Perceptually Unequal Packet Loss Protection by Weighting Saliency and Error Propagation

Hojin Ha, Jincheol Park, Sanghoon Lee, *Member, IEEE,* and Alan Conrad Bovik, *Fellow, IEEE*

***Abstract*—We describe a method for achieving perceptually minimal video distortion over packet-erasure networks using perceptually unequal loss protection (PULP). There are two main ingredients in the algorithm. First, a perceptual weighting scheme is employed wherein the compressed video is weighted as a function of the nonuniform distribution of retinal photoreceptors. Secondly, packets are assigned temporal importance within each group of pictures (GOP), recognizing that the severity of error propagation increases with elapsed time within a GOP. Using both frame-level perceptual importance and GOP-level hierarchical importance, the PULP algorithm seeks efficient forward error correction assignment that balances efficiency and fairness by controlling the size of identified salient region(s) relative to the channel state. PULP demonstrates robust performance and significantly improved subjective and objective visual quality in the face of burst packet losses.**

***Index Terms*—Forward error correction, human visual system, internet video, perceptual coding, unequal loss protection (ULP).**

I. Introduction

WITH THE EXPLOSIVE growth of multimedia environments, the robust transmission of video data has become an important requirement to enable smooth and seamless interaction with multimedia content [1], [2]. In error-prone environments, significant spatio-temporal dependencies in the video data may be lost owing to congestion, jitter, or delays over packet-erasure networks. This leads to substantial deterioration of received video quality from error propagation. To minimize visual quality degradation from packet losses, it is necessary to simultaneously consider the question of perceptual video quality [3]–[11] while accounting for error propagation effects arising from the video coding structure [13]–[21].

Manuscript received March 21, 2009; revised September 25, 2009; accepted January 21, 2010. Date of publication May 27, 2010; date of current version September 9, 2010. This work was supported by the Agency for Defense Development under Contract UD1000221D, Korea. This paper was recommended by Associate Editor H. Sun.

H. Ha is with the Digital Media and Communications Research and Development Center, Samsung Electronics, Yeongtong-gu, Suwon-si 443-373, Korea (e-mail: hojiniha@yonsei.ac.kr).

J. Park and S. Lee are with the Department of Electrical and Electronics Engineering, Yonsei University, Seoul 120-749, Korea (e-mail: dewofdawn@yonsei.ac.kr; slee@yonsei.ac.kr).

A. C. Bovik is with the Laboratory for Image and Video Engineering, Center for Perceptual Systems, University of Texas, Austin, TX 78712-1084 USA (e-mail:bovik@ece.utexas.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2010.2051368

Here, we present a packet loss resilience scheme that is based on an unequal loss protection (ULP) method that seeks to minimize perceptual distortions in the compressed video bit stream. This is accomplished by assigning unequal importance to different levels in the video coding structure using models of human visual sensitivity. We begin by quantifying the relative importance of video frames within each group-of-pictures (GOP) [13]–[15]. The motion compensation that is computed from the frames in each GOP causes the picture quality of a current reconstructed frame to be strongly dependent on the reconstructed version of its preceding frames. Generally, when packet losses occur earlier in a GOP, the reconstructed quality of following frames will be more severely compromised owing to the longer error propagation. In this sense, the perceptual importance of each frame descends from the first frame to the last frame in each GOP. We then define a procedure for incorporating perceptual weights into an ULP scheme [22], [23]. Similar approaches have been used to improve visual quality in other resource allocation schemes, by allocating more resources to perceptually important bit information via the use of visual saliency weights [3], [4]. In [6]–[9], a nonuniform spatial filtering law, called *foveation*, was employed to define spatial perceptual weights on coding macroblocks (MBs). Larger weights were applied near presumed visual fixation points, which were represented at high resolution, while lower weights were assigned to peripheral points. This process of foveal weighting attempts to match the nonuniform density of photoreceptors over the retina to achieve better visual quality. In that approach, a nonuniform foveation filtering method causes the local spatial bandwidth (LSB) to rapidly decrease with distance from the presumed fixation point(s).

Fig. 1 depicts the dependence of packet loss induced perceptual quality degradation on the error resilience scheme. Specifically, the degree of visual quality degradation that occurs in the 45th frame of the *Foreman* sequence due to a packet loss in the 30th frame. Fig. 1(a) and (b) shows the 30th and 40th reconstructed frames when a perceptually salient region is protected from packet loss in the 30th frame. It is assumed in this example that the point of fixation is on the face in the spatial center of the 30th frame, although this need not be the case. By contrast, Fig. 1(c) and (d) shows the same frames suffering from the same degradation, but without the perceptual weighting mechanism. In previous paper, it has been observed that higher perceptual video quality can be obtained by protecting those portions of the bitstream corresponding to salient regions from packet loss [14], [22], [23]. However, the

question remains as to how to achieve the minimum degree of visual quality degradation from packet loss for a given video coding algorithm, using limited channel resources. The general approach we take is that, for a given number of channel coding bits and a given video coding structure, formulate an optimization procedure that enables forward error correction (FEC) based on an appropriate perceptual weighting mechanism.

More specifically, we propose a performance metric based on both foveal weighting and on the temporal error propagation effect. The metric consists of two factors. One is the LSB obtained using a foveation filter model [7]–[11]. The other is called the perceptual weight on error propagation (PWEP). Using this metric, we develop an optimal FEC assignment algorithm which perceptually allocates channel coding resources. There have been related studies of error propagation modeling [24], [25]. However, these are computationally formidable when applied on a video server providing multiple concurrent video streams. Therefore, we developed a simple, alternative temporal error propagation model that requires much less computation. Of course, any type of error propagation modeling could be applied to our proposed scheme without difficulty. We chose to adopt a performance metric derived from the packet loss rate, and concentrating on the perceptual application of FEC in terms of fairness and efficiency.

This optimal allocation is defined in terms of efficiency and fairness as a function of the spatio-temporal weight carried in each packet. At high-packet loss rates, efficiency is given greater emphasis by allocating increased protection to localized salient regions. In this way, if degradation from packet loss occurs in less salient regions, higher quality can be still attained in more salient region(s). On the other hand, at low-packet loss rates, the size of the salient region(s) can be expanded. Fairness among data packets is fulfilled by allocating more bits to region(s) of low saliency. Thus, a tradeoff between efficiency and fairness is mediated based on the number of available channel coding bits and the channel status. In the simulations, it is shown that definite performance gains are achieved in terms of visual quality, efficiency and fairness, relative to conventional algorithms.

II. Related Work

Retransmission-based error control techniques such as automatic retransmission request have been shown to enhance the reliability of video transmission [26]. Nevertheless, simple techniques of this sort present limitations in real-time situations, owing due to delays arising from retransmitted packets. As an alternative, FEC deployed at the application layer yields a greater degree of efficiency. FEC can be adapted to variable bandwidths with reduced delay in wireless networks as well as in best-effort Internet networks. A number of researchers have proposed unequal FEC assignments to improve the quality of videos corrupted by packet loss [13]–[36]. For single layer videos, unequal protection can be conducted as a function of the coding type of each frame along the temporal axis. In [16] and [17], FEC codes were unequally assigned to I- and P-frames in each GOP according to the channel status. Unequal importance can also be assigned at the packet level.

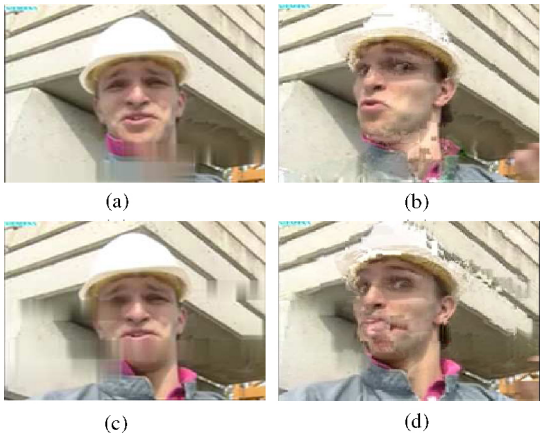

Fig. 1. Comparison of perceptual quality. (a) 30th frame using a perceptual ULP when a packet loss occurs in the frame. (b) 45th frame after error propagation from the 30th frame in (a). (c) 30th frame using a conventional ULP when a packet loss occurs in the frame. (d) 45th frame after error propagation from the 30th frame in (c).

For example, the packet header, motion information, and text information can be adapted to improve video quality in packet erasure networks [15]. In multilayered coding schemes, such as set partitioning in hierarchical trees, different degrees of importance can be assigned to the base and enhancement layers. By assigning unequal importance to the packets in different layers, unequal FEC schemes have been efficiently applied to multilayered coding [19], [20]. In [21], unequal error protection (UEP) is applied to MPEG-4 fine granular scalable (FGS) compressed video data using rate-distortion information for each layer. UEP schemes for multiple description coding (MDC) and for hybrid space-time coding have also been developed to achieve more robust video transmission [30], [31]. In recent years, source and channel rate allocation schemes have been deeply investigated for video communications [33]–[36]. A rate allocation scheme with a delay constraint was presented in [33]. The number of FEC codes is determined based on the network delay and the packet generation interval. The number of redundant packets is then allocated to attain a required packet loss ratio. In [34], a lower bound on the total transmission rate was computed by exploiting both source coding bits to attain minimum quality and channel coding bits to achieve the required packet loss ratio.

