Foveation-based Error Resilience and Unequal Error Protection over Mobile Networks

Sanghoon Lee, Chris Podilchuk, Vidhya Krishnan and Alan C. Bovik, *Fellow, IEEE*

Abstract

By exploiting new human-machine interface techniques, such as visual eyetrackers, it should be possible to develop more efficient visual multimedia services associated with low bandwidth, dynamic channel adaptation and robust visual data transmission. In this paper, we introduce foveation-based error resilience and unequal error protection techniques over highly error-prone mobile networks. Each frame is spatially divided into foveated and background layers according to perceptual importance. Perceptual importance is determined either through an eye tracker or by manually selecting a region of interest. We attempt to improve reconstructed visual quality by maintaining the high visual source throughput of the foveated layer using foveation-based error resilience and error correction using a combination of turbo codes and ARQ (automatic reQuest). In order to alleviate the degradation of visual quality, a foveation based bitstream partitioning is developed. In an effort to increase the source throughput of the foveated layer, we develop unequal delay-constrained ARQ (automatic reQuest) and rate compatible punctured turbo codes where the punctual pattern of RCPC (rate compatible punctured convolutional) codes in H.223 Annex C is used. In the simulation, the visual quality is significantly increased in the area of interest using foveation-based error resilience and unequal error protection; (as much as 3 dB FPSNR (foveal peak signal to noise ratio) improvement) at 40 % packet error rate. Over real-fading statistics measured in the downtown area of Austin, Texas, the visual quality is increased up to 1.5 dB in PSNR and 1.8 dB in FPSNR at a channel SNR of 5 dB.

This research was supported in part by Lucent Technologies, and by the Texas Advanced Technology Program.

Corresponding author: Professor Alan C. Bovik, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084 USA; Phone: (512) 471-5370; Fax: (512) 471-1225; E-mail: bovik@ece.utexas.edu.

Keywords: foveated video, error resilience, unequal error protection, wireless video, internet video.

I. Introduction

The ongoing explosion in wireless communications services has resulted in an increased interest and major research effort to adapt and extend multimedia applications and services into the wireless telecommunications domain. The next generation wireless services include broadcasting, multimedia messaging, videoconferencing, internet access including streaming video, security monitoring, navigation of remote vehicles through remote wireless visual interfaces and transmission of high-resolution medical images/video from in-transit emergency vehicles. In order to guarantee high quality visual services over constrained bandwidth and lossy channel conditions due to fading, it is necessary to develop more robust source and channel coding techniques.

In mobile channels, the bit error rate (BER) varies according to fading signal attenuation and it can be greater than 10^{-3} for several seconds. A key technical challenge of video transmission is to minimize perceptual (visual) errors at the receiver or maximize perceived image/video quality through effective source/channel coding design. In the H.263+, H.263++ and MPEG-4 standards, the following error resilience features are included : reference picture selection mode, independent segment decoding mode, data partitioning and RVLC (reversible variable length coding) [1][2][3][4]. Using data partitioning, the bitstream is divided into different layers or rearranged according to the syntax. Data partitioning attempts to divide the bitstream into different priorities according to the significance of the data in reconstructing the original video. Network protocols and appropriate channel coding algorithms can then be applied according to the priority of the different data streams. Usually, the error resilience schemes do not consider the gazing direction of the human eye. Error resilient features are based on maximizing the quality of the entire reconstructed image, without any special treatment for the area of interest. Thus, the objective quality criterion signal-to-noise ratio (SNR) has been widely used to measure the performance of error resilience schemes. However, since the effectiveness of a source/channel coding system is ultimately judged by the human observer at the receiver, it is advantagious to take advantage of the fact that the human eye senses visual information at varying resolutions, with the finest resolution at the fixation point. Thus, for point-to-point visual communications, the gaze direction of the end-user or selected region of interest can be used to improve the visual quality of the transmitted video. If the focused region is protected more effectively, the perceived visual quality can be greatly improved over highly error prone networks.

By taking advantage of the foveation property of the human visual system (HVS), improved video quality is possible using foveation-based compression methods [5][6]. We have recently introduced and systematically developed foveated video that results in increased compression performance [6], adaptation to human visual fixation and visual resolution [7], robust transmission characteristics [8][9][10], and efficient motion estimation/compensation [11].

In this paper, we introduce foveation-based error resilience and unequal error protection techniques over mobile networks [8][9]. For the source coding, error resilience features supported by H.263++ / MPEG-4 are included [2]. In addition, a *foveation-based bitstream partitioning* is developed, where each frame is spatially divided into foveated and background layers. Foveation-based data partitioning is based on spatial adaptation of video priority as opposed to the conventional data partitioning techniques which are based on the importance of the compressed bitstream in reconstructing the entire image. Since both layers are spatially independent, each layer can get channel protection according to the importance of each partition. For unequal error protection, we introduce two different types of transmission protocols : delay-constrained ARQ and hybrid delay-constrained ARQ (ARQ + FEC (forward error correction)). In these schemes, we assign unequal encoder queuing delay bounds to the layers for increasing the source throughput of the perceptually important video stream (foveated layer). In particular, rate compatible punctured turbo code is developed as the FEC method. We obtain varying code rates using the same encoder by employing the puncture patterns specified in H.223 Annex C [12]. This enables us to apply this coding technique to foveated video to obtain maximum source throughput.

II. Foveated visual communications

A. Foveation

In order to coincide with the sampling characteristics of the human eye, we propose that picture quality be assessed based on the distortion between the original image and the reconstructed image formed by the nonuniform sampling of the retinal neurons, instead of the image displayed on the monitor.

Several psychophysical studies have been aimed in this direction, using simple image presentations (e.g. sinusoidal gratings) [13], and expressing image quality in terms of the display parameters: resolution, contrast, luminance, display size, viewing distance and retinal eccentricity [14]. Assuming a linear system approximation for the response of visual neurons [15][16], a contrast-sensitivity function can be obtained, and subsequently, local image frequencies can be derived.

A foveated image may be interpreted as having position-varying local bandwidths. Once local spatial frequency information has been obtained relative to a foveation point, a variable-resolution image having the same information content as a foveated image can be obtained by applying a *foveation filter*, which consists of a bank of low pass filters having variable cut-off frequencies. There has been some research directed towards analyzing this spatially-varying information content [17][18]. One approach for analyzing such spatio-spectral signals is via the AM-FM transform [19], which generalizes the spatio-spectral representation across curvilinear coordinate mappings. In this approach, a foveated image can be mapped into an image that has been uniformly sampled, allowing analysis to proceed without considering the superimposed variable local bandwidth.

B. Foveated visual processing for multimedia applications

The recent research advances of foveation-based image and video compression [20][14][21][22] has coincided with recent improvements in the methods used for determining and tracking human fixation points. We expect that this synergy will lead to future foveated multimedia services, with applications in videoconferencing, video broadcasting and streaming, remote surveillance, internet news, and others. Even without eye tracking technology, foveation–based techniques are useful in appplications where an area of interest can be identified, either through manual selection or with object recognition algorithms such as face detection.

There are various mechanical and algorithmic methods that can be used to select fixation / foveation points in an image or video sequence. Simple interactive methods that are very effective include the use of a mouse or a touch screen. For images or video that do not change rapidly, determining the fixation point using these devices can be easily learned, although they are limited in more generic applications. In controlled applications containing objects of known characteristics, such as faces, the foveation point or region can be automatically selected and tracked by detecting and recognizing the facial shape, color, or motion characteristics[22]. Recently, very sophisticated mechanical eyetrackers have become commercially available that accurately track the direction of gaze of a human observer by detecting the motion of the eye through IR reflection (by detecting the pupil). These devices are effective in situations where the user is located in front of a terminal or other display device and within a prescribed expected physical location. One such successful system is described in [14], where a real-time visual communication system is demonstrated using an eyetracker device.

Fig. 1 depicts a simplified example for foveated visual communications over a mobile channel. Using an eye tracker, the gazing direction can be traced in real-time over end-to-end visual communications system. The foveation points are transmitted to the service-provider or the corresponding subscriber. After receiving the foveation points, foveation filtering is applied to the video sequence to remove irrelevant spatial redundancy and create a foveated video sequence. Fig 2 (a) shows the original image where the foveation point is indicated by “x” on the right eye; (b) is the foveated image after removing high frequency components from the periphery; (c) is the local bandwidths used to obtain the foveated image; (d) is the uniform version of the foveated image after mapping it into new coordinates system associated with the local bandwidths.

Newer models of eyetrackers have also been integrated into virtual reality goggles allowing for highly precise tracking. In addition, sophisticated tracking algorithms exist that can be used to augment the placement of foveation points from frame to frame. Other information, such as the audio stream, can also be used to determine the area of interest in a foveation–based system. These developments collectively make a compelling case that foveated image/video transmission and display systems for multimedia applications are worthy of investigation. In a foveated multimedia video communication system, the number of foveation points can vary. It is possible to have a single foveation point, multiple foveation points, or object-based foveation regions, such as the human face in videoconferencing, automobiles in a race or athletes in a sporting event.

