

Impact of Reference Distance for Motion Compensation Prediction on Video Quality

Yubing Wang^{1a}, Mark Claypool^b, and Robert Kinicki^b

^aEMC Corporation, 32 Coslin Drive, Southborough, MA 01729, USA

^bWorcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609, USA

ABSTRACT

Transmitting high-quality, real-time interactive video over lossy networks is challenging because data loss due to the network can severely degrade video quality. A promising feedback technique for low-latency video repair is Reference Picture Selection (RPS), whereby the encoder selects one of several previous frames as a reference frame for predictive encoding of subsequent frames. RPS can operate in two different modes: an optimistic policy that uses negative acknowledgements (NACKs) and a more conservative policy that relies upon positive acknowledgements (ACKs). The choice between RPS ACK mode and NACK mode to some extent depends upon the effects of reference distance on the encoded video quality. This paper provides a systematic study of the effects of reference distance on video quality for a range of video coding conditions. High-quality videos with a wide variety of scene complexity and motion characteristics are selected and encoded using H.264 with a bandwidth constraint and a range of reference distances. Two objective measures of video quality, PSNR and VQM, are analyzed to show that scene complexity and motion characteristics of the video determine the amount of degradation in quality as reference distance increases. In particular, videos with low motion degrade in quality more with an increase in reference distance since they cannot take advantage of the strong similarity between adjacent frames. Videos with high motion do not suffer as much with higher reference distance since the similarity between adjacent frames is already low. The motion characteristics also determine the initial quality under the bandwidth constraint. The data presented should be useful for selecting ACK or NACK mode or for modeling video repair techniques.

Keywords: RPS, Reference Distance, PSNR, VQM, H.264

1. INTRODUCTION

The growth in power and display capabilities of today's computers has enabled streaming video with a range of quality to be viewed by end-users. High-end users with modern desktop displays can watch videos in full-quality, wide-screen mode at their desk-top while low-end, mobile users with video-capable mobile phones can watch low resolution video on their mobile phone. The growth in computer technology has been matched by an equal the growth in capacity and connectivity of networks. Users on high-speed corporate and academic networks have had sufficient bandwidth to stream video for some time, but the pervasiveness of broadband networks has also given home users access to high-quality streaming video. Moreover, increasing bandwidth for digital cellular networks has enabled streaming video to mobile laptops, PDAs and even mobile phones. Despite the increase in network power and connectivity, many network connections still lose video data packets. Lost packets are especially detrimental to streaming video because of the dependency between video frames during encoding where one lost video packet can result in error propagation to many other video frames. While video clients that experience loss can use local concealment to visually cover up the loss, the ability to adequately repair video without feedback to the video server is limited. For video connections that have low end-to-end delay requirements between end hosts, such as for a video conference or interactive, thin-client desktops [1], feedback to the server to request retransmission of lost packets adds too much latency to the video session. Forward error correction [2] has also been used to ameliorate the effects of loss, but FEC requires additional data to be added to the video stream and FEC encoding and decoding can be somewhat complicated.

A promising feedback technique for low-latency video repair technique is Reference Picture Selection (RPS) [3, 4, 5], whereby the encoder selects one of several previous frames to use as a reference frame when encoding. Broadly, the reference frame can be the previous frame (called RPS NACK mode), as is done in typically video encoding, or the

¹wang_yubing@emc.com; phone 508-305-8849; claypool@cs.wpi.edu; phone 508-831-5409; rek@cs.wpi.edu; phone 508-831-6116

reference frame can be several frames back if the encoder waits for the receiver to confirm receipt of the frame (called RPS ACK mode). The choice between RPS ACK mode and RPS NACK mode depends upon the latency and loss pattern, but also upon the effects of reference distance on the encoded video quality. In particular, videos that do not significantly degrade with high reference distance will benefit from the use of RPS ACK over RPS NACK while videos that are sensitive to reference distance may choose RPS NACK or even some other repair techniques.

Although there have been numerous studies detailing the benefits to video quality for various repair techniques [5, 6, 7], these studies typically do not vary the reference distance during encoding. To the best of our knowledge, the effects of encoding distance on video quality have not been quantitatively studied. Thus, the choice between RPS NACK and RPS ACK is hard to determine and since RPS NACK uses an increased reference distance when packets are lost, predicting the video quality for any RPS case is difficult.

This paper provides a systematic study of the effects of reference distance on video quality for a range of video coding conditions. High-quality videos with a wide variety of scene complexity and motion characteristics are selected for baseline encoding. The videos are all encoded using H.264 [8, 9], an increasingly popularly deployed compression standard with support for RPS, with a bandwidth constraint and a range of reference distances. Two objective measures of video quality are used, the popular Peak Signal to Noise Ratio (PSNR), and the reportedly more accurate Video Quality Metric (VQM) [11]. Analysis shows that for both measures of quality, the scene complexity and motion characteristics determine the degradation of video with higher reference distances. In particular, videos with low motion degrade more with higher reference distance since they cannot take advantage of the similarity between adjacent frames. Videos with high motion do not suffer as much with an increase in reference distance since the similarity between frames is already low. The scene complexity determines the overall starting quality with a default, encoding reference distance of one and the bandwidth constraint. The data provided by this paper is useful for helping select the repair technique that maximizes video quality, particularly when choosing between RPS ACK and NACK modes. The data can also be useful for analytic modeling of the effects of latency (hence, reference distance) on video performance for various repair techniques [12, 13].