III. Perceptually Unequal Loss Protection (PULP)

A. Motivation

Fig. 2 shows the mechanics of PULP as compared to a conventional approach. If the available resources for video coding or transmission are plentiful, then we do not expect a performance improvement of the proposed scheme relative to conventional ones. However, when the resources are insufficient, then noticeably better performance can be attained by protecting perceptually important regions. The new approach balances a tradeoff between fairness and efficiency from the perspective of perceptual improvement. Fairness and efficiency mediate the visual quality by controlling the

size(s) of identified salient region(s). As the fairness level is increased, salient regions are increased in size, leading to improved visual quality of the reconstructed video. The conventional approach, shown in Fig. 2(b) employs equal perceptual weighting across the fairness levels. No spatial assignment of visual importance or salience is used in defining the fairness levels. Nevertheless, there are opportunities for incorporating perceptual relevance. For example, if regions that attract visual attention can be identified, then resources can be allocated to them, while also taking into account human contrast sensitivity when selecting the quantization level or the prediction block size. By comparison, Fig. 2(c) shows the proposed PULP framework, which adaptively configures each fairness level to the channel behavior. FEC assignments are based on perceptual weights, which are dynamically selected as a function of the channel state. The size of the salient region(s) is adaptively adjusted as a function of the fairness level and of the channel state. The larger the size of the salient region, the higher the fairness level is. For example, the fairness of level 1 is larger than that of level 0 in Fig. 2(c). When the packet loss rate increases, the size of the salient region is reduced, to improve efficiency by setting a low-fairness level. When the packet loss rate decreases, the size of the salient region is expanded by setting a high-fairness level.

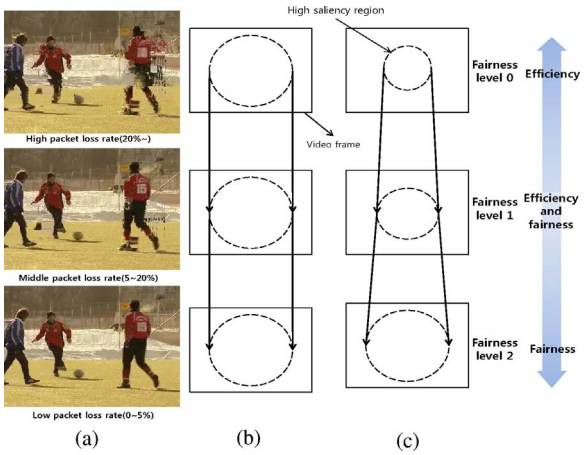

Fig. 2. Proposed PULP framework compared to conventional ULP. (a) Channel status. (b) Conventional ULP. (c) PULP.

B. Overview of the PULP Algorithm

Fig. 3 diagrams various essential aspects of PULP. Fig. 3(a) shows the flow of PULP. Raw video frames are first fed into the video encoding module. During the encoding process, the degree of degradation due to packet loss is estimated used a quality metric called PWEP. The packet loss rate is estimated using the Markov model in [46]. It can then be reported by the underlying protocol, such as the real time control protocol (RTCP) [32]. Furthermore, by deploying cross-layer cooperation, the channel signal-to-interference-plus-noise-ratio (SINR) can be measured using the pilot channel. From the SINR, the bit error rate and the packet error rate (PER) can then be estimated. If the PER information is fed back periodically to the end-user via RTCP packets, the QoS may be more reliably controlled. PWEP values are obtained using a foveal weighting model and a GOP-level hierarchical weighting model. The foveal weighting model calculates the LSB for each video packet. The LSB is decreased exponentially from the centers of each salient region, which are called *foveation points*. The exponential drop-off is such that, when a visual fixation falls on the salient region, the projection of the distribution of LSBs onto the retina will approximately match the nonuniform distribution of retinal photoreceptors [3]–[9].

Fig. 3(b) shows the reconstructed 35th frame of the video test clip "*Silent*" after applying perceptual weighting, where three salient regions were identified. The figure also depicts the foveal weighting model. Assume that the face and the left hand, both of which are in motion, are selected as a region of heightened visual interests. Picture-level perceptual weighting is allocated as a function of the spatial placement within the indicated iso-contours of the foveation-induced LSBs. In this example, video packets in region A are located in a highly salient region and are thus well protected. Video packets in region B are located in a low-saliency region and are less well protected. The spatial weighting is obtained for each frame in the GOP. In addition, the GOP-level hierarchical weighting model is used to identify regions that have unequal importance in the compressed video packets. Specifically, in each GOP the pictures have importance that descends with time relative to the first reference frame (I-frame), owing to the increasing severity of error propagation with elapsed time within the GOP. Using both of frame-level perceptual importance and GOP-level hierarchical importance, the PULP FEC assigner seeks to balance and optimize efficiency and fairness in order to achieve improved visual quality. The right portion of Fig. 3(c) depicts the architecture of the model-based FEC assignment algorithm. Video packets from the video encoder are assembled into blocks of packets (BOP) by the BOP assembler for each GOP. Since channel error propagation is terminated within each GOP, this assembling leads to improved FEC capacity. The bit stream is sequentially packetized without considering regions of interest. Reed–Solomon (RS) codes are used across packets for FEC in the face of packet loss in packet erasure networks [15], [39]. The (N, K) RS code has a code rate of K/N, where N packets are transmitted over the channel for K video packets. These N packets build a BOP, and this code rate can be adjusted for each BOP as a function of the unequal importance of the visual quality degradation and the channel state.

1) *Problem Formulation:* Let F_j denote the number of FEC packets assigned into BOP j, Then, the FEC assignment vector for the current GOP is $\vec{F} = [F_1, F_2, ..., F_J]$. The optimal FEC assignment vector $\vec{F}^*$ can be obtained by minimizing an appropriate performance metric $D(\vec{F})$ which is defined in the next section. For a given F_j in BOP j, we denote $\gamma(F_j)$ to be the packet loss rate after recovering with RS (N_j, K_j) codes. A two-state Markov model is used to model the packet loss rate [46]. If $Pr(m, N)$ is the probability of losing m packets among N packets, then the original data can be recovered if the number of lost packets is less than the number of protection packets. $\gamma(F_j)$ can be formulated as

$$\gamma(F_j) = \sum_{m=N_j-K_j+1}^{N_j} \Pr(m, N_j) \tag{1}$$

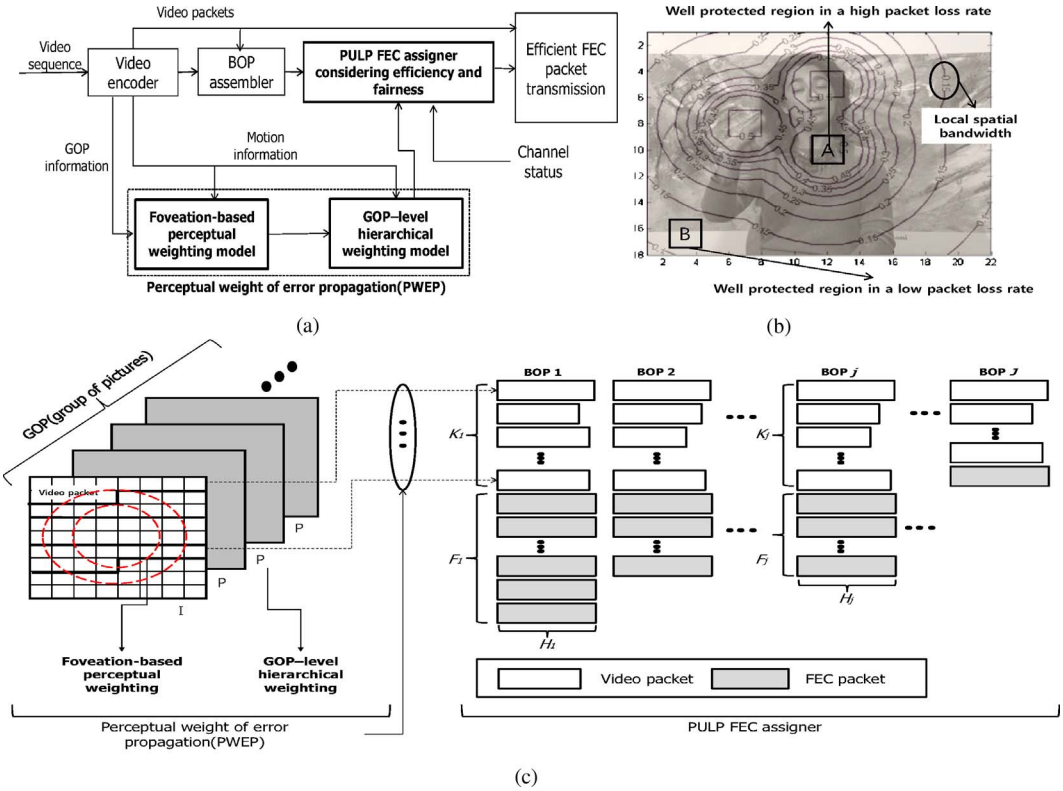

Fig. 3. Overview of the proposed VULP scheme. (a) Block diagram of the proposed VULP scheme. (b) Depiction of the foveation-based perceptual weighting model. (c) Architecture of the model-based FEC assignment algorithm.

An optimal FEC assignment vector $\vec{F}^*$ can be obtained by minimizing the spatio–temporal performance metric $D(\vec{F})$, which is the measure of quality degradation due to packet loss from the previous section. The problem of finding the optimal FEC assignment vector $\vec{F}^*$ is then

$$\min_{\vec{F}} \left[D(\vec{F}) \right] \tag{2}$$

subject to

$$\gamma(F_j) \leq \gamma(F_{\hat{j}}) \quad \text{if } j \leq \hat{j} \tag{3}$$

where j and $\hat{j}$ are BOP indices. The constraint (3) is referred as "descending priority" from the sorted performance metric in the GOP

$$\sum_{j=1}^{J} H_j \cdot F_j \leq B_{ch} \tag{4}$$

where J is the total number of BOPs for a given K. In constraint (4), B_{ch} is the available number of channel bits for constructing FEC packets. H_j is the length of FEC packets in BOP j as determined by $H_j = \max_{k=1,2,\ldots,K_j} \{h_{k,j}\}$ where $h_{k,j}$ is the length of the kth video packet in BOP j. For smaller video packets, filler bytes are used to equalize the length of the video packets before FEC encoding.

IV. Perceptual Weight of Error Propagation (PWEP)

Here, we describe a model-based performance metric for measuring the amount of visual quality degradation that occurs due to packet loss in a GOP.