Foveated visual communication systems can be divided into receiver driven, sender driven, and both sender/receiver driven types depending on who selects the foveation points. The eyetracker is one example of receiver driven foveation. Another simple human interactive receiver driven example is the manual selection of an object of interest using a mouse or touch screen. These receiver-driven foveation selection techniques can be effectively used for remote navigation, surveillance, or virtual reality training. Sender-driven foveation selection techniques include using a focused region indicator on a video camcorder to create an arbitrary foveation on the sequence as determined by the person shooting the video. Fig. 4 (a) shows an example when the region indicator is set to the rectangular region and (b) is the foveated version of the original image associated with the foveation pattern. Over mobile-to-home internet, the video bitstream can be delivered and stored in the computer at home. For video conferencing systems over mobile-to-home/home-to-home internet, the face can be assumed to be the gaze direction of the corresponding receiver and automatically tracked [22]. In video broadcasting, the foveation points or foveation region can be chosen by the service provider. A good example of a stored application that could greatly benefit by this is internet news, where the foveation points can be decided prior to broadcast.

C. Foveated visual quality criterion

Existing algorithms for the automated objective assessment of video quality have generally not taken into account the gazing patterns of observers, viz., the video quality is uniformly assessed over space. The most commonly-applied objective error measurements have been the mean square error (MSE) and the mean absolute error (MAE). Both of these have the advantage of simplicity and feasible real-time video implementation. However, both the MSE and the MAE are not highly correlated with subjective quality measurements.

In order to capture the spatially-varying response of the HVS (human visual system), a foveal weighting metric f_n^2 has been introduced [5][6]. The FMSE (*foveal mean square error*) was defined as:

$$\mathrm{FMSE} = \frac{1}{\sum_{n=1}^{N_p} f_n^2} \sum_{n=1}^{N_p} \left[a(\mathbf{x}_n) - b(\mathbf{x}_n)\right]^2 f_n^2, \tag{1}$$

and the FPSNR (*foveal peak signal-to-noise ratio*) as

$$\mathrm{FPSNR} = 10\,\log_{10} \frac{\max[a(\mathbf{x}_n)]^2}{\mathrm{FMSE}} \tag{2}$$

where N_p is the number of pixels in a picture, f_n is the local bandwidth at the n^{th} point, $a(\mathbf{x}_n)$ is the original image or the foveated image, and $b(\mathbf{x}_n)$ is the coded version of $a(\mathbf{x}_n)$. Using the FMSE, the nonuniform resolution of the CSF of the human visual system can be taken into account.

III. Foveation based error resilience video coding

A. Foveation based bitstream partitioning

The MPEG-4 standard defines two layers (object and background) for temporal scalability. Using a segmentation technique, the object layer is extracted from the original frame, and shape coding is used to compress the object shape. However, the computational complexity is increased due to the segmentation, and additional overhead is needed to represent the shape.

In foveation based temporal scalability, we define the foveated layer and the background layer following the basic structure of MPEG-4. For a given foveation point, the maximum detectable frequency at each pixel can be derived from human visual modeling. The frequency is a function of eccentricity (visual angle) and is converted into a local bandwidth in the discrete domain. After the frequency is converted into the local bandwidth in the discrete domain, the average local bandwidth $\bar{f}_n$ for the n^{th} macroblock can be obtained by

$$\bar{f}_n = \frac{1}{M} \sum_{i \in n} f_i \tag{3}$$

where f_i is the local bandwidth of the i^{th} pixel in the n^{th} macroblock as shown in Fig. 2 (c) and M is the number of pixels in a macroblock. Since $\bar{f}_n$ is a function of eccentricity, the average eccentricity of the n^{th} macroblock is also derived.

According to the average of local bandwidths in each macroblock, a frame can be spatially divided into the foveated layer and the background layer by setting a threshold for the local bandwidth. The foveated layer consists of macroblocks whose average local bandwidth is larger than a given threshold.

Fig. 2 (b) is the foveated image when the visual distance is 30 cm and the horizontal image size is 4.5 cm. In order to divide each frame into two layers, the thresholded local bandwidth is set to 0.35 (cycles/pixel). Thus, the macroblocks corresponding to $f_n > 0.35$ constructs the foveated layer as shown in Fig. 3 (a) and the rest of the macroblocks determine the background layer in Fig. 3 (b).

In this scheme, the encoder is able to determine whether each macroblock belongs to the foveation or background layer based on a specified foveation point. The segmentation is determined on a macroblock-level not pixel-level and can be implemented in real-time. In addition, the side information needed to label each macroblock is significantly less than the number of bits needed to represent the shape information in the conventional scheme.

B. Foveation based temporal scalability

The encoder compresses each layer independently and generates two different video bitstreams using *foveation based bitstream partitioning.* Since the two bitstreams are independent, we can control the temporal resolution for each layer depending on channel condition. The temporal resolution in the human visual system has been found to be proportional to the spatial resolution [23]. Based on the bitstream partitioning, it is possible to maintain higher source throughput and temporal resolution in the foveated layer compared to the background layer, which can prevent visual quality degradation from deep signal fading attenuation over mobile channels. Moreover, the high frequency reduction in the background leads to increasing the temporal resolution so that the relative contrast sensitivity can decrease compared to regular video. Thus, the human vision system perceives less coding distortion in foveated video coding.

C. Independent segment decoding, data partitioning and RVLC

The independent segment decoding and data partitioning proposed in the current standards (H.263+,H.263++,MPEG-4) can be adapted to the foveation-based error resilience [3][24]. Since two layers are spatially independent, the range of motion vectors is restricted to each layer such that no information outside each layer is used for prediction. Thus, the effect of channel errors does not spatially and temporally propagate into the other layer. The data partitioning splits the video bitstream into packets and rearranges bit information to reduce spatial error propagation. In this scheme, the bitstream in each layer is partitioned using a bitstream rearrangement to enable early detection of and recovery from errors [3]. Here, we use the same data partitioning in [24] for each layer. Reversible variable length codes (RVLCs) are used for the data compression of COD, MCBPC and motion differential vectors.

D. Reference region selection

Since the two layers are spatially independent, a previous layer delivered to the decoder without transmission errors can be used as a reference for preventing temporal error propagation. Similar to RPS (reference picture selection) defined in the standards, this mode is called RRS (reference region selection) where each layer is used as a reference instead of using a previous frame or a GOB (group of blocks).

The performance of RPS depends on the round-trip delay. If the round-trip delay increases, the amount of motion compensated errors increases and the coding efficiency decreases. The frame interval between the selected reference frame and the current frame depends on the round-trip delay. If the frame interval is over the number of frames stored in the encoder, it is necessary to send an intra frame or intra GOBs to minimize temporal error propagation.

The round-trip delay can be reduced by detecting channel errors in the data-link layer instead of the application layer (decoder). The bitstream is also layered into two separate bitstreams associated with the region, and a one bit feedback signal can notify the encoder which region is affected by channel errors.

E. Intra updating

For a non-feedback channel, the period of intra updating can be set to be constant depending on the channel condition. When a feedback channel is used, the updating period can be adaptively changed according to the feedback channel SNR.

In foveated video, the number of generated bits can be reduced by using foveation filtering. Thus, the traffic congestion due to the usage of intra frames can be alleviated at the encoder buffer while maintaining equivalent visual quality. In addition, the intra frame can be updated according to layers. Since unessential high frequency components in the background region can be removed by foveation filtering, motion compensated errors are effectively reduced from the background region. Therefore, the error propagation can be reduced by using more intra frames in the region of interest given a fixed, overall bandwidth.

F. Nonuniform synchronization

The local bandwidth is a line mapping ratio where a non-uniform foveated image is mapped into a uniform image over curvilinear coordinates as shown in Fig. 2 (b) and (c). Over the uniform domain, the area is unchanged near the center of the foveation point and decreases from the foveation point toward the periphery. The uniform distribution of synchronization markers over the uniform domain corresponds to the non-uniform distribution over the non-uniform foveated image. In proportion to the perceptual importance of the region, synchronization markers are uniformly distributed over curvilinear coordinates. In other words, we can place each synchronization marker in proportion to the local bandwidth. Let the sum of local bandwidths in a GOB be $\tilde{f}_s = \frac{1}{N_s}\sum_{n=1}^{M} \bar{f}_n$ where N_s is the total number of synchronization markers in a frame, M is the number of macroblocks in a frame and $\bar{f}_n$ is the average local bandwidth for the n^{th} macroblock. The number of macroblocks at the k^{th} GOB is determined by

$$m_k = \text{argmin}[m] \;\; \text{for minimizing} \;\; |\tilde{f}_s - \sum_{n=1}^{m} \bar{f}_n| \tag{4}$$

where the index n=1 means the first macroblock at the k^{th} GOB.

G. The performance of the layered coding in error-free environments

The foveation-based layered coding follows the same rules of standard-based regular coding except for dividing each frame into two macroblock-based regions. Using such layered coding brings two major coding restrictions. One is the range of motion vectors restricted to each layer and the other is the redefinition of GOB(group of blocks) within each spatially divided layer.

