No-Reference Image Quality Estimation for Motion JPEG 2000 Enabling Precise Estimation of PSNR Values

Kiyoshi NISHIKAWA and Hitoshi KIYA
Tokyo Metropolitan University, Hino 191-0065 Tokyo Japan
E-mail: knishikawa@m.ieice.org, kiya@sd.tmu.ac.jp Tel: +81-42-585-8423

***Abstract*—In this paper, we propose a no-reference (NR) image quality estimation method based on the novel quality layer construction of Motion JPEG 2000. Using the proposed method, we can precisely estimate the quality of images, in terms of the PSNR, those degraded due to network packet loss. Through the computer simulations, we show that the method achieves the maximum estimation error of about 1[dB] in PSNR. The proposed method maps a PSNR (or equivalently MSE) value to each layer so that we can easily estimate PSNR of received image by examining the number of layers received without data loss. The advantages of the proposed method are (i) we can estimate the PSNR or MSE from the code-stream before the decoding operation at the receiver-side and (ii) no additional calculation is required to implement at the receiver.**

***Index Terms*—No-reference image quality estimation, Motion JPEG2000, PCRD optimization**

I. Introduction

In this paper, we propose a no-reference (NR) image quality estimation method for Motion JPEG 2000 based on the novel quality layer construction given target PSNRs (peak signal to noise ratio) for JPEG 2000[1]. We show that the proposed method enables us to precisely estimate the image quality in terms of PSNR with the maximum estimation error of about 1[dB]. Note that we use the PSNR to measure the image quality because the proposed method enables the system to judge whether the received image satisfies the specified quality before decoding the stream.

The Motion JPEG 2000 is a video coding standard based on JPEG 2000 and is intended to be used in applications where high quality frame images are required, i.e., remote surveillance systems, or digital cinema[1]. When a video sequence is transmitted through the Internet, we cannot avoid degradation of the image quality due to the network packet loss. In some applications including those mentioned above, a precise estimation of the degradation of the received frames are required to maintain their quality at the receiver-side.

Image quality estimation without any information of the original one is called no-reference (NR) estimation method[2], [3], [4], [5]. The target of most of the existing NR methods for video transmission is MPEG or H.264/AVC, e.g., [4], [5], and we cannot apply them to Motion JPEG 2000. In [2], we proposed an NR method for the Motion JPEG 2000, to estimate the image degradation due to the packet loss. Although the method[2] is a kind of NR estimation for Motion JPEG 2000 and it provides precise estimation, it could only estimate the PSNR between the JPEG 2000 encoded image at the sender, and the received one. In other words, we could not estimate the true PSNR between the original reference image and the received one. Besides, it requires the information from the neighboring frames to estimate.

In this paper, we propose a novel NR image quality estimation method based on the quality layer construction given the target PSNRs for JPEG 2000. In the proposed method, each quality layer will be used to indicate the image quality in terms of PSNR value given as the target. In general, a quality layer is used to represent quality increment, however, there are no expression on how much increased in PSNR or equivalently MSE (mean square error) values. On the other hand, we use each layer to indicate the exact increment of the image quality in terms of PSNR values. This is done by the quality layer construction method proposed in this paper.

Using the proposed method, we have two advantages against the previous method of [2]. Namely, we can estimate the PSNR between the original reference image and the received one, and also, the estimation could be done without any information of original one, nor neighboring frames. The information required in the proposed method is the number of layers which are received without any loss due to the network packet loss and it could be obtained with the aid from the network protocol such as the RTP (real-time transport protocol) or SCTP (stream control transmission protocol). Besides, the proposed method could estimate the video quality before the received streams would be applied to JPEG 2000 decoder at the receiver. This means that we do not need to pay additional computational cost for quality estimation. Through the computer simulation, we show that the proposed method enables us to accurately estimate the image quality in terms of the PSNR, or MSE.

II. Preparation

Here, we give a brief review of the JPEG 2000 and the problem of image transmission using the Internet.

