05 | 2.40 | 1.88 |
| Terminator | 3.96 | 9.00 | 10.85 | 12.39 | 12.49 | 10.52 | 1.44 | 1.69 | 1.79 | 2.19 | 2.56 | 1.83 |
| Football_WC | 2.57 | 7.01 | 10.30 | 13.95 | 18.98 | 9.70 | 2.01 | 2.06 | 2.17 | 2.87 | 1.79 | 2.22 |
| Tonight Show | 6.01 | 7.88 | 9.97 | 12.80 | 19.49 | 8.47 | 1.61 | 1.43 | 1.91 | 1.83 | 3.26 | 1.56 |
| Lady and Tramp | 6.01 | 8.12 | 8.93 | 9.91 | — | 8.69 | 1.42 | 1.71 | 1.83 | 1.89 | — | 1.78 |

Table 13: The mean actual quality degradation and the mean absolute prediction error (in PSNR) between actual reconstructed quality and estimated quality using frame level analysis when the 3rd P-frame is lost.

| Video sequence | The mean quality reducion per activity level | | | | | | The mean absolute prediction error per activity level | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | All shots | 1 | 2 | 3 | 4 | 5 | All shots |
| Star Wars | 3.52 | 6.80 | 6.92 | 10.73 | 9.66 | 7.91 | 1.57 | 1.35 | 1.42 | 1.62 | 1.87 | 1.47 |
| Terminator | 3.01 | 7.85 | 9.73 | 11.25 | 11.21 | 9.40 | 1.02 | 1.24 | 1.41 | 1.68 | 3.22 | 1.42 |
| Football_WC | 1.64 | 5.95 | 9.07 | 12.47 | 18.42 | 8.51 | 3.38 | 1.82 | 1.89 | 2.57 | 1.39 | 1.99 |
| Tonight Show | 5.11 | 7.05 | 9.10 | 12.12 | 18.46 | 7.63 | 0.95 | 0.81 | 1.33 | 1.26 | 3.11 | 0.95 |
| Lady and Tramp | 4.84 | 7.18 | 8.00 | 8.98 | — | 7.74 | 0.99 | 1.27 | 1.37 | 1.40 | — | 1.32 |

Table 14: The mean absolute prediction error (in PSNR) between actual reconstructed quality and estimated quality using GoP level analysis when the 1st P-frame is lost.

| Video sequence | The mean absolute prediction error per activity level | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | All shots |
| Star Wars | 1.79 | 0.95 | 1.12 | 1.27 | 1.55 | 1.15 |
| Terminator | 0.59 | 0.82 | 1.07 | 1.35 | 2.02 | 1.06 |
| Football_WC | 2.59 | 1.23 | 1.40 | 2.05 | 1.12 | 1.46 |
| Tonight Show | 0.97 | 0.83 | 1.27 | 1.22 | 2.80 | 0.95 |
| Lady and Tramp | 0.59 | 0.80 | 0.89 | 1.03 | — | 0.87 |

Table 15: The mean absolute prediction error (in PSNR) between actual reconstructed quality and estimated quality using GoP level analysis when the 2nd P-frame is lost.

| Video sequence | The mean absolute prediction error per activity level | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | All shots |
| Star Wars | 1.61 | 0.86 | 1.02 | 1.15 | 1.46 | 1.05 |
| Terminator | 0.56 | 0.83 | 1.02 | 1.31 | 1.73 | 1.03 |
| Football_WC | 1.60 | 1.20 | 1.29 | 1.92 | 1.22 | 1.35 |
| Tonight Show | 1.02 | 0.75 | 1.16 | 1.09 | 2.67 | 0.86 |
| Lady and Tramp | 0.59 | 0.76 | 0.87 | 0.95 | — | 0.84 |

Table 16: The mean absolute prediction error (in PSNR) between actual reconstructed quality and estimated quality using GoP level analysis when the 3rd P-frame is lost.

| Video sequence | The mean absolute prediction error per activity level | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | All shots |
| Star Wars | 1.21 | 0.96 | 1.07 | 1.11 | 1.30 | 1.07 |
| Terminator | 0.63 | 0.83 | 1.05 | 1.21 | 2.09 | 1.03 |
| Football_WC | 1.02 | 1.13 | 1.23 | 1.73 | 0.69 | 1.26 |
| Tonight Show | 0.88 | 0.72 | 1.10 | 1.01 | 2.57 | 0.82 |
| Lady and Tramp | 0.80 | 0.80 | 0.83 | 0.88 | — | 0.83 |

traces that allow for the prediction of the reconstructed video quality when multiple frames are lost within a GoP. Another direction is to examine how the quality predictors can be improved by incorporating color-related content descriptors such as the color layout descriptors (frame level, GoP level, and shot level) as well as the camera movement descriptors, which characterize the zoom-in, zoom-out, panning, and tilting operations.

## ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation through Grant ANI-0136774. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

## REFERENCES

[1] R. Narasimha and R. Rao, "Modeling variable bit rate video on wired and wireless networks using discrete-time self-similar systems," in *Proceedings of IEEE International Conference on Personal Wireless Communications (ICPWC '02)*, pp. 290–294, New Delhi, India, December 2002.

[2] A. Bhattacharya, A. G. Parlos, and A. F. Atiya, "Prediction of MPEG-coded video source traffic using recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 8, pp. 2177–2190, 2003.

[3] F. H. P. Fitzek and M. Reisslein, "MPEG-4 and H.263 video traces for network performance evaluation," *IEEE Network*, vol. 15, no. 6, pp. 40–54, 2001.

[4] P. Seeling, M. Reisslein, and B. Kulapala, "Network performance evaluation using frame size and quality traces of single-layer and two-layer video: a tutorial," *IEEE Communications Surveys and Tutorials*, vol. 6, no. 3, pp. 58–78, 2004.

[5] S. Valaee and J.-C. Gregoire, "Resource allocation for video streaming in wireless environment," in *Proceedings of the 5th International Symposium on Wireless Personal Multimedia Communications (WPMC '02)*, vol. 3, pp. 1103–1107, Honolulu, Hawaii, USA, October 2002.

[6] A. Kanjanavapastit and H. Mehrpour, "Packet reservation multiple access for multimedia traffic," in *Proceedings of 10th IEEE International Conference on Networks (ICON '02)*, pp. 162–166, Singapore, 2002.

[7] A. B. Watson, J. Hu, and J. F. McGowan III, "Digital video quality metric based on human vision," *Journal of Electronic Imaging*, vol. 10, no. 1, pp. 20–29, 2001.

[8] Z. He and C. W. Chen, "End-to-end video quality analysis and modeling for video streaming over IP network," in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME '02)*, vol. 1, pp. 853–856, Lausanne, Switzerland, 2002.

[9] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: a review," *Proceedings of the IEEE*, vol. 86, no. 5, pp. 974–997, 1998.

[10] A. R. Reibman, V. A. Vaishampayan, and Y. Sermadevi, "Quality monitoring of video over a packet network," *IEEE Transactions on Multimedia*, vol. 6, no. 2, pp. 327–334, 2004.

[11] S. Kanumuri, P. Cosman, and A. R. Reibman, "A generalized linear model for MPEG-2 packet-loss visibility," in *Proceedings of 14th International Packet Video Workshop (PV '04)*, Irvine, Calif, USA, December 2004.

[12] Y. J. Liang, J. G. Apostolopoulos, and B. Girod, "Analysis of packet loss for compressed video: does burst-length matter?" in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03)*, vol. 5, pp. 684–687, Hong Kong, April 2003.

[13] M. Masry and S. S. Hemami, "Perceived quality metrics for low bit rate compressed video," in *Proceedings of IEEE International Conference on Image Processing (ICIP '02)*, vol. 3, pp. 49–52, Rochester, NY, USA, 2002.

[14] N. G. Duffield, K. K. Ramakrishnan, and A. R. Reibman, "Issues of quality and multiplexing when smoothing rate adaptive video," *IEEE Transactions on Multimedia*, vol. 1, no. 4, pp. 352–364, 1999.

[15] O. Verscheure, X. Garcia, G. Karlsson, and J.-P. Hubaux, "User-oriented QoS in packet video delivery," *IEEE Network*, vol. 12, no. 6, pp. 12–21, 1998.

[16] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications ", Recommendations of the ITU, Telecommunication Standardization Sector, approved in September 1999.