However, by applying foveation filtering, high frequency components away from the fixation point are selectively reduced or eliminated in a graded fashion. If the bit rate is maintained, then the reduced overhead can be reallocated to bits in the foveal region surrounding the fixation point, which greatly improves the picture quality. In P and B pictures, motion compensation errors can be also effectively reduced, because of the reduction of high frequencies in motion compensated macroblocks [11]. In addition, the lowpass filtered macroblock is less sensitive to changes in the QP than is the original macroblock since the Lagrange multiplier in the low-pass filtered macroblock also slowly varies according to the QP as compared to the original macroblock [6]. Utilizing such characteristics, a large QP may be used without degrading performance in the background layer.

Therefore, the performance degradation due to the restrictions of the layered coding can be alleviated by foveation filtering. Moreover, the visual performance can be improved by saving a large number of bits used to represent unnecessary high frequency information and reallocating them into visually fixated area. In error-free environments, it is also possible to switch the layered coding to a non-layered coding with foveation filtering. Then, the coding method is the same as the regular standard video coding except for using foveation filtering.

IV. Unequal error protection

A. Channel throughput at packet level transmission

One major difference between wireless and wireline channels is the bit error probability. When the bit error rate (BER) is low as typically found in a wireline channel, (e.g., 10^{-9}), the data loss can be largely ignored. The original video compression schemes were designed for such low BERs. In wireless mobile networks, channel errors can occur consecutively for a few seconds due to large-scale fading. Therefore, the service rate is no longer constant and depends on the channel status. When using an ARQ protocol, the service rate can be measured by counting the number of ACKs. Since the data is transmitted using frames (physical layer frames, not video frames), the channel throughput can be calculated using the frame error probability which is a function of the channel signal-to-noise ratio (SNR), the modulation scheme, the frame size and the method of error correction coding.

Let SNR_o be the average SNR within a cell. Then, the SNR signal $SNR(t)$ corresponding to the received signal $r(t)$ can be obtained by

$$SNR(t) = r_n^2(t) SNR_o \tag{5}$$

where $r_n(t)$ is a normalized version of $r(t)$. Here, the ideal $\pi/4$ QPSK is assumed to be used. Then, the bit error probability $P_b(t)$ becomes

$$P_b(t) = Q(\sqrt{2SNR(t)}) \tag{6}$$

where $Q(.)$ is the complementary error function defined as

$$Q(z) = \frac{1}{2\pi} \int_z^{\infty} e^{-x^2/2} dx. \tag{7}$$

Then, the frame error probability $P_e(t)$ without channel coding becomes

$$P_e(t) = 1 - [1 - P_b(t)]^{L_p} \tag{8}$$

where L_p is the number of bits in a frame. The above equations are based on the assumption that the fading variation is slow compared to the frame transmission time, i.e., the channel SNR during a frame transmission time is a constant. If the frame propagation time is shorter than the frame transmission interval, then the average number of transmitted bits in the frame at time t can be obtained by

$$\bar{R}(t) = [1 - P_e(t)] L_p. \tag{9}$$

B. Unequal delay-constrained ARQ

ARQ has been quite effective in dealing with lossy channel conditions. However, the retransmission of corrupted frames introduces additional delay, which might be unacceptable for real-time conversational or interactive services. The additional delay can cause temporal distortions due to variations in temporal resolution and video-audio synchronization failure (lip-sync problems). For effective real-time video services, the number of retransmission attempts has to be limited within a delay constraint. This is called *delay-constrained ARQ*. In this scheme, frames violating the delay constraint can be dropped from the encoder buffer which results in error propagation in the decoder. However, the reconstructed errors due to the frame loss propagate temporally. Thus, the delay-constrained ARQ must be combined with error resilience schemes to reduce the effect of error propagation.

Unequal delay-constrained ARQ is an example of combining both ARQ and error resilience. Using the foveation-based bitstream partitioning, foveated and background bitstreams are sequentially generated. The bitstreams are mapped into frames at the encoder buffer before transmission. Due to the frame retransmission, the queuing delay of each frame varies according to channel conditions. The encoder counts the queuing delay and drops frames whose queuing delay is over the prespecified delay constraints. Since the bitstream of the foveated layer contains more visual information, the queuing delay bound of the foveated layer should be larger than that of the background layer.

If facial information is transmitted more often than background information over a given frame loss rate, users can receive improved visual quality. Suppose that we assign a 100 msec delay bound for the foveated layer and 50 msec for the background layer. Then, the effective frame loss rate in the foveated layer will be less than the frame loss rate of the background layer because of the unequal delay constraint. The increased source throughput rate in the foveated layer can reduce the impact of frame loss on visual quality in the area of interest.

The detail description is the following. Let T_I be the transmission interval, $d_q[n]$ be the accumulated queuing delay at the n^{th} trial. Depending on the transmission status, $d_q[n]$ is updated by

$$\begin{aligned} d_q[0] &= 0 \\ d_q[n] &= d_q[n-1] + T_I, \quad \text{if the } (n-1)^{th} \text{ transmission is failed} \\ &= d_q[n-1], \quad \text{if the } (n-1)^{th} \text{ transmission is succeeded} \end{aligned}$$

For *delay-constrained ARQ*, the n^{th} frame is dropped when $d_q[n]$ is over a queuing delay bound Q_B. For unequal data protection, the bound is set differently according to the layer of the n^{th} frame. Let Q_B^f and Q_B^b be the delay bound of the foveated and background layers respectively. Then, the n^{th} frame is dropped if the queuing delay is greater than the bound, and $d_q[n]$ is updated by

$$\begin{aligned} d_q^{'}[n] &= d_q[n] - T_I, \quad \text{if } Q_B^f \leq d_q[n] \text{ and the } n^{th} \text{ frame is in the foveated layer} \\ &= d_q[n] - T_I, \quad \text{if } Q_B^b \leq d_q[n] \text{ and the } n^{th} \text{ packet is in the background layer} \end{aligned} \tag{10}$$

where $d_q^{'}[n]$ is the updated version of $d_q[n]$ after the n^{th} frame is dropped.

C. Adaptive channel coding

Let m be the m^{th} frame and $P_e(m)$ be the frame error probability of the m^{th} frame. The average number of transferred bits at the m^{th} frame is obtained by $\bar{R}(m) = [1 - P_e(m)]L_p = \bar{R}_s(m) + \bar{R}_c(m)$ where L_p is the number of bits in a frame and $\bar{R}_s(m)$ $[\bar{R}_c(m)]$ is the average source [channel] throughput at the m^{th} frame.

The optimal rate control problem is to find the state vector $\vec{Q}$ which minimizes the overall distortion $D(\vec{Q})$ subject to the rate constraint $R(\vec{Q}) \leq \bar{R}_s$. By introducing a Lagrange multiplier $\lambda \geq 0$, the constrained problem can be defined and solved. For λ ranging from 0 to ∞, an optimal quantization state vector $\vec{Q}^*$ is obtained which minimizes the Lagrangian cost function $J(\vec{Q}, \lambda)$ $= D(\vec{Q}) + \lambda R(\vec{Q})$ while satisfying the rate constraint. If $\lambda_1 < \lambda_2 < \lambda_3$, $R(\lambda_1) < R(\lambda_2) < R(\lambda_3)$ and $D(\lambda_1) > D(\lambda_2) > D(\lambda_3)$, then the distortion of the reconstructed image can be minimized at the maximum rate $\bar{R}_s$.

Suppose that the channel redundancy can be adapted for every frame. Then, the optimal number of parity-check bits a_m^* for the m^{th} frame is obtained by the following criterion.

Maximum source throughput criterion : The optimal number of parity-check digits a_m^* for the m^{th} frame is obtained by

$$a_m^* = \text{argmin}[\ a_m\ ] \text{ for } \max[\ \bar{R}_s(m)\ ] \text{ subject to } a_m + b_m = L_p \tag{11}$$

where a_m is the number of parity bits and b_m is the number of source bits at the m^{th} frame.

Then, $\bar{R}(m) = [1 - P_e(m, a_m^*)]L_p = \bar{R}_s(m) + \bar{R}_c(m)$ is the average number of transfered bits at the m^{th} frame.

The error correction capability depends on the number of parity bits, the number of shift registers, hard/soft decision and channel correlation. As the number of shift registers increases, the error correction capability also increases at the cost of computation overhead for decoding. The channel correlation depends on the mobile velocity or on multipath fading. In order to measure the error correction performance in a fading channel, Jake's model is used to measure the error correction capability according to mobile speed. For a given average mobile speed, a pair of source and channel bits is chosen based on the maximum source throughput criterion.

The channel redundancy can be adaptively changed with RCPC where one time convolutional code is punctured to get different rates. In H.223 Annex C, the source bits are coded using the rate 1/4 mother code and punctured to obtain the desired rate. Fig. 5 (a) shows the BER vs. the SNR over 5 different code rates of 8/8, 8/12, 8/16, 8/24 and 8/32. According to the maximum source throughput criterion, a channel parity-check digits a_m^* is chosen by (11) when $P_e(m)$ is given by

$$P_e(m) = 1 - [1 - P_b(m)]^{L_p}. \tag{12}$$

Fig. 5 (b) shows the performance parameters associated with a_m^* such as the normal source throughput $\bar{R}_s^*(m)/L_p$ and the optimal code rate in the range of 8/8 - 8/32 in the RCPC codes.

