$PSNR_{r,f}$-MOS_r: An Easy-To-Compute Multiuser Perceptual Video Quality Measure

Jing Hu, Sayantan Choudhury, and Jerry D. Gibson

Abstract—**In this paper, we propose a new statistical objective perceptual video quality measure $PSNR_{r,f}$-MOS_r. $PSNR_{r,f}$ is defined as the PSNR achieved by $f\%$ of the frames in each one of the $r\%$ of the transmissions over a network. This quantity has the potential to capture the performance loss due to damaged frames in a particular video sequence ($f\%$), as well as to indicate the probablity of a user experiencing a specified quality over the channel ($r\%$). The percentage of transmissions also has the interpretation as what percentage out of many video users who access the same channel, would experience a given video quality. A subjective experiment is conducted to establish a linear equation connecting $PSNR_{r,f=90\%}$ and MOS_r, the mean opinion score (MOS) achieved by $r\%$ of the transmissions. It is shown from this subjective experiment that $PSNR_{f=90\%}$ correlates much better with the delivered perceptual video quality than the average PSNR across all frames of a video, and is a good representation of perceptual quality of a video transmitted over networks with possible transmission errors.**

I. INTRODUCTION

Over the past two decades, digital video compression and communication have fundamentally changed the way we create, communicate and consume visual information. As a vital part of the fast advancing video technologies, perceptual quality measurement of video sequences has also attracted a significant amount of interest.

The perceptual video quality measures are usually divided into two categories: subjective measures and objective measures. Subjective video quality measures provide an ultimate measure of the viewers' satisfaction with a delivered video. They involve a large number of experiments on human subjects and therefore are expensive, time-consuming and can not be conducted in real time. The conventional and most commonly used objective video quality measure is the mean squared error (MSE) or equivalently the peak signal to noise ratio (PSNR) of the distorted videos. Although very easy to compute, MSE-PSNR is often criticized for correlating poorly to perceptual video quality. The recently proposed sophisticated perceptual video quality measures, such as those included in International Telecommunication Union (ITU) recommendations ITU-R

Jing Hu and Sayantan Choudhury were with the Department of Electrical and Computer Engineering, University of California, Santa Barbara. Jing Hu is now with the Digital Signal Processing Group, Cisco Systems and Sayantan Choudhury is now with the Consumer Systems and Technology, Sharp Laboratories of America. Jerry D. Gibson is with the Department of Electrical and Computer Engineering, University of California, Santa Barbara (emails: jinghu@cisco.com; {sayantan,gibson}@ece.ucsb.edu).

BT.1683 [1] and ITU-T J.144 [2], are based on comprehensive studies of the human vision system (HVS). They examine the perceptual impacts of the compression artifacts and are shown to perform better than the average PSNR across the video frames for a compressed video [3]. However, these sophisticated objective measures are computationally very intensive [4], [5] and normally can only be implemented by using proprietary software. These two features have largely limited their usage in both research and implementation.

These sophisticated objective measures are challenged further when we investigate the end-to-end quality of videos transmitted over ubiquitous networks. First, for video delivered over networks that might incur packet errors and packet losses, the quality degradation due to compression can be overwhelmed by the quality degradation caused by the possible transmission errors, even with error concealment at the video decoder. In this case, the subtle improvement in accuracy made by using the sophisticated objective measures becomes less significant compared to the large variance of the delivered video quality. Second, due to the design in network protocols and the nature of some transmission channels such as multipath fading in wireless local area networks (WLANs), video distortion caused by transmission errors is not deterministic. As a result, a video user's experience with a network of fixed topology and configuration is probabilistic; multiple video users accessing the same network at the same time could experience totally different video quality.

