Hindawi Publishing Corporation
EURASIP Journal on Wireless Communications and Networking
Volume 2008, Article ID 902592, 7 pages
doi:10.1155/2008/902592

Research Article

Content-Based Distortion Control Scheme for High-Quality Wireless Multimedia Services

Chang Sun,[1] Hong-Jun Wang,[1,2] Seoksoo Kim,[3] Young-Sil Kim,[4] Seung-youn Lee,[5] and Xin-Bo Yu[1]

[1] *School of Information Science and Engineering, Shandong University, Jinan 250100, China*
[2] *School of Electronic and Information Engineering, Tianjin University, Tianjin 300072, China*
[3] *Department of Multimedia Engineering, Hannam University, Daejeon 306791, South Korea*
[4] *Department of Computer Science and Information, Daelim College, Anyang 431715, South Korea*
[5] *Department of Electrical Information Control, Dong Seoul College, Seongnam 461714, South Korea*

Correspondence should be addressed to Chang Sun, changsun.ee@gmail.com

Received 24 January 2008; Accepted 15 April 2008

Recommended by Jong Hyuk Park

In recent years, the wireless mobile markets are witnessing an unprecedented growth. High-quality video service will be greatly needed as one of the hottest wireless multimedia services in the future generation wireless networks. In this paper, a novel content-based distortion control scheme is proposed to provide higher quality of the wireless video services. Our scheme adopts rate-distortion optimization techniques in state-of-the-art video coding standard H.264/AVC. In order to improve the subjective video quality in the process of encode, we create three visual distortion sensitivity models to minimize the perceptual distortion. We arrange more bits to visual distortion sensitive macroblocks during rate-distortion optimization process. The perceptual distortion in these regions is thus efficiently controlled with a relatively higher rate. Meanwhile, rate balance is achieved by allotting fewer bits to macroblocks that are perceptually less sensitive to distortion. Experiments results show that the subjective qualities of encoded video are improved without compromising PSNR.

1. INTRODUCTION

In recent years, the wireless mobile markets are witnessing an unprecedented growth. As the enormous increase of mobile device users, more wireless information services, and mobile commerce applications are demanded. The so-called future generation wireless networks (FGWNs) involve the concept that the next generation of wireless communications will be a major move toward ubiquitous wireless communications systems and high-quality wireless services [1]. With the promise of greater networking bandwidth and rapid advances in compression technologies, the quality of wireless-based multimedia services providing to mobile device users is likely to be further improved.

Video applications will be extensively used in future generation wireless multimedia services [2], which include applications such as multimedia messaging services (MMSs), packet-switched streaming services (PSSs), and packet-switched conversational services (PCSs), and so forth; video transmission over FGWN is a challenging task requiring high compression efficiency. H.264/AVC is the newest video coding standard that achieves much higher coding efficiency than that of the previous standards. This makes it an ideal video coder for future generation wireless multimedia services. Our work aims at offering high quality of service (QoS) in FGWN by improving subjective coding quality of the video coder without varying bitrate.

Rate-distortion optimization (RDO) is an effective technique used in H.264/AVC coder to achieve the best quality under certain rate constraints [3]. When frame-layer rate control is enabled, Lagrange multiplier λ was decided only by QP to minimize the mean absolute difference (MAD). However, the human visual system (HVS) does not perceive quality in the MAD sense [4]. As the ultimate video quality is judged by the HVS, it is wise to adapt the coding algorithm to the sensitivity of the human eyes.

Although there are extensive sophisticated perceptual coding techniques for audio coding and still image coding [5–7], attempts to incorporate perceptual considerations into video coding are short. In the past decades, several

practical methods in perceptual video coding are employed based on the concept of just noticeable distortion (JND) [8, 9]. The idea is that the HVS can tolerate certain amount of noise depending on its sensitivity to the source and type of noise for a given region in a given frame. It can be further used for bit allocation. In [10], a video bit allocation technique adopting a visual distortion sensitivity model for better rate-visual distortion coding control is proposed. Apart from these, in very recent literatures, some methods are suggested to adjust Lagrange multipliers by considering perceptual characters of the video content [11, 12]. Both of the RDO schemes in [11, 12] can effectively distinguish texture regions, edged regions, and flat regions of the encoded frame, and then arrange different λ to these regions according to their contents. Nevertheless, they just consider one perceptual feature to modulate λ. For achieving the better RDO performance, a more complete set of the HVS features should be taken into account.

