PACKET DROPPING FOR H.264 VIDEOS CONSIDERING BOTH CODING AND PACKET-LOSS ARTIFACTS

Yuxia Wang, Ting-Lan Lin, Pamela C. Cosman**

School of Information Engineering, Communication University of China, Beijing, China, 100024
*Dept. of ECE, University of California, San Diego, CA, USA, 92093-0407

ABSTRACT

In the context of both compression artifacts and packet loss artifacts, we use generalized linear models to predict VQM quality scores. Using a network-based model, a router can estimate the visual importance of each incoming packet and decide which packet to drop when congestion happens. Considering a wide variety of bit reduction rates, we perform packet dropping experiments for combinations of video streams and examine the effects of video contents and different bit rates. By comparing with randomly dropping B slices or B frames, we conclude that our model gives a good performance on objective evaluation of packet importance.

Index Terms— compression artifacts, H.264, packet loss, video quality evaluation, VQM

1. INTRODUCTION AND BACKGROUND

Compressed video streams transmitted over heterogeneous networks can experience visual quality impairments due to packet losses and compression artifacts. Effective methods of video quality assessment are necessary when designing or testing a system for transporting video on networks.

Many objective models for evaluation of quality degradation due to compression artifacts have been developed, such as SSIM [1], JND [2], and Winkler's Perceptual Distortion Metric [3]. However, for video transmission over networks, due to limited bandwidth or channel errors, video quality at the receiver can be highly affected by packet losses in addition to compression artifacts. So it is a challenging problem to evaluate the quality impairments produced by packet losses accurately and efficiently. VQM [4,5] is a standardized full-reference (FR) method of objectively measuring video quality considering both coding artifacts and transmission errors. It measures the perceptual effects of a broad range of quality impairments including blurring, jerky or unnatural motion, global noise, block distortion, color distortion and packet loss. It has been adopted by the ANSI as a U.S. national standard and as an international ITU Recommendation and has been shown to be better correlated with human perception than other full reference video quality metrics [6]. We are interested in building network-based models to predict VQM scores for compressed video with packet losses.

In our prior work [7,8,9,10], we modeled packet loss visibility under the assumption that there is no visible compression artifact both for MEPG-2 and H.264 coding standards. [7] proposed a generalized linear model to predict the probability that a packet loss will be visible to an average viewer. [9] gave a generalized linear model for video packet loss visibility that is applicable to different group-of-picture structures. The models in [7,8,9,10] were obtained based on data from subjective experiments which are reliable but expensive and time-consuming. In contrast, in [11], we focused on modeling of overall perceptual video quality based on the VQM quality scores for H.264 videos, in the context of both compression artifacts and packet loss artifacts. We proposed two network-based models that are fully self-contained at the packet level and are of low computational complexity.

In [12], we researched packet dropping algorithms for various bit reduction rates. We drop the least visible packets or frames using a network-based visibility model to achieve the required bit reduction rates. This work considered network losses without compression artifacts. In the current paper, we are interested in evaluating a new model for video packet importance considering both compression artifacts and packet losses. Simulation experiments are designed with different bit reduction rates to verify our model. Also we compare the performance using our dropping method with that of randomly dropping B slices or B frames.

The paper differs from our prior work as follows. Papers [7,8,9,10,12] all considered high quality compression where visible losses are due to packet drops and not compression artifacts. In [11], we considered lower quality compression; the current paper is an experimental validation of the model developed in [11] with videos at various bit rates subjected to various target levels of packet dropping. The rest of this paper is organized as follows. In Section 2, the factors used for our network-based model are given, and the method of model building is described using a generalized linear model (GLM) [13]. Section 3 presents the design of the packet dropping experiments. Section 4 presents simulation results for combinations of videos at various target packet dropping rates.

This work was supported by the Communication University of China and by the National Science Foundation.

2. NETWORK-BASED MODEL FOR PACKET IMPORTANCE

In this section, we introduce the network-based packet importance model. Most of this section is taken from our prior work [11], where we proposed two models predicting VQM$_B$ scores and ΔVQM scores respectively, as shown in Figure 1. VQM$_A$ is the VQM score computed between the original GOP and the compressed GOP that has been reconstructed without any packet losses. VQM$_B$ is the VQM score computed between the original GOP and the GOP that has both compression artifacts and a packet loss, including error propagation, if any, from that loss. ΔVQM = VQM$_B$ - VQM$_A$, represents the additional degradation of the VQM score of a GOP that comes from dropping a packet, beyond the effect that comes from compression alone. We focus on packet dropping experiments based on the model of ΔVQM in order to explore the quality degradation due to choices of packet drops.

