# Joint Spatial-Temporal Quality Improvement Scheme for H.264 Low Bit Rate Video Coding via Adaptive Frameskip

**Ziguan Cui, Zongliang Gan and Xiuchang Zhu**
Image Processing and Image Communication Lab, Nanjing University of Posts and Telecommunications
Nanjing, 210003 - China
[e-mail: {czg1982001, ganzongliang}@163.com, zhuxc@njupt.edu.cn]
*Corresponding author: Ziguan Cui

*Received September 24, 2011; revised December 21, 2011; January 7, 2012;*
*Published January 31, 2012*

## Abstract

Conventional rate control (RC) schemes for H.264 video coding usually regulate output bit rate to match channel bandwidth by adjusting quantization parameter (QP) at fixed full frame rate, and the passive frame skipping to avoid buffer overflow usually occurs when scene changes or high motions exist in video sequences especially at low bit rate, which degrades spatial-temporal quality and causes jerky effect. In this paper, an active content adaptive frame skipping scheme is proposed instead of passive methods, which skips subjectively trivial frames by structural similarity (SSIM) measurement between the original frame and the interpolated frame via motion vector (MV) copy scheme. The saved bits from skipped frames are allocated to coded key ones to enhance their spatial quality, and the skipped frames are well recovered based on MV copy scheme from adjacent key ones at the decoder side to maintain constant frame rate. Experimental results show that the proposed active SSIM-based frameskip scheme acquires better and more consistent spatial-temporal quality both in objective (PSNR) and subjective (SSIM) sense with low complexity compared to classic fixed frame rate control method JVT-G012 and prior objective metric based frameskip method.

***Keywords:*** Video coding, H.264, RC, adaptive frame skipping, perceptual quality

A preliminary version of this paper appeared in IEEE ICCT 2010, November 11-14, Nanjing, China. This version proposes a novel motion vector copy based frame inperpolation scheme and an adaptive frameskip threshold, and an integrated frame layer rate control scheme combined with adaptive frameskip is presented and more detailed experimental results are shown to validate the encoding performance. This work was supported by National Natural Science Foundation of China (60672134, 61071091) and Jiangsu Province Postgraduate Innovative Research Plan (CX10B_190Z).

**DOI: 10.3837/tiis.2012.01.024**

## 1. Introduction

H.264 as the latest video coding standard acquires higher compression efficiency and will be applied in various applications. Rate control (RC) schemes for video coding and transmission are adopted to regulate output bit rate to match channel bandwidth and simultaneously acquire optimal video quality. Conventional rate control methods usually adjust frame layer or marcoblock (MB) layer quantization parameter (QP) to meet desired target bit rate at fixed full frame rate, which can work well at high bit rate [1][2][3][4][5][6]. However, these methods often result in poor spatial quality of each coded frame at low bit rate and even cause passive frame skipping and sharp fluctuation of spatial-temporal quality due to buffer constraints. Instead of passive frame skipping, active or intentional frame skipping schemes based on video content and channel conditions can achieve more efficient coding performance and better recovery quality for the skipped frames at the decoder side.

So joint spatial-temporal quality optimization schemes are proposed to enhance overall coding performance by jointly adjusting encoding frame rate and QPs [7][8][9]. The basic idea is that the active content adaptive frame skipping is employed to skip trivial frames and the saved bits can be allocated to coded key frames to enhance their spatial quality, and the skipped frames can be well interpolated from adjacent high quality key frames at the decoder side to achieve constant frame rate. The key issues are the selection of position and number of to-be-skipped frames in video sequence adaptively, as well as the target bits and QPs of to-be-coded frames at given target bit rate and buffer constraints.

Until now, there are only a few literature concerning about variable frame rate control. In [10][11], Pan et al. decided the frame skipping factor by spatial (QP) and temporal (MV) quality of the video, and by the buffer status. Thammineni et al. first assessed the impact of new frame rate by actual coding trial frames and then decided how to change frame rate [12]. Jun et al. adaptively skipped frames by the objective similarity between adjacent frames measured by peak signal noise ratio (PSNR) [13]. Usach et al. considered the issue of bits allocation for group of pictures (GOP) layer after frame rate changes [14]. In [15], Liu and Kuo jointly determined the frames to be skipped and QPs of coded frames with consideration of temporal dependency in rate-distortion optimization sense, and proposed adaptively frame grouping method based on the differential of mean of absolute difference (MAD) to reduce complexity but still could not be used in real-time applications. Vetro and Lee proposed distortion estimation models for both coded and skipped frames, together with famous rate model to optimize coding performance with frameskip in rate-distortion sense [16][17]. Song and Kuo adjusted the encoding frame rate regularly for current sub-GOP based on the motion activity of previous sub-GOP which is expressed by the histogram of difference image (HOD) [18]. Yang et al. embedded an iterative motion-compensated frame interpolation (MCFI) and subjective quality based measurement module of interpolated frames in the encoding process to adaptively determine the number of frames to be skipped [19]. The method considers subjective property and thus enhances the overall video quality but has very high complexity and encoding delay due to iteratively encoding and interpolation of frames for quality evaluation. In [20], a novel selective frame discard method is proposed for 3D video over IP networks. The 2D video frames and the depth map frames are discarded symmetrically with additional consideration of the playback deadline, the network bandwidth, and the inter-frame dependency within the GOP. In [21], a cost-effective rate control scheme for streaming video is proposed based on buffer status to prevent buffer fullness and emptiness.