In this paper, we propose a new statistical objective video quality measure that (a) is as easy to compute as average PSNR; (b) is a good representation of perceptual quality of a video transmitted over networks with possible transmission errors; and (c) addresses the randomness in video quality delivered over a network. We call this multiuser perceptual video quality measure $PSNR_{r,f}$-MOS_r. $PSNR_{r,f}$ is defined as the PSNR achieved by $f\%$ of the frames in each one of the $r\%$ of the transmissions over a network. This quantity has the potential to capture the performance loss due to damaged frames in a particular video sequence ($f\%$), as well as to indicate the probablity of a user experiencing a specified quality over the channel ($r\%$). The percentage of transmissions also has the interpretation as what percentage out of many video users who access the same channel, would experience a given video quality. We further investigate the correspondence between $PSNR_f$ and perceptual video quality through a subjective experiment which results in a linear equation connecting $PSNR_{r,f=90\%}$ and MOS_r, the mean opinion score (MOS) achieved by $r\%$ of the transmissions. It is shown from

this subjective experiment that $PSNR_{f=90\%}$ correlates much better with the delivered perceptual video quality than the average PSNR across all frames of a video, while with no extra computatation. The MOS calculated from $PSNR_f$ is shown to be sufficient to indicate the perceptual quality of a delivered video sequence, without the huge computation required by the more sophisticated video quality measurements.

$PSNR_{r,f}$-MOS_r is motivated by a simulation of AVC/H.264 [6] coded videos over IEEE 802.11a WLANs [7] with multipath fading. In this simulation, it is observed that even when the average PSNR over all transmitted frames of a video with packet losses is reasonably high, PSNRs vary significantly across the video frames. Furthermore, the video quality varies dramatically across the different transmissions over the channel. This new mutliuser perceptual video quality measure, however, can be utilized in other video communication applications that are different from the simulation scenario.

II. MOTIVATION: AN AVC/H.264 VIDEO OVER 802.11A WLAN SIMULATION

To demonstrate the huge variances of the video quality across the different frames of a delivered video and across different transmissions of the same network, we simulate the transmission of AVC/H.264 coded videos over IEEE 802.11a WLANs. The details of the simulation setup are described below.

A. Simulation Setup

Video Codec: We choose the Baseline Profile of AVC/H.264 in its reference software [8] version JM10.1 with low delay and low computational complexity. Ninety frames each from a group of videos, representing different types of video content, are coded using combinations of group of picture sizes (GOPS) (10, 15, 30, 45 frames), quantization parameters (QP) (26 for fine quantization and 30 for coarse quantization) and payload sizes (PS) (small-100 bytes and large-1100 bytes). The remainder of the encoder parameters are optimally chosen in the encoder to yield the minimum source bit rate. We do not employ rate control schemes to dynamically choose QPs to compress the video sequences at a constant bit rate. Instead we use a constant QP throughout a video sequence. However, if a rate control scheme is to be adapted, the variance of the video quality will be even larger than the constant QP case to maintain a relatively stable bit rate.

IEEE 802.11a WLAN: We consider one-hop WLANs, in which case we limit our attention to the PHY, MAC and APP layers. In the medium access control (MAC) layer of IEEE 802.11, a cyclic redundancy check (CRC) is computed over the entire packet, and if a single bit error is detected, the packet is discarded. For data, a retransmission would be requested, however, for our particular video applications we do not request a retransmission, but rely on packet loss concealment. Each realization of the multipath delay profile corresponds to a certain loss pattern for that fading realization. Two hundred and fifty packet loss realizations are generated

for each combination of the chosen PHY data rate 6 Mbps, different average channel SNRs (3.5 dB for bad channel, 5 dB for average channel, 7 dB for good channel), and two video payload sizes (small–100 bytes and large–1100 bytes).

Multipath Fading Channel: The Nafteli Chayat model [9], an important indoor wireless channel model with an exponentially decaying Rayleigh faded path delay profile, is employed. The rms delay spread used is 50 nanoseconds, which is typical for home and office environments. In order to estimate the packet error rate under different channel conditions, we modify a readily available OFDM simulator for the IEEE 802.11a PHY [10]. Non-fading channels are also considered for comparison. Noise is modeled as additive white Gaussian noise (AWGN) for both the fading and non-fading cases. The decoding at the receiver is based on soft decision Viterbi decoding. We assume perfect synchronization and channel estimation.