The rest of this paper is organized as follows: Section 2 provides background information on technologies used in this paper; Section 3 provides the hypothesis for this study; Section 4 describes the methodology used to test the hypothesis; Section 5 analyzes the experimental results; and Section 6 draws conclusions and describes possible future research.

2. BACKGROUND

This section provides background information on technologies used in this paper. Section 2.1 reviews H.264, the video compression standard explored in this study, and reference picture selection (RPS), the feedback-based error control technique that depends upon reference picture distance. Section 2.2 introduces two video quality measurement metrics, Peak-Signal-to-Noise-Ratio (PSNR) and Video Quality Metric (VQM).

2.1 H.264

H.264 [8, 9] represents the state of art in video compression standards and is used throughout this study to encode/decode the video clips. The H.264 video codec supports a broad range of applications from low bit-rate Internet streaming applications to HDTV broadcast. An H.264 picture consists of macro-blocks, each containing 16x16 luminance samples, and two corresponding 8x8 chrominance samples. Within each picture, macro-blocks are arranged in slices, where a slice is a set of macro-blocks in raster scan order. The number of macro-blocks per slice need not be constant within a picture. Macro-blocks can be classified as one of three types: I (intra-coded), P (predictive-coded) and B (bidirectional predictive-coded). I-Blocks are encoded independently and contain all information required to decode the macro-block. P-Blocks are encoded using motion compensation techniques which are dependent on the previous I or P macro-blocks. B-Blocks further exploit motion compensation techniques by using information in the previous and following I or P macro-blocks. H.264 allows the encoder to select among several pictures that have been successfully decoded in the decoder as reference pictures for motion-compensated prediction [10]. However, as the reference distance increases, the coding efficiency may degrade since the similarities between the encoding frame and the reference frame decreases. A P-Block can be further divided into macro-block partitions; i.e. blocks of size 16x16, 16x8, 8x16 or 8x8 luminance blocks. By using these finer partitions for motion-compensated prediction, better prediction accuracy can be achieved.

2.2 Reference Picture Selection (RPS)

As mentioned in the Section 1, transmitting high-quality, real-time interactive video over lossy networks is challenging because data loss due to transmission errors can severely degrade video quality. Feedback-based error control techniques use information on the data sent by the decoder to adjust the coding parameters at the encoder to achieve better error resilience [5]. Reference Picture Selection (RPS) [3, 4, 5] allows the encoder to select one of several previous frames that have been successfully decoded as a reference frame for predictive encoding of subsequent frames. RPS can operate in two different modes. In negative acknowledgement (NACK) mode, when a transmission error is observed by the decoder, the decoder sends a NACK message for the erroneous frame, along with the number of a previously-received, correctly-decoded reference frame that can be used as a reference for prediction, to the encoder. The encoder then uses that frame as reference picture to encode the current frame. As illustrated in Figure 1, when the decoder observes that frame 4 has a transmission error, it sends a NACK to the encoder with an explicit request to use frame 3, which has been decoded correctly, for prediction. The encoder then uses frame 3 as a reference to encode frame 7 after receiving the NACK from the decoder.

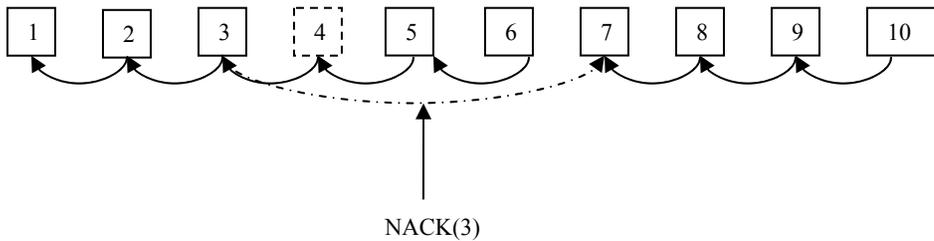


Fig. 1. Illustration of Encoding using RPS with NACK Mode. Frame 4 has a transmission error. The arrows indicate the selected reference frame.