We used six original video sequences with varied levels of detail and motion types, including both object and camera motion. Each one is encoded at three relatively low bit rates: 200, 300 and 400 kbps, so there are 18 videos in total. For modeling, the videos are compressed by the H.264 JM9.3 encoder in CIF resolution (352 by 288). The GOP structure is (IDR)BBPBBPBB with 15 frames per GOP. Rather than using a fixed quantization parameter, we use the default rate control of the encoder. The values of the quantization parameter can vary from frame to frame. One horizontal row of macroblocks (MBs) is packetized into one slice, so there are 18 slices in a frame. To explore the influence on video quality of each lost packet (each slice), we randomly drop one slice in each GOP. In a separate realization, a different single slice is chosen for dropping in that GOP. In total, there are 15 realizations generated for each GOP. The decoder conceals the lost slice using motion-compensated error concealment (MCEC) where the motion vector is estimated from surrounding blocks.

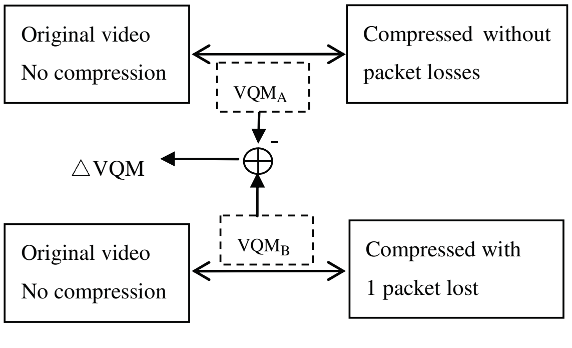


Fig. 1. Two computations of VQM scores.

2.1. Factors affecting video quality

In contrast to an encoder-based model, which can evaluate packet importance at the encoder with access to the original video, a network-based model must evaluate packet importance using only the compressed data. In [11], we aim to create a network-based model that is fully self-contained at the packet level and is of low computational complexity. Self-contained at the packet level means that a network node can evaluate the impact on visual quality of each individual packet, without having access to the original video or having access to any other packets in the stream.

The candidate factors for modeling associated with a packet (slice) are defined as follows:

(1) **TMDR**, representing time duration, is the maximum number of frames to which one packet loss can propagate. For the I frame of the GOP, which is the reference of the following P and B frames, TMDR=15 due to the length of a GOP. For B frames, TMDR=1 as they are non-reference frames. For P frames, TMDR is variable depending on its position in that GOP.

(2) **NAL_size** is the packet size in bits. Since a packet contains a fixed number of MBs, usually the NAL_size of a packet from I and P frames is much larger than that from B frames.

(3) **MeanQP** is the mean of the quantization parameter (QP) values of all MBs in the slice. Larger values of QP correspond to lower bit rates, and worse quality.

(4) **DevFromCenter** is the vertical distance from the slice to the center of the frame. Usually human perception is more sensitive to an artifact or glitch in the center of a picture [10].

(5) **RSENGY** refers to the energy of the residual after motion compensation, which can be calculated from the DCT coefficients for each MB. **MeanRSENGY and MaxRSENGY** are the average and maximum of RSENGY values of all MBs in the slice.

(6) **Motion related factors** are calculated from the motion vectors of all MBs in a given slice. MeanMotX and MeanMotY denote the average values of the motion vectors in the x and y directions. VarMotX and VarMotY are the variances of the motion vectors. We define **MotM** = $\sqrt{MeanMotX^2 + MeanMotY^2}$, and **VarM** = $VarMotX + VarMotY$. **MeanMotA** and **MaxMotA** are the mean and maximum of the phases of non-zero motion vectors.

In model-building, we also considered the interaction terms of the factors, but experiment results of modeling showed that there was not a significant improvement (about 2-3 percent) in correlation after adding those terms. For simplicity, we removed the interaction terms from our final model.

2.2. Building the model based on GLM

For model-building, we use a generalized linear model (GLM) with "identity" as the link function based on the

Order	Factors	Coefficients
Intercept	1	-1.44e-003
1	NAL_size	3.54e-006
2	TMDR	5.69e-004
3	log(MeanRSENGY)	7.73e-005
4	MotM	4.81e-004
5	MeanMotA	1.96e-004

Table 1. Factors in order of importance in the model predicting ΔVQM.

Combination	$S1$	$S2$
1	*news* 200	*news* 400
2	*template* 200	*template* 400
3	*template* 200	*news* 400
4	*news* 400	*mobile* 200

Table 2. Four combinations of videos for experiments.

distribution of the scores. As an initial exploration, we built many simple GLM models using each factor alone in predicting VQMA, VQMB and DeltaVQM separately. Thus we obtained a correlation value for each factor and we included the factor in the final model-building only when the statistical P value was less than 0.05.

When building the final model, we used ten-fold cross-validation to determine a model of the right size. Factors are added into a model in order of importance. We used the MATLAB function "sequentialfs" which performs sequential feature selection. It selects factors by importance from all the factors mentioned above, based on the mean squared error between predicted values and actual values. The selection proceeds until there is no improvement in prediction.