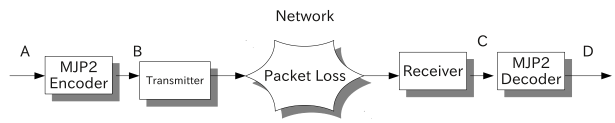

Fig. 1. Possible measurement points for estimating video quality

Proceedings of the Second APSIPA Annual Summit and Conference, pages 294–297,
Biopolis, Singapore, 14-17 December 2010.

Before that, we summarize the possible measurement points of image quality of networked video transmission. In Fig.1, we show a simplified diagram of a network video transmission system. We have four points, namely, A, B, C, and D, at which we have information on images. The images at A are the original ones, and those at D are received and possibly degraded ones after the decoding operation at the receivers. The images at B are encoded ones, and in this paper, each image is encoded by JPEG 2000. Finally, the point C shows that the coded images with degradation due to the loss of some network packets. The aim of NR methods in [4], [5] is to measure the degradation of images between A and D when they are encoded in MPEG or H.264. Besides, the method in [2] enables us to estimate the image quality between B and C for Motion JPEG 2000 systems. The proposed method, on the other hand, can estimate the image quality between A and C by using the information of quality layer construction of JPEG 2000 as described in the following.

A. JPEG 2000 and Motion JPEG 2000

In Fig.2, we show a typical flow of the JPEG 2000 encoding process. The input image is first decomposed into one or several components and each component is (optionally) partitioned into smaller independent rectangular blocks called tiles. Each tile is decomposed into its sub-bands by a discrete wavelet transform (DWT), and the wavelet coefficients of each sub-band are partitioned code-blocks. After being quantized, the wavelet coefficients in each code-block are independently entropy coded. Motion JPEG 2000 is an extension of the JPEG 2000 standard for video format. It is a intra-frame coding and utilizes the same coding engine as the JPEG 2000. Hence, the code-stream of Motion JPEG 2000 is formed as a continuous series of JPEG 2000 coded images.

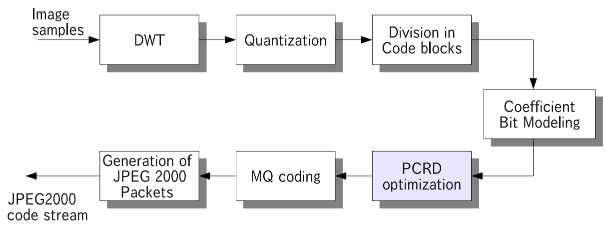

Fig. 2. JPEG2000 encoding process

B. PSNR Calculation

Next, we briefly show the expression of MSE, or PSNR using the wavelet-domain information. In this paper, we concentrate on the expression of them of a luminance component of a color image. Then, in the spatial domain, the MSE of an erroneous image with the size of $X \times Y$ is calculated as

$$\text{MSE}_{\text{spa}} = \frac{1}{XY}\sum_{x=1}^{X}\sum_{y=1}^{Y}\left\{f(x,y) - \hat{f}(x,y)\right\}^2, \tag{1}$$

where

$f(x,y)$	:	Luminance values of original image.
$\hat{f}(x,y)$	:	Luminance values of erroneous image.

Alternatively, we can express the MSE in the wavelet domain of a $U \times V$ transformed matrix as

$$\text{MSE}_{\text{wav}} = \frac{1}{UV}\sum_{r=1}^{N_r}\sum_{b}\left[\lambda_{r,b}^2 \times \sum_{u=1}^{U_{r,b}}\sum_{v=1}^{V_{r,b}}\left\{F_{r,b}(u,v) - \hat{F}_{r,b}(u,v)\right\}^2\right], \tag{2}$$

where

r	:	Decomposition level.
b	:	Subband (one of LL, LH, HL, and HH).
$F_{r,b}(u,v)$	:	Wavelet coefficients of original image.
$\hat{F}_{r,b}(u,v)$	:	Wavelet coefficients of erroneous image.
N_r	:	Total number of r.
$U_{r,b}$	:	Width at r and b.
$V_{r,b}$	:	Height at r and b.
$\lambda_{r,b}$	:	Weighting factor of the wavelet filter.

Note that the values of $\lambda_{r,b}$ are given for JP2 9/7 and 5/3 filters and are shown in [2] for example. The Eq. (2) is the essential equation of the proposed method.