In the RPS positive acknowledgement (ACK) mode, all correctly received frames are acknowledged and the encoder only uses acknowledged frames as reference pictures. Since the encoder has to use an older (compared with NACK mode) reference frame for prediction, the coding efficiency of RPS ACK degrades as the round-trip delay increases. On the other hand, using RPS ACK mode entirely eliminates error propagation since the encoder only uses acknowledged frames as reference frames. Figure 2 illustrates the use of RPS with ACK mode. In this example, if no transmission error occurs, the encoder receives an ACK for frame 1 while encoding frame 4, and thus uses frame 1 as a prediction reference to encode frame 4. For the same reason, the encoder uses frame 2 as reference picture for frame 5, and frame 3 for frame 6, and so on. However, if no ACK is received for frame 4, frame 7 uses acknowledged frame 3, instead of frame 4, as a reference frame.

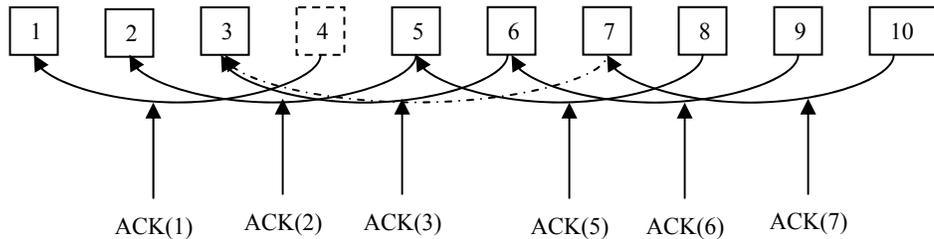


Fig. 2. Illustration of Encoding using RPS with ACK Mode. Frame 4 has a transmission error. The arrows indicate the selected reference frame.

2.3 Video Quality Measurement

There are two basic ways to evaluate video quality. One way is to have the videos viewed by a group of users for a subjective evaluation. Since having users view videos for evaluation is expensive and often impractical, several algorithms for predicting subjective video quality have been developed. Peak-Signal-to-Noise-Ratio (PSNR) continues

to be a popular evaluation of video quality [14]. PSNR is derived by setting the mean squared error (MSE) in relation to the maximum possible value of the luminance (for a typical 8-bit value this is $2^8 - 1 = 255$) as follows:

$$MSE = \frac{\sum_{i=1}^M \sum_{j=1}^N [(f(i, j) - F(i, j))]^2}{M \cdot N}$$

$$PSNR = 20 \cdot \log_{10} \left(\frac{255}{\sqrt{MSE}} \right)$$

Where $f(i,j)$ is the original signal at pixel (i, j) , $F(i, j)$ is the reconstructed signal, and $M \times N$ is the picture size. The result is a single number in decibels, typically ranging from 30 to 40 for medium- to high-quality video.

VQM [11] is developed by (ITS)² to provide an objective measurement for perceived video quality. VQM measures the perceptual effects of video impairments including blurring, uneven motion, global noise, block distortion and color distortion, and combines them into a single metric, a degradation from 0 to 1. Testing results show VQM has a high correlation with subjective video quality assessment and has been adopted by ANSI as an objective video quality standard. VQM can be computed using various models based on certain optimization criteria. These models include: (1) Television, (2) Videoconferencing, (3) General, (4) Developer, and (5) PSNR. In our study, we used the general model to evaluate video quality because H.264, the coding standard used in this study, can be used for a wide range of applications. The general model uses a linear combination of seven parameters to determine quality. Four parameters are based on features extracted from spatial gradients of the Y luminance component, two parameters are based on features extracted from the vector formed by the two chrominance components (CB, CR), and one parameter is based on contrast and absolute temporal information features, both extracted from the Y luminance component.

Both PSNR and VQM are used to evaluate video quality in this investigation. The expectation is that the relationship between video quality and reference distance revealed in this study should hold for both video quality metrics.

3. HYPOTHESIS

As the reference distance increases, the coding efficiency decreases since the similarities between the current frame and the reference frame decreases. If the network capacity is constrained, the video quality degrades as the coding efficiency drops. The degree of the coding efficiency degradation is affected by the video content. For instance, if a video sequence contains high motion scenes, then the similarities among adjacent frames are low. Thus, there are more macro-blocks within the video that must be intra encoded. On the other hand, if a video sequence contains low motion scenes, it is more likely the macro-blocks within the video can be inter-coded using motion-compensation predictions since the similarities among frames are high. Since the intra-coded macro-blocks are independent of the reference frames, the coding efficiency for pictures containing more intra-coded macro-blocks (high motion) degrades less with an increase in reference distance than those containing more inter-coded macro-blocks (low motion). Figure 3 depicts our hypothesis of the relationship between video quality and reference distance for videos with high motion and low motion. As shown in the figure, as the reference distance increases, the video quality degrades as the coding efficiency drops for both high-motion and low-motion videos. However, the video quality for high-motion videos degrades slower than those low-motion videos as the reference distance increases.