In this paper, we incorporate three important visual features into video coding and further establish three visual distortion sensitivity models. These models are then used for minimizing the perceptual distortion and helping to adjust λ adaptively in accordance with the video contents in the RDO process. Based on these models, λ for every MB of the encoded frame is further refined. The MBs with more preattentive features will get more distortion reduction by assigning a smaller λ. Rate balance is achieved by assigning a relatively larger λ to the MBs that are visually less sensitive to distortion so that more distortion is allowed without noticeable visual degradation in the decoded images.

The rest of the paper is structured as follows. Section 2 gives an overview of video services in FGWN. Section 3 presents the visual distortion sensitive models based on the HVS. The rate-distortion optimization scheme in H.264/AVC is briefly introduced in Section 4. Then in Section 5, we discuss the proposed content-based distortion control scheme. Experimental results are summarized in Section 6 followed by conclusions in Section 7.

2. VIDEO SERVICES IN FGWN

2.1. H.264/AVC—an ideal video coder for wireless multimedia services

Most current and future cellular networks, like GSM-GPRS, UMTS, or CDMA, contain a variety of packet-oriented transmission modes allowing transport of practically any type of IP-based traffic to and from mobile terminals, thus providing users with a simple and flexible transport interface [13]. The third generation partnership project (3GPP) has selected several multimedia codecs for the inclusion into its multimedia specifications [14]. To provide basic video service in the first release of the third generation (3G) wireless systems, the well-established and almost identical baseline H.263 and the MPEG-4 visual simple profile have been integrated. However, due to the transmitted data volume, the limited resources bandwidth, and the transmission power in wireless networks, compression efficiency is the main target for wireless video and multimedia applications in FGWN.

H.264/AVC is state-of-the-art coding standard created by joint video team (JVT). It adopts several technical strategies that can achieve much higher coding efficiency than that of the previous standards such as MPEG-2, H.263, H.263+, H.263++, and the MPEG-4. The main strategies used in H.264/AVC are multiframe motion-compensated prediction, adaptive block size for motion compensation, generalized B-pictures concepts, quartel-pel motion accuracy, intracoding utilizing prediction in the spatial domain, in-loop deblocking filters, and efficient entropy-coding methods [15]. All these advanced techniques significantly increase coding efficiency of H.264/AVC, which makes it the most ideal candidate for all wireless multimedia services in FGWN including MMS, PSS, and PCS. A simplified wireless video services system is illustrated in Figure 1.

2.2. Transmission of H.264/AVC video over FGWN

Video transmission for wireless terminals is likely to be a major application in FGWN systems and may be a key factor in their success. A sample protocol we consider for the transport of video over FGWN is shown in Figure 2. H.264/AVC video distinguishes between two different conceptual layers, the video coding layer (VCL) and the network abstraction layer (NAL) [15]. The VCL specifies an efficient representation for the coded video signal. The NAL of H.264/AVC defines the interface between the video codec itself and the outside world. It operates on NAL units which give support for the packet-based approach of most existing networks. As recommended in internet engineering task force (IETF), compressed video can be transported, multiplexed, and synchronized by using real-time transport protocol (RTP)/UDP/IP protocol stack. The radio link control (RLC) at data link layer provides a "best effort" level of reliability for data delivery, with a maximal number of retransmissions. The physical layer offers information transfer services to upper layers. One of the main services provided by the physical layer is the measurement of various quantities, such as physical-channel bit-error rate (BER), transport-channel block-error rate (BLER), transport-channel bitrate, bitrate of each transport channel, and so forth. All these measured information are reported to the higher layer for system performance analysis.

3. VISUAL DISTORTION SENSITIVITY MODELS

As the HVS makes final evaluations on the quality of the video, we seek to minimize the perceptual distortion based on the encoded video content rather than traditional MAD distortion in this paper. For achieving this goal, visual distortion sensitivity models should be established at first. Three models are presented in the following subsections.

3.1. Motion attention model

The HVS is more attentive and sensitive to the active regions within a video sequence. When there is a movement, the HVS tends to concentrate on the moving objects and ignore the static objects [16]. If the object is moving continuously, the

HVS can easily notice the moving object and predicts the position of the object in the future. Therefore, if the same MAD distortion occurs at an active region and at a static region, the visual distortion tends to be larger in the active region.