The relation between the PSNR of an 8-bit luminance component and the MSE are related by

$$\text{PSNR} = 10\,\log_{10}(255^2/\text{MSE}), \tag{3}$$

As this equation suggests that we use the terms MSE and PSNR interchangeably in this paper.

C. Network transmission via the Internet

When a code-stream is transmitted through the Internet, it will be divided into a sequence of chunks of data called network packets. Each network packet will be sent using one of the transport layer protocols. RTP (real-time transport protocol) is widely used transport layer protocol, constructed generally on UDP (user datagram protocol) for transmitting media streams, such as videos, audio, and so forth. In this paper, although the proposed method does not depend on the selection of the protocol, we assume that the code-stream will be sent using the RTP for simplicity.

Each RTP packet has a header part which contains some information required to reproduce the transmitted media data. For example, the RTP header contains the 'sequence number' field and by examining the values of this field of the received packets, we can easily know which packet was lost during the transmission. Hence, in this paper, it is assumed that we can know which part of the bit sequence was lost during the transmission based on the those header information.

D. Packet Loss and its effects

When some part of the code-stream is missing during the transmission, an incomplete code-stream will be received at the receiver and the corrupted code-stream cannot be decoded correctly, so that a perceptually degraded image will be reconstructed. Even in this situation, we might decode it by using the JPEG 2000 error resilience tools (ERT), however, most of the data following the lost part will be discarded to resynchronize the decoding process. Therefore, we should accurately estimate the quality of each frame in advance the playback of the sequence to maintain the required quality.

III. Proposed method

Here, we describe the proposed method. The proposed method is based on the information provided by the quality layer construction given target PSNRs.

A. Quality Layer Construction given Target PSNRs

At first, we describe the quality layer construction given target PSNRs for JPEG2000 in which we can specify the target PSNR for each layer. The method is a variable bit rate (VBR) encoding and utilizes the Eqs (2) and (3) to realize to achieves the target PSNRs.

In PCRD-opt algorithm, which is used in EBCOT algorithm, searches the optimum slope of the R-D curve which minimizes the distortion under the given target bit-rate[6]. When we construct the quality layers, we repeat this procedure for each layer. So, we can say that each layer provides the optimum quality at given target bit-rate.

On the other hand, in the propose method, we specify the target PSNR for each layer. Namely, instead of rate optimization, the proposed method uses the PSNR (or MSE) values to determine the truncation points in each layer. In Fig.3, we show the flow char of the proposed method. In the figure, we calculate and examine the PSNR values using Eqs (2) and (3) instead of examining the bit-rate of the generated code-stream.

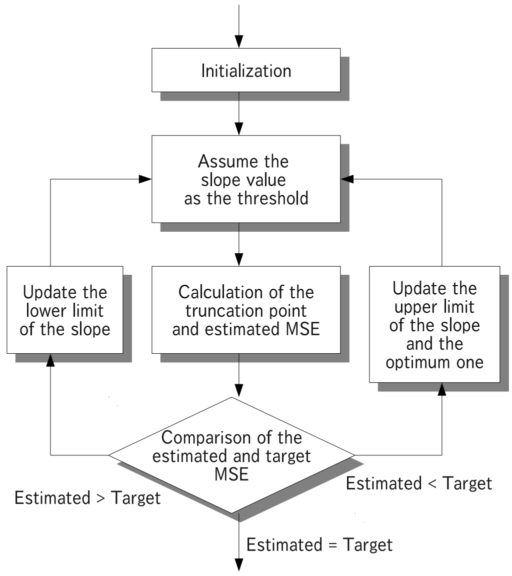

Fig. 3. Chart of the proposed quality layer construction given target MSEs

In Fig.4, we show an example of the relation between the layer construction and the PSNR of a frame. In this example, the target PSNR of i-th layer is shown as T_i [dB] and satisfies $T_i < T_{i+1}$ and there are no other restriction on T_i.