² The Institute for Telecommunication Science

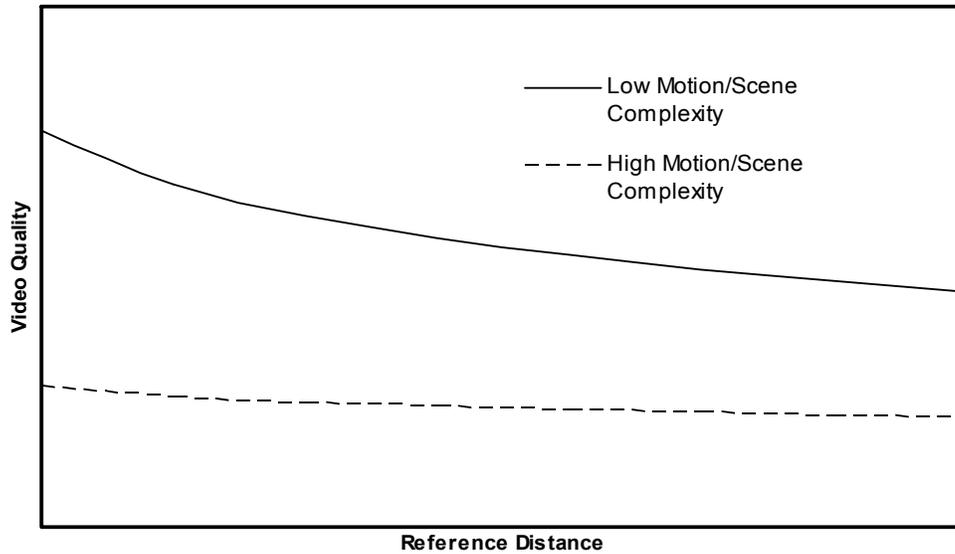


Fig. 3. Hypothesis of the relationship between Video Quality and Reference Distance for videos with high motion and low motion.

4. METHODOLOGY

In order to explore the relationship between video quality and reference distance, the following methodology was used:

- Select a set of video clips with a variety of motion content (see Section 4.1).
- Change reference distances for each selected video sequence (see Section 4.2).
- Encode the video clips using H.264 (see Section 4.3).
- Measure video quality using PSNR and VQM (see Section 4.4).
- Analyze the results (see Section 5).

4.1 Select Video Clips

A set of video clips with a variety of motion content are selected to determine the effects of reference distance over a wide range of videos. These video clips are all in YUV 4:2:0 formats which are widely used in the video research community. The picture resolutions are all common intermediate format (CIF, 352x288 pixels). Each video sequence contains 300 video frames with a frame rate of 30 frames/second (fps). The content of these video clips can be roughly categorized into three groups: high motion/scene complexity, medium motion/scene complexity and low motion/scene complexity. Table 1 provides an approximate content classification of each video clip, with an identifying name and a short description of the video content.

Table 1. Video Clips Used in the Experiments.

Video Clip	Motion	Description
Akiyo	Low	A news reporter talking
Container	Low	A container ship moving slowly
News	Low	Two news reporters talking
Silent	Medium	A person demonstrating sign language
Mom & Daughter	Medium	A mother and daughter talking
Foreman	High	A foreman talking
Mobile	High	Panning of toy train moving
Coastguard	High	Panning of a coastguard ship moving

Table 2 shows the fraction of P-Blocks in the encoded video clips. For the low-motion videos (Akiyo and Container), the majority of the macro-blocks are inter-coded (P-blocks), whereas for the high-motion videos (Foreman and Coastguard), only around half of the macro-blocks are inter-coded.

Table 2. The Fraction of the Inter Blocks for Different Video Clips

Video Clip	Fraction of the Inter Blocks (P-Blocks)
Akiyo	0.9666
Container	0.9246
News	0.8746
Silent	0.8637
Mom & Daughter	0.8423
Foreman	0.5947
Mobile	0.5722
Coastguard	0.5225

4.2 Changing Reference Distance

The main purpose of this study is to explore the relationship between video quality and reference distance. Thus, the encoder needs to be able to alter the reference distance in a controllable manner while encoding a video sequence. One way to achieve this is to modify the encoder to select the reference pictures specified by users. This approach is complicated as it involves modifying the encoder and may result in inaccurate measurements if done incorrectly. We take an alternative approach by changing the input video sequences. For instance, to use a reference picture which is two frames before the current frame instead of one frame before, the odd-number frames (1, 3, 5, ...) are extracted and then the even-number frames (2, 4, 6, ...). Both sequences are fed into the unmodified encoder, resulting in two video quality values. The video quality for the original video sequence is the average of these two quality values. The same approach is applied to other reference distances.

4.3 Encode/Decode

H.264 is used for video compression to encode/decode the video clips. The H.264 encoder/decoder used by this study is the Joint Model (JM 10.2)³ developed by the Joint Video Team (JVT) which consists of experts from ITU-T VCEG and ISO/IEC MPEG. In this study, the following settings are applied to all experiments:

³ <http://iphome.hhi.de/suehring/tml/download/>

- Since our study mainly explores how changing reference distance affects the quality of P-frames, primarily used in videoconferences, all the video frames are encoded as either P-frames or I-frames, and no B-frames are used in the experiments.
- Under a bit-rate constraint, changing reference distance affects encoding efficiency and thus video quality. For fair comparison, the same bit-rate constraint is imposed for all experiments.
- H.264 supports multiple reference picture motion compensation which allows the encoder to select among several pictures that have been decoded and stored in the decoder. Since in our study the reference distance is between the encoding frame and a previously encoded frame, only one single reference frame is used in our experiments.