From above analysis, we can see that prior methods mostly adjust encoding frame rate based on objective quality property in video sequence such as MAD, PSNR, HOD and motion activity, so the skipped frames may not be subjectively trivial. Subjective quality based frame skipping is only considered in [19], but its complexity is extremely high and therefore it is hard to be applied extensively especially for real-time applications. Another important issue with most prior frameskip methods is that the recovery quality of skipped frames at the decoder side is not well considered or irrelevant when designing the frameskip strategy at the encoder side, which causes poor recovery quality of skipped frames and temporal quality fluctuation of the entire sequence. To address the two issues, in this work we first propose an improved MV copy based frame interpolation scheme to recover the skipped frames with high quality and motion continuity, and then propose an active content adaptive frame skipping scheme with low complexity, the subjectively trivial frames are adaptively skipped based on structural similarity (SSIM) measurement [22] between the original frame and the interpolated frame by MV copy scheme at the encoder side, and thus the overall video quality for both objective (PSNR) and subjective (SSIM) is improved simultaneously.

The rest of this paper is organized as follows. Section 2 first introduces the SSIM-based subjective property briefly, and proposes an improved MV copy frame interpolation method and SSIM-based content adaptive frame skipping scheme. Section 3 proposes the integrated frameskip scheme combined with the classic JVT-G012 frame layer RC. Section 4 presents the experimental results and analysis compared with classic fixed frame rate control method JVT-G012 and objectve metric based frameskip method. Section 5 concludes the paper.

## 2. SSIM-based Content Adaptive Frame Skipping

### 2.1 Structural Similarity Index

Traditional objective video quality evaluation and distortion measurement are based on pixel pointwise absolute signal error such as mean square error (MSE) or PSNR, which is used extensively due to clear physical meanings and low complexity. However, we all know that PSNR is not proportional to subjective quality because it does not consider human visual system (HVS) characteristics. Thus how to develop a subjective quality evaluation method by fully exploiting HVS characteristics and use it to optimize video coding performance has become an important research issue in recent years. Several works have been done on this subject. In [22], Wang et al. proposed the structural similarity (SSIM) index to measure subjective similarity between the original image and the distorted image under the assumption that HVS is more sensitive to the extracted structural information in a scene than absolute signal error. SSIM compares two images from three aspects (luminance, contrast, structure) and then combines the three parts to generate overall similarity measure. The formula can be expressed as follows:

$$\begin{aligned} SSIM(x,y) &= l(x,y)\cdot c(x,y)\cdot s(x,y) \\ &= \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \end{aligned} \tag{1}$$

where $x$ and $y$ are two images for comparison, $\mu_x$, $\mu_y$ and $\sigma_x$, $\sigma_y$ are the mean luminance and the standard deviation of $x$ and $y$, $\sigma_{xy}$ is the covariance of $x$ and $y$, $C_1$ and $C_2$ are small constants to avoid instability when denominator is very close to zero.

Note that the SSIM index should be applied locally rather than globally [22], so the overall SSIM index of a frame is calculated as the mean of all local windows such as non-overlapped 8×8 square blocks [23] over the whole frame. SSIM index has the boundary: $SSIM(x,y)\leq 1$ , and more close to 1 the value is means the two images are more similar in subjective quality. Due to low computation complexity and more accordant with HVS, SSIM has been embedded in latest H.264 reference software (JM) [23] to assess the subjective quality of encoded video.

## 2.2 Improved MV Copy Frame Interpolation Scheme

The skipped frames during encoding should be interpolated by using adjacent coded key frames to achieve constant frame rate at the decoder side. There are several frame interpolation methods such as forward frame repetition, bidirectional frame interpolation, and motion compensated frame interpolation (MCFI) [24]. To be simple and not introduce delay, an improved frame recovery scheme considering frameskip based on 4×4 block MV copy frame error concealment (EC) method [25] is proposed in this work to interpolate the skipped frames with high quality and motion continuity at the decoder side.

Due to motion continuity existing in adjacent frames of local video sequences extensively, the MV fields of adjacent frames also have intensive similarity. MV copy based frame EC method for fixed frame rate coding over lossy networks assumes that the missing frame has the same MV field as its previous frame, and then recover it using the copied block-based MVs and reference indices.

Considering the effect of frameskip during encoding, we improve MV copy based method from two aspects to estimate the 4×4 block MV field of skipped frames accurately. First, due to multiple reference frames allowed in H.264, the MVs in a coded frame may have different reference indices. In order to fully utilize the information of most adjacent frame when interpolating the skipped ones, all the 4×4 block MVs of the previous coded frame are rescaled to the most adjacent reference frame according to temporal distance. For example, as shown in **Fig. 1**, the coded key frames are $[f_{n-3},f_{n-2},f_n,f_{n+3}]$ and the skipped frames are $[f_{n-1},f_{n+1},f_{n+2}]$. The reference frames of $f_n$ are $f_{n-3}$ and $f_{n-2}$, so the MVs reference to $f_{n-3}$ should be first rescaled to $f_{n-2}$ with a multiplier (n-(n-2))/(n-(n-3))=2/3 when estimating the MV field of $f_{n+1}$.