[17] S. Jeannin and A. Divakaran, "MPEG-7 visual motion descriptors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 720–724, 2001.

[18] J. S. Boreczky and L. A. Rowe, "Comparison of video shot boundary detection techniques," in *Storage and Retrieval for Still Images and Video Databases IV*, vol. 2664 of *Proceedings of SPIE*, San Jose, Calif, USA, pp. 170–179, 1996.

[19] K. A. Peker and A. Divakaran, "Framework for measurement of the intensity of motion activity of video segments," *Journal of Visual Communication and Image Representation*, vol. 15, no. 3, pp. 265–284, 2004.

[20] R. V. Hogg and A. T. Craig, *Introduction to Mathematical Statistics*, Macmillan, New York, NY, USA, 5th edition, 1995.

[21] E. L. Lehmann and H. J. M. D'Abrera, *Nonparametrics: Statistical Methods Based on Ranks*, Prentice-Hall, Englewood Cliffs, NJ, USA, rev. edition, 1998.

**Osama A. Lotfallah** is a Postdoctoral Research Associate in the Department of Computer Science and Engineering of Arizona State University since January 2005. He received his B.S. and Master's degrees from the School of Computer Engineering at Cairo University, Egypt, in July 1997 and July 2001, respectively. During his Master's study, he was working as Teacher Assistant in the Computer Science Department of Cairo University. He received his Ph.D. degree in electrical engineering from Arizona State University, in December 2004, under the supervision of Prof Sethuraman Panchanathan. He was actively involved in the teaching and research activities in the field of digital signal processing. He was also an active Member of the Video Traces Research Group of Arizona State University (http://trace.eas.asu.edu). His research interest is in the fields of advanced video coding, digital video processing, visual content extraction, and video streaming, with a focus on adaptive video transmission schemes. He has two provisional USA patents in the field of content-aware video streaming. He is a regular reviewer of many international conferences in the field of visual communication as well as periodical journal and magazines in the field of multimedia and signal processing.

**Martin Reisslein** is an Associate Professor in the Department of Electrical Engineering at Arizona State University (ASU), Tempe. He received the Dipl.-Ing. (FH) degree from the Fachhochschule Dieburg, Germany, in 1994, and the MSE degree from the University of Pennsylvania, Philadelphia, in 1996. He received his Ph.D. in systems engineering from the University of Pennsylvania in 1998. During the academic year 1994–1995, he visited the University of Pennsylvania as a Fulbright scholar. From July 1998 through October 2000, he was a Scientist with the German National Research Center for Information Technology (GMD FOKUS), Berlin, and a Lecturer at the Technical University Berlin. From October 2000 through August 2005, he was an Assistant Professor at ASU. He is the Editor-in-Chief of the IEEE Communications Surveys and Tutorials and has served on the Technical Program Committees of IEEE Infocom and IEEE Globecom. He maintains an extensive library of video traces for network performance evaluation, including frame size traces of MPEG-4 and H.263 encoded video, at http://trace.eas.asu.edu. He is corecipient of the Best Paper Award of the SPIE Photonics East 2000—Terabit Optical Networking Conference. His research interests are in the areas of Internet quality of service, video traffic characterization, wireless networking, and optical networking.

**Sethuraman Panchanathan** is a Professor and Chair of the Computer Science and Engineering Department as well as the Interim Director of the Department of Biomedical Informatics (BM), Director of the Institute for Computing & Information Sciences & Engineering, and Director of the Research Center on Ubiquitous Computing (CUbiC) at Arizona State University, Tempe, Arizona. He has published over 200 papers in refereed journals and conferences. He has been a Chair of many conferences, program committee member of numerous conferences, organizer of special sessions in several conferences, and an invited panel member of special sessions. He has presented several invited talks in conferences, universities, and industry. He is a Fellow of the IEEE and SPIE. He is an Associate Editor of the IEEE Transactions on Multimedia, IEEE Transactions on Circuits and Systems for Video Technology, Area Editor of the Journal of Visual Communications and Image Representation, and an Associate Editor of the Journal of Electronic Imaging. He has guest edited special issues in the Journal of Visual Communication and Image Representation, Canadian Journal of Electrical and Computer Engineering, and the IEEE Transactions on Circuits and Systems for Video Technology.