A combination of ARQ and FEC is called *hybrid ARQ*. Since the delay-constrained ARQ scheme takes into account the unequal video data protection, it is not necessary to use different channel redundancy for each layer. In the algorithm, each frame is protected according to its priority by assigning different queuing delay bounds, not different channel coding rates.

D. Rate compatible punctured turbo codes

Turbo codes introduced in [25] are in essence parallel concatenated convolutional codes. The encoder consists of two recursive systematic binary convolutional encoders. The encoders are identical and they both have M memory elements. The first encoder operates on the data directly while the second encoder operates on the interleaved data.

The decoder uses an iterative, suboptimal soft decoding rule where each constituent RSC (recursive systematic code) is decoded separately. They share the bit likelihood information in an iterative fashion. They use the BCJR algorithm [26] in which the MAP (maximum a-posteriori) algorithm is used. The logarithm of the LIR (Likelihood ratio) output by one decoder is converted to a priori probability to be used by the other decoder. The goal of the MAP algorithm is to obtain the log of the ratio of the APP (*a posteriori probability*) associated with each information bit.

Normally, the error correcting coding consists of selecting a code with a specific rate and correction capability to match the protection requirement of all the data to be transmitted. A number of times, as in the case of a wireless channel, the data transmitted requires different protection requirements, as the channel characteristics are time varying. Thus, a flexible encoding and decoding scheme is extremely useful in such cases. The rate compatibility restriction ensures that all the code bits of a high rate code are used by a lower rate code.

To design the turbo encoder, we used two RSC coders of memory length 4. A schematic representation of the same is shown in Fig. 6 (a). In this case, we obtain an output of 4 bits from each encoder i.e. the rate of the individual encoders is 1/4. The turbo encoder which is formed by the parallel concatenation of two such RSC encoders has an unpunctured rate of 1/7 where we send the systematic bit and 3 parity bits corresponding to the non-interleaved data bits and 3 parity bits corresponding to the interleaved data bits. A schematic representation of such a decoder is shown in Fig. 6 (b). A random interleaver is used to interleave the data bits.

In particular, we used the following values for the generator polynomials to generate a rate 1/7 turbo encoder. The generator polynomials used are $G0 = 10011$, $G1 = 11011$, $G2 = 10101$, $G3 = 11111$. To adaptively change the number of parity bits depending on the relative importance of the source bits, we use the puncture patterns specified in Annex C of H.223. While using the puncture patterns, the parity bits corresponding to the interleaved data bits and the non-interleaved data bits are alternately sent. This ensures that both the decoders receive a fair amount of correct channel bits and thus the iterative decoding scheme followed yield good performance. We use 8 iterations for the decoding process.

Various rates of turbo codes given by $8/8 \ldots 8/32$ were implemented using the puncture pattern specified in H.223 Annex C. When the frame length used is 640 bits and the channel signaling mode is BPSK, as expected, the performance of the turbo codes significantly improves with decreasing code rate. We get a BER of 10^{-5} at a SNR of around $4.5dB$ for a code rate of 8/24 which is very close to the simulation result specified in [27] where a BER of 10^{-5} was achieved at a SNR of approximately $4dB$ with a helical interleaver. For a code rate of 1/4, we get a BER of 10^{-5} at $3.5dB$. This, as expected is better than that in [27] as we use a lower rate.

As expected, we notice that the BER for a rate 8/32 code is significantly lesser than that for a rate 8/8 code. We consider the performance of turbo codes by performing the iterative process 8 times. This causes some amount of delay when compared to regular rate punctured convolutional codes. However, we notice that the performance of the turbo codes is significantly better than that of the RCPC codes and thus we observe that inspite of the latency involved in the iterative decoding process, the turbo codes are more effective than RCPC codes. Fig. 7 shows the performance comparison between the turbo vs. RCPC codes. The normalized source throughput of the turbo code is relatively higher than that of the RCPC code at low channel SNR as the channel SNR is decreasing.

E. Target bit estimation based on channel signal decomposition

In mobile channels, the received signal $r(t)$ can be decomposed into large-scale fading $l(t)$ and Rayleigh fading $s(t)$ and expressed as $r(t) = l(t) \times s(t)$. The signal $l(t)$ can be extracted by averaging the local power of $r(t)$ [28][29].

Let $r(m)[l(m)]$ be the instantaneously fading signal of $r(t)[l(t)]$ at the m^{th} frame, and $\mathbf{r}(m)[\mathbf{l}(m)]$ be the power of $r(m)[l(m)]$, respectively, measured in decibels (dB). Let the channel statistics be the same during the N_f sampling period which corresponds to the time interval 20 λ to 40 λ. In [28], the large-scale fading signal $\mathbf{l}(m)$ is obtained by :

$$\mathbf{l}(m) = \frac{1}{N_f} \sum_{i=-N_f/2}^{N_f/2-1} \mathbf{r}(m+i). \tag{13}$$

In a real system, we only know the power of previous signals $\mathbf{r}(m)$ for $-N_f/2 \leq m \leq -1$. Therefore, $\mathbf{l}(m)$ is approximately obtained by :

$$\mathbf{l}(m) \approx \frac{2}{N_f} \sum_{i=-N_f/2}^{-1} \mathbf{r}(m+i). \tag{14}$$

Based on the value of $\mathbf{l}(m)$, a pair of source and channel codes is determined. Based on the signal power $\mathbf{l}(m)$ in (14), a_m^* can be obtained by (11).

Fig. 8 shows the fading signal decomposition after obtaining the large scale fading SNR using (14). The decoder sends the signal to the encoder system as a feedback signal.

After predicting the channel throughput over the target frame rate, we then use it as the target rate for the next frame. Let N_a be the target number of frames corresponding to one video frame. The average number of transmitted bits over the next N_a frames is the estimate based on the average channel throughput of previous $N_f/2$ frames:

$$\tilde{R}_T = \frac{2N_a}{N_f} \sum_{m=-N_f/2}^{-1} \bar{R}_s[\mathbf{r}(m)] \tag{15}$$

where $\bar{R}_s[\mathbf{r}(m)]$ is the average source throughput for the given power $\mathbf{r}(m)$.

V. Simulation results

A. Real fading channel in the downtown area of Austin

In order to measure the error-resilience performance over fading channels, real fading statistics collected in the downtown area of Austin, Texas are used in our simulations [30]. Fig. 9 (a) shows a simplified map in the downtown Austin area. The base station is set up at the corner of Congress and 7th street with an 18 meter high transmitter mounted on top of the van. The car was driven on every street and the continuous wave signal was measured at 1 $msec$ interval. The size of the area covered by the measurement is approximately 400 m in diameter. The carrier frequency is 1.9 GHz. Fig. 9 (b) shows a part of fading signal attenuation.

We model the channel characteristics of the downtown area as a single square cell and construct imaginary infinite cells, where each cell is assumed to have the same channel attenuation measured in the downtown area. Thus, the channel characteristics between adjacent cells are mirrored with respect to the cell boundaries.

B. Simulation parameters

To demonstrate foveation-based error resilience, we use the QCIF *carphone* (176×144) with 4:2:0 chrominance format, reference frame rate 30 fps, one skip frame, target transmission rate 64 $Kbps$, and target frame rate 15 with H.263++ encoded video. We choose a temporal frame rate of 10 $msec$ and a packet size of $L_p = 640$ bits. For the regular video, packets whose queuing delay is over 100 $msec$ are dropped from the encoder buffer. For the foveated video, we assign 100 $msec$ for the foveated layer and 50 $msec$ for the background layer as the maximum queuing delay in the encoder buffer. The transmission delay bound is set to 10 $msec$ for real time communications and the queuing delay bound varies according to the traffic in equation (10).

C. Regular vs. foveation-based error resilience

In order to measure the performance gain due to foveation based error resilience alone, foveation filtering is not used for the video compression. For both foveated and regular videos, each frame is encoded with a fixed quantizer step size, which is adapted at each frame to adjust to a given target transmission rate. As the concealment technique, corrupted macroblocks are replaced by corresponding macroblocks in the previous frame for both methods. For the regular video, the error resilience schemes proposed in the H.263++ standard are employed.

Fig. 10 shows the reconstructed pictures before and after error concealment when the PER (packet error rate) is 40 %. In the regular video, the effect of packet loss is uniformly distributed over the frame. In the foveated video, the foveated layer has higher temporal resolution so that better visual quality can be maintained. Fig. 10 (c) demonstrates high visual quality compared to Fig. 10 (d) because of high source throughput of the foveated layer. In Fig. 10 (c), blocking artifacts due to temporal replacement errors of error concealment are considerably reduced from the face. A live demonstration of a foveated and H.263++ compressed video stream at 40 % packet loss is available for perusal at the following website [31].

Table I shows the average picture quality of the reconstructed image sequences. After error concealment, the FPSNR (PSNR) of the foveated video is improved up to 3(1.5) dB compared to the regular video.

D. Error resilience using hybrid delay-constrained ARQ, RCPC and turbo codes

As one of the FEC methods, the RCPC code in H.223 Annex C is implemented. Using the *maximum source throughput criterion*, the number of channel parity bits is determined by the feedback channel SNR. Here, we use the same rate control and error concealment method as the subsection V-C. Also, the following methods are used for performance comparison.

