Real-time Video Transmission over WCDMA Systems

Yu Sun[1], Ishfaq Ahmad[2] and Dongdong Li[3]

[1]Department of Computer Science, University of Central Arkansas, Conway, AR 72035
[2]Department of Computer Science & Engineering, University of Texas at Arlington, Arlington, TX 76010
[3]Department of Electrical Engineering, University of Texas at Arlington, Arlington, TX 76010

{yusun@uca.edu, iahmad@cse.uta.edu, ldd@ieee.org}

Abstract – **This paper proposes a joint source-channel rate control algorithm for real-time video transmissions over WCDMA systems. The algorithm encompasses a region-based bit allocation strategy and a macroblock-based segmentation technique. To enhance the video quality while saving bits, the algorithm effectively allocates more bits to important parts of a video frame, and also adapts to time-varying WCDMA channels. Simulation results demonstrate that the proposed algorithm effectively improves perceptual quality, reduces frame skipping and, thereby, maintains the motion continuity in wireless video transmissions.**

I. Introduction

With the increasing bandwidth in the wideband code division multiple access (WCDMA) systems and rapidly growing demand for video communications, the wireless video transmission has become feasible and received much attention recently. Since video transmissions are very sensitive to channel burst errors, it is obligatory for the video encoder to protect the video data from channel errors. Automatic repeat request (ARQ) schemes have been used as an efficient mechanism to control packet errors in video transmissions in the presence of a feedback channel. Retransmissions are only required during periods of poor channel conditions, thus the channel becomes a variable bit-rate channel (VBR) with the throughputs depending on the channel conditions [1]. However, rate control (RC) schemes in MPEG-4 and H.263 are only optimized for constant bit-rate channels, but not for VBR channels. Due to the limited channel throughput and time-varying characteristics of wireless channels, rate control for wireless video transmission is a challenging task, as it has to jointly decide encoding parameters and estimate current channel conditions in order to optimize the encoding performance.

To improve the coding efficiency and ensure good subjective qualities of frames in the very low bit-rate video coding, region-based coding is usually exploited to code the regions of interest (ROI) more accurately than the rest of the video content. Some research works [2-4] adopted a heuristic approach to decide the quantization parameters (QPs) for different regions in a frame, for example: an initial QP for a frame or macroblock (MB) is chosen first. Then the initial QP is directly decreased by a factor to obtain a finer quantizer for ROI, which results in more bits used in coding ROI; or it is increased by a factor to acquire a coarser quantizer for non-ROI and thus fewer bits are used in coding non-ROI. These factors are heuristically set to constants, and the contents of regions are not taken into consideration. The direction of the QP adjustments is correct, but these algorithms lack a quantitative method to perform bit allocation among different regions, this may cause improper QPs and unreasonable bits used for different regions.

This paper proposes a region-based rate control algorithm to transmit video in real-time over variable WCDMA channels. By using a fast segmentation method, the algorithm divides a frame into ROI and non-ROI, then it adopts the most effective criteria "coding qualities" as a quantitative factor to directly control bit allocation among different regions, and employs a dynamic priority factor to further adjust bit allocation, aiming at improving ROI's visual quality and reducing frame skipping under time-varying channels.

The rest of this paper is organized as follows: In section II, we present the segmentation method. In section III, we show how to establish the channel model in WCDMA systems to estimate the variable channel-bandwidth. Section IV describes the rate control algorithm. Section V includes the simulation results. Finally, section VI concludes the paper.

II. Fast Segmentation of Moving Regions

Since the human visual system (HVS) is more sensitive to the moving regions, it is worthwhile to sacrifice the perceptual quality of the still regions while enhancing that of the moving regions. Moving regions are classified as the foreground (ROI) while still regions are regarded as the background (non-ROI). To detect changes between two successive frames, each frame is smoothed by a low-pass filter to reduce high frequency noises, then the difference between the current frame and previous frame is computed. We can calculate the difference mask DM_t as

$$DM_t(i,j) = \begin{cases} 1, & \left| I_t(i,j) - I_{t-1}(i,j) \right| > Thr_t \\ 0, & else \end{cases} \quad (1)$$

where (i,j) represents the coordinate of the processed pixel, I_t and I_{t-1} represent the current and previous frames respectively, and Thr_t is a threshold for DM_t given by

$$Thr_t = \frac{1}{M \cdot N} \sum_{i=1}^{M} \sum_{j=1}^{N} \left| I_t(i,j) - I_{t-1}(i,j) \right|, \quad (2)$$