In order to determine the active regions, an error image $E(x, y, n)$ at location (x, y) in the nth MB in the encoded frame is generated as

$$E(x,y,n) = \begin{cases} 1, & \text{if } \left| I(x,y,n) - I'(x,y,n) \right| > T, \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where I and I' represent the pixel intensity of the current and the previous pictures, respectively, and T is a predetermined intensity threshold ($T = 5$ in our scheme). The motion attention factor of the nth MB in the encoded frame $F_{\text{Motion}}(n)$ is defined as

$$F_{\text{Motion}}(n) = \frac{1}{MS^2} \sum_{x=0}^{MS-1} \sum_{y=0}^{MS-1} E(x,y,n), \tag{2}$$

where MS is the MB size ($MS = 16$ in our scheme). We can find that a larger F_{Motion} corresponds to the active MB, whereas a smaller F_{Motion} indicates the inactive MB.

3.2. Position model

The HVS is sensitive to the central regions of the image. Human will be more concentrated on the central regions than the regions near the boundary of the image. It is due to intensity and light-sensitive cone cell distribution on the retina of the eyes. The intensity distribution is Gaussian at the center of the fovea on the retina [17]. At the center, more light intensity is received and more region information is obtained. The image projected by the central regions of the image is near the center of the retina. For the regions faraway from the center, the light intensity is lower compared with the central part. Therefore, the sensitivity of the central region is higher. Similar to the motion attention case, when the same MAD distortion occurs at the central and the boundary regions, the visual distortion is larger in the central regions.

In this paper, we obtain the position information by employing a 2D Gaussian function. MB is the basic unit in this subsection. The position factor of the nth MB in the encoded frame $F_{\text{Position}}(n)$ is calculated as

$$F_{\text{Position}}(n) = \exp\left(- \frac{(x_n - x_M)^2 + (y_n - y_M)^2}{2 \times \sigma^2} \right), \tag{3}$$

where (x_n, y_n) are the coordinates of the nth MB in the encoded frame; (x_M, y_M) are the midpoint coordinates of the image, that is, $x_M = IW/2$, $y_M = IH/2$, where IW and IH represent the width and height of the image, respectively. σ is the Gaussian standard deviation, whose value decides the descending speed of the position factor from the image center to the neighboring areas. We obtain σ by

$$\sigma = \min\left[\frac{IW}{2 \times MS}, \frac{IH}{2 \times MS} \right], \tag{4}$$

where MS is the MB size ($MS = 16$ in our scheme). We can find that a larger F_{Position} corresponds to the central MB, whereas a smaller F_{Position} indicates the boundary MB.

3.3. Texture structure model

In addition to the active regions and central regions, the HVS is also sensitive to the regions with edges and curves and the flat regions [12]. Conversely, the HVS can tolerate large distortion in random texture regions [12]. Therefore, if the same MAD distortion occurs at an edged (or flat) region and at a random texture region, the visual distortion tends to be larger in the edged (or flat) region.

In this paper, we adopt the texture structure model based on the gradient-based methods developed in [18]. In the gradient-based approach, the pixel gradient vectors in an image are described by the gradient vector of the nth MB in the encoded frame $[G_x(x, y, n), G_y(x, y, n)]^T$, which can be simplified by using Sobel operators [19] as (5) and (6). These operators take the first derivative of the input image, and they have the advantages of enabling both a differencing and a smoothing effect

$$\begin{aligned} G_x(x,y,n) = {} & I(x-1,y+1,n) + 2 \times I(x,y+1,n) \\ & + I(x+1,y+1,n) - I(x-1,y-1,n) \\ & - 2 \times I(x,y-1,n) - I(x+1,y-1,n), \end{aligned} \tag{5}$$

$$\begin{aligned} G_y(x,y,n) = {} & I(x+1,y-1,n) + 2 \times I(x+1,y,n) \\ & + I(x+1,y+1,n) - I(x-1,y-1,n) \\ & - 2 \times I(x-1,y,n) - I(x-1,y+1,n), \end{aligned} \tag{6}$$

where $I(x, y, n)$ represents pixel intensity at location (x, y) in the nth MB in the current encoded frame. The complex number representation of the gradient vectors is squared before averaging. The complex representation of the squared gradient vectors $[G_{sx}(x, y, n), G_{sy}(x, y, n)]^T$ is

$$\begin{aligned} G_{sx} + jG_{sy} &= (G_x + j \cdot G_y)^2 \\ &= (G_x{}^2 - G_y{}^2) + 2j \cdot G_x G_y. \end{aligned} \tag{7}$$