- *reg.+ARQ* : error resilience in the H.263++, delay-constraint ARQ
- *reg.+RCPC+ARQ* : error resilience in the H.263++, hybrid delay-constraint ARQ
- *fov.+ARQ* : foveation-based error resilience, delay-constraint ARQ
- *reg.+RCPC+ARQ* : foveation-based error resilience, hybrid delay-constraint ARQ

Fig. 11 shows the reconstructed video quality measured in PSNR and FPSNR over the real fading channel statistics in the downtown, Austin Texas. It is shown that both PSNR and FPSNR are highest in the method *fvt.+RCPC+ARQ* where the PSNR is increased in the range of 0.3 - 1 dB over channel SNR range of 5 - 15 dB compared to the method *reg.+RCPC+ARQ*. In Fig. 12 (a), the number of skip frames is relatively reduced in the foveation-based error resilience. In Fig. 12 (b), the picture quality along temporal axis is shown. The hybrid ARQ combined with the foveation-based error resilience demonstrates better performance than other methods. Since the face region in the *carphone* sequence has relatively high motion and is focused, the foveation based error resilience is able to improve both PSNR and FPSNR by increasing source throughput of the region and reducing concealment errors.

Fig. 13 shows the reconstructed picture quality using the turbo code developed in subsection IV-D where the following methods were used.

- *TB+fvt.seq+fvt.resil* : turbo code, foveated video sequence, foveation-based error resilience
- *TB+fvt.seq+reg.resil* : turbo code, foveated video sequence, error resilience in the H.263++
- *TB+reg.seq+fvt.resil* : turbo code, regular video sequence, foveation-based error resilience

- *RCPC+reg.seq+fvt.resil* : RCPC code, regular video sequence, foveation-based error resilience

Using the turbo code, higher visual quality can be maintained at lower channel SNR. The foveation error resilience demonstrates high reconstructed visual quality. At channel SNR 5 dB, the visual quality is increased up to 1.5 dB in PSNR and 1.8 dB in FPSNR compared to the regular error resilience. Using the foveated image sequence, it is possible to eliminate much unessential high frequency DCT information, which results in significantly reduced coding overhead. The saved bits could also be assigned to points near foveation providing higher visual quality.

VI. Conclusions

Generally, error resilience techniques have attempted to reduce the effect of data loss on reconstructed picture quality based on the objective quality criterion *signal-to-noise ratio* (SNR). For point-to-point visual communication over internet/wireless networks, it is desirable to develop more robust error resilience techniques for a given limited bandwidth. By exploiting the properties of human foveation, it may be possible to protect visual information in the area of interest more robustly, resulting in overall increased visual quality.

In this paper, we introduced foveation-based error resilience and unequal error protection schemes. For source coding, the degree of visual importance at each pixel can be quantified based on foveation modeling of the HVS. Visually detectable spatial frequencies are expressed by local bandwidths in the discrete domain. Using a threshold value, the macroblocks are divided into two layers : the foveated layer and the background layer. In order to take advantage of channel adaptation, the perceptually significant foveated layer is encoded followed by the background layer. In order to increase the source throughput of the foveated layer, different delay bounds are assigned for each layer in order to take advantage of ARQ. In addition, the source throughput is increased by using a rate compatible punctured turbo code whose punctured pattern is the same as the RCPC in H.223 Annex C. Using a feedback channel, the channel coding rate is adapted according to the channel conditions.

In the simulations, the visual quality of the reconstructed frames is measured over real fading statistics collected in the downtown area of Austin, Texas. The performance comparison has been done using the following factors : foveation-based source coding, turbo channel coding, and foveation filtering. In particular, it is observed that foveation-based error resilience can significantly increase the visual quality compared to regular video protection over highly error-prone mobile networks.

References

[1] R. Talluri, "Error-resilient video coding in the ISO MPEG-4 standard," *IEEE Commun. Mag.*, pp. 112–119, June 1998.

[2] "Description of error resilient core experiments," tech. rep., ISO/IEC JTC1/SC29/WG11 N1646 MPEG97, April 1997.

[3] "MPEG-4 video verification model version 13.0," tech. rep., ISO/IEC JTC1/SC29/WG11, May 1999.

[4] "Draft text of recommendation H.263 version 2 ("h.263+") for decision," tech. rep., ITU Telecom. Standardization Section of ITU, Sep. 1997.

[5] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video quality assessment," *IEEE Trans. Multimedia*, March 2002.

[6] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video compression with optimal rate control," *IEEE Trans. Image Processing*, vol. 10, pp. 977–992, July 2001.

[7] S. Lee, A. C. Bovik, and B. L. Evans, "Efficient implementation of foveation filtering," in *Proc. Texas Instruments DSP Educator's Conference*, (Houston, TX), Aug. 1999.

[8] S. Lee, C. Podilchuk, and A. C. Bovik, "Foveation-based error resilience for video transmission over mobile networks," in *Proc. IEEE International Conference on Multimedia and Expo*, (New York City), July 2000.

[9] S. Lee, C. Podilchuk, V. Krishnan, and A. C. Bovik, "Unequal error protection for foveation-based error resilience over mobile networks," in *Proc. IEEE Int'l. Conf. Image Proc.*, (Vancouver, Canada), Sep. 2000.

[10] S. Lee, A. C. Bovik, and Y. Y. Kim, "Low delay foveated visual communications over wireless channels," in *Proc. IEEE ICIP'99*, (Kobe, Japan), Oct. 1999.

[11] S. Lee and A. C. Bovik, "Motion estimation and compensation for foveated video," in *Proc. IEEE ICIP'99*, (Kobe, Japan), Oct. 1999.

[12] "Multiplexing protocol for low bit rate multimedia mobile communication over highly error-prone channels," tech. rep., ITU T Rec. H.223 Annex C, Feb. 1998.

[13] J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Trans. Inform. Theory*, vol. 20, pp. 525–536, July 1974.

[14] W. S. Geisler and J. S. Perry, "A real-time foveated multiresolution system for low-bandwidth video communincation," in *SPIE Proceedings*, vol. 3299, 1998.

[15] M. S. Banks, A. B. Sekuler, and S. J. Anderson, "Peripheral spatial vision: limits imposed by optics, photoreceptors, and receptor pooling," *J. Opt. Soc. Amer.*, vol. 8, pp. 1775–1787, Nov. 1991.

[16] T. L. Arnow and W. S. Geisler, "Visual detection following retinal damage: Predictions of an inhomogeneous retino-cortical model," in *SPIE Proceedings:Laser-Inflicted Eye Injuries*, vol. 2674, pp. 119–130, 1996.

[17] J. J. Clark, M. R. Palmer, and P. D. Lawrence, "A transformation method for the reconstruction of functions from nonuniformly spaced samples," *IEEE Trans. on Acoust., Speech, Signal Processing*, vol. 33, pp. 1151–1165, Oct. 1985.

[18] Y. Zeevi and E. Shlomot, “Nonuniform sampling and antialiasing in image representation,” *IEEE Trans. on Signal Processing*, vol. 41, pp. 1223–1236, March 1993.

[19] M. S. Pattichis and A. C. Bovik, “AM-FM expansions for images,” in *Proc. European Signal Processing Conf.*, (Trieste, Italy), Sep. 1996.

[20] S. Lee and A. C. Bovik, “Very low bit rate foveated video coding for H.263,” in *Proc. IEEE ICASSP’99*, (Phoenix, AZ), pp. VI3113 – VI3116, Mar. 1999.

[21] A. Basu and K. J. Wiebe, “Enhancing videoconferencing using spatially varying sensing,” *IEEE Trans. Syst., Man, Cybern. -part A : systems and humans*, vol. 28, pp. 137–148, March 1998.

[22] S. Daly, K. Matthews, and J. Ribas-Corbera, “Visual eccentricity models in face-based video compression,” in *Proc. of SPIE, (Human Vision and Electronic Imaging IV)*, vol. 3644, (San Jose), Jan. 1999.

[23] B. A. Wandell, *Foundations of Vision.* Sunderland, MA: Sinauer Associates, Inc, 1994.

[24] M. Luttrell and J. D. Villasenor, “Proposal for data partitioning annex to H.263,” tech. rep., ITU-T Documentation Q15-G-13, Monterey, Feb. 1999.

[25] C. Berrou, A. Glavieux, and P. Thitimajshima, “Near shannon limit error correcting coding and decoding : Turbo-codes,” in *Proc. IEEE ICC*, (Geneva, Switzerland), pp. 1064–1070, May 1993.

[26] L.R.Bahl, J.Cocke, F. Jelinek, and J.Rajiv, “Optimal decoding of linear codes for minimising the symbol error rate,” *IEEE Trans. Inform. Theory*, pp. 284–287, 1974.

[27] Eric.K.hall and Stephen.G.Wilson, “Design and analysis of turbo codes on rayleigh fading channels,” *IEEE J. Selected Areas Commun.*, vol. 16, pp. 160–173, April 1998.

[28] W. C. Y. Lee, “Estimate of local average power of a mobile radio signal,” *IEEE Trans. Vehicular Technology*, vol. VT-34, pp. 22 – 27, Feb. 1985.