A. Salient Point Selection

Selection of salient points is designed under the following assumption: The HVS often directs more attention to moving objects than to a stationary background [10], [11], [42], [44]. The HVS deals unequally with incoming visual information according to selective focal attention. This selectivity implies that if the human gaze is directed toward any specific locations in a video, then that observer is less likely to notice defects in other areas of the video [43]. In PULP, available MB saliency information, such as velocity magnitude, and motion partition, can be used to define salient MBs. This can be used in conjunction with standard video formats, such as H.264/AVC, which provide inter and intra prediction modes to obtain improved coding performance [45]. The MB partition information for the (i, j)th MB in the kth picture is defined as $P_k(i, j)$ and it is calculated by using rate-distortion optimization (RDO) for each MB. An example of MB partition information is showed in Fig. 4(a) by using the 13th frame in the '*Soccer*' test video clip. Small partition sizes are generally identified with detailed regions or the edges of objects. Large partition sizes are usually associated with monotonous, stationary image regions. For intra picture

coding, a similar rule can be applied. Prediction modes that deploy a small block partition size are usually used to represent detailed regions. Larger partition sizes are used to coding homogeneous areas.

In addition to $P_k(i, j)$, we utilize the velocity magnitude to define salient areas. The motion intensity of (i, j)th MB in the kth picture is defined as $I_k(i, j)$ and it is calculated by $I_k(i, j) = \sqrt{MV_x^k(i, j)^2 + MV_y^k(i, j)^2}$ where $MV_x^k(i, j)$ and $MV_y^k(i, j)$ represent the horizontal and vertical direction velocity magnitude for the (i, j)th MB in the kth frame, respectively. Using $P_k(i, j)$ and $I_k(i, j)$, candidate salient points can be selected. The detailed decision procedure is described in Fig. 4(b). If $P_k(i, j)$ is small, it is probable that the MB contains details or information-bearing edges. Such MBs are taken to be part of salient regions. However, in the case that an MB is a part of the background, then $P_k(i, j)$ may not be large enough to report a possible saliency. Thus, after filtering out MBs using $P_k(i, j)$ we use the additional step of $I_k(i, j)$. If $I_k(i, j) > 0$, the MB is regarded as part of a salient region. If $I_k(i, j)$ is too large, the viewer may not perceive such a rapid change, or might only obtain a limited amount of information [42], [44]. This explains the use of a variable threshold on $I_k(i, j)$ for selecting salient points in the kth frame. The variability depends the global mean of velocity magnitudes in the frame (denoted by σ_k) which accounts for egomotion. Assuming that each frame is divided into $M \times N$ MBs, σ_k then is calculated as $\sigma_k = \frac{1}{M \cdot N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I_k(i, j)$.

All other MBs excluding the selected salient MBs are treated as nonsalient MBs. Based on $I_k(i, j)$, $P_k(i, j)$, and σ_k, we describe the decision algorithm for selecting salient points in Step 1 and 2.

Step 1) Calculate $I_k(i, j)$, $P_k(i, j)$ and σ_k for the (i, j)th MB in the kth frame.

Step 2) We define a binary function $A_k(i, j)$ as an indicator of whether or not the (i, j)th MB in kth frame belongs to a salient region. $A_k(i, j) = 1$ means a salient point, while $A_k(i, j) = 0$ means others.
By using $I_k(i, j)$, $P_k(i, j)$ and σ_k, we determine whether $A_k(i, j)$ is 1 or 0 as follows:

$$A_k(i, j) = \begin{cases} 1, & \text{if } (P_k(i, j) < \text{MODE\_}16 \times 16 \text{ and } \\ & \quad 0 < I_k(i, j) < \sigma_k) \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

A result of the proposed algorithm is shown in Fig. 4(c).

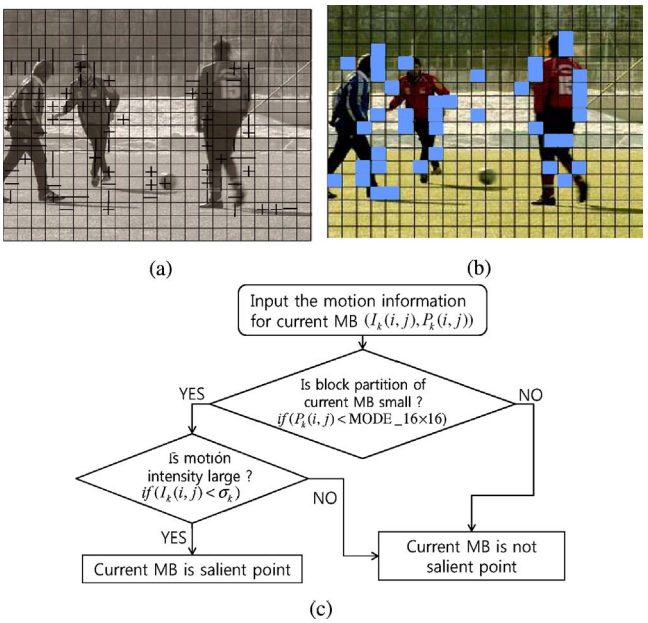

Fig. 4. Salient point selection for the 13th frame in the *Soccer* sequence. (a) A result of the MB block partition ($P_k(i, j)$) obtained from H.264/AVC. (b) Salient MB selection algorithm. (c) Selected salient MBs.

B. Foveation-Based Perceptual Weighting

Since video images are intended for human viewers, it is presumed that the point of visual fixation falls somewhere on the displayed video. It is also a reasonable assumption that visual input is dominated by the response of the cones (photopic vision), since the central dominant photoreceptors are highly responsive to bright objects, such as a glowing display monitor. The point on an object or monitor surface that projects light onto the center of the fovea, presuming that the gaze is fixed, is termed a point of visual fixation. At certain locations in the video stream where it is deemed likely that the human gaze will fall at a given point in space and time, the video will be either represented at a higher resolution than other locations, and possibly given another kind of priority, such as increased error resilience. Such locations in the video stream will be referred to as *foveation points*. In the vicinity of a foveation point, the video is represented with a high-spatial resolution which falls off systematically away from the foveation point [except near other fixation point(s)]. In this way, the video presentation is made to have high resolution where the observers' visual fixations are known or predicted to be placed. There has been useful work done on determining the visual resolution response (contrast sensitivity) as a function of the placement of the stimulus on the retina relative to the fovea, which is known as the *retinal eccentricity* [41], [42], [44].

For any given point $\vec{x} = (x_1, x_2)$ (pixels) in an image or video frame, the eccentricity (e) can be found by assuming that the position of the foveation point $\vec{x}^f = (x_1^f, x_2^f)$ (pixels) in the image plane and the viewing distance u from the eye to the image of size W(pixels) are known. The distance from $\vec{x}$ to $\vec{x}^f$ is $d(\vec{x}, \vec{x}^f) = \sqrt{\left(x_1 - x_1^f\right)^2 + \left(x_2 - x_2^f\right)^2}$(pixels). The eccentricity is $e(u, \vec{x}) = \tan^{-1}(\frac{d(\vec{x}, \vec{x}^f)}{Wu})$ [8].

For a given eccentricity, $e(u, \vec{x})$, the local spatial cut-off frequency (cycle/degree) in $\vec{x}$, w_c is defined in the sense that any higher frequency component beyond it is less visible or invisible. By setting the maximum possible contrast sensitivity to 1.0, w_c is calculated as follows:

$$w_c(e(u, \vec{x})) = \frac{e_2 \ln(\frac{1}{CT_0})}{\alpha(e(u, \vec{x}) + e_2)} \left(\frac{\text{cycles}}{\text{degree}}\right) \tag{6}$$

where CT_0 is a minimum contrast threshold, e_2 is a half-resolution eccentricity constant, and α is a spatial frequency

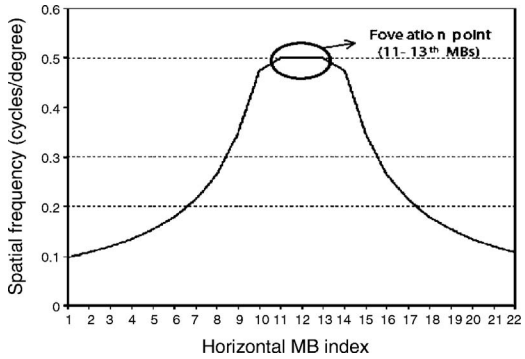

Fig. 5. Distribution of normalized $w_c^{n,k}(u)$ of the 5th horizontal MBs for W=1024 and u=30 cm.

decay constant. The fitting parameters given in [41] are $a = 0.106$, $e_2 = 2.3$, and $CT_0 = 1/64$.

In a displayed digital image, the maximum effective resolution is limited by the display's visual resolution r (pixels/degree), which is approximately

$$r \approx Wu \frac{\pi}{180} \left(\frac{\text{pixels}}{\text{degree}} \right). \tag{7}$$

Based on the sampling theorem, the highest displayed Nyquist frequency is half the display resolution from (7)

$$w_d(u) = \frac{r}{2} \approx Wu \frac{\pi}{360} \left(\frac{\text{cycles}}{\text{degree}} \right). \tag{8}$$

Combining (6) and (8), the local foveal cutoff frequency for a given location $\vec{x}$ is

$$w_c(u, \vec{x}) = \min(w_c(e(u, \vec{x})), w_d(u)). \tag{9}$$

In addition, the cutoff frequency of the kth MB in the nth frame, $w_c^{n,k}(u)$ can be expressed as the average value of the cutoff frequencies in the macroblock

$$w_c^{n,k} = \text{avg}(w_c(u, \vec{x})) \tag{10}$$

where the $\vec{x}$ are the pixels in the kth MB.

As an example, Fig. 5 shows the distribution of normalized $w_c^{n,k}(u)$ of the 5th horizontal MBs for W=1024 and u=30 cm. It may be observed that larger weights are assigned to salient regions. These weights decrease as $w_c^{n,k}(u)$ is decreased exponentially as a function of the distance from the foveation point. This process of foveation and weighting makes it possible to eliminate visual redundancies from nonsalient regions in order to improve coding efficiency. For brevity, $w_c^{n,k}(u)$ is expressed by $w_c^{n,k}$ to eliminate the dependence on u. Based on $w_c^{n,k}$, the perceptual weighting of the ith video packet of the jth BOP, $\mu^{i,j}$ can be calculated as

$$\mu_{i,j} = \sum_{k \in S(p_{i,j})} w_c^{n,k} \tag{11}$$

where $p_{i,j}$ is the ith video packet of the jth BOP and $S(p_{i,j})$ is the set of MBs in $p_{i,j}$.