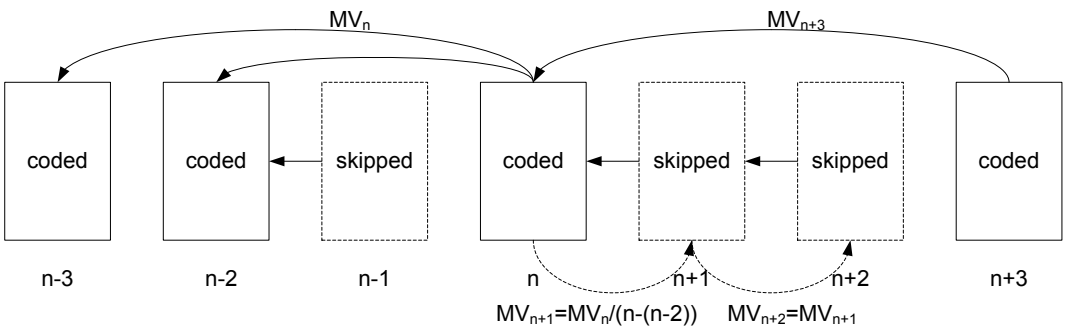



**Fig. 1.** Improved MV field estimation method with frameskip

Second, after all the MVs have been rescaled to the most adjacent reference frame, all the MVs in coded frame should be further rescaled to the immediate previous frame if there are

skipped frames between current coded and previous coded frame when estimating the MVs of next skipped frame. Also as shown in **Fig. 1**, there is a skipped frame $f_{n-1}$ between $f_n$ and $f_{n-2}$, all the MVs of $f_n$ should be further rescaled to reference $f_{n-1}$ with a distance multiplier (n-(n-1))/(n-(n-2))=1/2 when estimating the MV field of $f_{n+1}$. Then we can interpolate $f_{n+1}$ using the copied rescaled 4×4 block MVs and reference frame $f_n$ by block-based motion compensation. Similarly, we can interpolate the skipped frame $f_{n+2}$ using the copied MVs.

The improved MV copy frame interpolation scheme considering frameskip has three advantages. First, it does not involve any filter operation and thus has very low complexity. Second, it does not introduce any delay, which is expected by video applications. Finally, it preserves motion continuity among frames better, which is very pleasant to HVS.

### 2.3 SSIM-Based Content Adaptive Frame Skipping

Prior heuristic frame skipping methods are mostly based on objective quality metrics, which are not subjectively optimized and without considering the recovery quality of the skipped frames at the decoder side. To address the two problems, we first assume current frame is skipped and recover it using the improved MV copy frame interpolation scheme as described above, and then utilize SSIM-based subjective quality assessment to perform content adaptive frame skipping in this work.

Concretely, we first assume current frame $f_n$ is skipped, and estimate its 4×4 block MV field from previous coded frame. Then we can get the interpolation frame $\overline{f}_n$ by its copied MVs and reference frame $\hat{f}_{n-1}$. Thereafter, we can calculate the SSIM value $SSIM(f_n, \overline{f}_n)$ between current original frame $f_n$ and the corresponding interpolation frame $\overline{f}_n$ to assess the subjectively recovery quality of $f_n$ if it is skipped and recovered by MV copy scheme. $SSIM(f_n, \overline{f}_n)$ is also the index of subjective similarity and motion correlation between $f_n$ and $\hat{f}_{n-1}$, and the larger the value is means more probable to be skipped $f_n$ is, because $f_n$ can be well recovered from $\hat{f}_{n-1}$ based on copied MV field and thus reckoned as subjectively trivial.

Then the actual average SSIM value of previously frames in a sliding window with size $K$ is calculated as a reference, the formula is as follows:

$$\overline{SSIM}_{SW(K)} = \frac{1}{K}\sum_{k=1}^{K} SSIM(f_{n-k}) \tag{2}$$

where SSIM($f_{n-k}$) is the actual SSIM value of $f_{n-k}$ no matter whether it is encoded or skipped. Namely if $f_{n-k}$ is encoded, the SSIM is calculated between the original and its reconstructed frame; or if it is skipped, the SSIM is calculated between the original and the interpolated frame. That is:

$$SSIM(f_{n-k}) = \begin{cases} SSIM(f_{n-k}, \hat{f}_{n-k}), & if\ (f_{n-k}\ is\ encoded) \\ SSIM(f_{n-k}, \overline{f}_{n-k}), & if\ (f_{n-k}\ is\ skipped) \end{cases} \tag{3}$$

where $\overline{f}_{n-k}$ is the interpolated frame based on the improved MV copy scheme.

The size $K$ of sliding window reflects how fast the frame skipping scheme responses to locally temporal subjective quality change in a video. A large $K$ means slow response while a small $K$ means fast response. In our experiment, $K$=2 is used to obtain a tradeoff between algorithm robustness and fast response.

The current frame $f_n$ will be skipped if the following condition is satisfied:

$$SSIM(f_n, \bar{f}_n) \geq \overline{SSIM}_{SW(K)} \times T_s \tag{4}$$

where the threshold $T_s$ is empirically set as 0.98 as its initial value and will be adjusted frame by frame according to current buffer fullness and target buffer level in our scheme.