0-7803-8794-5/04/$20.00 © 2004 IEEE

where M and N denote the line and pixel number in a frame respectively. The moving ratio $r_t(B_m)$ of B_m (the m[th] MB) is calculated as: the summation of $DM_t(i,j)$ for each pixel in B_m divided by B_m's pixel number. The threshold MB_Thr_t is set to

$$MB\_Thr_t = k \cdot (\sum_{m=1}^{N_{MB}} r_t(B_m) / N_{MB}), \tag{3}$$

where N_{MB} is the number of MBs in a frame, k is a constant and is empirically chosen to be 1.4. If $r_t(B_m)$ is larger than MB_Thr_t, B_m is selected as the moving MB; otherwise, B_m is the still MB.

Finally, we sort all moving MBs by $r_t(B_m)$ in decreasing order. If $r_t(B_m)$ of an isolated moving MB is in the last 40% of the sorted list, it is merged to the still region since it is not a heavily moving MB among all of the moving MBs. An isolated still MB is combined to the moving region by the same way.

III. WIRELESS CHANNEL MODEL

The Gilbert two-state Markov model [5] is adopted to model the wireless channel. The model has two states, good (packet error free) state S_0 and bad (packet error) state S_1, and states transfer between each other at a certain probability. The channel state transition probability matrix is given by P with the element $P_{ij}=P[S(k)=S_j|S(k-1)=S_i]$, $i,j \in \{0,1\}$, which is the transfer probability from the state S_i at time t_{k-1} (i.e. $S(k-1)$) to the state S_j at time t_k (i.e. $S(k)$) [6]. Without loss of generality, assuming that the channel state $S(k-m)$ at time t_{k-m} is known, the probability that the channel in state S_j at time t_k is denoted as $\pi_j(k|S(k-m))$. Therefore, the state probability vector at time t_k is $\pi(k|S(k-m))=[\pi_0(k|S(k-m)), \pi_1(k|S(k-m))]$. Since the channel state at time t_{k-m} is known to be $S(k-m)$, the state probability vector at time t_{k-m} is initialized by

$$\pi_j(k-m \mid S(k-m)) = \begin{cases} 1, & S(k-m) = S_j \\ 0, & otherwise \end{cases} \quad j \in \{0,1\}. \tag{4}$$

Thus, the state probability vector at time t_k can be derived from the state probability vector at time t_{k-m} by using

$$\pi(k \mid S(k-m)) = \pi(k-m \mid S(k-m))P^m. \tag{5}$$

The probability of packet error-free at time t_k is $\pi_0(k \mid S(k-m))$.

Notice that even more accurate packet error probability can be obtained by taking more channels states. However, since our proposed rate control can well adapt to the channels and the video is transmitted in the unit of frame, the accuracy of the channel modeling is not a major problem.

IV. RATE CONTROL ALGORITHM

1) Estimate Initial Target Bits: It is impossible to calculate remaining bits and frames in real time video communications, since the total sequence is not known at the moment of the encoding. We set the initial target bits T_t for a frame to R/F_r, where R and F_r are the channel and frame rates, respectively.

2) Adjust the Target Bits Based on the Buffer Fullness: To get more accurate target bit estimation, the initial bit target is further refined based on the buffer fullness. Here, we adopt our proposed Proportional-Integral-Differential (PID) buffer control technique [7]. The PID buffer adjusting factor is computed by

$$PID_t = K_p \cdot (E_t + K_i \cdot \int_0^t E_\tau \cdot d\tau + K_d \cdot \frac{dE_t}{dt}) \quad with \quad E_t = B_s / 2 - B_{f,t} \tag{6}$$

where B_s is the buffer size, E_t is the error deviation between the target buffer fullness ($B_s/2$) and the current buffer fullness $B_{f,t}$; K_p, K_i and K_d are the Proportional, Integral and Differential control parameters respectively, and are empirically set to 0.1, 0.25 and 0.3 correspondingly in the simulations. Then the target bits can be further adjusted by $T_t:=T_t + PID_t$.