Packet Corruption and Loss Concealment: Each compressed bit stream is corrupted based on the packet loss patterns generated by the multipath fading channel and then reconstructed in the AVC/H.264 decoder with its nominal packet loss concealment (PLC) scheme. Different PLC schemes will have an impact on the concealed video quality and there exists an exhaustive literature proposing different error concealment techniques. However, only a few simple schemes are commonly used in practical applications [11]. As a baseline, we apply the basic PLC method integrated in AVC/H.264 reference software [8]. This PLC method recovers the missing MBs in an I frame through spatial interpolation and the missing MBs in a P frame by searching and copying the most likely MBs in the correctly received reference frames. The previous frame is copied when the whole frame is lost. This method is shown to be effective in both PSNR and perceptual quality [12].

B. Simulation Results

We obtain a PSNR for each frame and each packet loss pattern, for a combination of the codec and channel parameters. Only the PSNR of the luminance component of the video sequences is considered and the peak signal amplitude picked in this paper is 255. Figure 1 plots the PSNRs of each frame of the video silent.cif coded at QP = 26 and 30, GOPS = 15, PS = 100 for 100 realizations of the multipath fading channel of average SNR 7 dB and AWGN channel of SNR 3 dB, respectively, when PHY data rate 6 Mbps is used. The thick lines in each plot represent the average PSNRs across the 100 channel realizations. This average should be slightly different than the PSNR calculated from averaging the MSEs. In practice, however, there is no significant difference between the two definitions [13].

It is clear in Fig. 1 that even for the same video, coded using the same parameters for the same average channel SNR, the quality of the delivered video in terms of PSNR varies significantly across different channel realizations. The plots in Fig. 1 are typical for all of the videos and codec parameters we tested. PSNRs also can vary dramatically from
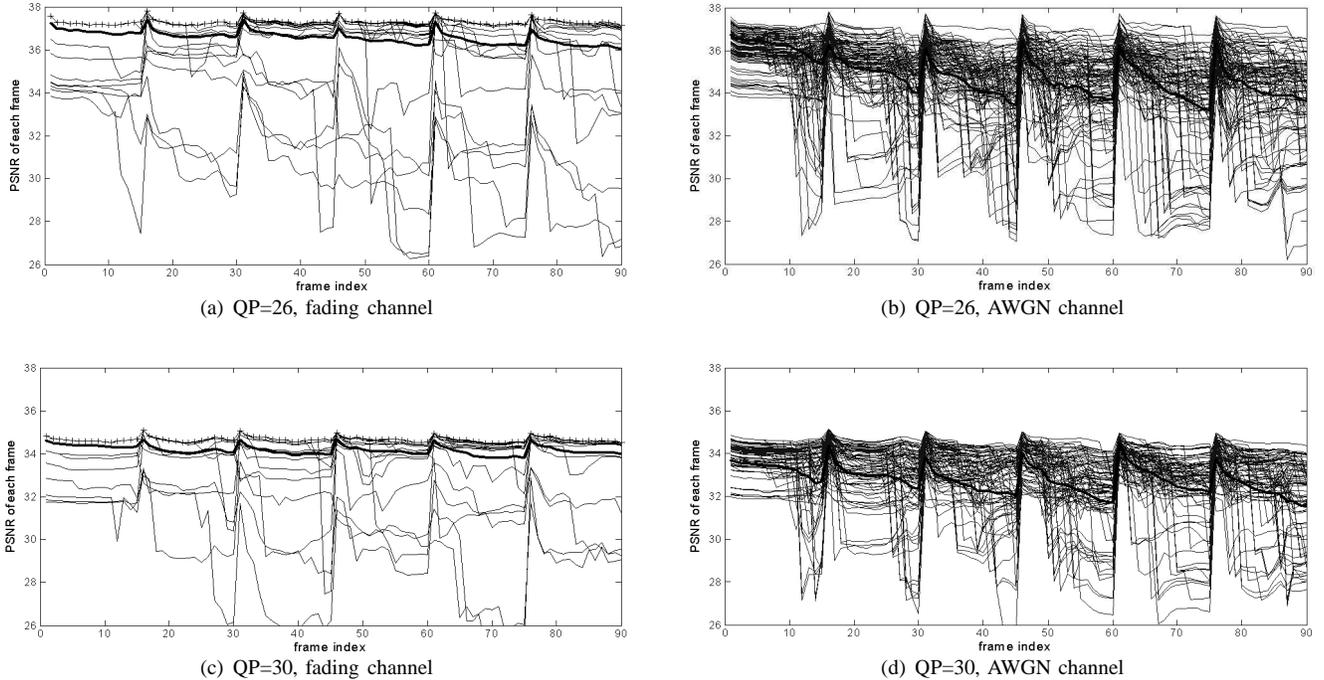
Fig. 1. PSNRs of each frame of the video silent.cif over 100 realizations of both multipath fading channels and AWGN channels. The thick lines in each plot represent the average PSNRs across the 100 channel realizations