### 2.4 Buffer Fullness Based Forced Frame Skipping and Threshold Update

Buffer fullness should be maintained at safe levels over the whole encoding process, because buffer overflow will cause the encoded bit stream to be forced dropped and this will result in a huge waste of encoding resource such as bits and time. Note that there is a significant difference between buffer overflow and active frame skipping, the former will drop encoded bit stream while the latter can save bits and time from skipped frames. Buffer underflow should also be avoided because it will cause a waste of buffer and bandwidth resource, and the decoder will have to wait until bits of next frame comes, which degrades spatial-temporal video quality severely.

So buffer fullness based frame skipping is also employed in this work to avoid buffer overflow and underflow. We set two buffer fullness boundaries, the lower threshold $T_1$ is 10% of buffer size and the upper threshold $T_2$ is 90% of buffer size. If the current buffer fullness ($b_c$) is below the lower threshold $T_1$, the current frame will be forcibly encoded regardless of SSIM value. Similarly if $b_c$ is above the upper threshold $T_2$, the current frame will be forcibly skipped regardless of SSIM value. If $b_c$ is between $T_1$ and $T_2$, we think $b_c$ is in the safe levels and whether to skip the current frame will be decided by our adaptive SSIM-based scheme.

The higher $b_c$ is relative to the predefined target buffer level ($Tbl$), the more probable the current frame will be skipped, and vice versa. So when $b_c$ is between $T_1$ and $T_2$, we correspondingly adjust $T_s$ by a linear function of the difference between $b_c$ and $Tbl$. The linear function is as follows:

$$T_s = 0.98 \times [1 - 0.025 \times (b_c / B_s - Tbl / B_s)] \tag{5}$$

where $B_s$ is the used buffer size. When $b_c$ is above $Tbl$, $T_s$ is adjusted downwards to increase the possibility of frameskip. While $b_c$ is below $Tbl$, $T_s$ is correspondingly adjusted upwards to decrease the possibility of frameskip.

## 3. Integrated Frame Skipping Scheme Combined with JVT-G012

### 3.1 Framework of Proposed Adaptive Frame Skipping Scheme

Our active frame skipping scheme can be combined with any rate control algorithm for H.264 video coding. In this work, our scheme is employed with classic JVT-G012 frame layer rate control to validate encoding performance, and the overall flowchart is summarized in **Fig. 2**. Note that in our scheme, the initial I frame and the last P frame of each GOP are forced to encode. Considering motion continuity and the residual between frames will become large as

temporal distance increases, which will degrade the coding performance, the maximal number of successive frame skipping is set to 3. The maximal QP change between adjacent coded P frames is set to 4 considering frameskip instead of 2 in original fixed frame rate control method JVT-G012.

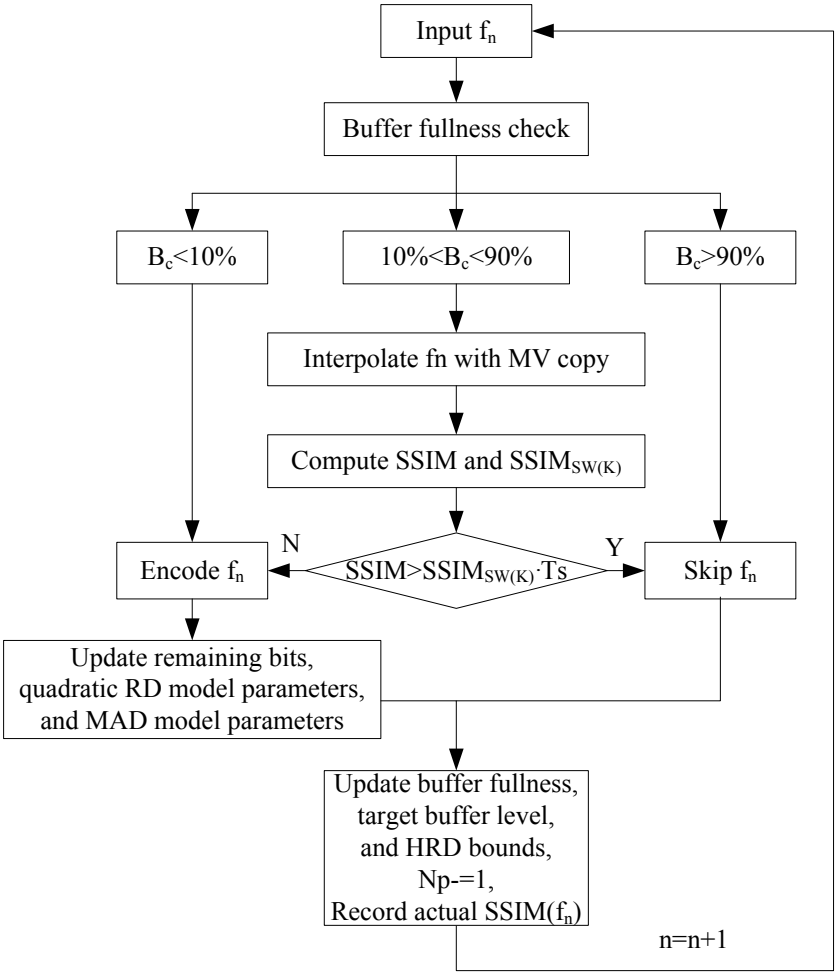



**Fig. 2.** Flowchart of integrated RC scheme with adaptive frame skipping

### 3.2 GOP Layer Rate Control

GOP layer rate control performs total target bits allocation for current GOP and remaining bits update after encoding a frame using the following formulas [1]:

$$T_r(n_{i,0}) = \frac{b_r}{f_r} \times N_{gop} - B_c(n_{i-1,N_{gop}}) \tag{6}$$

$$T_r(n_{i,j}) = T_r(n_{i,j-1}) - b(n_{i,j-1}) \tag{7}$$

where $b_r$ is the target bit rate, $f_r$ is the predefined frame rate, $N_{gop}$ is the number of frames in a GOP, $B_c(n_{i-1,N_{gop}})$ is the actual buffer occupancy at the beginning of current GOP, $b(n_{i,j-1})$ is the actual number of bits of previous coded frame.

### 3.3 Frame Layer Rate Control with Adaptive Frame Skipping

The detailed step by step frame layer rate control with our proposed adaptive frame skipping scheme is described as follows:

Step 1: If current frame is the initial I frame of a GOP, the frame is forced to encode.

Step 1.1: Rate control initialization: Compute $T_r(n_{i,0})$ with (6), initialize the number of successive frameskip $f_s$=0, initialize the number of remaining non-coded P frames $N_p$=$N_{gop}$-1, initialize current buffer fullness $B_c(n_{1,1})$=0 for the first GOP and $B_c(n_{i+1,0}) = B_c(n_{i,N_{gop}})$ for other GOPs, initial delay offset is 0.8×$B_s$, the lower bound and upper bound of target bits of each frame considering the hypothetical reference decoder (HRD) are initialized as: $L(n_{i,1})=b_r/f_r$, $U(n_{i,1})=InitialDelayOffset$.

Step 1.2: The starting QP is determined by bits per pixel (*bpp*) for the first GOP and by the average QP of all coded P frames excluding the skipped frames in the previous GOP for other GOPs. The formula is as follows:

$$QP_{st} = \begin{cases} QP_0, \textit{ for the first GOP} \\ \dfrac{Sum_{pc,QP}}{N_{pc}} - \min\{2, \dfrac{N_{gop}}{15}\}, \textit{ for other GOPs} \end{cases} \tag{8}$$

where $QP_0$ is empirically determined by *bpp* [1], $Sum_{pc,QP}$ and $N_{pc}$ are the sum of QP and the total number of all actually coded P frames in the previous GOP excluding the skipped ones.

Step 1.3: After encoding a frame, update $T_r(n_{i,j})$ with (7), and update buffer fullness and HRD bounds as follows:

$$B_c(n_{i,j+1}) = B_c(n_{i,j}) + b(n_{i,j}) - \frac{b_r}{f_r} \tag{9}$$

$$L(n_{i,j+1}) = L(n_{i,j}) + \frac{b_r}{f_r} - b(n_{i,j}) \tag{10}$$

$$U(n_{i,j+1}) = U(n_{i,j}) + (\frac{b_r}{f_r} - b(n_{i,j})) \times \omega \tag{11}$$

where $\omega$=0.9 is used in our experiment.

Step 2: Determine $f_s \in \{0,1,2,3\}$ and the next P frame need to be actually encoded by our proposed adaptive frameskip scheme. Then update the following items:

$$N_p -= f_s \tag{12}$$

$$B_c(n_{i,j+1}) = B_c(n_{i,j}) - \frac{b_r}{f_r} \times f_s \tag{13}$$

$$L(n_{i,j+1}) = L(n_{i,j}) + \frac{b_r}{f_r} \times f_s \tag{14}$$

$$U(n_{i,j+1}) = U(n_{i,j}) + (\frac{b_r}{f_r} \times f_s) \times \omega \tag{15}$$

Step 3: For the first P frame need to be actually encoded of current GOP by our scheme, its QP is the same as $QP_{st}$. After encoding, update $N_p$-=1, update $T_r(n_{i,j})$ with (7), and update buffer fullness and HRD bounds with (9), (10), and (11). And the initial target buffer level is reset as follows:

$$Tbl(n_{i,f_s^0+2}) = B_c(n_{i,f_s^0+2}) \tag{16}$$

where $f_s^0$ is the number of frames skipped between I frame and the first actually encoded P frame, $B_c(n_{i,f_s^0+2})$ is the buffer occupancy after the first actually encoded P frame. And the target buffer level for the subsequent P frames which are needed to be actually encoded by our scheme considering frameskip is determined by:

$$Tbl(n_{i,j+1+f_s}) = Tbl(n_{i,j}) - (f_s + 1) \times \frac{Tbl(n_{i,f_s^0+2})}{N_p - 1 - f_s^0} \tag{17}$$

Step 4: For other P frames need to be actually encoded by our scheme, the QPs are computed by quadratic R-D model as in [1] with some modifications due to frameskip.