3) Compute the Number of Retransmitted Bits per Frame Interval: First, we assume the length of the encoded frame r to be L_r, then $N_r=L_r/L_p$ packets have been transmitted in a packet size of L_p bits. Second, the average error packet ratio of the previous L_{rp} frames is calculated by

$$r_{avg}(L_{rp}) = \sum_{r=1}^{L_{rp}} \sum_{i=1}^{N_r} P_e(r,i) \Big/ \sum_{r=1}^{L_{rp}} N_r, \tag{7}$$

where $P_e(r,i)$ equals to 1 when the i^{th} packet in the r^{th} frame is in error, otherwise 0. If $r_{avg}(L_{rp})$ is less than a threshold value W_{th}, the current state of the wireless channel $S(0)$ is said to be good. Otherwise, the channel is said to be in the bad state. Third, using the current channel state and (1), we can estimate the channel state for the next L_{rf} frames. The average number of packets per frame-interval to be transmitted can be calculated by $L_c=(R/F_r)/L_p$. Therefore, we need to generate channel states for the next $N_{rf}=L_{rf}L_c$ packets as follows: To obtain a smooth estimation of the channel state, the average probability of a error-free transmission of the m^{th} packet given the current channel state $S(0)$ is derived based on (5) as

$$p_{avg}(m \mid S(0)) = \frac{1}{m} \sum_{i=1}^{m} \pi_0(i \mid S(0)). \tag{8}$$

Following a commonly accepted statistical approach [6], we generate a uniform distributed random variable α in the interval [0,1] and compared it with the $P_{avg}(m|S(0))$ to decide the channel state of the m^{th} packet. This approach can dynamically characterize the channel state. If α is greater than $P_{avg}(m|S(0))$, the channel of the m^{th} packet is said to be bad and $P_e(m)=1$; otherwise, $P_e(m)=0$. Therefore, the total retransmitted bits RTB in the current frame can be calculated by

$$RTB = \frac{1}{L_{rf}} \sum_{m=1}^{N_{rf}} P_e(m) \times L_p. \tag{9}$$

4) Adjust the Target Bits Based on RTB: During the retransmissions of error packets, the video data in the encoder buffer are not transmitted. Therefore, due to the reduced channel throughput, the encoder buffer fills up quickly which may cause the RC algorithm to skip frames or significantly reduce the bits allocated to each frame [1]. In order to counteract the buffer fill-up in the future, the frame target is decreased by $T_t := T_t - RTB$.

5) Distribute the Target Bits: To avoid large perceptual quality differences between the foreground (F) and background (B), the algorithm sets weights for them. The larger the weight is, the more target bits should be allocated to the corresponding region. To directly control bit allocation, the most effective criterions, foreground PSNR ($PSNR_{F,t}$) and background PSNR ($PSNR_{B,t}$), are employed in the weight adjustment. In addition, priority is also used to assist target bit allocation so as to express application requirements. We adopt the background as a referential base, its weight $W_{B,t}$ is always 1.0. Let $W_{F,t}$ to be the weight for the foreground at time t, its initial value is 1.0. U_F is the foreground's priority, U_F >0 (dB) means a higher priority while U_F <0 (dB) corresponds to a lower priority. At time t-1, if ($PSNR_{F,t-1}$-U_F) is lower than $PSNR_{B,t-1}$, the algorithm improves $W_{F,t}$, thus the foreground obtains more target bits and achieves a higher quality, and vice visa. $W_{F,t}$ is updated as

$$W_{F,t} = W_{F,t-1} \cdot e^{\left(\frac{PSNR_{B,t-1} - PSNR_{F,t-1} + U_F}{\theta}\right)}, \qquad (10)$$

here, the tuning factor θ is selected to 4 empirically. We have chosen the normalized weight, size and variance as coding complexity measurements in distributing target bits between the foreground and background. The numbers of target bits for the foreground ($T_{F,t}$) and background ($T_{B,t}$) are allocated by

$$T_{C,t} = \frac{NW_{C,t} \cdot (NMB_{C,t} \cdot NVAR_{C,t})}{NW_{F,t} \cdot (NMB_{F,t} \cdot NVAR_{F,t}) + NW_{B,t} \cdot (NMB_{B,t} \cdot NVAR_{B,t})} \cdot T_t \qquad C \in \{F, B\} \qquad (11)$$

where $NMB_{F,t}$ and $NMB_{B,t}$ are the numbers of MBs in the foreground and background respectively, normalized by the total number of MBs in a frame; $NVAR_{F,t}$ and $NVAR_{B,t}$ are the normalized foreground and background variances of the motion-compensated residual frame respectively; $NW_{F,t}$ is the normalized weight of the foreground and $NW_{B,t}$ is that of the background.