TABLE I
NORMALIZED LOCAL SPATIAL FREQUENCY, $N_s(m, n)$ IN THE 4×4 DCT DOMAIN

Item(m,n)	0	1	2	3
0	0.01	0.13	0.25	0.38
1	0.13	0.18	0.28	0.40
2	0.25	0.28	0.35	0.45
3	0.38	0.40	0.45	0.50

C. GOP-Level Hierarchical Weighting

To quantify the temporal propagation effects of packet loss on video quality, we use the length of the possible error propagation for each video packet. For example, a packet loss of the first frame causes a much more severe impact on the quality of the reconstructed sequence than a packet loss in one of the frames near the end ending. This simple method of assessing frame quality loss due to error propagation also has the virtue of simplicity and low complexity. Let $f_{i,j}$ and $\lambda_{i,j}$ be the frame index in a GOP, and the length of error propagation of the ith video packet of the jth BOP, respectively. Then, $\lambda_{i,j}$ is given by

$$\lambda_{i,j} = G + 1 - f_{i,j}. \tag{12}$$

Using (11) and (12), we then define the perceptual weight of error propagation (PWEP), $\chi_{i,j}$, which combines effects of both spatial and temporal video quality degradation

$$\chi_{i,j} = \mu_{i,j} \cdot \lambda_{i,j}. \tag{13}$$

V. OPTIMAL FEC ASSIGNMENT IN PULP: EFFICIENCY AND FAIRNESS

The PWEP for each video packet can be obtained from (13) by using the spatial and temporal weighting principles outlined in the preceding. In order to minimize visual quality degradation as a function of the perceptual weighting, the proposed FEC assignment is adjusted as a function of the size of the salient region(s) and the channel status.

A. PWEP-FL(l)

For simplicity, denote the fairness level l as FL(l). The PWEP in (14) is specifically determined, as a function of the channel status, to achieve a desirable tradeoff between efficiency and fairness using the FL(l) algorithm. We term the proposed performance metric PWEP-FL(l). In the FL(l) algorithm, video packets having large LSB are protected from packet loss by adding more protection bits, and vice-versa. The lower the fairness level the algorithm obtains, the more unfair the video packets having a low LSB will be. Therefore it is important to carefully determine FL(l) to maintain an appropriate modicum of fairness for each channel state. Given the FL(l), a threshold on the cutoff frequency is determined. If $w_{n,k}$ exceeds the threshold, then the kth MB becomes a part of the salient region, and so on.

Fig. 6(a) and (b) shows the mechanics of the proposed PULP framework as a function of the channel status. In this

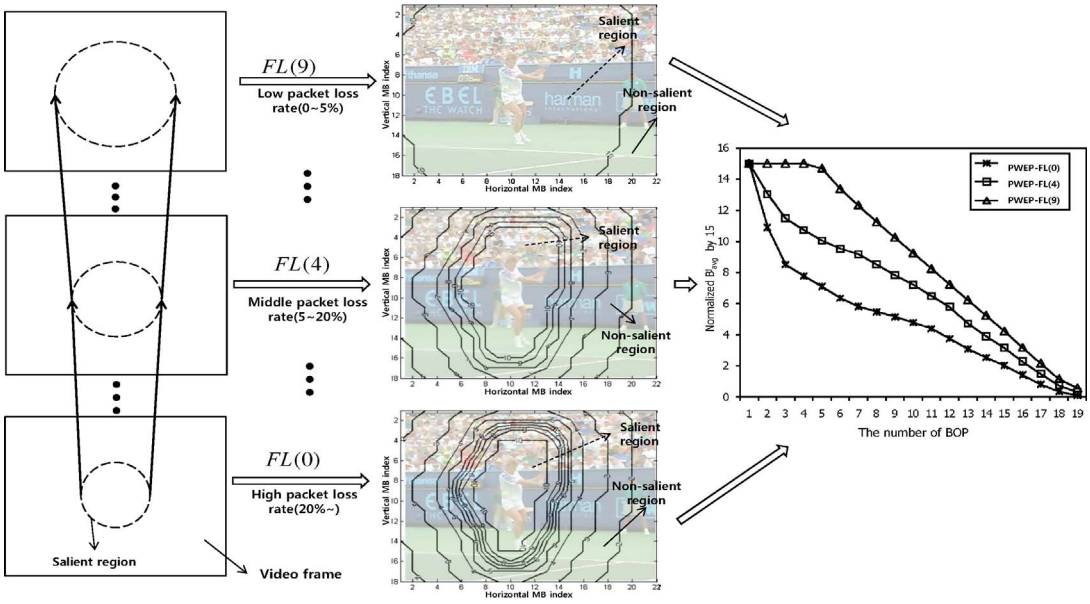

Fig. 6. PULP framework. (a) Illustration of variation of the saliency size across the fairness level, FL(l). (b) Distribution of perceptual weighting in the FL(l) algorithm as a function of the channel status. (c) Distribution of normalized PWEP-FL(l) for each BOP.

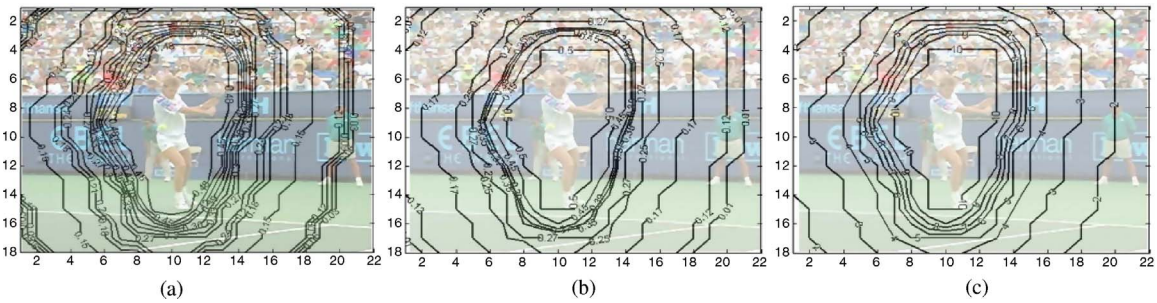

Fig. 7. For the 17th *Stefan* test sequence, shown are (a) contours of $w_c^{n,k}$, (b) contours of $\hat{w}_c^{n,k}$ from $w_c^{n,k}$, (c) contours of $v_{n,k}$ from $\hat{w}_c^{n,k}$.

example, the 17th frame of the "*Stefan*" test sequence is utilized. In Fig. 6(a), the lowest fairness level 0 is assigned by setting FL(0). By maintaining the smallest salient region, it is possible to maximally protect perceptual quality within the salient region against a high-packet loss rate. For a moderate packet loss rate, the fairness level is increased by assigning an intermediate value of the LSB to, for example, FL(4), leading to enlargement of the salient region. Finally, at a low-packet loss rate, increased fairness can be assured by assigning a low threshold on the cutoff frequency. For example, for FL(9), the size of the salient region is noticeably larger. The LSB of those MBs lying within the salient region is set to the highest value of 0.5 in the discrete frequency domain. To decide the size of the salient region(s), $w_c^{n,k}$ from (10) is mapped onto a discrete level of frequency sensitivity, which varies with the frequency indices of the transform coefficients, the coefficient magnitudes, and the block luminances [3], [10], [11]. Let m and n be the indices of 2-D transform coefficients in a block. Then, the normalized local spatial frequency (cycle/degree) can be expressed as $N_s(m,n) = \frac{1}{N}\sqrt{m^2+n^2}$ where $N_s(m,n)$ is normalized by 0.5 in [10]. As shown in Table I, 10 values of $N_s(m,n)$ are used to control the size of the salient region(s). For this purpose, $w_c^{n,k}$ is quantized as a function of the value of $N_s(m,n)$, so that the perceptual weighting of each MB is modified and the size of the salient region(s) is controlled.

Let $\hat{w}_c^{n,k}$ be the quantized version of $w_c^{n,k}$, where $\hat{w}_c^{n,k}$ is mapped into the nearest discrete value of $N_s(m,n)$. Fig. 7(a) and (b) shows the distribution of $w_c^{n,k}$ and $\hat{w}_c^{n,k}$. The values of $\hat{w}_c^{n,k}$ are mapped onto integer values denoted $v_{n,k}$ in the range [0, 9], as shown in Fig. 7(c). The relationship between $v_{n,k}$ and $\hat{w}_c^{n,k}$ is defined by a weighting function $\varphi_{n,k}$ as follows:

$$\hat{w}_c^{n,k} = \varphi_{n,k}(v_{n,k}) \tag{14}$$

where $v_{n,k}$ is an index used in Table I. If $v_{n,k} = 9$, the associated MB obtains the highest discrete value $\hat{w}_c^{n,k} = 0.5$. At the other extreme, if $v_{n,k} = 0$, then the lowest value $\hat{w}_c^{n,k} = 0.01$ is assigned to the MB.