4.4 Measure of Video Quality

PSNR and VQM are used as metrics to measure video quality. The purpose of using two different quality metrics is to investigate whether different quality metrics have different relationships between video quality and reference distance. The PSNR measurement is conducted by JM as it reports the resulting PSNR for each video sequence being encoded. VQM is not reported by JM, so we have to use a VQM measurement tool named VQM-PC, downloaded from the VQM web site⁴. This VQM tool takes the original and the processed video clips as input and measures the video quality of the processed video clips relative to the original video clips. The resulting VQM score is in the range of (0, 1), where 0 represents no impairment and 1 represents the maximum impairment.

5. ANALYSIS

Chapter 3 hypothesizes that the video quality degrades as the reference distance increases, and the video quality for high-motion videos degrades slower than those low-motion videos as the reference distance increase. A series of experiments are conducted as described in Chapter 4. Figure 4 depicts PSNR versus reference distances for eight video clips with different content. The bit-rate constraint for this experiment is 4.8 Mbps. As shown in Figure 4, as the reference distance increases, the PSNR for all the video clips degrade. However, the degrees of quality degradations of the eight videos are different. Akiyo shows the steepest quality degradation: as the reference distance is increased from 1 to 8, the PSNR drops from 48.06 db to 43.74 db. Coastguard shows the slowest quality degradation: as the reference distance is increased from 1 to 8, the PSNR drops from 35.47 db to 33.8 db.

Figure 5 depicts the trendlines and equations for Akiyo, Mom & Daughter, and Coastguard. As Figure 5 shows, the curves can be well described using the logarithmic function: $y = a \ln(x) + b$, where a is the gradient of the logarithmic function, determined by the motion of a video clip, and b is the y-intersection of the logarithmic function, determined by the scene complexity of a video clip. The coefficients of the logarithmic functions for all the curves shown in Figure 4 are presented in Table 3. Table 3 shows that as the amount of motion increases in the video clip, the gradient (a) of the quality degradation decreases, and as the scene complexity increases, the intersect (b) decreases. Note that Mobile has the most complex scene among the eight video clips. The R-Squared values for all the logarithmic functions are also presented in Table 3.

⁴ http://www.its.bldrdoc.gov/n3/video/vqmdownload_US.htm

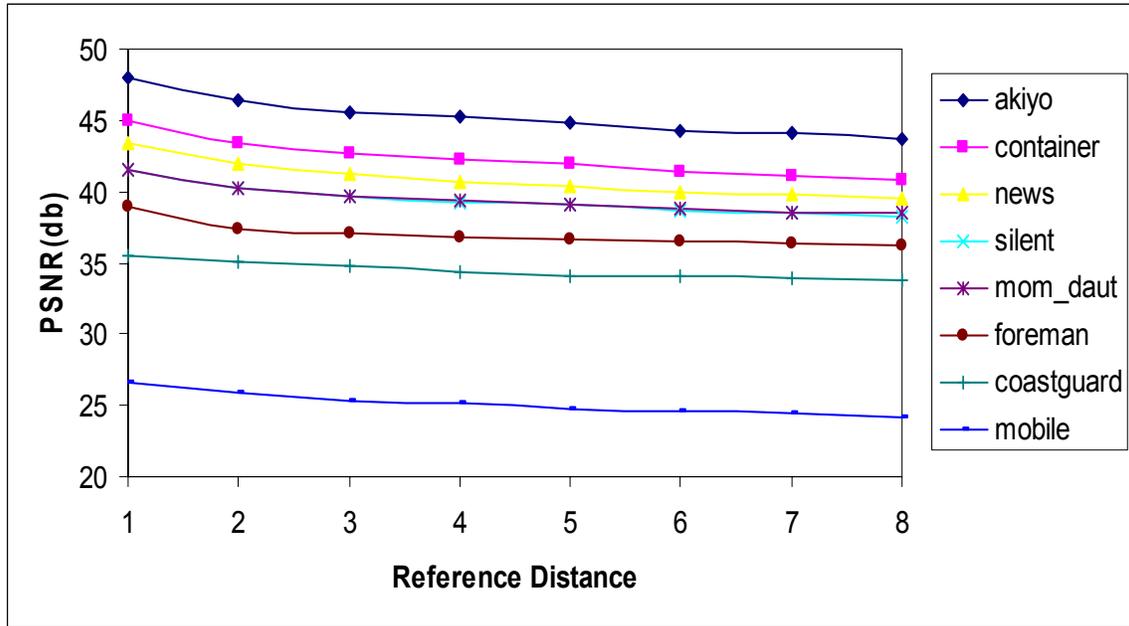


Fig. 4. PSNR vs. Reference Distance for Video Clips with different content characteristics

