社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE

JPEG2000 ビデオ･ストリーミングの為の知覚品質に基づく QoS 制御方式

Anna YAMAGUCHI† Stefan WINKLER‡ Riaz ESMAILZADEH† 中川 正雄†

†慶應義塾大学理工学部情報工学科 〒223-8522 横浜市港北区日吉 3-14-1
‡Genista Corporation, #206B Telok Ayer Street, 06841 Singapore

E-mail: {ania, riaz, nakagawa}@nkgw.ics.keio.ac.jp

あらまし この本稿では情報圧縮規格である Motion-JPEG2000 のビデオ知覚品質 QoS 制御方式に新しい方式を提案する。今日に於けるすべてのネットワーク（インターネット・移動又は無線 LAN）ビデオ QoS は殆どの場合、Peak Signal-to-Noise Ratio（PSNR）ノイズ割合最高点を測定尺度として使われている。しかしながら、PSNR は実際の人による視覚を考慮しておらず、QoS の信頼性のある予告は出来ない。人による観察はデジタル・ビデオにおいてスムースさ、ブロック歪、ぼやけなどの異なった種類のゆがみを知覚する。私達は知覚品質 QoS の根本的な定義を考慮し、ランダムビットエラーの存在に於ける知覚されたビデオ品質を最大化する研究報告をする。

キーワード 知覚品質、QoS、ビデオ、ストリーミング、JPEG2000、移動通信

Perceptual Quality of Service Control for Wireless Video Streaming

Anna YAMAGUCHI† Stefan WINKLER‡ Riaz ESMAILZADEH† and Masao NAKAGAWA†

†Dept. of Information and Computer Science, Keio University
3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan
‡Genista Corporation, #206B Telok Ayer Street, 06841 Singapore

E-mail: {ania, riaz, nakagawa}@nkgw.ics.keio.ac.jp

Abstract This paper reports on perceptual quality of service (PQoS) investigation for the Motion-JPEG2000 video compression. In all today’s networks (Internet, mobile or Wireless Local Area Networks), video quality is measured mostly using peak signal-to-noise ratio (PSNR). PSNR, however, does not take into account human vision and thus cannot be a reliable predictor of perceived visual quality. Human observers will perceive different kinds of distortions in digital video, like jerkiness, blockiness, and blurriness. We consider a more fundamental definition of quality, PQoS, and report on a method which maximizes the perceived video quality in the presence of random bit errors, as these conditions are expected in the context of wireless transmission.

Keyword Perceptual Quality, Quality of Service, Video, JPEG2000, Streaming, Mobile Communications

1. Introduction

With increasing Internet traffic congestion, the demand and provision of quality of service (QoS) has been a topic of active research. QoS refers to the ability to ensure consistent, predictable delivery of a service, satisfying customer requirements. One of the most significant challenges for IP-based networks, which have traditionally provided only best-effort service, has been to provide some type of service guarantees for different types of traffic. This has been a particular challenge for streaming video applications [1], which often require a significant amount of reserved bandwidth to be useful.

In all today's networks (Internet, mobile or Wireless Local Area Networks), video QoS is measured mostly using peak signal-to-noise ratio (PSNR). PSNR, however, does not take into account human vision and thus cannot be a reliable predictor of perceived visual quality [2], [3]. Human observers will perceive different kinds of distortions in digital video, like jerkiness, blockiness, and blurriness. We consider a more fundamental definition of quality – perceptual quality of service (PQoS), and report on a method that maximizes the perceived video quality in the presence of random bit errors.

The proposed method uses an adaptive source-channel coding to achieve a target PQoS. Our simulation results illustrate that there are optimal combinations of source-channel coding which can deliver a desired level of quality. Furthermore, the optimal combination is different for each video, because it depends on the content. We believe the best way of controlling video delivery is through considering user-based perceptual metrics, such as defined in this paper. Optimal combinations of source and channel coding can then be made to deliver a desired level of quality.

The rest of the paper is organized as follows: Section 2 describes gives a brief overview of the JPEG2000 standardization. The proposed technique and test materials are presented in Section 3. Simulation-based evaluation and an observation of PQoS achievement follow in Section 4. Finally, Section 5 concludes the paper with observations and remarks.

2. Motion-JPEG2000

A set of error resilient tools has been included in the JPEG2000 standard [4] to reduce the impact of transmission errors on compressed images over noisy channels, e.g. hierarchical structure of a bit-stream or independent coding of consecutive frames. However, these tools only detect where errors occur, conceal the erroneous data, and resynchronize the decoder. They do not correct transmission errors. Bit errors can still affect the coded information. In particular, if a packet header, which is the most important part of the codestream, is corrupted, the code-block contributions from that packet's body cannot be correctly recovered. For these reasons, they are not sufficient in the context of wireless transmissions.

To be widely adopted for wireless multimedia applications, JPEG 2000 has to be robust to transmission errors. To address this issue, the JPEG committee has established in December 2001 a new work item, JPEG 2000 Wireless (JPWL), as Part 11 of the standard.