[29] W. C. Y. Lee, “Elements of cellular mobile radio systems,” *IEEE Trans. Vehicular Technology*, vol. VT-35, pp. 48 – 56, May 1986.

[30] H. Ling, “Wireless channel modeling,” in *http://ling0.ece.utexas.edu/comm/ comms.html*, 1997.

[31] S. Lee and A. C. Bovik, “Foveated video demonstration,” in *http://pineapple.ece.utexas.edu /class/Video/demo.html*, 1999.

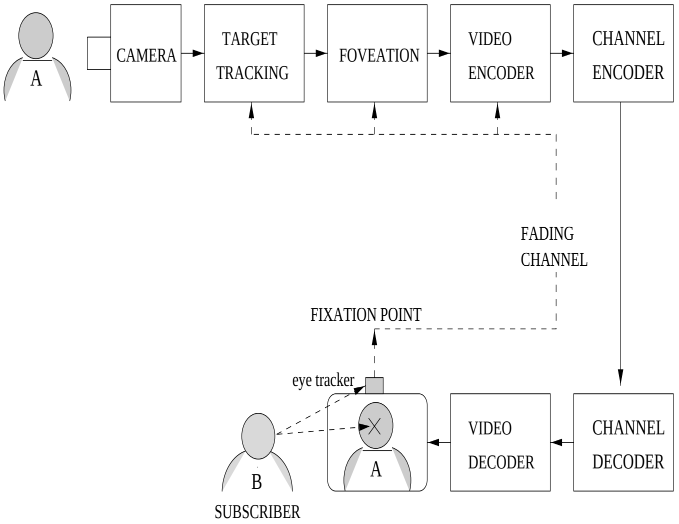
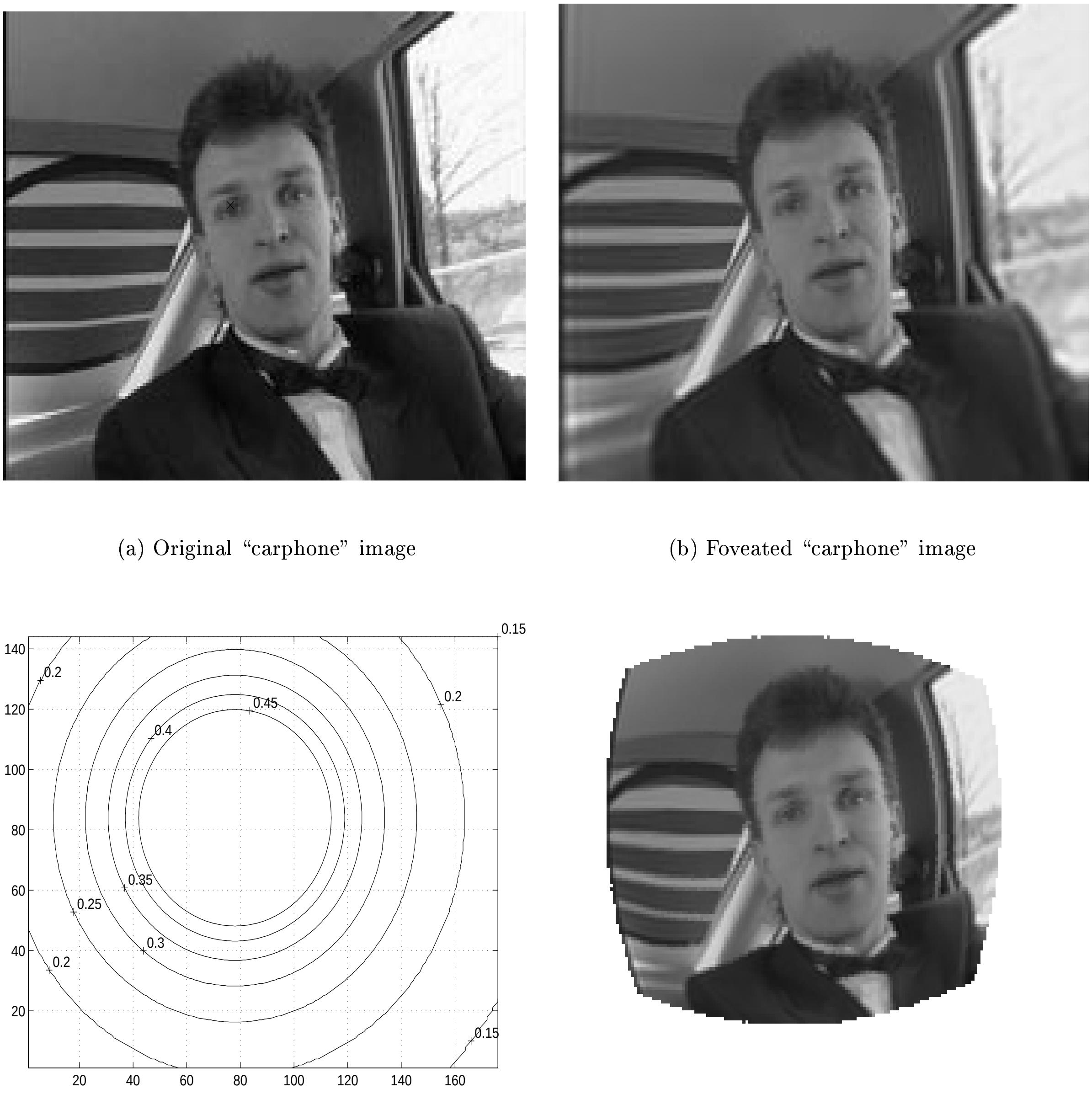


Fig. 1. Foveated visual communications

TABLE I

THE AVERAGE QUALITY OF RECONSTRUCTED IMAGE SEQUENCES

name		R, no C	F, no C	R, C	F, C
carphone	PSNR(dB)	17.60	16.46	26.37	27.85
	FPSNR(dB)	18.43	18.75	25.52	28.45



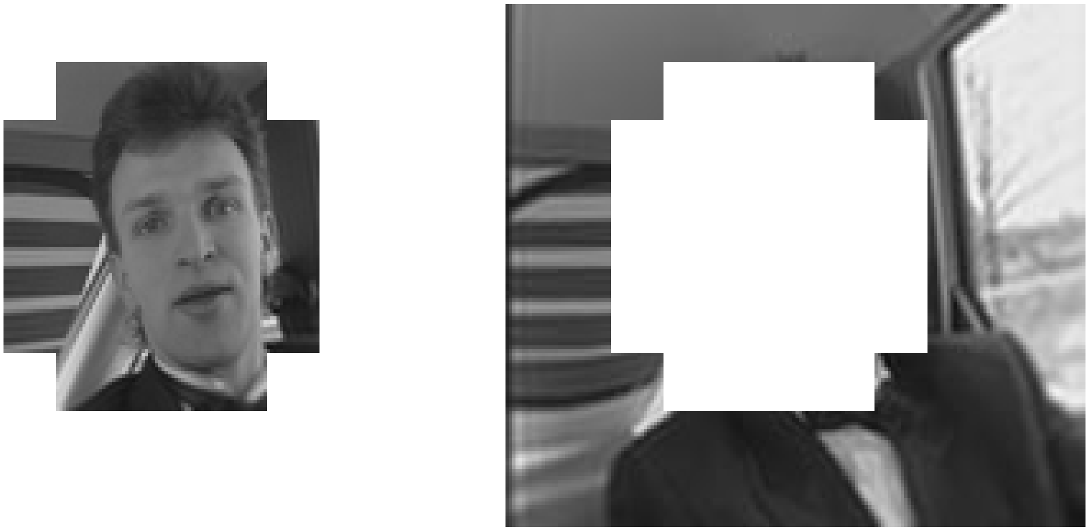
(a) Original “carphone” image

(b) Foveated “carphone” image

(c) Local bandwidth

(d) Foveated image over curvilinear coordinates

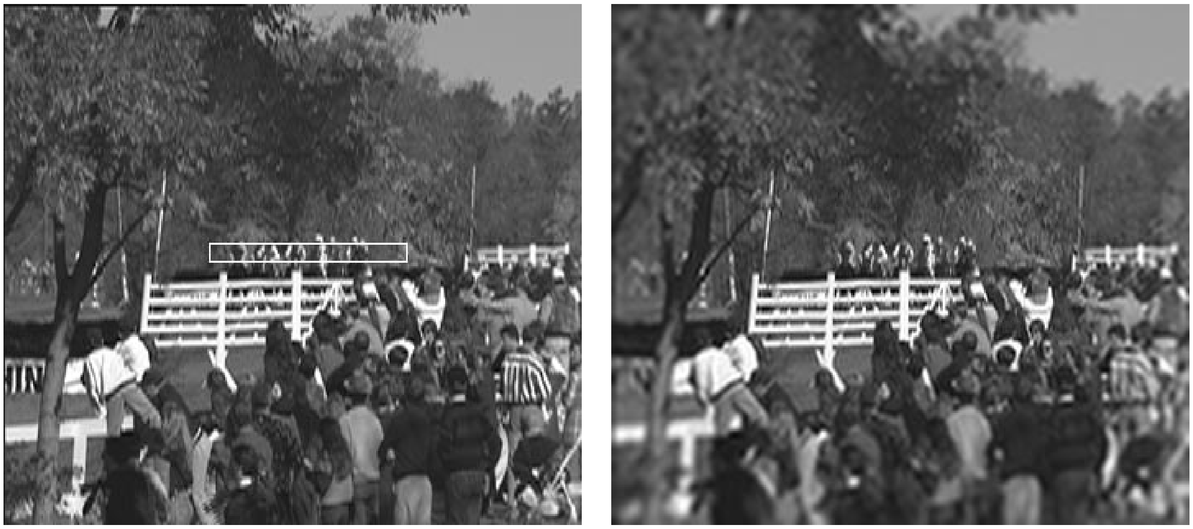
Fig. 2. Original image v.s. foveated image