Table 1 gives the factors in order of importance. NAL_size is the most important factor which makes the deviance decrease significantly, and the correlation between the predicted ΔVQM scores and the actual ΔVQM scores increases to 0.6043 by only using one factor (NAL_size) [11]. NAL_size is directly proportional to ΔVQM which means that larger NAL_size of the lost packet corresponds to higher (worse) predicted ΔVQM score. The time duration TMDR is the second most important factor to predict ΔVQM because packet loss in I or P frames leads to error propagation which influences the value of ΔVQM to a large extent. We can see NAL_size, TMDR, log(MeanRSENGY) and MotM included in the model are all with positive coefficients, as one would expect, indicating that larger values of the factor leads to larger drops in VQM quality scores (ΔVQM). But the last three factors in the model only bring a slight improvement in performance [11].

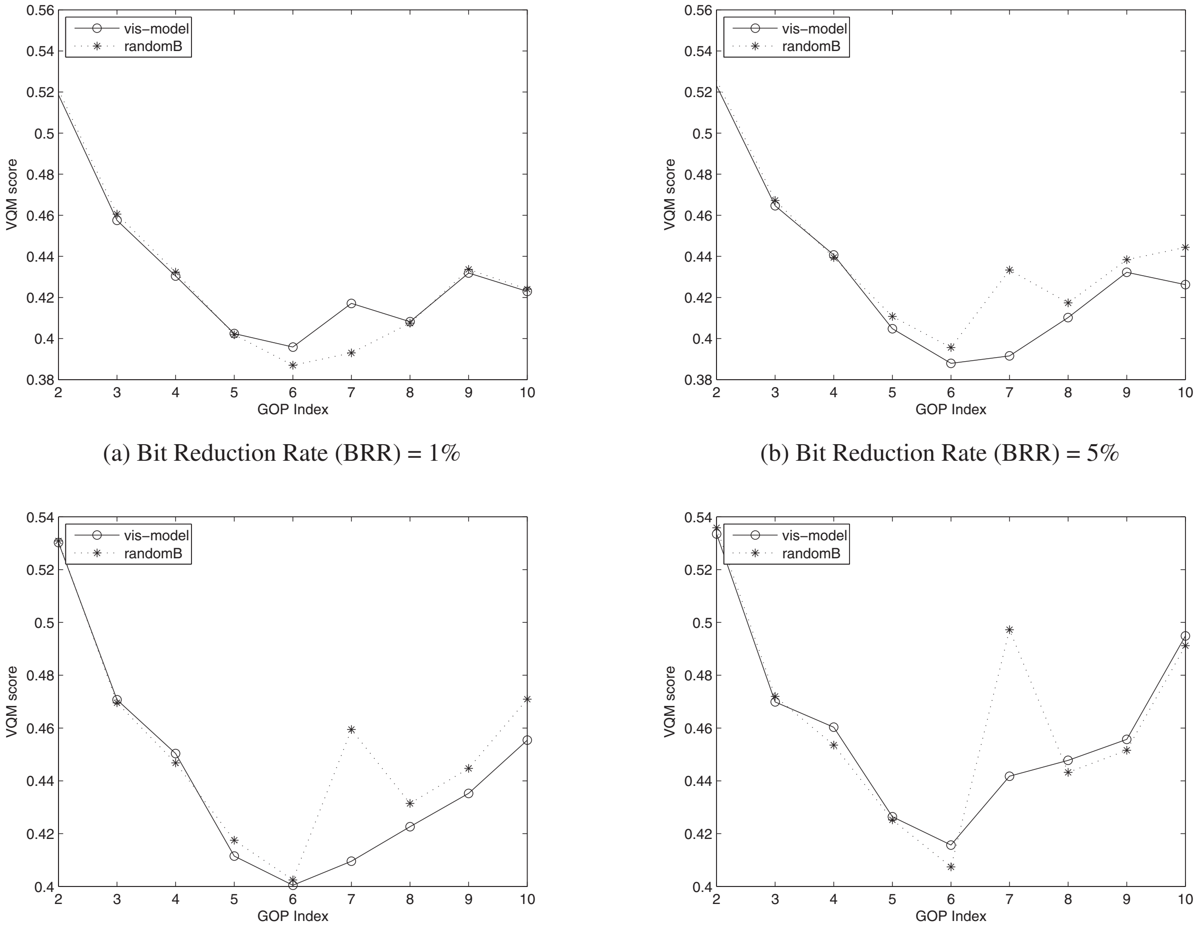
3. PACKET DROPPING FOR TWO VIDEO STREAMS

In this section, we design a packet dropping scheme to explore the performance of our model. For each incoming packet to the router, the model predicts the degradation to the VQM score if that packet is lost, using the information within one packet (NAL in H.264), and the implementation doesn't need any information from the pixel domain. Therefore, we can reduce the computation complexity by having only partial decoding of the streams which is very important for a realistic application.