The average squared gradient can be calculated by averaging in local neighborhood with a window size of W,

$$\begin{aligned} G_{xx} &= \sum_W G_x{}^2, \\ G_{yy} &= \sum_W G_y{}^2, \\ G_{xy} &= \sum_W G_x \cdot G_y. \end{aligned} \tag{8}$$

In our scheme W is set to 16. The coherence factor of the nth MB in the encoded frame $F_{\text{Coh}}(n)$ using the squared gradient method can be calculated as

$$F_{\text{Coh}}(n) = \frac{\sqrt{(G_{xx} - G_{yy})^2 + 4G_{xy}{}^2}}{G_{xx} + G_{yy}}. \tag{9}$$

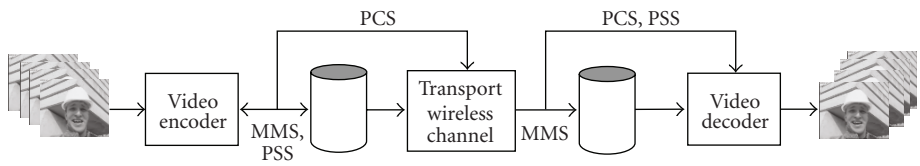

Figure 1: A simplified wireless video services system in FGWN.

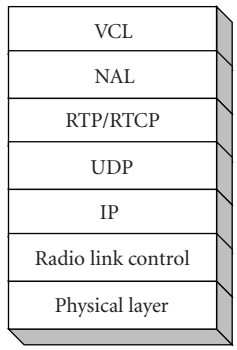

Figure 2: A sample protocol stack for video transmission over FGWN.

The value of F_{Coh} can provide important information in classifying image into texture, edged, and flat regions. A F_{Coh} value of 1 refers to all squared gradient vectors are in the same direction, which implies that the neighborhood edges are consistently pointing in the same direction and this is related to MB with strong edges. On the other hand, a F_{Coh} value of 0 indicates that the squared gradient vectors are equally distributed in all directions, which implies that the neighborhood edges are scattered in all directions and there are no edges, that is, corresponding to a flat MB. A coherence factor value near 0.5 implies the random texture MB. So F_{Coh} is a nonlinear function of the texture structure characters of the MB.

4. RATE-DISTORTION OPTIMIZATION SCHEME IN H.264/AVC

The purpose of this section is to briefly review rate-distortion optimization scheme in H.264/AVC, so that this paper is self-contained and to provide theoretical foundation of our scheme described in the next section.

H.264/AVC coder adopts optimization using Lagrange multipliers to determine a set of coding parameters in the rate-distortion sense under some coding constraints [3]. For one image region, the coder decides an optimal set of coding parameters to minimize the distortion D subject to a rate constraint R. The Rate-distortion model is based on the following Lagrangian formulation [3]:

$$\min\{J\}, \quad \text{where } J = D + \lambda R. \tag{10}$$

The Lagrange cost J is minimized for a given Lagrange multiplier $\lambda\,(\lambda \geq 0)$. The multiplier controls the tradeoff between distortion and rate for different coding modes and motion vectors.

In the motion estimation stage, RDO minimizes

$$J_{\text{Motion}} = D_{\text{DFD}} + \lambda_{\text{Motion}} R_{\text{Motion}}, \tag{11}$$

where D_{DFD} is the motion compensation prediction error, λ_{Motion} is the Lagrange multiplier, and R_{Motion} is the number of bit for coding the motion vectors.

In the mode decision stage, RDO minimizes

$$J_{\text{Mode}} = D_{\text{REC}} + \lambda_{\text{Mode}} R_{\text{REC}}, \tag{12}$$

where D_{REC} is the distortion between the reconstructed and the original MBs, λ_{Mode} is the Lagrange multiplier, and R_{REC} is the total bits for mode information, motion vectors and the transform coefficients. λ_{Mode} is empirically determined by

$$\lambda_{\text{Mode}} = 0.85 \cdot 2^{(QP-12)/3}. \tag{13}$$

And λ_{Motion} is calculated by

$$\lambda_{\text{Motion}} = \sqrt{\lambda_{\text{Mode}}}. \tag{14}$$

For an MB in the encoded frame, assuming that RD cost $J_1 = D_1 + \lambda_{\text{Mode}} R_1$ and $J_2 = D_2 + \lambda_{\text{Mode}} R_2$, respectively, where $D_1 < D_2$, $R_1 > R_2$, and $J_1 < J_2$ we can easily obtain

$$\lambda_{\text{Mode}} < \frac{D_2 - D_1}{R_1 - R_2}. \tag{15}$$

If we increase λ_{Mode}, then (6) may no longer be satisfied so that the better mode changes from mode 1 to mode 2 (i.e., $J_2 < J_1$). Therefore, a larger λ_{Mode} corresponds to higher D and lower R; the converse is true for lower λ_{Mode} either. This implies that changes in the λ_{Mode} influence the D and the R of the resulting encoded video, which is the theoretical basis for the following proposed scheme.