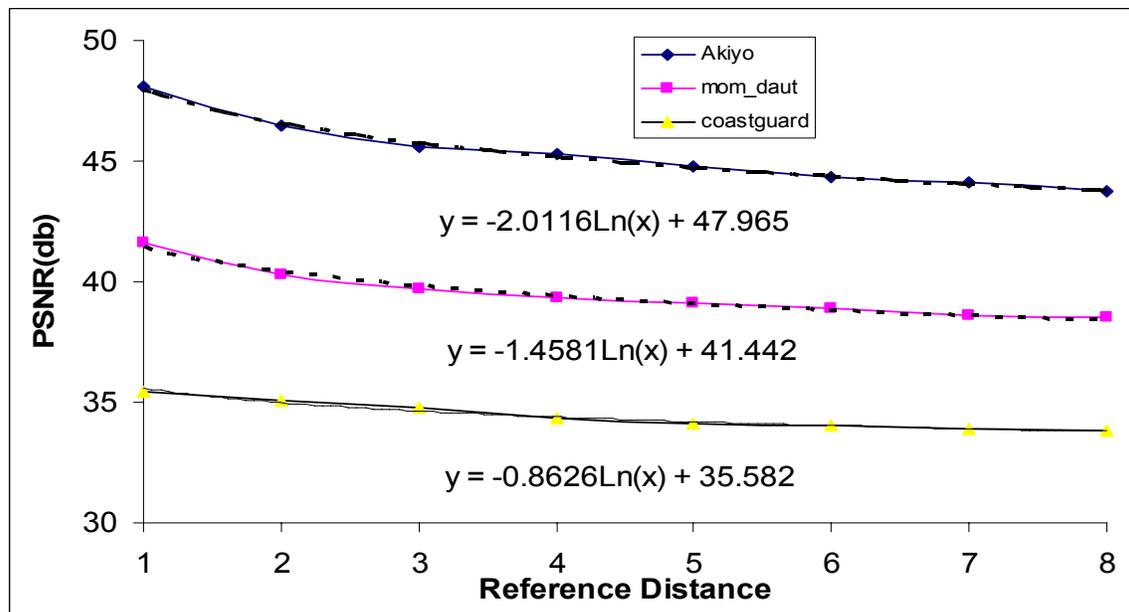


Fig. 5. Trendlines and Equations for Akiyo, Mom & Daughter, and Coastguard

Table 3. The Coefficients that Describe the Relationship between PSNR versus Reference Distance

Video Clips	Gradient (a)	Intersect (PSNR) (b)	R-Squared
Akiyo	-2.0116	47.965	0.9953
Container	-1.9023	44.838	0.9948
News	-1.8556	43.295	0.9984
Silent	-1.5283	41.41	0.9929
Mom & Daughter	-1.4581	41.442	0.9904
Foreman	-1.1681	38.511	0.9265
Mobile	-1.1553	26.663	0.9754
Coastguard	-0.8626	35.582	0.9975

We further test the hypothesis by using VQM to examine whether the results using PSNR hold for VQM. Figure 6 depicts VQM versus reference distance for eight video clips, the same clips used in the PSNR experiment. The bit-rate constraint for this experiment again is 4.8 Mbps. We use (1-VQM) as the quality metric for better comparisons with PSNR (i.e. higher values are better). Notice, by adopting this quality metric, 1 represents the best quality and 0 represents the worst. Figure 6 shows the same trend as Figure 4: as the reference distance increases, the video quality (1-VQM) degrades. Akiyo shows the steepest quality degradation: as the reference distance is increased from 1 to 8, the quality drops from 0.972 to 0.890. Coastguard shows the slowest quality degradations: as the reference distance is increased from 1 to 8, the quality drops from 0.843 to 0.831.

Figure 7 depicts the trendlines and equations for Akiyo, Mom & Daughter, and Coastguard using VQM. As Figure 7 shows, the lines can be well described using linear functions: $y = ax + b$, where a is the gradient of the linear function, determined by the motion of a video clip, and b is the y-intersection of the linear function, determined by the scene complexity of a video clip. The coefficients of the linear functions for all the lines shown in Figure 6 are presented in Table 4. Table 4 shows that, as the amount of motion increases in the video clip, the gradient (a) of the quality degradation decreases, and as the scene complexity increases, the intersect (b) decreases. The R-Squared values for all the linear functions are also presented in Table 4.