one frame to another in the same processed video sequence. In Figs. 1(a) and 1(c), the realizations that have no packet loss overlap and form the lines marked with "+". For the AWGN channel, all realizations have similar packet loss rates (PLR). However, because of the prediction employed in video coding, it is shown in Figures 1(b) and 1(d) that the realizations of similar PLR can generate completely different concealed video quality. This suggests that the average PSNR across all the frames and all the realizations is not a suitable indicator of the quality a video user experiences with a network with possible packet losses.

III. DEFINITION OF $PSNR_{r,f}$ AND ITS CORRESPONDENCE TO PERCEPTUAL QUALITY OF MULTIPLE USERS MOS_r

As shown in Section II and in particular in Fig. 1, for video communication over networks with possible packet errors, the PSNRs of the delivered videos vary significantly across the video frames and across the different realizations of the channel. In order to capture the distribution of the distortion across the video frames and channel uses, in this section we propose a statistical PSNR based video quality measure, $PSNR_{r,f}$, which is defined as the PSNR achieved by $f\%$ of the frames in each one of the $r\%$ of the realizations. Parameter r captures the reliability of a channel over many users and can be set as a number between 0% to 100% according to the desired consistency of the user experience.

The proposal of using $PSNR_f$, i.e., the lowest PSNR achieved by $f\%$ (usually set as a majority) of the frames in a single video sequence, to measure the perceptual quality of a single video sequence is based on three observations that are recognized by researchers in video quality assessment [5]: 1) the frames of poor quality in a video sequence dominate human viewers' experience with the video; 2) however, if only a very small portion of the video frames are of poor quality, the quality drop due to these few frames are not perceivable by the human viewers ; 3) when the PSNRs are higher than a threshold, increasing PSNR does not correspond to an increase in perceptual quality that is already excellent at the threshold.

To confirm these observations, i.e., to study the correlation between $PSNR_f$ and the perceptual quality of videos, as well as to find a suitable range for the parameter f, a subjective experiment is designed and conducted. Stimulus-comparison methods [14] are used in this experiment, where two video sequences of the same content were presented to the subjects side by side and were played simultaneously. The video on the left is considered to be of perfect quality while the video on the right is compressed and then reconstructed with possible packet loss and concealment. Three naive human subjects are involved in this experiment. They are asked to pick a number representing the perceptual quality of the processed video compared to the perfect video from the continuous quality scale shown in Figure 2. Fifty video pairs were tested and 20% of them appear twice in this experiment to test the consistency of the subjects' decisions.