6) Macroblock-Level Rate Control: The target bits for coding the i^{th} macroblock (MB_i), $T_{MB,i}$, can be allocated by:

$$\mathrm{T_{MB,i}} = (VAR_{MB,i} \,/ \sum_{k=i}^{NMB_{F,t}} VAR_{MB,k}) \cdot T^i_{F,t}, \quad MB_i \in foreground,$$
$$\mathrm{T_{MB,i}} = (VAR_{MB,i} \,/ \sum_{k=i}^{NMB_{B,t}} VAR_{MB,k}) \cdot T^i_{B,t}, \quad MB_i \in background, \qquad (12)$$

where $VAR_{MB,i}$ is the variance of the motion-compensated residual MB_i, $T_{F,t}^{\ i}$ and $T_{B,t}^{\ i}$ are the currently remaining available target bits for the foreground and background respectively when coding MB_i, initially $T^i_{F,t} = T_{F,t}$ and $T^i_{B,t} = T_{B,t}$. The MB-layer RC of MPEG-4 [8] is then used to compute the QP for MB_i and encode MB_i.

7) Update Buffer Fullness: The buffer fullness is updated after encoding, the number of bits used for encoding the current frame and the retransmission bits are added to the previous buffer level. Meanwhile, the number of bits to be output from the buffer per encoding time, R/F_r, is decreased from the buffer fullness. If the buffer occupancy exceeds 80% of the buffer size, the encoder skips the next frame [8, 9].

V. SIMULATION RESULTS

To evaluate the performance of the proposed algorithm, we compare it with the MPEG-4 Q2 scheme [8,9] under two sets of channel conditions: Set 1 (CCS1): walking speed (3 km/hour) and a user data rate 32 Kbps; Set 2 (CCS2): walking speed (3 km/hour) and a user data rate 64 Kbps. A WCDMA channel simulator is used to generate the bit-error pattern under CCS1 and CCS2, which is setup using the corresponding channel conditions and parameters listed in Table I. The packet sizes are 320 bits for CCS1 and 640 bits for CCS2. The transition probabilities of the Gilbert tow-state Markov model are obtained from the WCDMA channel simulator as follows: For CCS1, P_{00}=0.97538, P_{01}=0.02462, P_{10}=0.30367, P_{11}=0.69633, which corresponds to an average packet burst error length of 3.3 packets and a packet error rate of 0.075. For CCS2, P_{00}=0.96024, P_{01}=0.039759, P_{10}=0.17154, P_{11}=0.82846, which corresponds to an average packet burst error length of 5.8 packets and a packet error rate of 0.19. Packet error threshold W_{th} is empirically set to 0.2. Both the number of past L_{rp} frames and the future L_{rf} frames are empirically chosen to be 3.

To meet the low-delay requirement in video transmissions, a small buffer size $0.125 \bullet R$ is used in both MPEG-4 Q2 and our algorithms [8,9], frame skipping occurs when the buffer fullness is above $0.1 \bullet R$ [8,9]. When the channel is in the good state, we dynamically set U_F to 0 to obtain balanced coding qualities between the foreground and background; when the channel state is bad, in order to ensure the foreground's visual quality, we set U_F to 2 to improve the foreground's priority. Assume all erroneous packets are retransmitted successfully in a single attempt. Since a skipped frame is represented in the decoded sequence by repeating the previously coded frame according to MPEG-4 core experiments, the PSNR of a skipped frame is computed by using the previous encoded frame [9].

Because bit-rates are diminished in the background, the overall average PSNR values of the proposed algorithm in Table II are little lower than those of MEPG-4 Q2. To evaluate the picture quality more properly in the region-based RC algorithm, the foreground and background PSNRs are used here. In Table II, the foreground PSNRs of our algorithm are much higher than those of MEPG-4 Q2, while the background PSNRs of our algorithm are lower. Therefore, our algorithm can apparently enhance the visual quality in the foreground, at the expense of quality degradation on the background. Meanwhile, our algorithm can reduce frame skipping effectively when compared with MPEG-4 Q2. The bit count ratio of the foreground or background, defined as the individual used bit count normalized by the total used bit count of the whole sequence, is also shown in Table II. Using our algorithm, the bit count ratio of the foreground increases a lot, this indicates our bit allocation method is effective. As a result, the visual quality of the foreground has been enhanced.