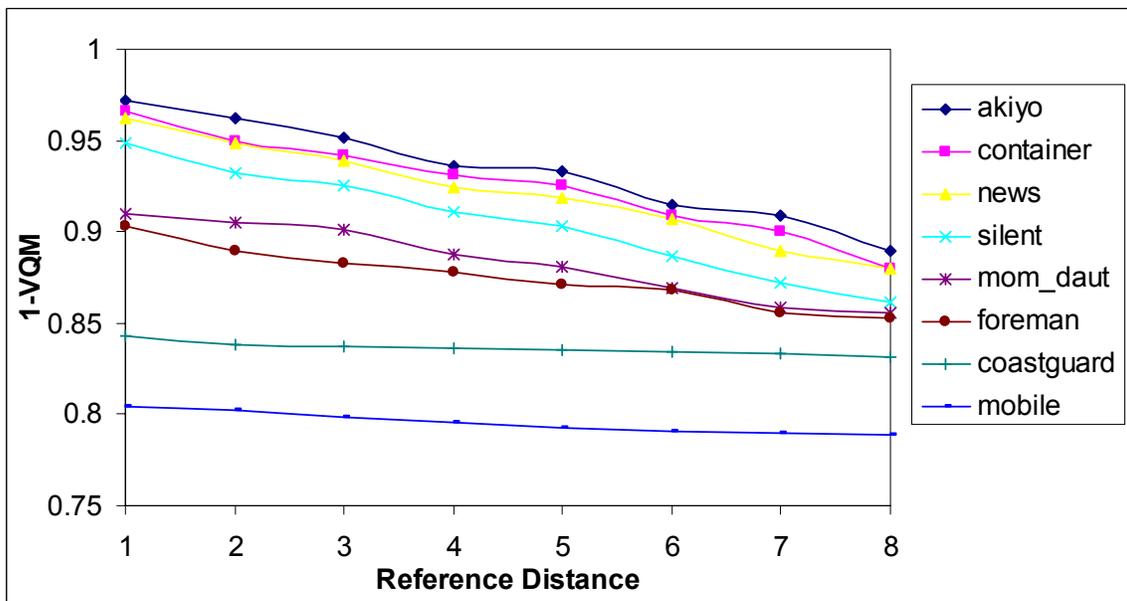


Fig. 6. VQM vs. Reference Distance for Video Clips with Different Content

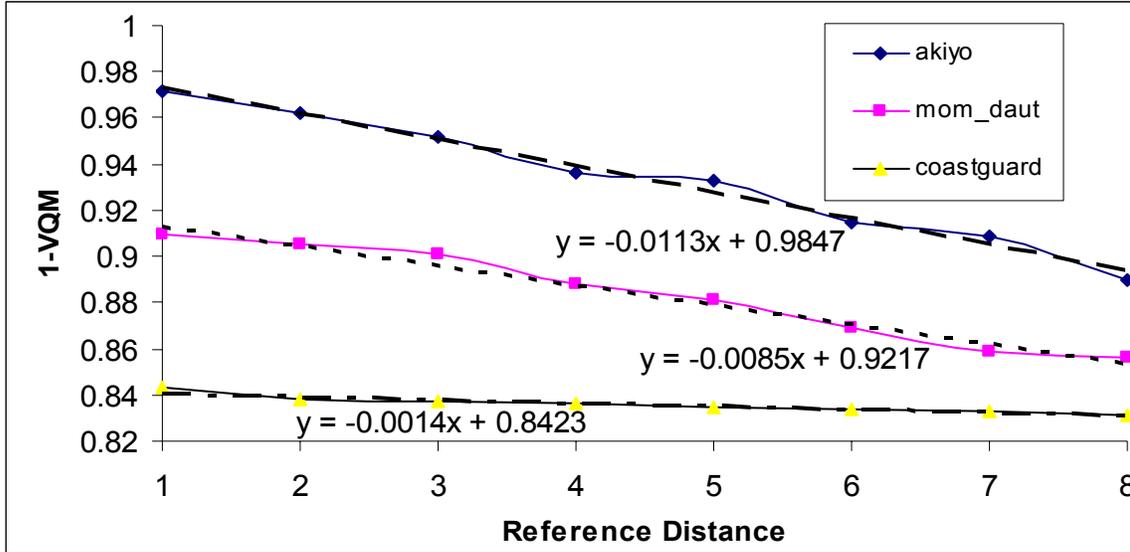


Fig. 7. The Trendlines and Equations for Akiyo, Mom & Daughter, and Coastguard

Table 4. The Coefficients that Describe the Relationship between (1-VQM) vs. Reference Distance

Video Clips	Gradient (a)	Intersect (1-VQM)	R-Squared
Akiyo	-0.0113	0.9847	0.9869
Container	-0.0114	0.9766	0.9848
News	-0.0115	0.9732	0.9931
Silent	-0.0124	0.9606	0.9937
Mom & Daughter	-0.0085	0.9217	0.9821
Foreman	-0.0068	0.9059	0.9779
Mobile	-0.0022	0.8055	0.9076
Coastguard	-0.0014	0.8423	0.9671

6. CONCLUSION AND FUTURE WORK

Reference Picture Selection allows the encoder to select one of several previous frames that have been successfully decoded as a reference frame for predictive encoding of subsequent frames. By using only the successfully decoded frames as reference frames for motion-compensated prediction, RPS reduces the error propagation due to data loss and achieves better error resilience. However, since the encoder may have to use an older reference frame for prediction, the coding efficiency decreases as the reference distance increases since the similarity between the encoding frame and the reference frame decreases. If the network capacity is constrained, the video quality degrades as the coding efficiency drops. It is thus important to explore the relationship between video quality and reference distance to understand how changing the reference distance affects the performance of RPS.