5. PROPOSED CONTENT-BASED DISTORTION CONTROL SCHEME

In the current H.264/AVC standard, RDO is carried out on frame level without consideration of the video contents. However, in this paper, a content-based distortion control scheme is proposed that can adjust Lagrange multiplier λ adaptively on MB level based on the perceptual models created in Section 3. According to the analysis in Section 4, larger λ should be assigned to MBs that are visually less sensitive to distortion so that the lower R can be achieved with the relative higher D. At the same time, smaller λ should be assigned to the perceptual distortion sensitive MBs to get the lower D with the increased R. Therefore, we arrange

smaller λ to the active, central, edged, and flat MBs to control distortion in these regions, whereas rate balance is realized by arranging larger λ to the inactive, boundary, and random texture MBs to permit distortion in these perceptually less sensitive regions. The proposed scheme is carried out as follows.

At first, we adjust λ in terms of the three visual distortion sensitivity models. In the motion attention model, from $F_{\text{Motion}} = 0$ to $F_{\text{Motion}} = 1$, it suggests a gradual change of perceptual characteristics from an inactive MB to an active MB and increasingly less distortion could be tolerated. We define the adaptive Lagrange multiplier of the motion attention model in the nth MB of the encoded frame as $\lambda_{\text{Motion}}(n)$, which is a mapping function that gives linearly decreasing Lagrange multiplier to provide increasingly lower D with the increase of F_{Motion}:

$$\lambda_{\text{Motion}}(n) = (\alpha_1 \cdot F_{\text{Motion}}(n) + \beta_1)\lambda. \tag{16}$$

Here, we assign the lower bound of $\lambda_{\text{Motion}}(n)$ to be 0.5λ and the higher bound 1.2λ, where λ is the original Lagrange multiplier in H.264/AVC standard, which is calculated by (2) or (3). Therefore, α_1 and β_1 should be -0.7 and 1.2, respectively.

As to the position model, from $F_{\text{Position}} = 0$ to $F_{\text{Position}} = 1$, it indicates a gradual change of perceptual characteristics from a boundary MB to a central MB and increasingly less distortion could be tolerated. We define the adaptive Lagrange multiplier of the position model in the nth MB of the encoded frame as $\lambda_{\text{Position}}(n)$, which is an exponential mapping function that gives exponentially decreasing Lagrange multiplier to provide increasingly lower D with the increase of F_{Position}:

$$\lambda_{\text{Position}}(n) = (\alpha_2 \cdot e^{\beta_2 \cdot F_{\text{Position}}(n)})\lambda. \tag{17}$$

We assign the lower bound of $\lambda_{\text{Position}}(n)$ to be 0.5λ and the higher bound 1.2λ, and α_2 and β_2 are decided as 1.2 and -0.875, respectively.

In the texture structure model, from $F_{\text{Coh}} = 0$ to $F_{\text{Coh}} = 0.5$, it represents a gradual change of perceptual characteristics from a flat MB to a random texture MB and increasingly more distortion could be tolerated. Besides, from $F_{\text{Coh}} = 0.5$ to $F_{\text{Coh}} = 1$, the perceptual characteristics gradually change from a random texture MB to an edged MB and less distortion could be tolerated. The adaptive Lagrange multiplier of the texture structure model in the nth MB of the encoded frame $\lambda_{\text{Coh}}(n)$ is calculated as

$$\lambda_{\text{Coh}}(n) = \begin{cases} (\alpha_3 \cdot F_{\text{Coh}}(n) + \beta_3)\lambda, & \text{if } F_{\text{Coh}} < 0.5, \\ (\alpha_4 \cdot F_{\text{Coh}}(n) + \beta_4)\lambda, & \text{otherwise.} \end{cases} \tag{18}$$

From $F_{\text{Coh}} = 0$ to $F_{\text{Coh}} = 0.5$, it is a linear mapping function giving increasingly higher D with the increase of F_{Coh}; whereas from $F_{\text{Coh}} = 0.5$ to $F_{\text{Coh}} = 1$, it offers increasingly lower D with the increase of F_{Coh}. The lower and higher bound of $\lambda_{\text{Coh}}(n)$ are also set to be 0.5λ and 1.2λ, respectively. So α_3, β_3, α_4, and β_4 are determined as 1.4, 0.5, -1.4, and 1.9, respectively.