B. Network transmission of the proposed method

In this paper, for simplicity, we assume to segment the code-stream into network packets at the end of each layer. In Fig.5, an example of this packtization is show. Although this figure illustrate one-to-one relation between a network packet and a layer, when we apply the method to actual applications, some of layers should be divided into several packets due to their sizes. On the other hand, in order to decrease the number of network transmission, we could packetize several small layers into a network packet. We should note that the proposed estimation method does not have any assumption on the construction of network packets so that we assume the simplest segmentation here.

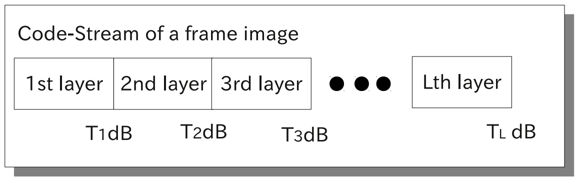

Fig. 4. An example of the relation between the layer construction and the PSNR of a frame. T_i shows the target MSE for the i-th layer.

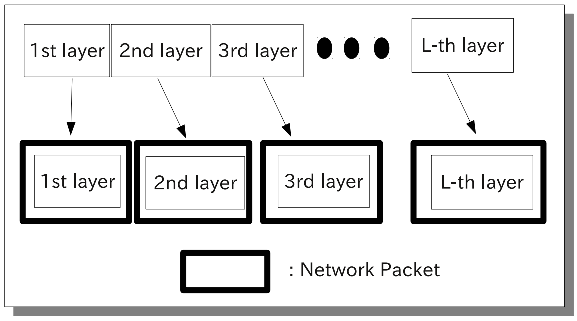

Fig. 5. An example of the relation between the layer construction and network packets

The mapping information of PSNR values to each layer is assumed to be known at both sides of transmission in advance. Hence, the receiver side could estimate the quality of the received images from the received code-stream by examining the number of layers those received without data loss.

C. Receiver side NR image quality estimation

The receiver-side estimates the image quality of each frame using the received information only. The data sequence will be reconstructed with some parts lost due to the packet loss. As we mentioned above, the location of lost parts are assumed to be known from the information of the protocol header and the lost parts are zero-padded. We select to use zero-padding instead of other possible operations such as interpolation from the surrounding pixels because the purpose of our method is to measure the degradation of image quality. From the zero-padded code-stream, we check the number of received layers and then transform it to the PSNR values. We can thus estimate the quality of the received images in terms of the PSNR. It should be emphasized that the proposed method enables us to precisely estimate the PSNR value without decoding the code-stream, and it does not require additional calculation. In Fig.6, we show an example of the proposed estimation from the reconstructed code-stream. In this example, the layers from the first through R-th were received without lost part, and in this case, the estimated PSNR value would be T_R [dB].

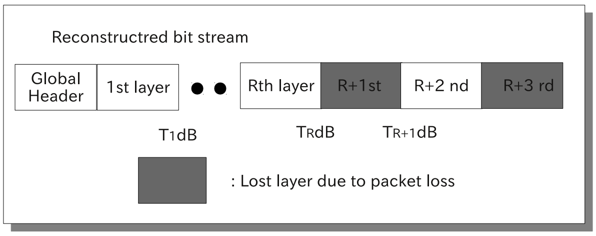

Fig. 6. An example of estimation of PSNR value from the reconstructed code-stream. In this example, T_R[dB] is the estimated value.

IV. Simulation Results

In order to verify the effectiveness of the proposed method, we provide some results of computer simulations in this section.

A. Accuracy of layer construction given target PSNRs

Here, we confirm the accuracy of the layer construction given target PSNR values. In this simulation, we set the target values T_i of i-th layer as $T_i = i + 19$ [dB], i.e., from the first layer, $T_1 = 20$, $T_2 = 21$, $T_3 = 22$ and so on. We apply the method to the image 'Lena' with 8-bit gray scale. The number of wavelet decomposition was 4, and 9/7 filter was used. In Fig.7, we show the results of simulation. From the Figure, we can confirm that the actual PSNRs almost agree with the target values. Please note that we can estimate the PSNR values by counting the number of layers.