Figure 3 plots the opinion scores given by the three subjects in circles ($'o'$), dots ($'\cdot'$) and crosses ($'+'$). Of the 50 tested videos, 18 are silent.cif, reconstructed from different levels of packet losses. They are arranged from left to right with ascending average of three subjects' opinion scores. The same
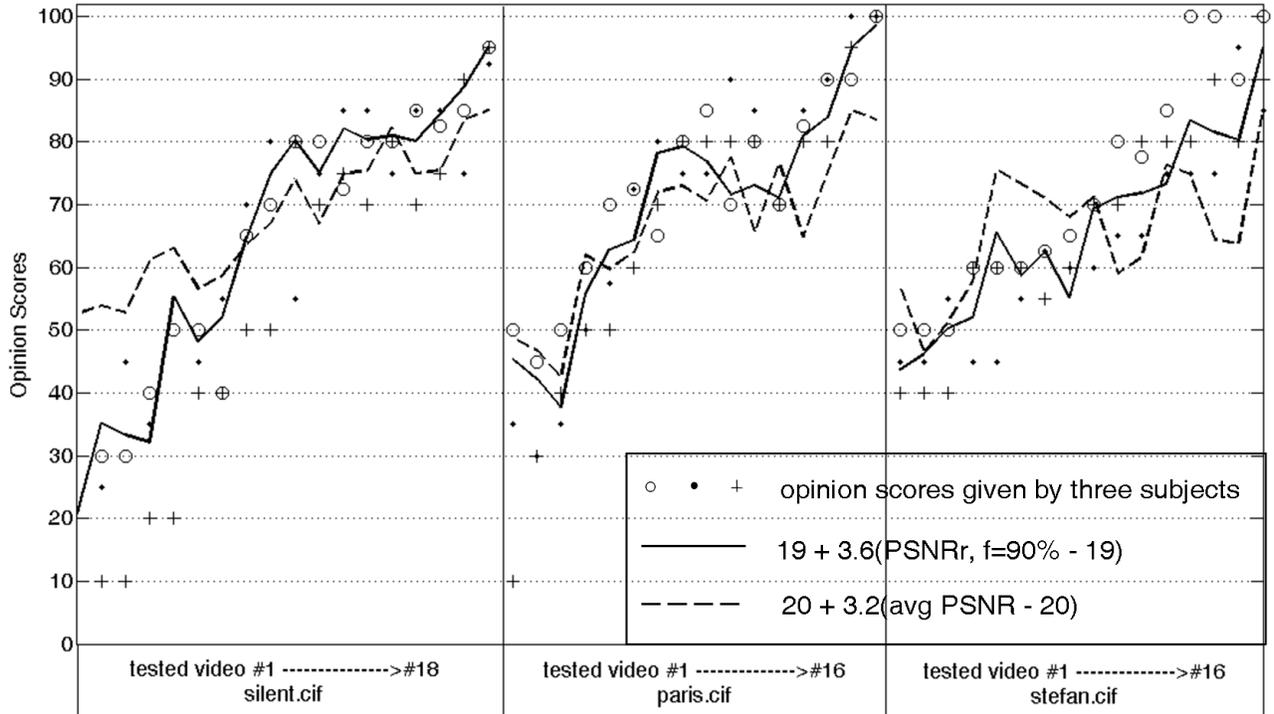
Fig. 3. The opinion scores given by the three subjects and the the best linear mappings of $PSNR_{f=90\%}$ and average PSNR. Of the 50 tested videos, 18 are silent.cif, 16 are paris.cif and 16 are stefan.cif, reconstructed from different levels of packet losses. They are arranged from left to right with ascending average of three subjects' opinion scores.

Fig. 2. Perceptual video quality scale in MOS

is done for the 16 videos of paris.cif and the 16 videos of stefan.cif. For each tested video, the PSNRs are calculated for each frame, from which both average PSNR across all frames and $PSNR_f$ with f as any value can be further calculated. Since the PSNRs and the opinion scores are of different scales, in order to compare them, the average PSNR and $PSNR_f$ with f ranging from 0.5 to 0.99 are mapped to the opinion scores through linear functions which yield the minimum mean square errors in the fit.