In this study, a series of experiments are conducted to reveal how the change of reference distance affects video quality. A set of video clips with a variety of motions are selected for study, and the video sequences are shuffled to change the reference distances. For each reshuffled video sequence, an H.264 encoder encodes the sequence and measures video quality with PSNR and VQM.

From analysis of the experimental results, the relationship between video quality and reference distance can be determined. Both PSNR and VQM video quality degrade as reference distance increases. The degree of the video quality degradation is affected by the video content. The video quality for videos with high motions tends to degrade slower than that for those videos with low motion. This is largely because high-motion videos have a much larger number of inter-coded macro-blocks (P-blocks) and are thus less sensitive to the change of reference distance than the low-motion videos. Although these findings hold for both PSNR and VQM, the characterizations of the relationship between video

quality and reference distance are different. While the relationship between PSNR and reference distance can be characterized using a logarithmic function, with VQM as the video quality metric, the same relationship can be characterized using a linear function.

Our study selects a diverse yet still somewhat limited set of videos. Future work could be to continue our current work with a broader set of videos. The work in this paper did not explore how the relationship between video quality and reference distance is affected by capacity constraints. Understanding these relationships enables further investigation of the behavior of RPS under different network capacity constraints.

REFERENCES

1. Davy De Winter, Pieter Simoens, Lien Deboosere, Filip DeTurck, Joris Moreau, Bart Dhoedt, and Piet Demeester. "A Hybrid Thin-Client Protocol for Multimedia Streaming and Interactive Gaming Applications", In *Proceedings of the Network and Operating System for Digital Audio and Video Workshop (NOSSDAV)*, Newport, RI, USA, May 2006.
2. J.C. Bolot, S. Fosse-Parisis, and D. Towsley. "Adaptive FEC-Based Error Control for Internet Telephony". In *Proceedings of IEEE INFOCOM*, Mar, 1999.
3. S. Fukunaga, T. Nakai, and H. Inoue. "Error Resilient Video Coding by Dynamic Replacing of Reference Pictures", In *Proceedings of IEEE Global Telecommunications Conf. (GLOBECOM)*, London, U.K., vol.3, Nov. 1996, pp.1503–1508.
4. Y.Tomita, T.Kimura, and T. Ichikawa. "Error Resilient Modified Inter-frame Coding System for Limited Reference Picture Memories", In *Proceedings of Int. Picture Coding Symp. (PCS)*, Berlin, Germany, Sept. 1997, pp.743–748.
5. B. Girod, and N. Färber. "Feedback-Based Error Control for Mobile Video Transmission". In *Proceedings of IEEE*, vol. 97, no. 10, pp. 1707-1723, Oct. 1999.
6. Y. Wang, Q. Zhu. "Error Control and Concealment for Video Communication: A Review", *Proceedings of the IEEE*, Vol. 86, No. 5, May 1998.
7. Y. Wang, S. Wenger, J. Wen, and A. K. Katsaggelos, "Error Resilient Video Coding Techniques," *IEEE Signal Proc. Mag.*, Vol. 17, pp. 61-82, July 2000.
8. I. E. G. Richardson. "H.264 and MPEG-4 Video Compression: Video Coding for Next Generation Multimedia", *Wiley*, ISBN 0-470-84837-5, 2004.
9. Joint Video Team of ITU-T and ISO/IEC JTC 1. "Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC)," document JVT-G050r1, May 2003; technical corrigendum 1 documents JVT-K050r1 (non-integrated form) and JVT-K051r1 (integrated form), March 2004; and Fidelity Range Extensions documents JVT-L047 (non-integrated form) and JVT-L050 (integrated form), July 2004.
10. T. Wiegand, X. Zhang, and B. Girod, "Long-Term Memory Motion-Compensated Prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 9, No. 1, pp. 70-84, Feb. 1999.
11. M. Pinson and S. Wolf. "A New Standardized Method for Objectively Measuring Video Quality," *IEEE Transactions on Broadcasting*, Vol. 50, No. 3, pp. 312-322, Sept. 2004.
12. K. Mayer-Patel, L. Le, and G. Carle. "An MPEG Performance Model and Its Application to Adaptive Forward Error Correction", In *Proceedings of ACM Multimedia*, Dec. 2002.
13. Huahui Wu, Mark Claypool, and Robert Kinicki. "A Model for MPEG with Forward Error Correction and TCP-Friendly Bandwidth", In *Proceedings of Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV)*, Monterey, California, USA, June 2003.
14. VQEG. "Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment", [Online] at <http://www.vqeg.org/>, Mar. 2000.