Table I

Specifications of the WCDMA simulator

Multiple Access	WCDMA
Modulation	QPSK
Carrier Frequency	1.9 GHz
Chip Rate	4.096 Mcps
CRC	16 bits
Coding	convolution code rate 1/3, constraint length 9
Interleaving	10 ms
Diversity	2 antenna branches
Bit Error Rate	10^{-3}

PSNR curves in Fig. 1(a) & 1(b) show that the foreground PSNRs of our algorithm are higher than those of MPEG-4 Q2, but our background PSNRs are lower. They further exhibit that our algorithm obtains smoother foreground and background qualities among frames. In addition, our buffer fullness curve in Fig. 1c is more stable. Fig. 2 indicates our algorithm can effectively detect the moving region and enhance its perceptual quality, the foreground including face and hand parts is clearer than that of MPEG-4 Q2. Although it may introduce some degradation on the background, it is almost invisible to human perception.

Due to the limited bandwidth, video sequences transmitted under wireless channels normally have low motion, such as head and shoulder scenarios. Since higher bandwidths are available by the current WCDMA technique, we also study the effect of transmitting some medium or fast motion sequences under the 64 Kbps channel (CCS2). The results in Table III demonstrate again that our algorithm effectively reduces the number of frame skipping and improves the perceived quality for the foreground. Furthermore, the overall PSNRs of the proposed algorithm are little higher than those of MEPG-4 Q2. This may due to the large number of frame skipping in MPEG-4 Q2, which causes overall PSNR degradation.

VI. CONCLUSIONS

In conjunction with a fast segmentation method, we proposed a region-based rate-control algorithm for real time video transmissions over WCDMA channels. Unlike traditional bit allocation methods used in the region/content-based rate control, the proposed algorithm exploits the most effective criteria “coding qualities” to quantitatively control bit allocation among different regions. Compared with the MPEG-4 rate control algorithm, the proposed algorithm enhances the visual quality for ROI, reduces the number of frame skipping, and improves the motion smoothness.

REFERENCES

[1] Supavadee Aramvith, I-Ming Pao, and Ming-Ting Sun, “A Rate-Control Scheme for Video Transport over Wireless Channels,” *IEEE Transactions On Circuits and Systems for Video Technology,* VOL. 11, No. 5, pp. 569-580, May 2001.

[2] Chu n-Hung Lin and Ja-Ling Wu, “Content-Based Rate Control Scheme for Very Low Bit-Rate Video Coding,” *IEEE Trans. On Consumer Electronics*, Vol.43, No.2, May 1997.

[3] S. Aramvith, H. Kortrakulkij, et al., “Joint Source-Channel Coding using Simplified Block-Based Segmentation and Content-based Rate-Control for Wireless Video transport,” in *Proc. of International Conference on Information Technology: Coding and Computing 2002*, Las Vegas, pp. 71-76, April 2002.

[4] C. Lin, Y. Chang, et al., “Low-complexity face-assisted coding scheme for low-bit-rate video telephony,” *IEICE Trans. Information and Systems*, VOL. E86-D, pp.101-108, Jan. 2003.

[5] M. Zorzi, R. R. Rao, and L. Milstein, “On the accuracy of a first-order Markov model for data transmission on fading channels,” in *Proc. ICUPC’95*, pp. 160-164.

[6] L. Garcia, *Probability and Random Processes for Electrical Engineering*, 2nd ed. Addison-Wesley 1994.

[7] Yu Sun, Ishfaq Ahmad, et al., "Synchronous and Asynchronous Multiple Object Rate Control For MPEG-4 Video Coding," *Proceedings of IEEE International Conference on Image Processing (ICIP 2003)*, Barcelona, Spain, Sep. 2003, Vol. 3, pp. 801-804.

[8] Hung-Ju Lee, Tihao Chiang, and Ya-Qin Zhang, “Scalable Rate Control for MPEG-4 Video,” *IEEE Trans. On Circuits and Systems for Video Technology*, VOL. 10, pp.878-894, Sep. 2000.