The main functionality of the JPWL system is to protect the codestream against transmission errors. The protection technique modifies the codestream to make it more resilient to errors, e.g. by adding redundancy or interleaving the data. Since bit errors in the packet header have a more devastating impact on image quality that error in the packet body [5], one way to reduce or eliminate the effect of error prone channels is the use of forward error correction (FEC) [6], [7].

We focus on video sequences coded with the ISO Motion-JPEG2000 video-coding algorithm [8]. Recent studies have shown that Motion JPEG2000 is very well suited for video transmission over wireless channels [9], [10].

3. Proposed Technique

The proposed technique is based on applying convolutional codes to perform the recovery, and controlling PQoS of the video by changing the rate of applied convolutional codes. A combination of source and channel coding rates is used to achieve a desired PQoS level.

As test material we used color video sequences with a frame size of 352 x 240 pixels that were selected from the set of test sequences used by the Video Quality Experts Group (VQEG)[1]. They cover different kinds of content, from almost static scenes (“news”) to strong motion (“football”). We found that the amount and visibility of distortions depend strongly on the video content. The images are compressed using the Kakadu v4.0.3 JPEG2000 software codec [11]. However, when the header contains too many bit errors, the decoder may not be able to decode the received bit-stream. For this reason, when not applying convolutional codes, we protected the header bytes by not introducing any bit errors there.

For analyzing the quality of the video, we use Genista’s *Video PQoS™* software[2]. It is an application for the measurement of artifacts affecting the perceptual quality of digital video that takes into account the video content. The measured impact of the artifacts is then combined into a prediction of Mean Opinion Score (MOS), i.e. an estimate of average viewer ratings on overall video quality. MOS is computed on a scale from 1 (unacceptable) to 5 (excellent) [10], [12].

Jerkiness is a perceptual measure of motion that does not look smooth. A primary cause of perceived jerkiness is the dropping of frames. Other factors are network congestion, packet loss, reduced frame rate. Jerkiness can also be introduces by the encoder dropping or repeating entire frames. Perceived jerkiness is measured in percent. 100% jerkiness for a frame is the case when a frame has been repeated. Lower levels can be perceived when sub-regions of the image appear to be moving in a jerky way.

For this paper, we analyzed MOS and jerkiness of the processed video using Genista’s full-reference metrics.

4. Simulation Results

In this section, we present the performance evaluation.

The source coding rates used in this work are 0.25, 0.5, 0.75 and 1 bits per pixel (bpp). Our results of predicted Mean Opinion Score with and without applying convolutional code for “news” and “football” video sequences are shown in Figure 1 and Figure 2, respectively.
The thick line shows the envelope for the highest PQoS level which can be delivered considering channel bit errors. The total transmission rate for sending video to the receiver is 2 Mb/s. At first, the source encoder outputs 1 bpp, and no channel coding is performed. Next, the source encoder outputs 0.75 bpp, and a rate 3/4 convolutional encoder is used. Finally, when BER > 0.01, the source coding rate 0.5 bpp with convolutional encoder and the code rate 1/2 is used. It can be

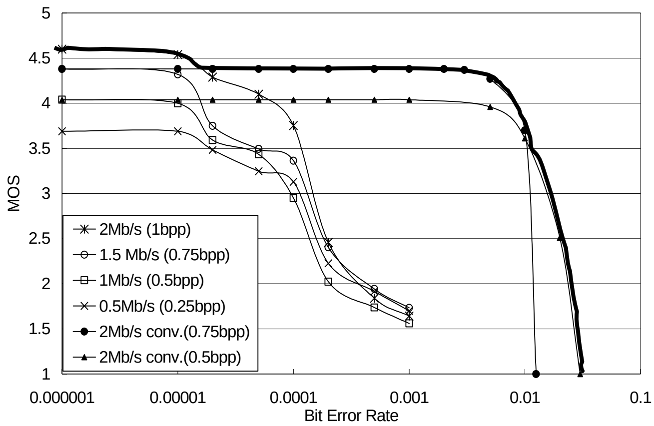

Fig. 1. Predicted Mean Opinion Score with and without applying convolutional code with code rate 3/4 and 1/2 for video sequence “news”.

---

[1] See http://www.vqeg.org/

[2] See http://www.genista.co.jp/ for more information.

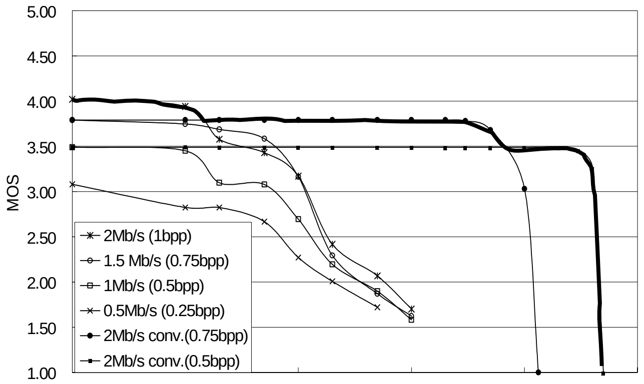

Fig. 2. Predicted Mean Opinion Score with and without applying convolutional code with code rate 3/4 and 1/2 for video sequence “football”.

seen that under good channel conditions, better PQoS is obtained with no coding (the first part of thick line). However, for BER > 0.00002 PQoS can be significantly improved using adaptive source and channel coding as shown by the thick envelope.