Step 4.1: Target bits allocation before encoding: After updating the target buffer level and actual buffer fullness for current frame considering frameskip, the target bit is allocated as:

$$\tilde{f}(n_{i,j}) = \frac{b_r'}{f_r} + \gamma \times (Tbl(n_{i,j}) - B_c(n_{i,j})) \tag{18}$$

Meanwhile, after updating the remaining bits and the number of non-coded P frames of current GOP considering frameskip, the target bit is allocated as:

$$\hat{f}(n_{i,j}) = \frac{T_r(n_{i,j})}{N_p} \tag{19}$$

The target bit is a weighted tradeoff of $\tilde{f}(n_{i,j})$ and $\hat{f}(n_{i,j})$:

$$f(n_{i,j}) = \beta \times \hat{f}(n_{i,j}) + (1 - \beta) \times \tilde{f}(n_{i,j}) \tag{20}$$

where $\gamma$=0.5 and $\beta$=0.5 are used in our experiment.

Finally, the target bit is bounded by the two updated bounds considering frameskip to be confirmed with the HRD constraints:

$$\begin{cases} f(n_{i,j}) = \max\{L(n_{i,j}), f(n_{i,j})\} \\ f(n_{i,j}) = \min\{U(n_{i,j}), f(n_{i,j})\} \end{cases} \quad (21)$$

Step 4.2: Predict the MAD value of current frame by linear MAD model as follows:

$$MAD_c = a_1 \times MAD_{pc} + a_2 \quad (22)$$

where $MAD_{pc}$ is the actual MAD value of previous coded P frame, $a_1$ and $a_2$ are two parameters of the linear model and will be updated after actually encoding a P frame.

Step 4.3: Compute $QP_c$ for current coded frame by quadratic R-D model and then bounded by maximal QP change compared to the QP of previous actually coded P frame ($QP_{pc}$):

$$QP_c = \min\{QP_{pc} + 4, \max\{QP_{pc} - 4, QP_c\}\} \quad (23)$$

And the final QP is further bounded by the allowed QP range [0, 51] in H.264.

Step 4.4: After encoding, update $N_p$-=1, update $T_r(n_{i,j})$ with (7), and update buffer fullness and HRD bounds with (9), (10), and (11). Moreover, the quadratic R-D model parameters and the linear MAD model parameters should also be updated using linear regression technique after actually encoding two P frames of current GOP.

Step 5: Then go to step 2 until all frames in current GOP are processed.

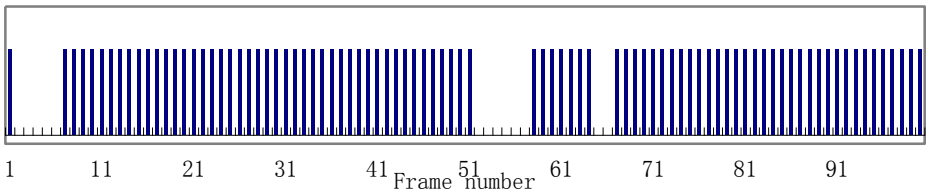
Step 6: Then go to Step 1 until all the GOPs of the video sequence are processed.

## 4. Experimental Results and Analysis

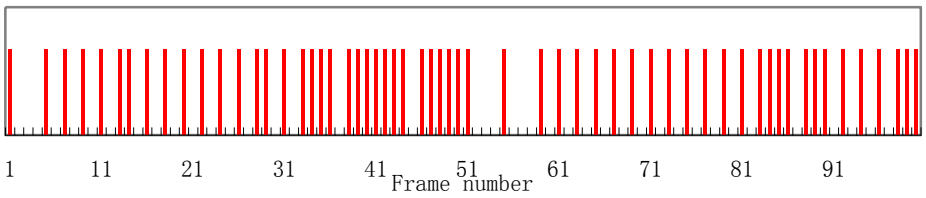
We implemented the proposed SSIM-based adaptive frameskip scheme, the objective metric based frameskip scheme [11] and the passive buffer-based frameskip scheme [1] in JM10.2 to evaluate the coding performance based on JVT-G012 frame layer RC. Note that JVT-G012 only describes buffer-based frameskip method, but the frame skipping step is not actually implemented in original JM10.2 reference software. Our experiments were conducted to the first 100 frames of several standard QCIF and CIF test sequences with YUV4:2:0 format at various target bit rate in baseline profile. The original frame rate is 30fps and the number of reference frame is 1. GOP coding structure is IPPP with GOP length 50. The motion search range is 16 and motion estimation (ME) accuracy is 1/4 pixel. The rate distortion optimization (RDO) is enabled both in mode decision and ME. Buffer size ($B_s$) is set as $b_r$/6 for low bit rate and low delay situations. Note that the skipped frames are all recovered from previous encoded frames by our improved MV copy based frame interpolation scheme in calculation of the average PSNR and SSIM for Y component of the whole sequence at the decoder side.

**Fig. 3** shows the coded key frames in akiyo QCIF sequence by JVT-G012 and the proposed scheme at 32kbps. Each bar indicates a coded key frame. Obviously our scheme skips subjectively trivial frames adaptively and uniformly across the whole sequence while JVT-G012 skips frames successively only by buffer fullness regardless of the content of frames. Similar test results for CIF sequence can also be observed in **Fig. 4**, in which container CIF sequence are encoded at 128kbps.

**Fig. 5** and **Fig. 6** show the PSNR and SSIM comparisons of the reconstructed akiyo QCIF sequence at 32kbps and reconstructed container CIF sequence at 128kbps by JVT-G012 and our scheme. Though our scheme does not aim to improve PSNR, actually our scheme acquires better spatial-temporal quality both in PSNR and SSIM index, and especially the subjective quality improvement measured by SSIM is more significant compared to PSNR improvement, which proves that our scheme acquires more pleasant perceptual quality.