(a) Foveated layer

(b) Background layer

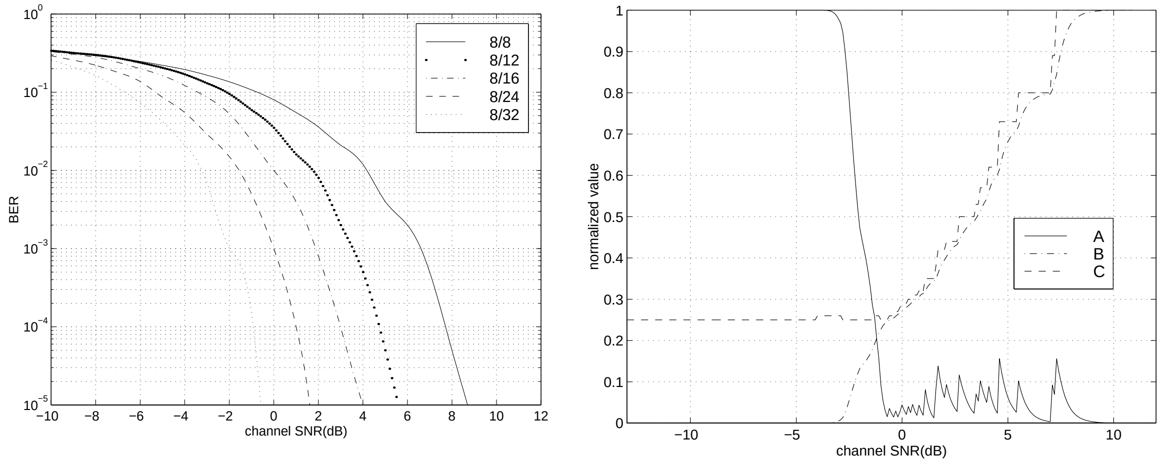
Fig. 3. Foveated layer vs. background layer



(a) Original “horse race” image

(b) Foveated “horse race” image

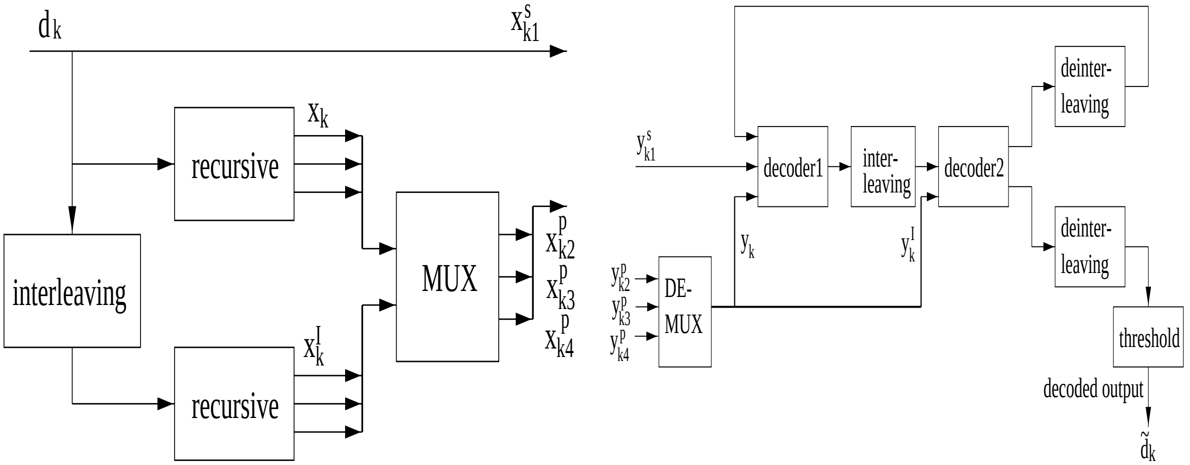
Fig. 4. Original image v.s. foveated image



(a) BER vs. code rate over channel SNR at mobile speed 50 Km/h

(b) A : packet error rate, B : normalized source throughput, C : code rate (from 8/32 to 8/8

Fig. 5. BER vs. code rate and the channel parity bit selection



(a) The encoding scheme

(b) The decoding scheme

Fig. 6. The turbo encoding/decoding schemes

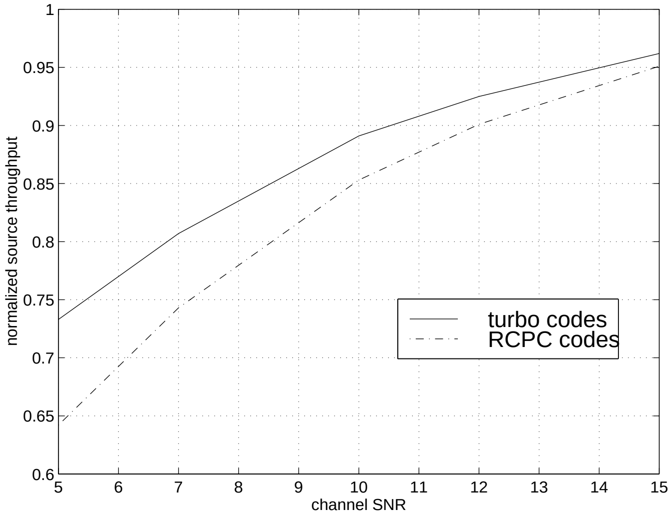


Fig. 7. Normalized source throughput : turbo vs. RCPC codes

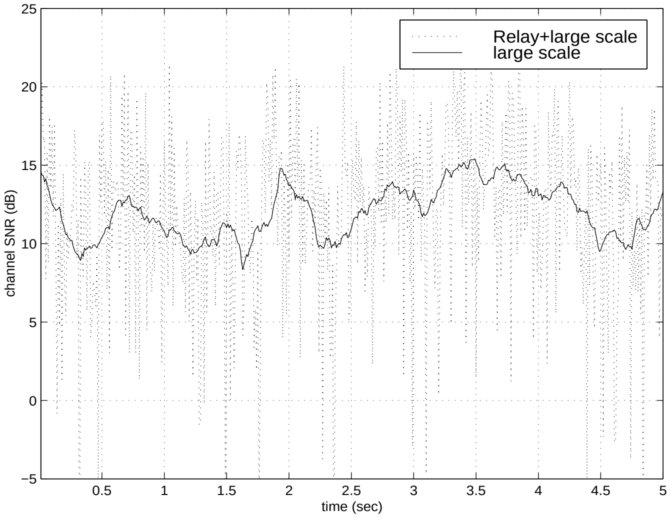
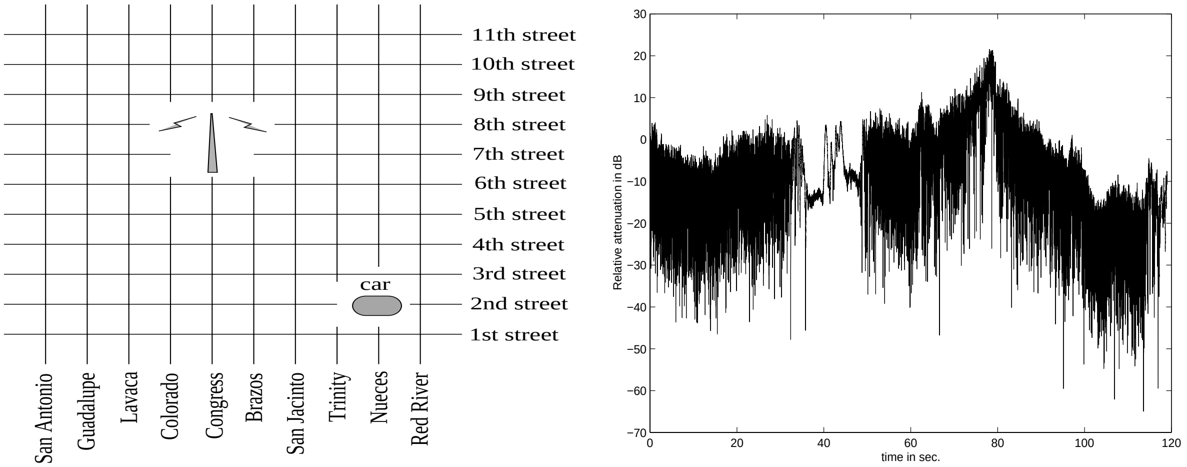