Different from [12], we build this model to predict the video quality affected by packet losses at fairly low bit rates (200, 300 and 400 kpbs for CIF resolution), so there are in general visible compression artifacts. Thus we would like to see whether this model is effective for videos at different bit rates. We assume two video streams (S1 and S2) coming to the router simultaneously which are coded with different bit rates. The size of the buffer is assumed to be large enough to hold 2 GOPs worth of bits, one GOP from each stream. The bit reduction rate (BRR) is the percentage of bits that need to be dropped of the buffered packets to alleviate the congestion. Given a certain BRR for two GOPs, a buffer can drop packets according to the ΔVQM scores from our packet importance model or can drop packets randomly from B slices or B frames, when possible. Generally, with I, P and B frames in a H.264 video stream, the dropping of packets from I and P frames will result in much worse influence on the quality of the reconstructed video than those from B frames. Therefore, for both our model and random dropping, we drop B packets first, and, when running out of B packets to be dropped, P packets are selected.

Three videos are selected for our experiment: *news*, *template* and *mobile*, which are a subset of the 18 videos used for modeling. *News* is of slow movement and fixed camera, *template* is of zooming, and *mobile* is of higher motion and panning. We devise the experiment in two conditions: one is transporting two videos at different bit rates but with the same content; the other one is transporting two videos with different content and different bit rates. Table 2 gives the details of various combinations of the videos.

We perform packet dropping within two GOPs (one from each video) at one time. We consider that the router's outgoing link is not able to accommodate the incoming bits, and the router must therefore drop bits to achieve a target BRR. We consider five different BRRs: 0.5%, 1%, 5%, 10% and 15%. BRR is the ratio of the size of dropped bits to the total bits of two GOPs. Note that BRR can be very different from packet loss rate (PLR). For example, 15% BRR can result in dropping most of the B packets in one GOP, which means the PLR can be higher than 50%. For each BRR, we drop the packets using two methods: one drops the packet with smallest predicted ΔVQM score among all the packets (slices) in



(a) Bit Reduction Rate (BRR) = 1%

(b) Bit Reduction Rate (BRR) = 5%

(c) Bit Reduction Rate (BRR) = 10%

(d) Bit Reduction Rate (BRR) = 15%

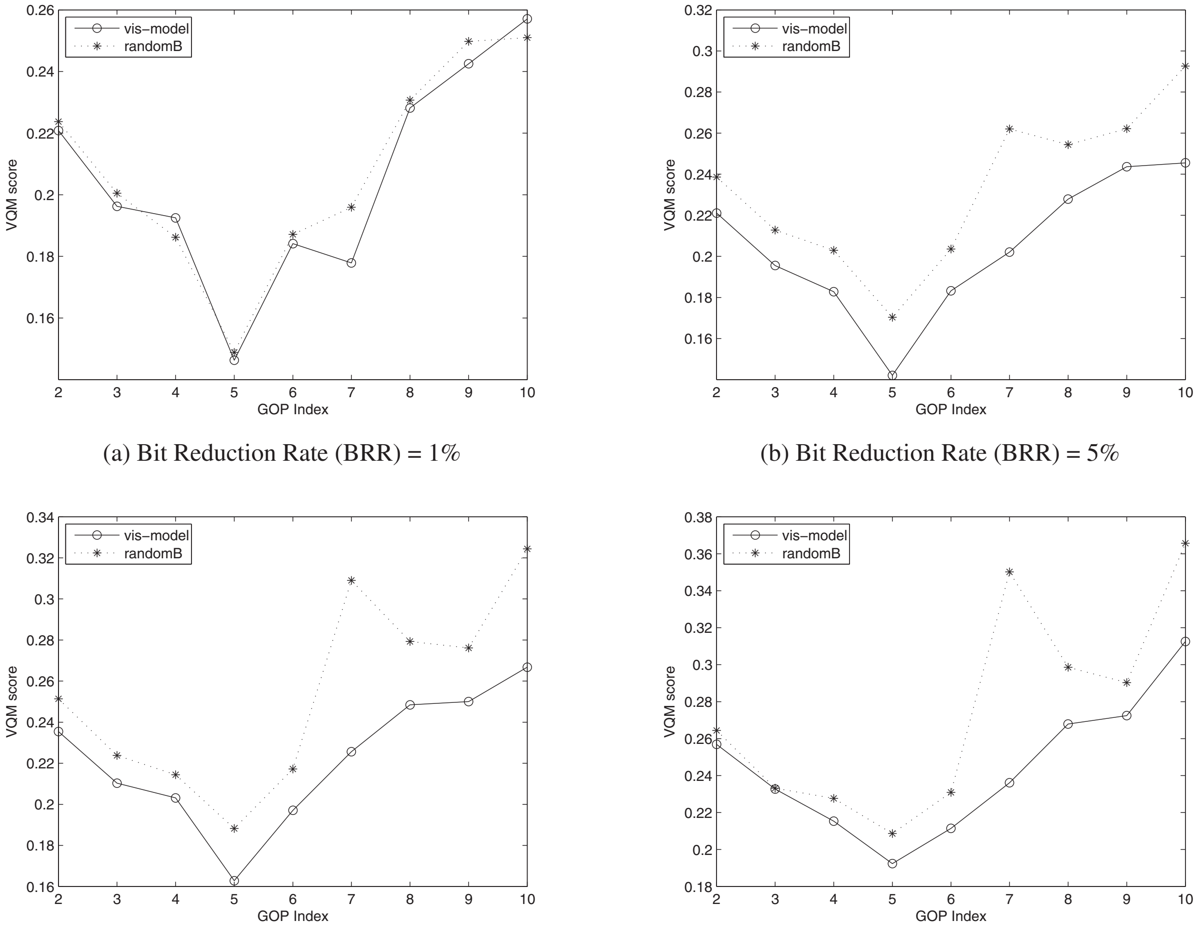
Fig. 2. VQM scores vs. GOP index for *news* at 200 kbps when it is sharing the buffer with *news* at 400 kbps.