We find that among all the values of f we investigate, $PSNR_f$ with $f = 90\%$ correlates to the opinion scores the best, whose linear mapping is plotted as solid lines in Fig. 3. We also plot the best linear mapping of average PSNRs in dashed lines for comparison. As seen from these curves, $PSNR_{f=90\%}$ correlates significantly better than average PSNR, to the perceptual quality for all three videos that are given in circles ($'o'$), dots ($'\cdot'$) and crosses ($'+'$) for each video. The average PSNR underestimates the quality at high quality level and overestimates the quality at low quality level. This is because average PSNR treats all frames equally. At high quality level, however, only a few frames with

relatively lower quality bring down the average PSNR but do not affect the perceptual quality. At low quality level, on the other hand, there are frames with extremely bad quality which affect the overall video quality significantly while the average PSNR is still quite high. This subjective experiment shows that $PSNR_{r,f}$ can serve as an effective video quality measure, and that f should be set around 90% for medium video frame rates, such as 15 fps used in this paper. In this case the linear mapping from $PSNR_{r,f=90\%}$ to MOS_r, the mean opinion score (MOS) achieved by $r\%$ of the transmissions, is

$$MOS_r = 19 + 3.6(PSNR_{r,f=90\%} - 19). \qquad (1)$$

IV. APPLICATIONS

The new multiuser perceptual video quality measure $PSNR_{r,f}$-MOS_r is composed of two parts. $PSNR_{r,f}$ focuses on the distribution of the video quality across the video frames and channel uses, while MOS_r also provides guidance on the perceptual quality across different users. The MOS in MOS_r can be calculated from $PSNR_{f=90\%}$ using Eq. (1). The proposal of $PSNR_{r,f}$-MOS_r is motivated by the AVC/H.264 coded video over IEEE 802.11a WLAN simulation, but this measure is independent of the simulation setup and can be exploited in different video communication systems.

Here we briefly discuss an example of how $PSNR_{r,f}$-MOS_r can be used. In Fig. 4 we plot the MOS_r of three

videos coded by AVC/H.264 (using QP = 26, GOPS = 10, PS = 100 bytes) and transmitted over an 802.11a WLAN with a PHY data rate of 6 Mbps at average channel SNRs of 5 and 7 dB, respectively. In the 7 dB channel (the three curves on the right) for example, if all users are assumed to be communicating the same type of videos and an 80% consistency in user experience is desired, i.e., r=0.8, the videoconferencing users (silent.cif) experience an MOS over 80 out of 100 which corresponds to an "excellent" video quality with regard to the scale plotted in Fig. 2; the news watchers (paris.cif) experience a "good" video quality (a MOS of 74 out of 100), but the sports fans only receive "bad" quality videos, corresponding to a MOS of 30 out of 100, which is 40 to 50 points lower than those of the other two groups of users. This information can then be utilized for link adaptation, system performance evaluation, or system design purposes. For example, depending on the type of videos a specific communication system targets, a lower PHY data rate might need to be used, if one is available, in order to achieve a good user experience.
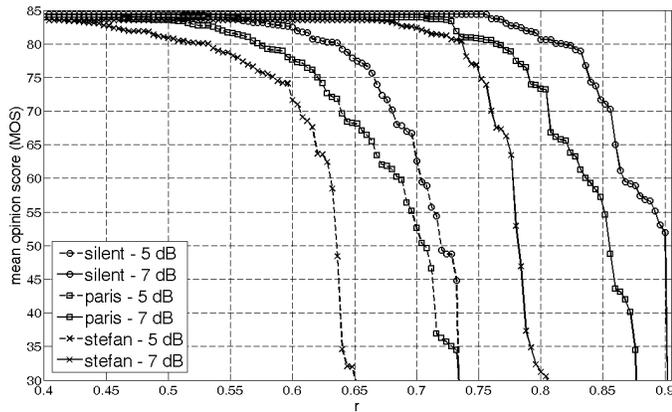
Fig. 4. MOS_r of three videos coded by AVC/H.264 using QP = 26, GOPS = 10, PS = 100 bytes and transmitted over 802.11a WLAN with a PHY data rate of 6 Mbps at average channel SNRs 5 and 7 dB