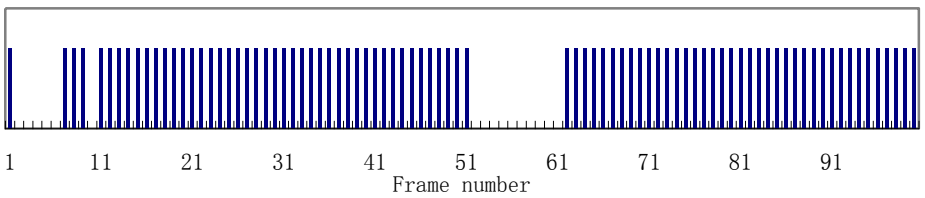



**(a)**

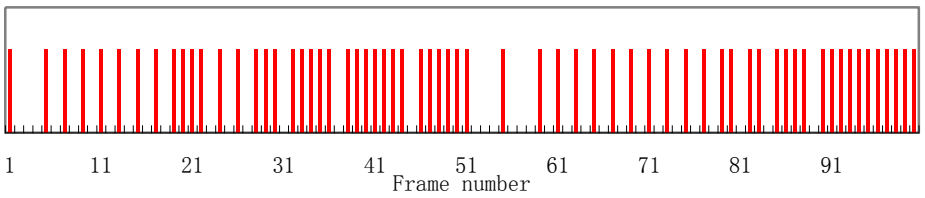



**(b)**

**Fig. 3.** Coded key frames in akiyo_qcif at 32kbps by **(a)** JVT-G012 and **(b)** proposed scheme

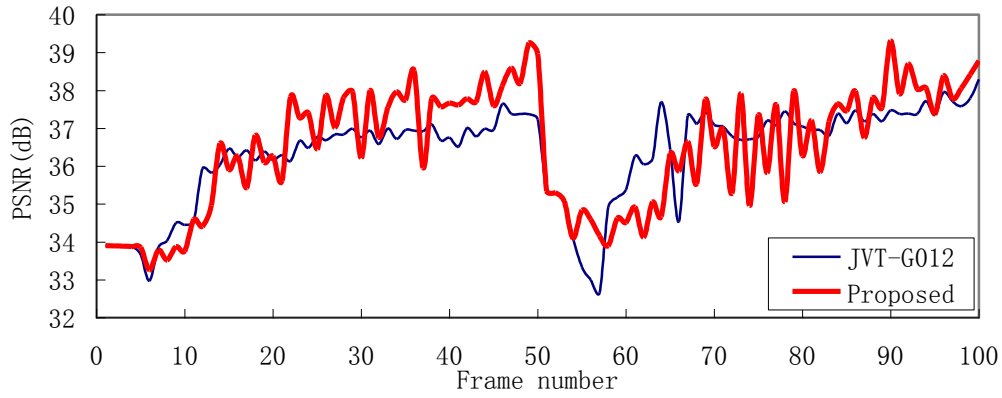



**(a)**

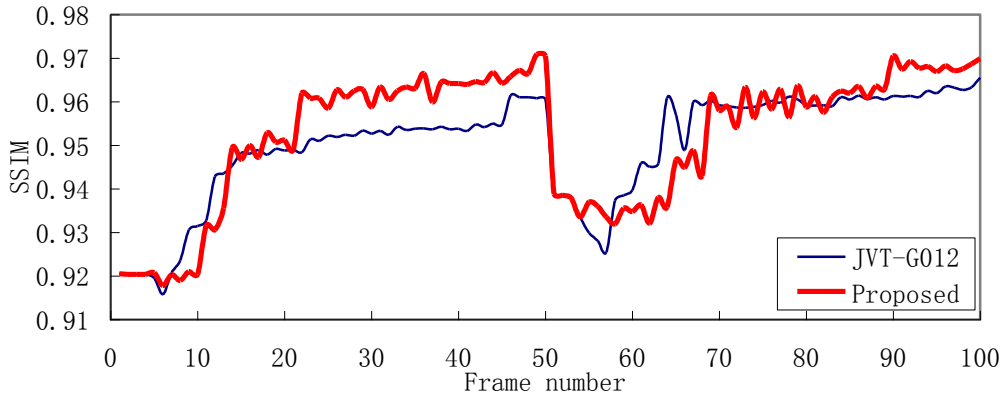



**(b)**

**Fig. 4.** Coded key frames in container_cif at 128kbps by **(a)** JVT-G012 and **(b)** proposed scheme

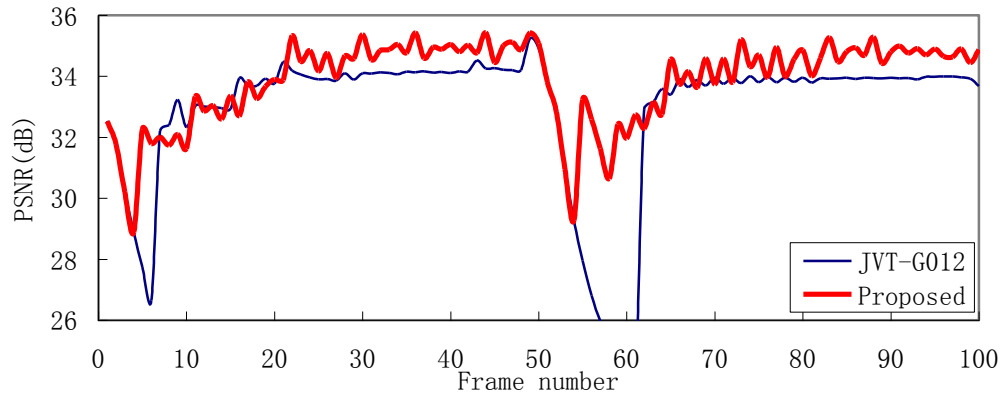



**(a)**

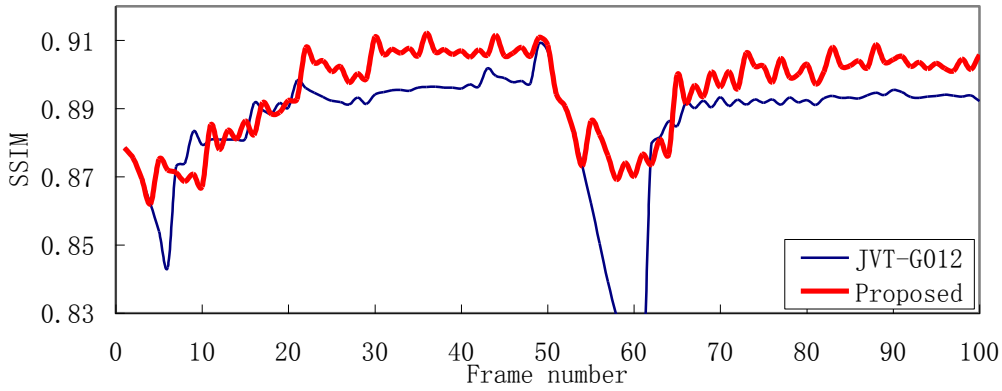



**(b)**

**Fig. 5.** Reconstructed frames quality comparison for akiyo_qcif at 32kbps at the decoder side by **(a)** objective quality (PSNR) and **(b)** subjective quality (SSIM)




**(a)**




**(b)**

**Fig. 6.** Reconstructed frames quality comparison for container_cif at 128kbps at the decoder side by **(a)** objective quality (PSNR) and **(b)** subjective quality (SSIM)

**Fig. 7** shows the subjective quality comparison of reconstructed 90th frame of coastguard CIF sequence at 128kbps at the decoder side by JVT-G012, objective metric based frameskip scheme [11] and our SSIM-based frameskip scheme. From which we can see that our scheme acquires better and more natural subjective effect, and the reconstructed frame by our scheme is more similar to the original frame perceptually. Please check the person in small boat and the windows of the building. The subjective quality improvement proves the effectiveness of our SSIM-based frameskip scheme.

**(a)** **(b)**

**(c)** **(d)**

**Fig. 7.** Subjective quality comparison of reconstructed 90th frame of coastguard_cif at 128kbps by **(a)** JVT-G012, **(b)** objective metric based scheme, **(c)** proposed scheme, and **(d)** the original frame

More experimental results between our proposed scheme and JVT-G012 for QCIF test sequences with