the two GOPs until the required BRR is obtained. That is, we choose the packet which has the least influence on video quality when it is lost. The other method randomly drops packets from B frames from the two GOPs, regardless of which stream they are from, until the required BRR is obtained. We do 50 realizations of the random dropping to observe the average performance (VQM score).

The lossy videos after packet dropping are decoded by FFMPEG [14], in which the error concealment algorithm differs from MCEC which was used when modeling. In [12], there is a detailed explanation of the error concealment algorithm for different modes of losses. In short, for the MBs which are estimated to be intra coded, FFMPEG takes a weighted average of the uncorrupted neighboring blocks for error concealment, and for inter coded MBs, it performs bi-directional motion estimation to conceal the MBs. Once the dropping is performed for the two GOPs, the FFMPEG decoding and error concealment are run. Then VQM is calculated separately for S1 and S2 to obtain the video quality score for each lossy GOP. Note that here the VQM scores are computed between the original GOP and the actual decoder reconstruction, after loss and concealment. Note that because VQM may not count quality degradations in the last few frames of a sequence, and because we compute VQM for each GOP separately, we only drop packets from the first 12 frames in a GOP, that is we do not perform packet-dropping for the last 3 frames of a GOP, since those losses would get discounted.

4. EXPERIMENTAL RESULTS

Figures 2 and 3 present the actual VQM scores versus GOP index for the same video *news* at two bit rates of 200 kpbs