Suppose that there are L fairness levels. For a given fairness level l, the perceptual weighting of each MB is fixed by $\varphi_{n,k}(\hat{v}_{n,k}(l))$ from (14), where $\hat{v}_{n,k}(l)$ indicates the modified value of $v_{n,k}$ from the fairness level l, which is calculated as

$$\hat{v}_{n,k}(l) = \begin{cases} L, & v_{n,k} \geq L - l \\ v_{n,k} + (L - l), & \text{otherwise} \end{cases} \tag{15}$$

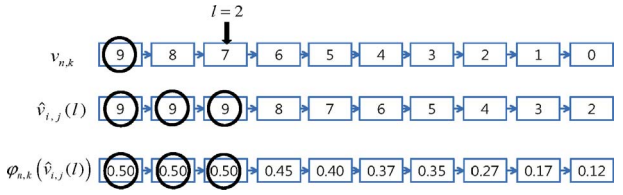

Fig. 8. Variation of $v_{n,k}$,$\hat{v}_{n,k}(l)$, and $\varphi_{n,k}(\hat{v}_{n,k}(l))$ when l =2.

where $L = 9$ in the implementation. This means that $\varphi_{n,k}(\hat{v}_{n,k}(l))$ is decreased in proportion to the distance from the center of the salient region, and is shifted as a function of the fairness level l. For example, when l=2, the variation of $v_{n,k}$, $\hat{v}_{n,k}(l)$, and $\varphi_{n,k}(\hat{v}_{n,k}(l))$ is depicted in Fig. 8. Using (14) and (15), $\hat{v}_{n,k}(2)$ is 9 until $v_{n,k}$ is 7, and thereafter is decremented in units. $\varphi_{n,k}(\hat{v}_{n,k}(l))$ shows the value of $\hat{w}_c^{n,k}$ obtained from $\hat{v}_{n,k}(l)$. Using the FL(l) algorithm, the perceptual weighting of the ith video packet of the jth BOP, which is denoted $\hat{\mu}_{i,j}(l)$, is found by including the fairness level l in $\mu_{i,j}$ in (11), and expressed using (14) and (15) as follows:

$$\hat{\mu}_{i,j}(l) = \sum_{k \in S(p_{i,j})} \varphi_{n,k}(\hat{v}_{n,k}(l)). \tag{16}$$

Finally, the performance metric PWEP-FL(l), $\hat{\chi}_{i,j}(l)$ is found by using (16) from (13)

$$\hat{\chi}_{i,j}(l) = \hat{\mu}_{i,j}(l) \cdot \lambda_{i,j}. \tag{17}$$

The average value of PWEP-FL(l) for the FEC assignment of the jth BOP is then calculated as

$$B^j_{\text{avg}}(l) = \frac{1}{K_j} \sum_{i=1}^{K_j} \hat{\chi}_{i,j}(l) \tag{18}$$

where K_j is the number of video packets in the jth BOP.

Fig. 6(c) depicts the distribution of B^j_{avg} as a function of FL(l) with $l = 0, 4, 9$. The slope of B^j_{avg} for PWEP-FL(0) is steeper than for PWEP-FL(9) or PWEP-FL(4), since the size of the salient region(s) are reduced by setting a low-fairness level, when protecting the visual quality in a high-packet loss rate environment. As the fairness level is increased, then the slope of B^j_{avg} becomes reduced, when improving the visual quality of expanded salient region(s) in a low-packet loss rate environment.

B. Optimal FEC Assignment

Based on the weighting for each BOP in (18), the PULP FEC assigner performs an optimal FEC assignment to minimize perceptual degradations, subject to a given protection redundancy and depending on the channel state. Expressed in terms of the packet loss rate in (18), and the average value of PWEP-FL(l) in (18), the spatio-temporal performance metric in (2) becomes

$$D(\vec{F}, l) = \sum_{j=1}^{J} B^j_{\text{avg}}(l) \cdot \gamma(F_j). \tag{19}$$

Using the definition of $D(\vec{F}, l)$, an optimal FEC assignment vector, $\vec{F}^*$ is found using a local hill-climbing search algorithm as in the following Steps 1–7.

Step 1) l_{curr} and l_{prev} represent the current and previous fairness levels, respectively. Initially, l_{curr} and l_{prev} are set to 0. Thus, the smallest saliency region is initially used for searching the optimal FEC assignment as a function of the channel state. For given l_{curr} and l_{prev}, the average distortions from (19) are denoted as $D(\vec{F}, l_{\text{curr}})$ and $D(\vec{F}, l_{\text{prev}})$ which are initially set to high values.

Step 2) Following compression of the video sequence within a GOP, video packets are generated. The perceptual weights $\hat{\chi}_{i,j}(l_{\text{curr}})$ are calculated by using (17). Each packet is then sorted to construct BOPs ordered by $\hat{\chi}_{i,j}(l_{\text{curr}})$. For a given K in RS(N, K), the collection of J BOPs is partitioned as shown in Fig. 3(c). The value $B^j_{\text{avg}}(l_{\text{curr}})$ associated with BOP j is calculated to allow the assignment of FEC packets.

Step 3) The number of FEC packets for BOP j in the face of a burst packet loss can be initially set to be $F_j^{\text{init}} = \left\lfloor \left(B_{ch} \cdot H_j\right) \Big/ \sum_{i=1}^{J} H_j \right\rfloor$ where J is the maximum number of BOPs. Then $D(\vec{F}^{\text{init}}, l_{\text{curr}})$ is calculated by using (19).

Step 4) Next, $\vec{F}^{\text{best}}$ and $\vec{F}^{\text{start}}$ are defined as the best FEC assignment in the GOP level and the starting point for the FEC assignment in the BOP level, respectively. The algorithm seeks $\vec{F}^{\text{best}}$ at the GOP level. The initial values of $\vec{F}^{\text{best}}$ and $D(\vec{F}^{\text{best}}, l_{\text{curr}})$ are $\vec{F}^{\text{init}}$ and $D(\vec{F}^{\text{init}}, l_{\text{curr}})$, respectively. Also, the initial value of $\vec{F}^{\text{start}}$ is set to a zero vector. $\vec{F}^{\text{start}}$ is replaced by $\vec{F}^{\text{best}}$ and the algorithm proceeds to Step 5.

Step 5) At the BOP level, the algorithm seeks the value of $\vec{F}$, denoted by $\vec{F}^{\text{temp}}$ that achieves an optimal FEC assignment in the sense of minimizing $D(\vec{F})$. $F_j^{\text{temp}} \in \vec{F}^{\text{temp}}$ is assumed to fall in the interval $[-\Delta_j, \Delta_j]$, where Δ_j is the search distance for BOP j which is determined by $\Delta_j = \max\left\{ \left\lfloor \varepsilon \cdot B^j_{\text{avg}} \right\rfloor, 1 \right\}$. This means that Δ_j I determined relative to the degree of the importance of the visual quality degradation from packet loss. If r is the determined value in the interval $[-\Delta_j, \Delta_j]$, then F_j^{temp} is updated as follows $\vec{F}^{\text{temp}} = \vec{F}^{\text{start}}$, $F_j^{\text{temp}} = F_j^{\text{temp}} + r$.

Step 6) We calculate $D(\vec{F}^{\text{temp}}) = \sum_{j=1}^{J} B^j_{\text{avg}} \cdot \gamma(F_j^{\text{temp}})$.

Step 7) If $D(\vec{F}^{\text{temp}}) < D(\vec{F}^{\text{best}})$ and $\sum_{i=1}^{J} F_j^{\text{temp}} < B_{ch}$, then $D(\vec{F}^{\text{best}})$ and $\vec{F}^{\text{best}}$ are replaced by $D(\vec{F}^{\text{temp}})$ and $\vec{F}^{\text{temp}}$, respectively. If $D(\vec{F}^{\text{temp}}) \geq D(\vec{F}^{\text{best}})$, the number of FEC packets in each BOP are adjusted to minimize $D(\vec{F}^{\text{temp}})$ in the interval $[-\Delta_j, \Delta_j]$, then the algorithm returns to Step 5. If the search range falls outside the interval $[-\Delta_j, \Delta_j]$, then BOP j is shifted into the next BOP using $j = j + 1$, and the algorithm returns to Step 5. If the index j of the BOP reaches the last value J, the algorithm goes to Step 8.

Step 8) If $\vec{F}^{\text{best}}$ is equal to $\vec{F}^{\text{start}}$, $\vec{F}^*$ is replaced by $\vec{F}^{\text{best}}$ and the FEC assignment process goes to

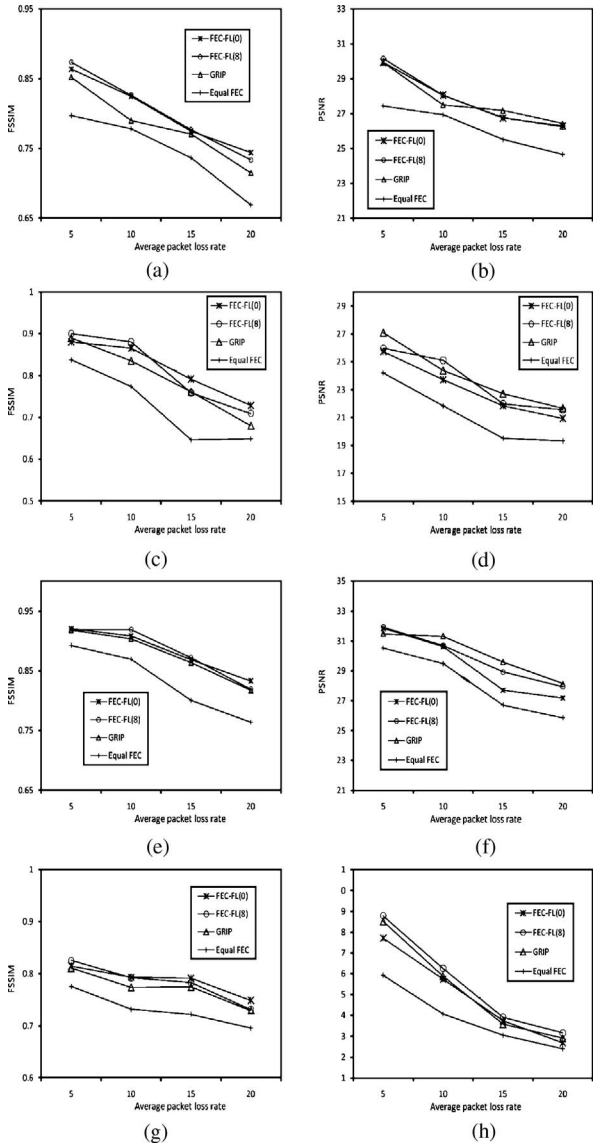

Fig. 9. Performance comparison of average FSSIM and PSNR between PULP and conventional ULP algorithms for a FEC ratio of 15%. (a) FSSIM of *City*. (b) PSNR of *City*. (c) FSSIM of *Stefan*. (d) PSNR of *Stefan*. (e) FSSIM of *Silent*. (f) PSNR of *Silent*. (g) FSSIM of *Soccer*. (h) PSNR of *Soccer*.