B. Image quality estimation of a video sequence

Next, we show the results of simulations of applying the proposed method to image quality estimation of a video sequence. We applied the method to the video sequence 'Elephants Dream (1920x1080)' with 8-bit gray scale. The number of wavelet decomposition was 4, and 9/7 filter was used. In this simulation, fifty frames were selected to be processed. We simulated a transmission through IP based network with packet loss rate at 5% and burst-type loss was assumed. Target values for each layer are same as those of IV-A. In Fig.8, the results are shown. In this figure, we compared the estimated PSNRs, and the actual PSNRs, which are obtained by directly calculated using Eq.(1) and Eq.(3). From the comparison, we can confirm that the proposed method enables us to accurately estimate the actual values at the maximum around ± 1 dB.

TABLE I
Error in PSNR estimation at packet loss rate 5%

	Average	Maximum	Minimum
Error [dB]	0.35	1.09	0

V. Conclusion

In this paper, we proposed a novel NR image quality estimation method based on the quality layer construction given target PSNRs for JPEG2000. The proposed method enables us to accurately estimate the image quality in terms of PSNR, or MSE. We showed that it does not require information of the reference images from the server-side, nor those from neighboring images. Besides, we do not need any additional calculations to implement the proposed method at the receiver side. Through the computer simulations, it was shown that the proposed method can accurately estimate the PSNR of the received image within ± 1 dB error.

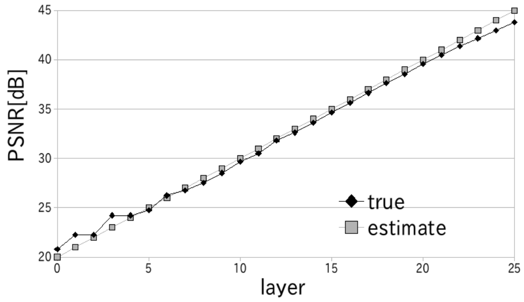

Fig. 7. Relation between the target and the actual PSNRs

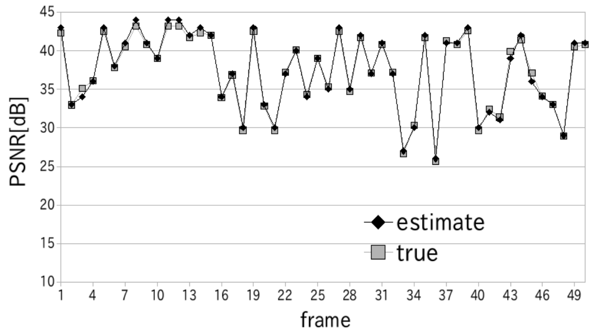

Fig. 8. Relation between the target and the actual PSNRs. We sent 'Elephants Dream' video sequence through a simulated network of packet loss ratio 5%. 'Estimated PSNR' shows the estimated values obtained by the proposed method. The errors between Estimated and Actual were within ± 1 dB.

References

[1] Peter Schelkens, Athanassios Skodras, and Touradj Ebrahimi, Eds., *The JPEG 2000 Suite*, Wiley, 20009.

[2] Kiyoshi Nishikawa, Khairul Munadi, and Hitoshi Kiya, "No-reference PSNR estimation for quality monitoring of Motion JPEG2000 video over lossy packet networks," *IEEE Transactions on Multimedia*, vol. 10, pp. 637–645, June 2008.

[3] Amy R. Reibman, Vinay A. Vaishampayan, and Yegnaswamy Sermadevi, "Quality monitoring of video over a packet network," *IEEE Transactions on Multimedia*, vol. 6, pp. 327–334, Apr. 2004.

[4] Osamu Sugimoto, Sei Naito, Shigeyuki Sakazawa, and Atsushi Koike, "Objective perceptual video quality measurement method based on hybrid no reference framework," in *Proc. of IEEE ICIP 2009*, Nov. 2009, pp. 2237–2240.

[5] Matteo Naccari, Marco Tagliasacchi, Fernando Pereira, and Stefano Tubaro, "No-reference modeling of the channel induced distortion at the decoder for H.264/AVC video coding," in *Proc. of IEEE ICIP 2008*, Oct. 2008, pp. 2324–2327.

[6] David Taubman, "High performance scalable image compression with EBCOT," *IEEE Transactions on Image Processing*, vol. 9, no. 7, pp. 1158–1170, July 2000.