(a) Bit Reduction Rate (BRR) = 1%

(b) Bit Reduction Rate (BRR) = 5%

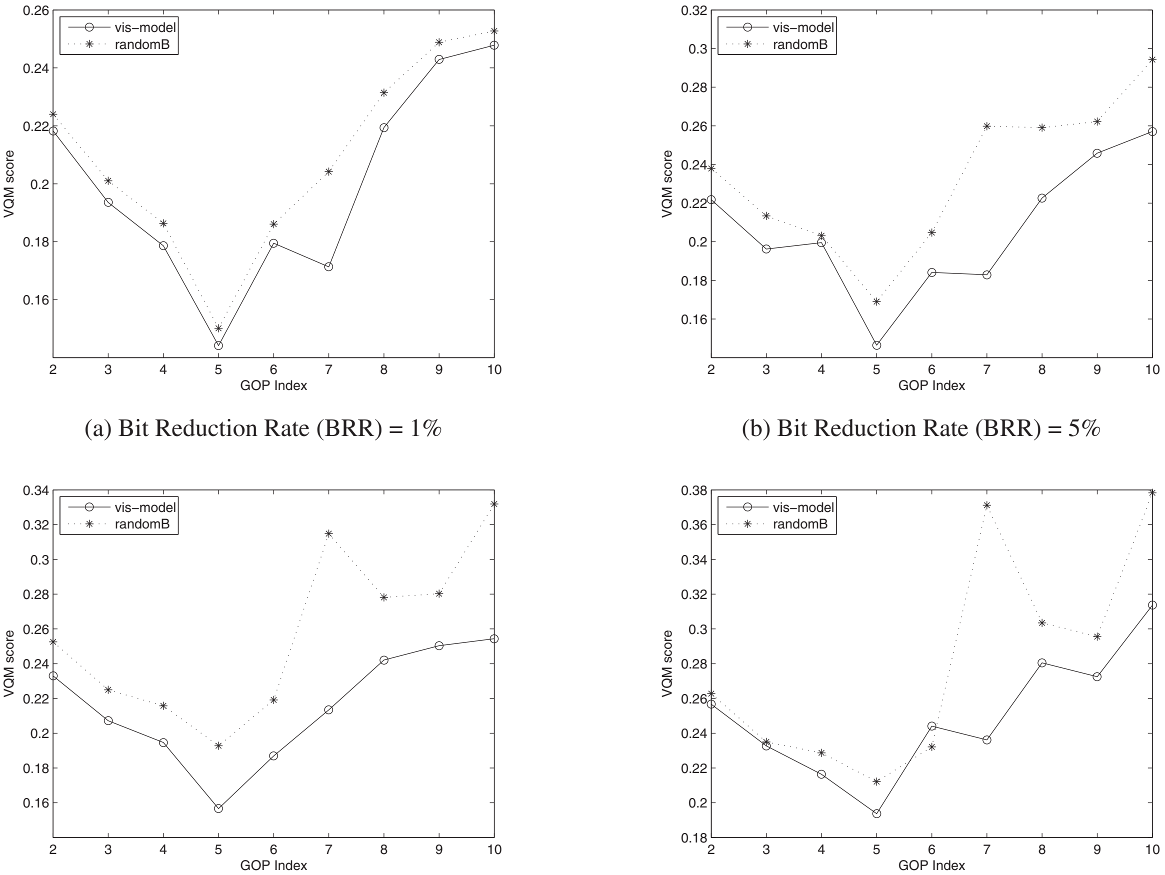
(c) Bit Reduction Rate (BRR) = 10%

(d) Bit Reduction Rate (BRR) = 15%

Fig. 3. VQM scores vs. GOP index for *news* at 400 kbps when it is sharing the buffer with *news* at 200 kbps.

and 400 kpbs (combination 1 in Table 2) when they share the buffer. Figure 2 shows the results for the 200 kbps version, and Figure 3 shows the results for the 400 kbps version. Each plot shows the results both for randomly dropping (randomB) and our packet importance model (vis-model). We display the results for BRR=1%, 5%, 10% and 15%. We exclude BRR=0.5% as it has a very similar performance with BRR=1%. From Figure 2 we see that the VQM scores mostly belong to [0.4 0.5], while the scores in Figure 3 usually lie in [0.2 0.3]. That is reasonable because the two videos are encoded at different bit rates and the compression artifacts play an important role in video quality scores. For the same video content, lower bit rates mean higher compression, so higher scores. Note that the y-axis scale changes as we look at (a), (b), (c) and (d), because higher BRR leads to higher VQM scores (worse quality) and more variance. As shown both in Figure 2 and Figure 3, when BRR is low, such as BRR=1%, the difference between the two dropping methods is quite small, but with increasing BRR, the performance of our model is significantly better than random dropping. This is because the quality degradation from dropping only a few packets (low BRRs) is usually masked by the compression artifacts, so the VQM scores are close. When more packets need to be dropped (higher BRR target) then the difference between random dropping and smarter dropping becomes more pronounced.

The VQM scores versus GOP index for different videos (*news* and *mobile*) at 400 kpbs and 200 kpbs (combination 4 in Table 2) are given in Figures 4 and 5. From Figure 5 we can see that, for all BRRs, the VQM scores obtained from the two methods are very close and there is almost no improvement in quality scores using our model. But there is significant improvement for the video *news* as shown in Figure 4. The reason is that the video clip *mobile* contains a lot of detail and



(a) Bit Reduction Rate (BRR) = 1%

(b) Bit Reduction Rate (BRR) = 5%

(c) Bit Reduction Rate (BRR) = 10%

(d) Bit Reduction Rate (BRR) = 15%

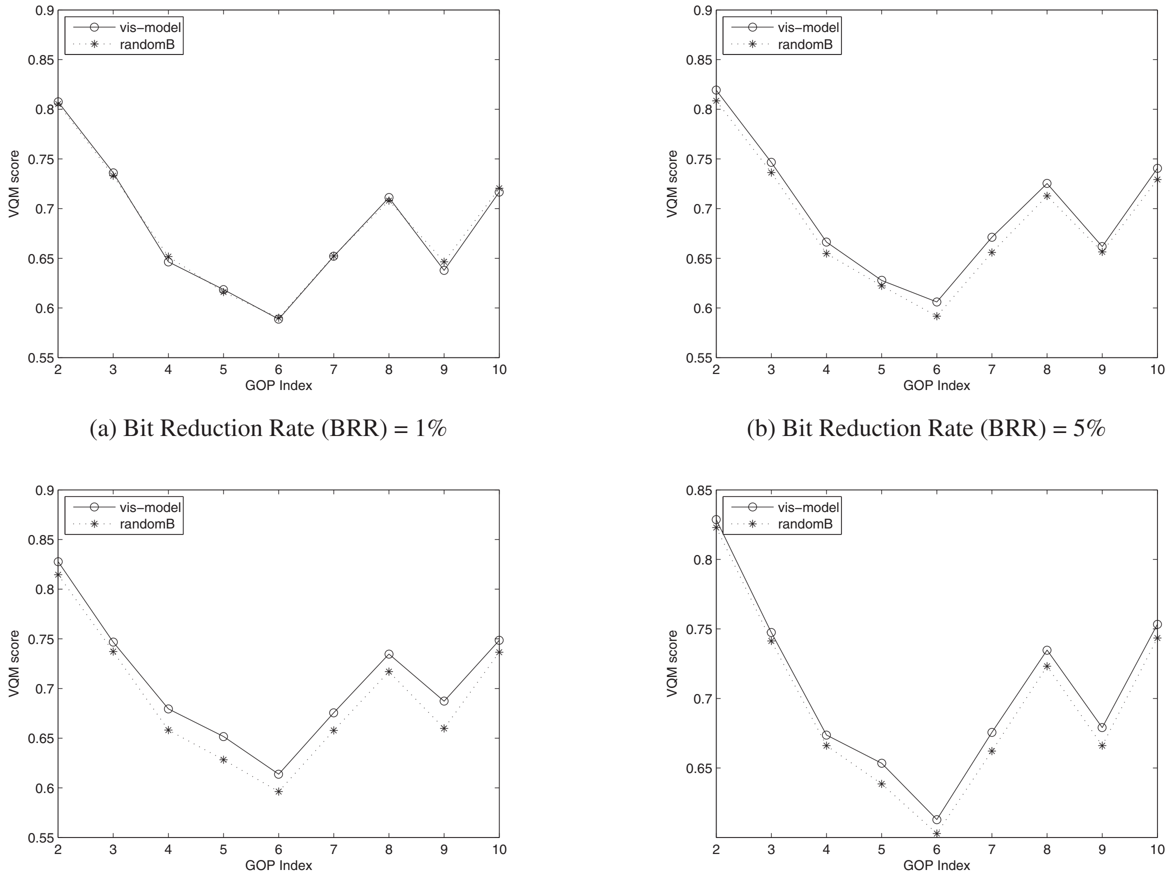
Fig. 4. VQM scores vs. GOP index for *news* at 400 kbps when it is sharing the buffer with *mobile* at 200 kbps.