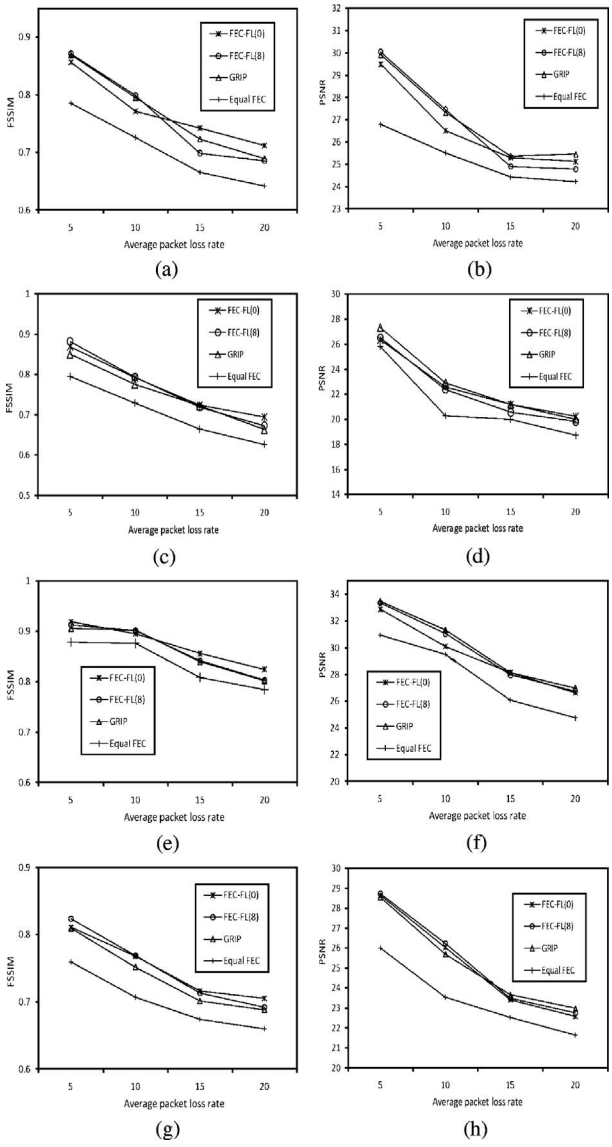

Fig. 10. Performance comparison of average FSSIM and PSNR between PULP and conventional ULP algorithms for a FEC ratio of 5%. (a) FSSIM of *City*. (b) PSNR of *City*. (c) FSSIM of *Stefan*. (d) PSNR of *Stefan*. (e) FSSIM of *Silent*. (f) PSNR of *Silent*. (g) FSSIM of *Soccer*. (h) PSNR of *Soccer*.

the next step. Otherwise, the algorithm jumps to Step 4.

Step 9) If $D(\vec{F}^*, l_{\text{curr}})$ is lower than $D(\vec{F}^*, l_{\text{prev}})$, we increase the fairness level by 1: $l_{\text{curr}} + 1$. This means that the size of the salient region is enlarged to adapt to the channel state. Then, $D(\vec{F}^*, l_{\text{prev}})$ is updated: $D(\vec{F}^*, l_{\text{curr}})$ and processing proceeds to Step 2. Otherwise, the FEC assignment process is terminated.

It can be seen that the computational complexity of the proposed FEC assignment algorithm depends on the choice of ε.

VI. Simulation Results

In order to evaluate the performance of the proposed FEC scheme, extensive experiments under various test conditions were conducted. Four CIF video sequences *City*, *Stefan*, *Silent*, and *Soccer,* were used. The number of frames for each sequence is 81 and the frame rate is 30 f/s. The initial quantization parameter is set to be 35. The videos were encoded using the H.264 reference software [45]. In the encoding configuration, the RDO mode and the loop filter were enabled. The content-based adaptive binary arithmetic coding option was enabled and variable block sizes with a search range of 32 were utilized for block motion estimation. The length of each GOP was selected to be 15 and the packet size 1280 bits. These sequences were encoded at a constant bit rate. A two-state Markov channel model described in [46] was used to model the packet loss with an average burst length of L_B and an average packet loss rate of P_B. The simulation parameters are shown in Table II. The error concealment scheme for H.264 [8] in the reference software [45] was applied at the decoder side. The simulations were conducted

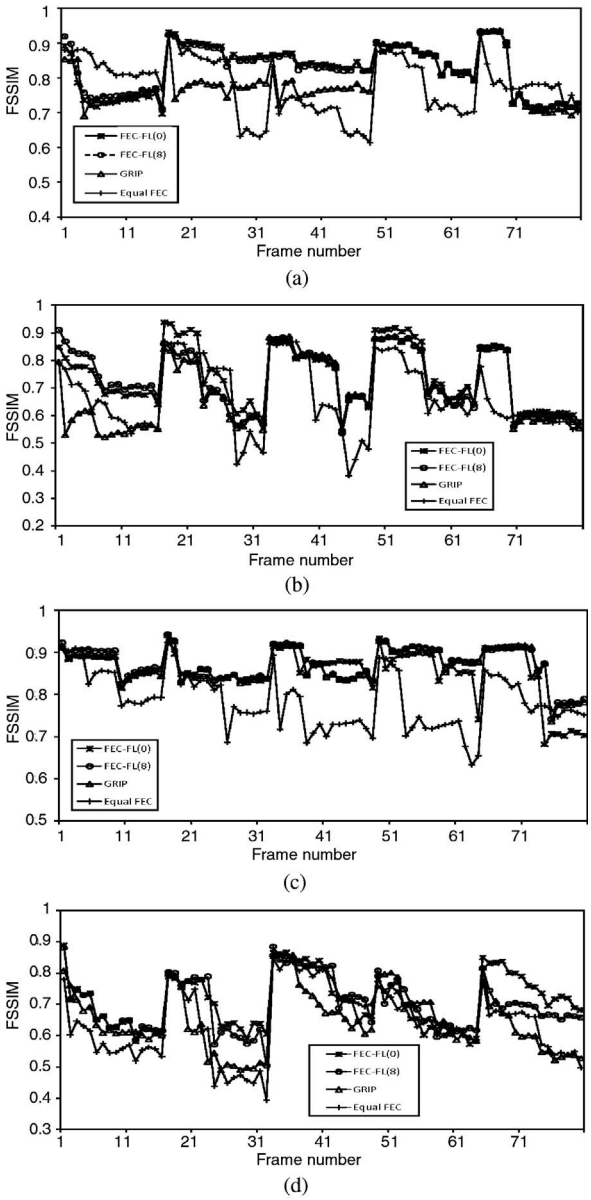

Fig. 11. Performance comparison of frame-by-frame FSSIM between PULP and conventional ULP algorithms for a FEC ratio of 5%–15% in *City*. (a) FEC ratio=15% and Pb=10%. (b) FEC ratio=15% and Pb=20%. (c) FEC ratio=5% and Pb=10%. (d) FEC ratio=5% and Pb=20%.

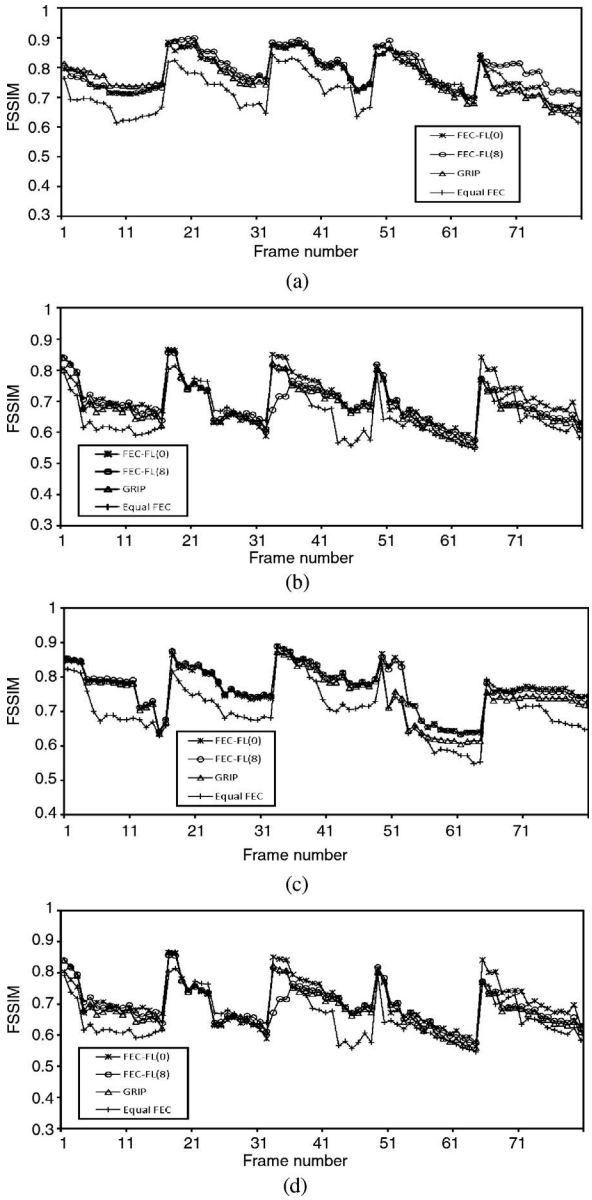

Fig. 12. Performance comparison of frame-by-frame FSSIM between PULP and conventional ULP algorithms for a FEC ratio of 5%–15% in *Soccer*. (a) FEC ratio=15% and Pb=10%. (b) FEC ratio=15% and Pb=20%. (c) FEC ratio=5% and Pb=10%. (d) FEC ratio=5% and Pb=20%.

TABLE II
SIMULATION PARAMETERS

Sequence Name	*Stefan*	*City*	*Soccer*	*Silent*
FEC ratio (%)	5 and 15	5 and 15	5 and 10	5 and 10
L_B	2	2	1	3
P_B (%)	from 5 to 20			
K in $RS(N, K)$	16			

over the 20 different random channel loss patterns, and their results averaged.

The foveal weighting model was configured using a block size of 4×4 to evaluate $\hat{w}_c^{n,k}$. The parameter ε in PULP was set to 1.0. The simulation results were analyzed from two perspectives: objective perceptual quality and subjective perceptual quality.