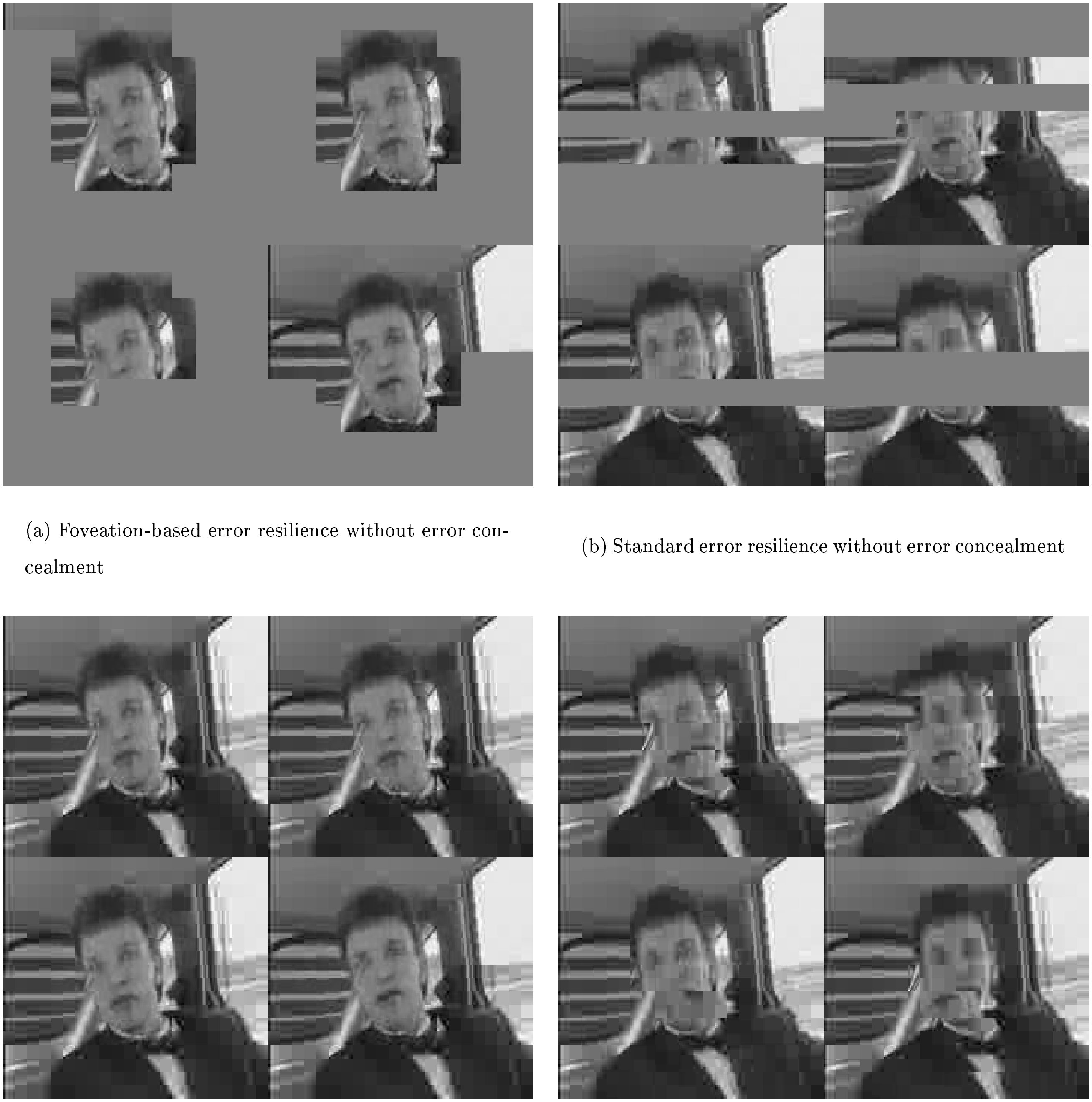
Fig. 8. Fading signal decomposition



(a) Signal attenuation measurement

(b) Real fading attenuation

Fig. 9. Fading channel measurements in the downtown area of Austin, Texas



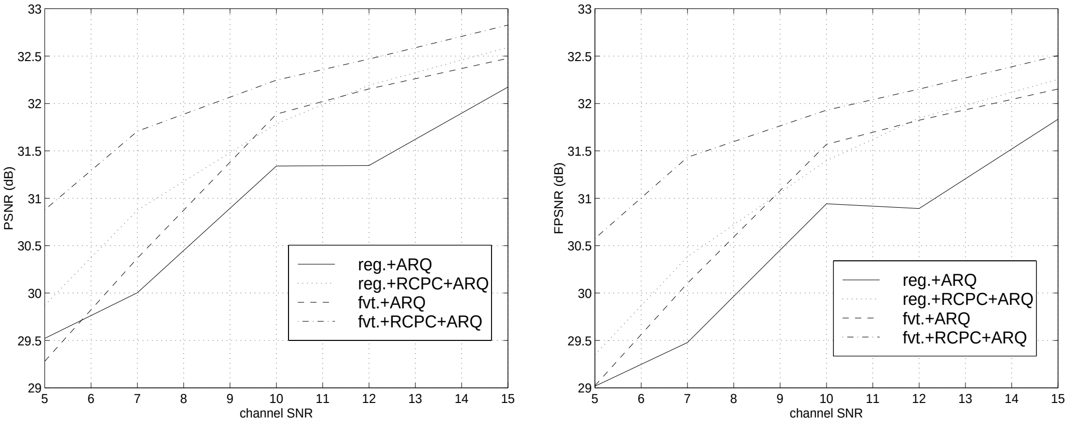
(a) Foveation-based error resilience without error concealment

(b) Standard error resilience without error concealment

(c) Reconstructed picture quality after error concealment

(d) Reconstructed picture quality after error concealment

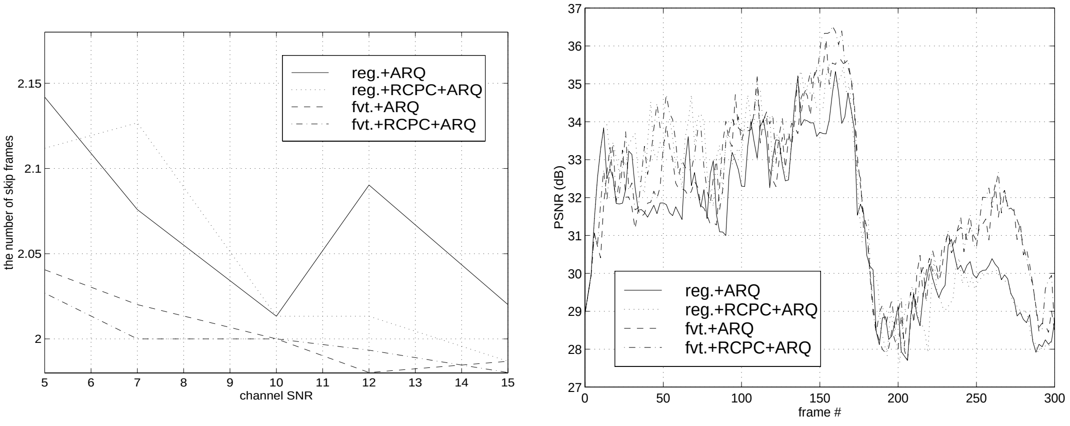
Fig. 10. Reconstructed pictures



(a) PSNR vs. channel SNR

(b) FPSNR vs. channel SNR

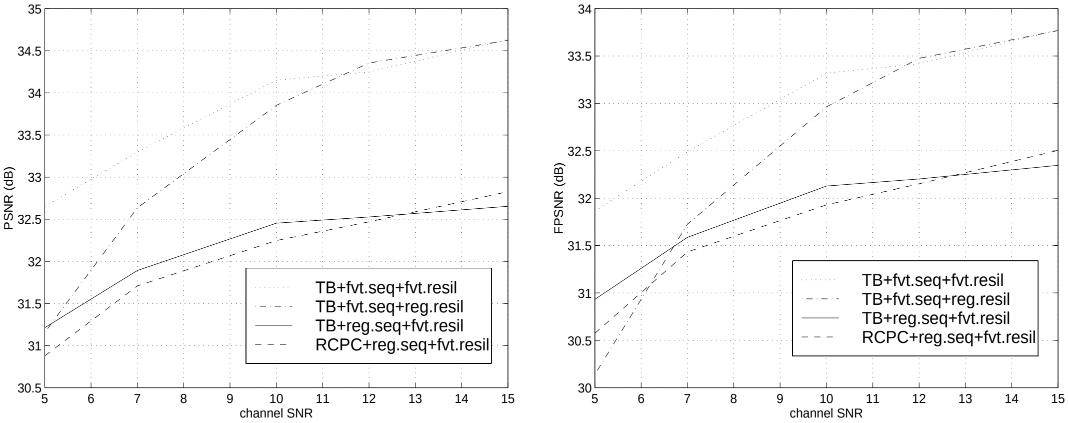
Fig. 11. The visual quality measurements according to unequal error protection methods



(a) The number of skip frames vs. channel SNR

(b) PSNR vs. frame number at channel SNR 10 dB

Fig. 12. Temporal resolution and picture quality along temporal axis



(a) PSNR vs. channel SNR

(b) FPSNR vs. channel SNR

Fig. 13. Reconstructed picture quality using the turbo code