motion, which leads to more compression artifacts at the bit rate of 200 kbps. Thus the degradation due to packet losses can be very small for it. So given the same degree of loss, the ΔVQM scores predicted by our model for the clip *mobile* are usually smaller than those from the second clip (*news* at 400 kbps). Therefore, when we consider two GOPs at the same time, more packets will be dropped from *mobile* than from *news* for a certain BRR. Compared to random dropping, we achieve a very good performance for one video while achieving a comparable performance for the other video.

The advantages of visibility-based dropping for two competing video streams are most clearly shown in Figures 6 and 7, which depict the VQM scores averaged over GOPs versus BRR for two combinations of videos (1 and 3) in Table 2. Both videos are better off compared to random B packet dropping. As before, the advantage of using our visibility-based dropping is more evident at higher BRRs. Figure 7 shows the VQM scores for different video contents with different bit rates, from which we get similar positive results with those in Figure 6. It can be concluded that even though the model was built on single packet loss for each GOP, it can be used where multiple packet dropping is necessary.

5. CONCLUSION

We considered the video quality degradation produced by both compression artifacts and packet losses. A network-based model for predicting the objective VQM quality scores was developed in [11] using factors which can be extracted from individual packets. The extraction of some factors does involve some processing of the payload, but in the future this processing may not be prohibitive, because the video packet does not need to be fully decoded (there is no inverse DCT and no pixel-level reconstruction). We note that this approach does not allow the contents to be encrypted, since some processing of the payload is required. We used this



(a) Bit Reduction Rate (BRR) = 1%

(b) Bit Reduction Rate (BRR) = 5%

(c) Bit Reduction Rate (BRR) = 10%

(d) Bit Reduction Rate (BRR) = 15%

Fig. 5. VQM scores vs. GOP index for *mobile* at 200 kbps when it is sharing the buffer with *news* at 400 kbps.

model to measure the visual importance of packets incoming to a router. The predicted additional degradation of VQM scores is used by the router to perform intelligent packet dropping. We validate this model by performing packet dropping experiments for multiple combinations of video streams at two different bit rates. Experiment results show that (a) for various bit reduction rates, our model has an advantage over the method of randomly dropping packets from B frames, and the improvement is significant especially for high BRRs. (b) For both videos with the same content and those with different contents coming to the router, our model outperforms the method of randomly dropping packets from B frames.

6. REFERENCES

[1] Z. Wang et al., "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Trans. on Image Processing*, vol.13, Apr. 2004.

[2] ATIS, "Objective Perceptual Video Quality Measurement using a JND-Based Full Reference Technique," *Alliance for Telecommunications Industry Solutions Technical Report*, T1.TR.PP. 75-2001, 2001.