A. Objective Visual Quality Evaluation

We evaluated PULP algorithm over different channel states. Two fairness levels l = 0 and 8 were considered for high and low-packet loss rates. Due to the randomness of such a channel, 100 different runs of the simulation were conducted using different packet loss rates ranging from 5% to 20%. We investigated how well PULP adapted to channel variations as compared to other ULP schemes.

1) *FEC-FL(0):* Using FL(0) and PWEP-FL(0), the FEC assignment was performed for pure efficiency.
2) *FEC-FL(8):* Using FL(8) and PWEP-FL(8), the FEC assignment was performed to achieve intermediate performance between efficiency and fairness.

(a) (b) (c) (d) (e)

Fig. 13. Subjective quality comparison on the 15th frame of the *Stefan* test video clip (P_B = 5%). (a) Original video clip. (b) FEC-FL(0). (c) FEC-FL(8). (d) GRIP. (e) Equal FEC.

3) *GOP and Resynchronization Integrated Protection (GRIP):* The ULP scheme in [15] was performed using the length of error propagation as the performance metric, without using perceptual weights in each video packet. The resynchronization weighting scheme was not considered.
4) *Equal FEC:* The ULP scheme allocates FEC packets without considering either the packet loss rate or the perceptual significance of error propagation. If FEC-FL(9) runs ULP without considering the packet loss rate in the FEC assignment, then this scheme becomes the same as the equal FEC.

To evaluate the quality of foveated video, the so-called foveal-peak signal-to-noise ratio (PSNR) was developed in [7]. The foveal-PSNR is defined by weighting the LSB relative to the mean square error (MSE). This performance metric has been demonstrated to be a good objective quality measurement tool for predicting subjective quality of foveated images. Here, we introduce foveal-SSIM (FSSIM) in a manner similar to foveal-PSNR, but replacing the MSE with SSIM [29]. To compute FSSIM on each frame, the quantized LSB $\hat{w}_c^{n,k}$ of the nth frame is applied to the SSIM index similar to [29]

$$FSSIM_n = \frac{\sum_{k=1}^{M} SSIM(o_{n,k}, d_{n,k}) \cdot \hat{w}_c^{n,k}}{\sum_{k=1}^{M} \hat{w}_c^{n,k}} \tag{20}$$

where M is the number of MBs in a frame, and $o_{n,k}$ and $d_{n,k}$ are the kth matched MBs of the nth original and distorted frames.

The LSB varies for each window, frame and sequence over the spatial and temporal axes. Therefore, the final score over the video is obtained using the simple pooling

$$FSSIM = \frac{\sum_{n=1}^{T} FSSIM_n \cdot \hat{w}_{\text{sum}}^{n}}{\sum_{n}^{T} \hat{w}_{\text{sum}}^{n}} \tag{21}$$

where T is the number of frames in the sequence, and $\hat{w}_{\text{sum}}^{n} = \sum_{n=1}^{T} \hat{w}_c^{n,k}$.

In addition to conducting a perceptually relevant examination using the FSSIM index, we also use the traditional (but perceptually questionable) average PSNR to evaluate the error. Figs. 9 and 10 compare the average FSSIM and PSNR values for the conventional ULP and PULP algorithms, using the FEC ratios of 5 % and 15 %. At the low-packet loss rate, the FEC-FL(8) scheme exhibits higher FSSIM values than do FEC-FL(0), GRIP and Equal FEC by 0.01–0.05 over all the test video frames. Relative to rate, a fair FEC assignment was performed for each BOP by enlarging the size of the salient region, leading to improved objective quality. Conversely, at the high-packet loss rate, the FEC-FL(0) scheme achieved graceful degradation as measured by FSSIM in the range P_B = 15%–20%, while the GRIP and Equal FEC schemes resulted in steep degradation as measured by FSSIM. The gradual decrease of FSSIM in the proposed FEC scheme is due to the better protection from the packet loss, of those video packets having a large impact on visual quality, by maintaining a small salient region. On the other hand, the PSNR comparison shows that the GRIP scheme without visual weighting delivers higher PSNR values than does the proposed PULP algorithm. In particular, at a high-packet loss rate, since the smallest size of the salient regions is set by the proposed algorithm, the difference in PSNR values is largest within the feasible range of the packet loss rate. However, the subjective visual quality comparison in the next subsection makes it clear that the proposed FEC algorithm yields better visual quality than does the conventional FEC algorithm.

Figs. 11 and 12 plot the frame-by-frame FSSIM using the FEC assignment scheme on the first 80 frames of the two test sequences, "*City*" and "*Soccer*" using FEC ratios of 5% and 15%. For each packet loss rate, it can be seen that the salient regions are better protected by FEC-FL(0) and FEC-FL(8) than those by the other protection schemes. Thus, spatial and temporal error propagation can be effectively alleviated using the proposed FEC assignment algorithm.

B. Subjective Quality Comparison

To conduct subjective quality comparisons, we utilized a video test clip reconstructed using the benchmark methods with average packet loss rates of 5% and 20%, respectively. Fig. 13 shows the 15th frame of the "*Stefan*" test video clip. Fig. 13(a) is the original video clip, while Fig. 13(b)–(e) are the reconstructed clips using FEC-FL(0), FEC-FL(8), GRIP, and Equal FEC, respectively, at a packet loss rate of P_B = 5%. It may be observed that improved subjective quality was delivered by the FEC-FL(8) scheme as compared to the GRIP, Equal FEC or FEC-FL(0) schemes. Thus, it may be deduced that fairness is more important than efficiency toward improving subjective video quality, by maintaining a wider range of salient regions given a low-packet loss rate.

Fig. 14 depicts the subjective quality comparison at the high-packet loss rate of P_B = 20% using the 15th frame of the "*Soccer*" test video clip. Since FEC-FL(8) allocates FEC codes to video packets having a wide range of sizes of the salient region, noticeable degradations of subjective quality occur

(a) (b) (c) (d) (e)

Fig. 14. Subjective quality comparison on the 15th frame of the *Soccer* test video clip(P_B = 20%). (a) Original video clip. (b) FEC-FL(0). (c) FEC-FL(8). (d) GRIP. (e) Equal FEC.

in the reconstructed image. It is noticeable that FEC-FL(0) improves the subjective video quality more effectively than the fairness scheme at the high-packet loss rate. In the GRIP scheme, evident perceptual degradation occurs in the middle of the frame, owing to the lack of FEC codes. Conversely, FEC-FL(0) effectively inhibits perceptual degradations in areas of identified perceptual importance, by allocating channel coding resource to those video packets.

VII. Conclusion

We proposed a new PULP algorithm appropriate for operation in a packet erasure network. To enable adaptation to the nonuniform resolution of the visual photoreceptors, we developed a simple and efficient performance metric, called PWEP. The proposed PULP scheme was developed based on two essential objectives, namely, enforcing efficiency and fairness across various channel states to improve visual quality. To mediate the tradeoff between efficiency and fairness, we proposed a FEC algorithm with a variable fairness level, FEC-FL(l) to allocate resources in order to manage the FEC codes for each video packet. The simulation results show that PULP algorithm achieves higher foveal-SSIM scores than conventional algorithms. It was demonstrated that PULP adapts well to dynamic channel environments, yielding good control of QoS, which is vital for achieving high quality and reliable video communication.

References

[1] M. R. Civanlar, A. Luthra, S. Wenger, and W. Zhu, "Introduction to the special issue on sreaming video," *IEEE Trans. Circuits Syst. Video Tech.*, vol. 11, no. 3, pp. 265–268, Mar. 2001.

[2] Y. Wang and Q. Zhu, "Error control and concealment for video communication: A review," *Proc. IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.

[3] B. Ciptpraset and K. R. Rao, "Human visual weighting progressive image transmission," in *Proc. ICCS*, Nov. 1987, pp. 1040–1044.

[4] B. Girod, "What's wrong with mean-square error?" in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: M.I.T. Univ. Press, 1993.

[5] A. B. Watson, J. Hu, and J. F. McGowan, III, "DVQ: A digital video quality metric based on human vision," *J. Electron. Imag.*, vol. 10, no. 1, pp. 1164–1175, Aug. 1997.

[6] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video quality assessment," *IEEE Trans. Multimedia*, vol. 4, no. 1, pp. 129–132, Mar. 2002.

[7] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video compression with optimal rate control," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 977–992, Jul. 2001.

[8] Z. Wang and A. C. Bovik, "Embedded foveation image coding," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1397–1410, Oct. 2001.

[9] Z. Wang, L. Lu, and A. C. Bovik, "Foveation scalable video coding with automatic fixation selection," *IEEE Trans. Image Process.*, vol. 12, no. 2, pp. 243–254, Feb. 2003.

[10] C. Tang, "Spatiotemporal visual consideration for video coding," *IEEE Trans. Multimedia*, vol. 9, no. 2, pp. 231–238, Feb. 2007.

[11] J. Chen, J. Zheng, and Y. He, "Macroblock-level adaptive frequency weighting for perceptual video coding," *IEEE Trans. Consumer Elec.*, vol. 53, no. 2, pp. 775–781, May 2007.

[12] K. Yang, C. C. Guest, K. E. Maleh, and P. K. Das, "Perceptual temporal quality metric for compressed video," *IEEE Trans. Multimedia*, vol. 9, no. 7, pp. 1528–1535, Nov. 2007.

[13] P. Frossard and O. Verscheure, "AMISP: A complete content-based error resilient scheme," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 11, no. 9, pp. 989–998, Sep. 2001.

[14] A. E. Mohr, E. A. Riskin, and R. E. Ladner, "Unequal loss protection: Graceful degradation of image quality over packet erasure channels through forward correction," *IEEE J. Select. Areas Commun.*, vol. 18, no. 6, pp. 819–828, Jun. 2000.

[15] X. Yang, C. Zhu, Z. G. Li, X. Lin, and N. Ling, "An unequal packet loss resilience scheme for video over the internet," *IEEE Trans. Multimedia*, vol. 7, no. 4, pp. 753–765, Aug. 2005.