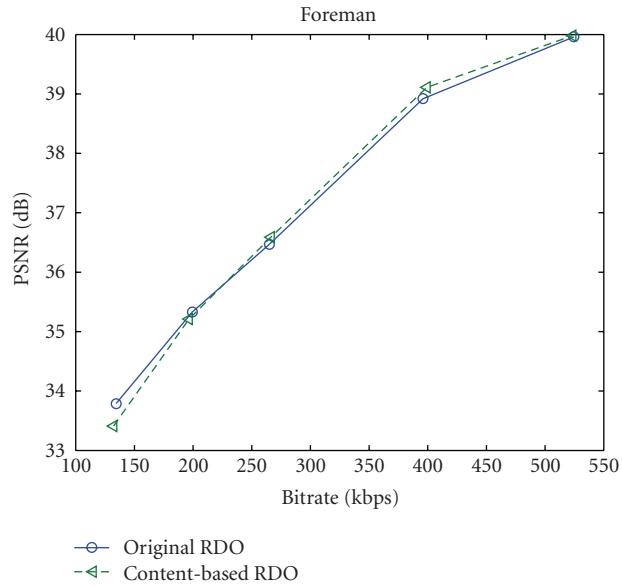

Figure 3: Comparison of PSNR results between the reference software and the proposed scheme for "Foreman."

Then the final adaptive Lagrange multiplier of the nth MB of the encoded frame $\lambda_{\text{MB}}(n)$ is formulated as

$$\lambda_{\text{MB}}(n) = w_1 \cdot \lambda_{\text{Motion}}(n) + w_2 \cdot \lambda_{\text{Position}}(n) + w_3 \cdot \lambda_{\text{Coh}}(n), \tag{19}$$

where w_1, w_2, and w_3 are the weights for $\lambda_{\text{Motion}}(n)$, $\lambda_{\text{Position}}(n)$, and $\lambda_{\text{Coh}}(n)$, and satisfy $w_1 + w_2 + w_3 = 1$; $0 \le w_1, w_2, w_3 \le 1$. w_1, w_2, and w_3 empirically decided as 0.3, 0.3, and 0.4, respectively.

Finally, we use (20) instead of (10) to minimize the MB-level Lagrange cost J_{MB} in the encoded frame:

$$\min\{J_{\text{MB}}\}, \quad \text{where } J_{\text{MB}} = D + \lambda_{\text{MB}} R. \tag{20}$$

6. EXPERIMENTAL RESULTS

We have tested the proposed algorithm on the H.264/AVC reference software JM 12.2 [20]. The encoded sequences are CIF versions of "Foreman," "Bus," "Mobile," and "Football" at 30 fps. For all the tests, context adaptive binary arithmetic coding (CABAC) and RDO are enabled. The in-loop filter is also enabled. The first frame is I frame and the others are P frames. All other parameters such as Hadamard transform and the numbers of reference frames are carefully selected to be equivalent. Because of the high throughput ability of FGWN and high fidelity video services required by prospective customers, we set comparatively high bitrates in our experiments, which are between 120 kbps and 550 kbps.

The PSNR results are summarized in Figures 3, 4, 5, and 6, where illustrate the PSNR comparisons of the results from the reference software and the proposed algorithm for the encoded sequences. Figures 7 and 8 show the subjective performance of the proposed scheme for "Foreman" and "Football" sequences, respectively. They are also compared with the results of the reference software.

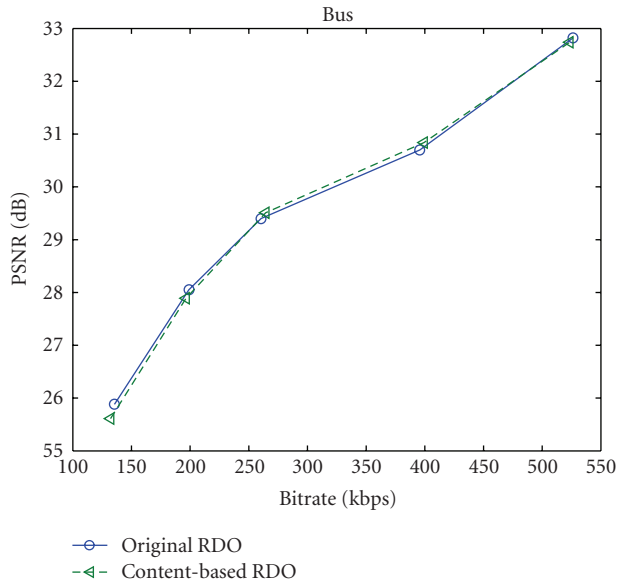

Figure 4: Comparison of PSNR results between the reference software and the proposed scheme for "Bus."