[9] MPEG-4 video verification model V8.0, ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Associated Audio MPEG97/N1796, July 1997, Stockholm, Sweden.

Table II

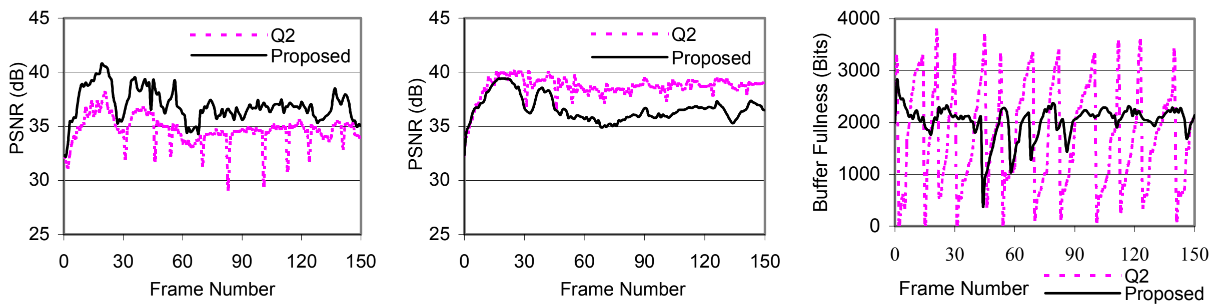
Simulation results under the WCDMA channel condition CCS1 (32 Kbps, 10 fps, 150 frames, QCIF)

Video Sequence	Algorithm	Bit count ratio (%)		# Skipped Frames	Average PSNR (dB)		
		Foreground	Background		Foreground	Background	Overall
Mother_ Daughter	Q2	46%	54%	12	34.66	38.68	37.36
	Proposed	82%	18%	0	36.99	36.63	36.70
Silent Voice	Q2	61%	39%	14	32.50	34.51	33.85
	Proposed	77%	23%	0	33.90	33.42	33.51
Salesman	Q2	42%	58%	18	32.18	35.37	34.54
	Proposed	62%	38%	0	33.98	34.20	34.05

Table III

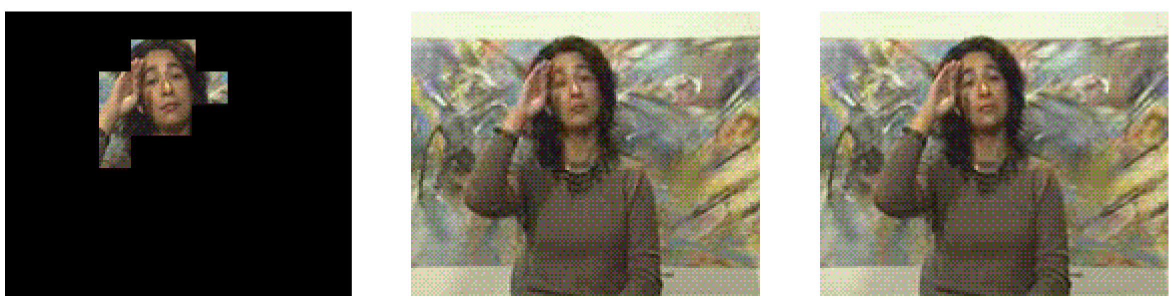
Simulation results under the WCDMA channel condition CCS2 (64 Kbps, 10 fps, 150 frames, QCIF)

Video Sequence	Algorithm	Bit count ratio (%)		# Skipped Frames	Average PSNR (dB)		
		Foreground	Background		Foreground	Background	Overall
Stefan	Q2	57%	43%	29	23.22	28.60	25.70
	Proposed	85%	15%	0	25.31	26.66	25.92
Foreman	Q2	43%	57%	26	31.76	34.30	33.25
	Proposed	67%	33%	0	34.40	33.40	33.61



(a) Foreground PSNR (b) Background PSNR (c) Buffer Occupancy

Fig.1. PSNR and buffer curves for the *Mother_Daughter* sequence on the WCDMA channel condition CCS1.



(a) Segmentation Result (Original Frame) (b) MPEG-4 Q2 (Reconstructed Frame) (c) Proposed (Reconstructed Frame)

Fig. 2. Subjective results of the 23th frame of the "Silent Voice" on the WCDMA Channel Condition CCS1.