[16] J. Albanese, J. Edmonds, M. Ludy, and U. Horn, "Peioiry encoding transmission," *IEEE Trans. Inform. Theory*, vol. 42, no. 6, pp. 1737–1744, Nov. 1996.

[17] F. Hartanto and H. R. Sirisena, "Hybrid error control mechanism for video transmission in the wireless IP networks," in *Proc. Select. Paper 10th IEEE Workshop Local Metropolitan Area Netw.*, 2001, pp. 126–132.

[18] M. Andronico and O. Verscheure, "Performance analysis of priority encoding transmission of MPEG video streams," in *Proc. Global Telecommun. Conf.*, Nov. 1996, pp. 267–271.

[19] J. Kim, R. M. Mercereau, and Y. Altunbasak, "Error-resilient image and video transmission over the internet using unequal error protection," *IEEE Trans. Image Process.*, vol. 12, no. 2, pp. 121–131, Feb. 2003.

[20] K. Stuhlmuller, M. Link, and B. Girod, "Robust internet video transmission based on scalable coding and unequal error protection," *Signal Process. Image Comm.*, vol. 15, no. 1, pp. 99–94, 1999.

[21] L. Zhuo, K.-M. Lam, and L. Shen, "Channel-adaptive error protection for streaming stored MPEG-4 FGS over error-prone environments," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 16, no. 5, pp. 649–654, May 2006.

[22] S. Lee, C. Podilchuk, V. Krishnan, and A. C. Bovik, "Foveation-based error resilience and unequal error protection over mobile networks," *Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology*, vol. 34, pp. 149–166, May 2003 (special issues on multimedia communications).

[23] L. Ning, Q. Zi-Yi, and M. Ai-Dong, "Perceptual optimized error-resilient H.264 video streaming system over the best-effort internet," in *Proc. PDCAT*, Dec. 2005, pp. 1039–1043.

[24] K. Stuhlmuller, N. Farber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE J. Select. Areas Commun.*, vol. 18, no. 6, pp. 1012–1032, Jun. 2000.

[25] X. K. Yang, C. Zhu, Z. G. Li, X. Lin, G. N. Feng, S. Wu, and N. Ling, "Unequal loss protection for robust transmission of motion compensated video over the internet," *Signal Process.: Image Commun.*, vol. 18, no. 3, pp. 157–167, Mar. 2003.

[26] N. Feamster and H. Balakrishnan, "Packet loss recovery for streaming video," in *Proc. IEEE Int. Packet Video Workshop*, Apr. 2002, pp. 1–11.

[27] Z. Wang, A. C. Bovik, and L. Lu, "Why is image quality assessment so difficult?" in *Proc. IEEE Int. Conf. Acoust. Speech Signal Proc.*, vol. 4. May 2002, pp. 3313–3316.

[28] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*. San Mateo, CA: Morgan and Claypool Publishers, 2006.

[29] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structual distortion measurement," *Signal Process. Image Commun.*, vol. 19, no. 2, pp. 121–132, Feb. 2004.
[30] H. Zheng, C. Ru, C. W. Chen, and L. Yu, "Video transmission over MIMO-OFDM system: MDC and space-time coding-based approaches," *Advances Multimedia*, vol. 2007, article 61491, pp. 1–8, 2007.
[31] C. H. Kuo, C. S. Kim, and C. C. J. Kuo, "Robust video transmission over wideband wireless channel using space-time coded OFDM systems," in *Proc. WCNC*, vol. 2. Mar. 2002, pp. 931–936.
[32] S. Wenger, Y.-K. Wang, and T. Schierl, "Transport and signaling of SVC in IP networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1164–1173, Sep. 2007.
[33] S. W. Yuk, M. G. Kang, B. C. Shin, and D. H. Cho, "An adaptive redundancy control method for erasure-code-based real-time data transmission over the internet," *IEEE Trans. Multimedia*, vol. 3, no. 3, pp. 366–374, Sep. 2001.
[34] Z. G. Li, C. Zhu, N. Ling, X. K. Yang, G. N. Feng, S. Wu, and F. Pan, "A unified architecture for real-time video-coding system," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 13, no. 6, pp. 472–486, Jun. 2003.
[35] Y. Ming and W. H. Yuan, "A rate control scheme for H.264 video under low-bandwidth channel," *J. Zhejiang Univ.-Sci. A*, vol. 7, no. 6, pp. 990–995, 2006.
[36] Y. Liu, Z. G. Li, and Y. C. Soh, "A novel rate control scheme for low-delay video communication of H.264/AVC standard," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 17, no. 1, pp. 68–78, Jan. 2007.
[37] J. Lee, "Rate-distortion optimization of parameterized quantization matrix for MPEG-2 encoding," in *Proc. ICIP*, Oct. 1998, pp. 383–386.
[38] H. Yin, "Adaptive quanitization in perceptual MPEG video encoders," in *Proc. PCS*, Apr. 2006, pp. 3–14.
[39] B. Girod, K. Stuhlmuller, M. Link, and U. Horn, "Packet loss resilient internet video streaming," in *Proc. SPIE VCIP*, Jan. 1999, pp. 833–844.
[40] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.
[41] W. S. Geisler and J. S. Perry, "A real-time foveated multiresolution system for low-bandwidth video communication," in *Proc. SPIE*, vol. 3299. 1998, pp. 1–13.
[42] J. G. Robson and N. Graham, "Probability summation and regional variation in contrast sensitivity across the visual field," *Vis. Res.*, vol. 21, no. 3, pp. 409–418, 1981.
[43] L. Itti, "Quantifying the contribution of low-level saliency human eye motivement in dynamic scenes," *Vis. Cognit.*, vol. 12, no. 6, pp. 1093–1123, Aug. 2005.
[44] M. S. Banks, A. B. Sekuler, and S. J. Anderson, "Peripheral spatial vision: Limited imposed by optics, photoreceptors and receptor pooling," *J. Opt. Soc. Amer.*, vol. 8, no. 11, pp. 1775–1787, 1991.
[45] *H.264/AVC Software Coordination* [Online]. Available: http://iphome.hhi.de/suehring/tml/index.htm
[46] E. O. Elliott, "A model of the switched telephone network for data communications," *J. Bell. Syst. Tech.*, vol. 44, no. 1, pp. 98–109, 1965.

Hojin Ha received the B.S. degree in control and instrumentation engineering from Myongji University, Yongin, Korea, in 1998, the M.S. degree in control and instrumentation engineering from Hanyang University, Ansan, Korea, in 2000, and the Ph.D. degree in electrical and electronic engineering from Yonsei University, Seoul, Korea, in 2009.

Since 2000, he has been a Research Engineer with the Digital Media and Communications Research and Development Center, Samsung Electronics, Suwon, Korea. His research interests include multimedia communications, multimedia signal processing, and peer-to-peer networking.

Jincheol Park was born in Korea in 1982. He received the B.S. degree in information and electronic engineering from Soongsil University, Seoul, Korea, in 2006, and the M.S. degree in electrical and electronic engineering from Yonsei University, Seoul, Korea, in 2008. He is currently pursuing the the Ph.D. degree with the Wireless Network Laboratory, Yonsei University.

His current research interests include wireless multimedia communications and video quality assessment.

Sanghoon Lee (M'05) was born in Korea in 1966. He received the B.S. degree from Yonsei University, Seoul, Korea, the M.S. degree from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, and the Ph.D. degree from the University of Texas, Austin, all in electronic engineering, in 1989, 1991, and 2000, respectively.

From 1991 to 1996, he was with Korea Telecom, Seocho-gu, Seoul, Korea. In 1999, he was with Bell Laboratory, Lucent Technologies, Murray Hill, NJ, and worked on wireless multimedia communications. From 2000 to 2002, he worked on developing real-time embedded software and communication protocols for 3G wireless networks with Lucent Technologies. Since 2003, he has been with the faculty of the Department of Electrical and Electronics Engineering, Yonsei University, where he is currently an Associate Professor. His current research interests include 4G wireless networks, 3G W-CDMA/CDMA networks, multihop sensor networks, wireless multimedia communications, and image/video quality assessments.

Dr. Lee is an Associate Editor of *Journal of Communications and Networks* and IEEE TRANSACTIONS ON IMAGE PROCESSING.

Alan Conrad Bovik (S'80–M'81–SM'89–F'96) received the B.S., M.S., and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Champaign, in 1980, 1982, and 1984, respectively.

He is currently the Curry/Cullen Trust Endowed Professor with the University of Texas, Austin, where he is the Director of the Laboratory for Image and Video Engineering Center for Perceptual Systems. He has published over 450 technical articles in his fields and holds two U.S. patents. His current research interests include image and video processing, computational vision, digital microscopy, and modeling of biological visual perception.

Dr. Bovik is the author of the *Handbook of Image and Video Processing* (Amsterdam, The Netherlands: Elsevier, 2005, 2nd ed.) and *Modern Image Quality Assessment* (San Mateo, CA: Morgan and Claypool, 2006). He has received a number of major awards from the IEEE Signal Processing Society, including the Education Award in 2007, Technical Achievement Award in 2005, Distinguished Lecturer Award in 2000, and Meritorious Service Award in 1998. He is also a recipient of the Distinguished Alumni Award from the University of Illinois at Urbana-Champaign in 2008, IEEE Third Millennium Medal in 2000, and two journal paper awards from the International Pattern Recognition Society in 1988 and 1993. He is a Fellow of the Optical Society of America, Society of Photo-Optical Instrumentation Engineers. He has held positions in numerous professional society activities, including the Board of Governors of the IEEE Signal Processing Society from 1996 to 1998, the Editor-in-Chief of the IEEE TRANSACTIONS ON IMAGE PROCESSING from 1996 to 2002, an Editorial Board Member of PROCEEDINGS OF THE IEEE from 1998 to 2004, the Series Editor for Image, Video, and Multimedia Processing, Morgan and Claypool Publishing Company from 2003 to present, and the Founding General Chairman of the First IEEE International Conference on Image Processing, Austin, in 1994. He is a registered Professional Engineer in the state of Texas and is a frequent consultant to legal, industrial, and academic institutions.