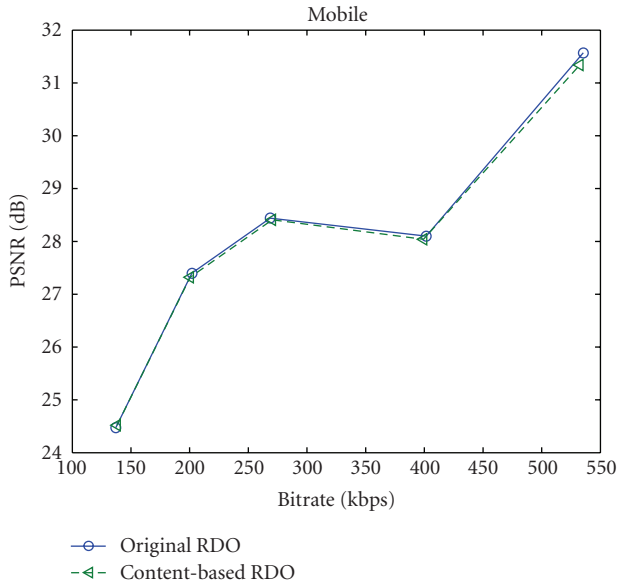

Figure 5: Comparison of PSNR results between the reference software and the proposed scheme for "Mobile."

For all the sequences tested, the average PSNRs are almost the same. At some rate points, the proposed algorithm even outperforms the reference method. This is because although the proposed scheme allows more local PSNR losses in the inactive regions, boundary regions, and random texture regions; the local loss is compensated by gains in PSNR in the active regions, central regions, edged regions, and flat regions, in which the restraint of distortion leads to a gain in perceptual quality. On the other hand, the increase in distortion in inactive regions and random texture region is much less visually noticeable. For instance, for

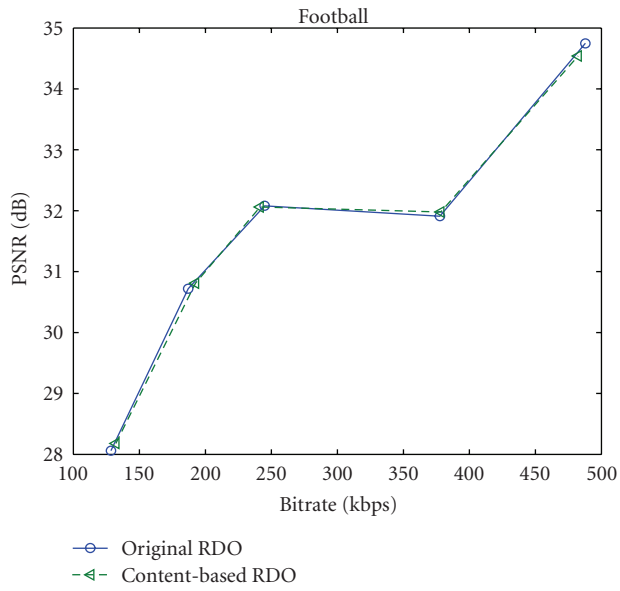

Figure 6: Comparison of PSNR results between the reference software and the proposed scheme for "Football."

(a) Reference software (b) Proposed scheme

Figure 7: Decoded frames of "Foreman."

"Foreman" sequence, there is a slight drop in PSNR at some rate points. This is because we deliberately emphasize distortion reduction in the active, central, edge, and flat regions such as the facial region, while ignore distortions in the nonactive background regions. The PSNR drop at some bitrates is mainly due to the drop in the background, and thus it is not easily noticeable. However, we can easily find that the perceptual quality of the detailed region (such as the wrinkles on facial area) is significantly improved by comparing the two encoded frames in Figure 7. We can also observe the similar improvement in the detailed region of "Football" sequence in Figure 8, where the moving football player in the central area of the image is more visually legible in the decoded frame of our proposed scheme. Although not be demonstrated, there are also subjective quality improvements achieved in the tested "Bus" and "Mobile" sequences.