Figure 3 shows predicted average levels of jerkiness for the video sequence “football”. Curves are plotted as a function of the bit error rate for source-channel coding with and without using convolutional codes. We can see that without applying convolutional codes, already for BER > 0.0001 the level of jerkiness in motion video is increasing and achieves over 20 % even for BER < 0.001. However, our proposed technique of using an adaptive source-channel coding allows accomplishing very low level of jerkiness even for high BER.

5. Conclusions

We presented a new technique to deliver Motion-JPEG2000 video sequences with high perceptual quality of service. The proposed method uses an adaptive source-channel coding to achieve a target PQoS. Our simulation results illustrate that there are optimal combinations of source-channel coding which can deliver a desired level of quality. We believe the best way of controlling video delivery is through considering user-based perceptual metrics, such as defined in this paper. Optimal combinations of source

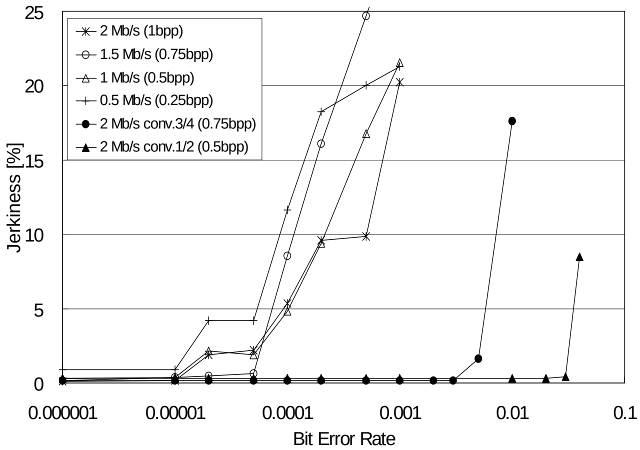

Fig. 3. Predicted average levels of jerkiness for video sequence “football” with and without applying convolutional code with the code rates 3/4 and 1/2.

and channel coding can then be made to deliver a desired level of quality.

References:

[1] D. Quaglia and J.C. De Martin, “Adaptive Packet Classification for Constant Perceptual Quality of Service Delivery of Video Streams over Time-Varying Networks,” Proc. of IEEE Int. Conf. on Multimedia & Expo (ICME), Baltimore, vol. 3, pp. 369-372, Jul 2003.

[2] A. R. Prasad, R. Esmailzadeh, S. Winkler, T. Ihara, B. Rohani, B. Pinguet and M. Capel, “Perceptual Quality Measurement and Control: Definition, Application and Performance,” Proc. 4th International Symposium on Wireless Personal Multimedia Communications, pp.547–552, Denmark, Sept. 2001.

[3] S. Dixit and R. Prasad, “Wireless IP and Building the Mobile Internet,” Artech House, 2003.

[4] JPEG 2000 Part I: Final Draft International Standard (ISO/IEC FDIS15444-1), ISO/IECJTC1 /SC29/WG1 N1855, Aug. 2000.

[5] V. S. Sachez and M. K. Mandal, “Adaptive Unequal Channel Protection for JPEG2000 Images”, Proc. of Indian Conference on Computer Vision,

Graphics and Image Processing, India, Dec. 2002.

[6] A. Perkis, “On the Importance of Error Resilience in Visual Communications Over Noisy Channels,” Circuits Systems Signal Processing, vol. 20, no. 3, pp. 415-446, 2001.

[7] I. E. G. Richardson, “Video Codec Design,” John Wiley & Sons, Apr. 2003.

[8] JPEG2000 Part 3: Motion JPEG2000, International Standard (ISO/IEC JTC 1/SC 29/WG 1 N2539, 15444-3), March 2002; available for download from: http://www.jpeg.org/jpeg2000/CDs15444.html

[9] F. Dufaux and T. Ebrahimi, “Motion JPEG2000 for Wireless Applications,” 1st. International Workshop on JPEG2000, Lugano, July 2003.

[10] S. Winkler and F. Dufaux, “Video Quality Evaluation for Mobile Applications,” Proc. SPIE/IS&T Visual Communication and Image Processing, vol. 5150, pp. 593–603, Switzerland, July 2003.

[11] Kakadu software implementation of the JPEG 2000 standard, available for download from: http://www.kakadusoftware.com/

[12] P. Marziliano, F. Dufaux, S. Winkler and T. Ebrahimi, “Perceptual Blur and Ringing Metrics: Application to JPEG2000,” Signal Processing: Image Communication, vol. 19, no. 2, pp.163–172, Feb. 2004.