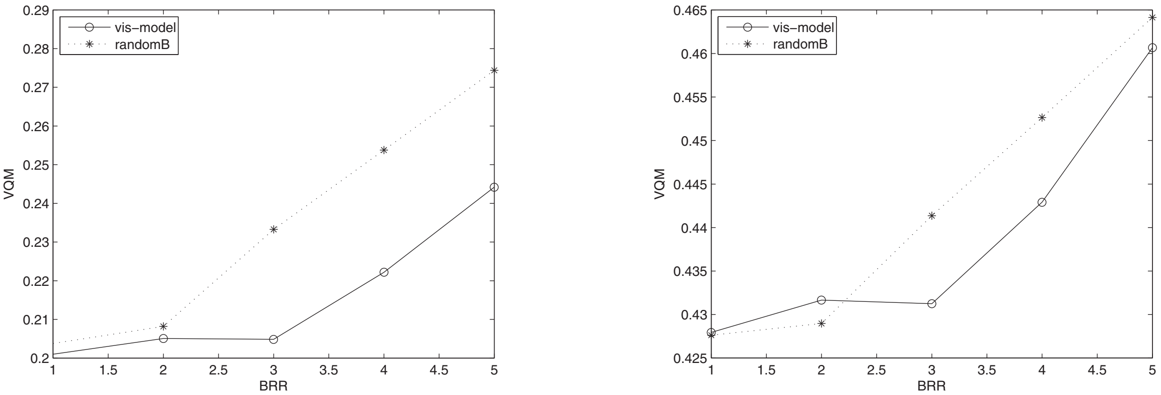
[3] S. Winkler, "A perceptual distortion metric for digital color video," *in Proc. SPIE Human Vis. Electron. Imag.*, San Jose, CA, 1999, vol. 3644, pp. 175-184.

[4] http://www.its.bldrdoc.gov/n3/video/index.php.

[5] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality, " *IEEE Trans. on Broadcasting*, vol. 50, pp. 312-322, Sep. 2004.

[6] M. H. Loke, E. P. Ong, W. Lin, Z. Lu, and S. Yao, "Comparison of video quality metrics on multimedia videos, " *IEEE ICIP*, Oct. 2006.

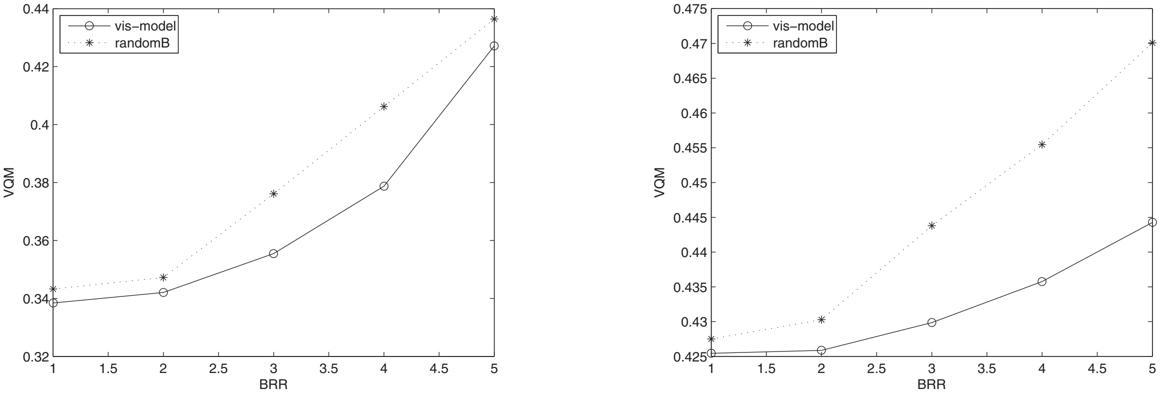
[7] S. Kanumuri, P. C. Cosman, A. R. Reibman, and V. Vaishampayan, "Modeling Packet-Loss Visibility in



(a) *news* at 400 kbps

(b) *news* at 200 kbps

Fig. 6. Average VQM scores vs. BRR for the combination of *news* at 200 kbps and *news* at 400 kbps.



(a) *template* at 400 kbps

(b) *news* at 200 kbps

Fig. 7. Average VQM scores vs. BRR for the combination of *template* at 400 kbps and *news* at 200 kpbs.

MPEG-2 Video," *IEEE Trans. on Multimedia*, vol. 8, pp. 341-355, Apr. 2006.

[8] S. Kanumuri, S. G. Subramanian, P. C. Cosman, and A. R. Reibman, "Packet-loss Visibility in H.264 Videos Using a Reduced Reference Method," *IEEE ICIP*, Oct. 2006.

[9] T.-L. Lin, S. Kanumuri, Y. Zhi, D. Poole, P. Cosman, and A. Reibman, "A Versatile Model for Packet Loss Visibility and its Application to Packet Prioritization," *IEEE Trans. on Image Processing*, Vol. 19, pp. 722-735, Mar. 2010.

[10] T.-L. Lin and P. Cosman, "Network-based packet loss visibility model for SDTV and HDTV for H.264 videos," *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2010.

[11] Y. Wang, T.-L. Lin and P. Cosman, "Network-based model for video packet importance considering both compression artifacts and packet losses," *IEEE Globecom 2010*, accepted.

[12] T.-L. Lin, J.Shin and P. Cosman, "Packet dropping for widely varying bit reduction rates using a network-based packet loss visibility model," *2010 Data Compression Conference*, 2010.

[13] P. McCullagh and J.A. Nelder, "Generalized Linear Models, 2nd Edition, " *Chapman and Hall*, 1989.

[14] The Official website of FFMPEG : http://ffmpeg.org/.