$PSNR_{r,f}$-MOS_r can be adapted partially. If a sophisticated objective video quality measure is desired despite its high cost, it can certainly be used to calculate the MOS in MOS_r and let the statistics of the MOS's be captured by the percentage of channel realization r that achieves each MOS value. On the other hand, when a video is to be transmitted over a reliable network with negligible packet error, or only video coding (i.e., no video transmission) is under investigation, the statistics of video quality across communication channel realizations becomes irrelevant. In this case, $PSNR_{r,f}$-MOS_r can be collapsed into $PSNR_f$-MOS, i.e., the MOS of a single compressed video sequence can be calculated from $PSNR_{f=90\%}$ using Eq. (1).

V. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a new statistical objective perceptual video quality measure $PSNR_{r,f}$-MOS_r. $PSNR_{r,f}$

is defined as the PSNR achieved by $f\%$ of the frames in each one of the $r\%$ of the transmissions over a network. This quantity has the potential to capture the performance loss due to damaged frames in a particular video sequence ($f\%$), as well as to indicate the probablity of a user experiencing a specified quality over the channel ($r\%$). The percentage of transmissions also has the interpretation as what percentage out of many video users who access the same channel, would experience a given video quality. We further investigate the correspondence between $PSNR_f$ and perceptual video quality through a subjective experiment which results in a linear equation connecting $PSNR_{r,f=90\%}$ and MOS_r, the mean opinion score (MOS) achieved by $r\%$ of the transmissions. It is shown from this subjective experiment that $PSNR_{f=90\%}$ correlates much better with the delivered perceptual video quality than the average PSNR across all frames of a video, while with no extra computatation.

Future work includes more subjects in the subjective experiment to construct a nonlinear relationship between the opinion scores MOS_r and $PSNR_{r,f}$.

REFERENCES

[1] ITU Recommendations, "Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a full reference," *ITU-R BT.1683*, Jun. 2004.

[2] ——, "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference," *ITU-T J.144*, Mar. 2004.

[3] V. Q. E. Group, "The Quest for Objective Methods: Phase II, Final Report," http://www.its.bldrdoc.gov/vqeg/, Aug. 2003.

[4] T. N. Pappas and R. J. Safranek, "Perceptual criteria for image quality evaluation," *Handbook of Image & Video Processing (A. Bivok eds.), Academic Press*, 2000.

[5] Z. Wang, H. R. Sheikh, and A. C. Bovik, "Objective video quality assessment," *The Handbook of Video Databases: Design and Applications (B. Furht and O. Marque, eds.), CRC Press*, pp. 1041–1078, Sep. 2003.

[6] ITU-T and ISO/IEC JTC 1, "Advanced video coding for generic audio-visual services," 2003.

[7] IEEE Std. 802.11-1999, Part 11, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications," *ISO/IEC 8802-11:1999(E)*, 1999.

[8] "H.264/AVC software coordination - reference software JM10.1," *http://iphome.hhi.de/suehring/tml/*, 2006.

[9] N. Chayat, "Tentative criteria for comparison of modulation methods," *IEEE P802.11-97/96*, Sep. 1997.

[10] R. van Nee and R. Prasad, *OFDM for Wireless Multimedia Communications.* Artech House, Jan. 2000.

[11] T. Stockhammer and W. Zia, "Error-resilient coding and decoding strategies for video communication," *Multimedia over IP Wireless Networks, M. van der Schaar and P. A. Chou eds. Elsevier / Academic Press, publishers*, Mar. 2007.

[12] F. Chiaraluce, L. Ciccarelli, E. Gambi, and S. Spinsante, "Performance evaluation of error concealment techniques in h.264 video coding," *Picture Coding Symposium*, no. 1, pp. 163–166, 2004.

[13] K. Stuhlmuller, N. Farber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, Jun. 2000.

[14] "Methodology for the subjective assessment of the quality of television pictures," *ITU-R Recommendation BT.500*, 2002.