In addition, we have performed subjective evaluation by inviting 30 people from different background to be human observers for our experiment. The test sequences are in CIF resolution. And the viewing distance is set at three picture heights (3H). At the end of the evaluation, 26 of the observers

(a) Reference software (b) Proposed scheme

Figure 8: Decoded frames of "Football."

said that there are obvious improvements in video quality. Only 4 of them said that the improvements are negligible or not noticeable.

7. CONCLUSIONS

High-quality wireless multimedia services will be greatly demanded by considerable mobile device users in FGWN. In this paper, a content-based distortion control scheme is proposed using rate-distortion optimization technologies and HVS characters to further enhance the quality of wireless video services. The new scheme builds up three visual distortion sensitivity models based on video content and adaptively adjusts the value of the Lagrange multiplier at the MB level according to these visual models. Further perceptual tuning of the Lagrange multiplier could effectively restrain distortion in the visually sensitive regions. Experimental results show that the subjective quality in most perceptually prominent regions is improved with no loss in PSNR.

REFERENCES

[1] R. Berezdivin, R. Breinig, and R. Topp, "Next-generation wireless communications concepts and technologies," *IEEE Communications Magazine*, vol. 40, no. 3, pp. 108–116, 2002.

[2] H.-B. Lim, "Beyond 3G: issues and challenges," *IEEE Potentials*, vol. 21, no. 4, pp. 18–23, 2002.

[3] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 688–703, 2003.

[4] S. Winkler, "Perceptual distortion metric for digital color video," in *Human Vision and Electronic Imaging IV*, vol. 3644 of *Proceedings of SPIE*Proceedings of SPIE, pp. 175–184, San Jose, Calif, USA, January 1999.

[5] L. Hanzo, F. C. A. Somerville, and J. P. Woodward, *Voice Compression and Communications: Principles and Applications for Fixed and Wireless Channels*, Wiley-IEEE Press, New York, NY, USA, 2001.

[6] A. B. Watson, "Visual optimization of DCT quantization matrices for individual images," in *Proceedings of the AIAA Computing in Aerospace 9*, pp. 286–291, American Institute of Aeronautics and Astronautics, San Diego, Calif, USA, October 1993.

[7] D. Taubman and M. W. Marcellin, *JPEG 2000: Image Compression Fundamentals, Standards and Practice*, Kluwer Academic Publishers, Boston, Mass, USA, 2002.

[8] C.-H. Chou and C.-W. Chen, "A perceptually optimized 3-D subband codec for video communication over wireless channels," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 2, pp. 143–156, 1996.

[9] Y.-J. Chin and T. Berger, "A software-only videocodec using pixelwise conditional differential replenishment and perceptual enhancements," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 3, pp. 438–450, 1999.

[10] C.-W. Tang, C.-H. Chen, Y.-H. Yu, and C.-J. Tsai, "Visual sensitivity guided bit allocation for video coding," *IEEE Transactions on Multimedia*, vol. 8, no. 1, pp. 11–18, 2006.

[11] C.-J. Tsai, C.-W. Tang, C.-H. Chen, and Y.-H. Yu, "Adaptive rate-distortion optimization using perceptual hints," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '04)*, vol. 1, pp. 667–670, Taipei, Taiwan, June 2004.

[12] Y. Sun, F. Pan, and A. A. Kassim, "Perceptually adaptive rate-distortion optimization for variable block size motion alignment in 3D wavelet coding," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, vol. 2, pp. 929–932, Philadelphia, Pa, USA, March 2005.

[13] T. Stockhammer, M. M. Hannuksela, and T. Wiegand, "H.264/AVC in wireless environments," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 657–673, 2003.

[14] "Multimedia Messaging Service (MMS); Media Formats and Codecs," 3GPP technical specification 3GPP TR 26.140.

[15] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T, ITU-T Recommendation H.264, "Advanced video coding for generic audiovisual services," March 2005.

[16] E. Kowler, *Eye Movements and Their Role in Visual and Cognitive Processes*, Elsevier Science, New York, NY, USA, 1990.

[17] C. W. Oyster, *The Human Eye: Structure and Function*, Sinauer Associates, Sunderland, Mass, USA, 1999.

[18] A. M. Bazen and S. H. Gerez, "Systematic methods for the computation of the directional fields and singular points of fingerprints," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 905–919, 2002.

[19] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley, Reading, Mass, USA, 1992.

[20] *JM Reference Software version 12.2.*, http://iphome.hhi.de/suehring/tml/download/.