different motion activity (low: container, grandma; moderate: foreman, news; high: carphone, salesman) at various target bit rate are shown in **Table 1**. From which we can see clearly that, the PSNR and SSIM for all test sequences are consistently improved greatly compared to JVT-G012, especially at low bit rate situations (32kbps). For example, for container sequence with low motion, the average PSNR gain is 0.6 dB while the average SSIM gain is 0.0036. For foreman sequence with moderate motion, the average PSNR gain is 0.61 dB while the average SSIM gain is 0.0161. And for carphone sequence with high motion, the average PSNR gain is 0.49 dB while the average SSIM gain is 0.0067. These results prove that our scheme is valid for sequences with different motion degree, and also indicate that PSNR is not proportional to SSIM. And SSIM gain is more desired because it is more relevant to HVS, which is the effort of this work.

**Table 1.** Performance comparisons for QCIF sequences at various target bit rate

| Sequence | Target bit rate (kbps) | Actual bit rate (kbps) | | Actual encoding frame number | | PSNR(dB) | | | SSIM | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | JVT-G012 | Proposed | JVT-G012 | Proposed | JVT-G012 | Proposed | Gain | JVT-G012 | Proposed | Gain |
| Container | 32 | 31.99 | 32.40 | 78 | 57 | 33.53 | 34.13 | **0.60** | 0.9258 | 0.9294 | **0.0036** |
| | 48 | 48.18 | 48.01 | 88 | 63 | 35.38 | 35.72 | **0.34** | 0.9364 | 0.9402 | **0.0038** |
| Grandma | 32 | 32.65 | 32.52 | 82 | 56 | 34.59 | 34.89 | **0.30** | 0.9137 | 0.9205 | **0.0068** |
| | 48 | 49.11 | 50.31 | 91 | 66 | 36.10 | 36.29 | **0.19** | 0.9359 | 0.9377 | **0.0018** |
| Foreman | 32 | 32.35 | 32.28 | 69 | 68 | 28.14 | 28.75 | **0.61** | 0.8492 | 0.8653 | **0.0161** |
| | 48 | 48.06 | 48.36 | 89 | 78 | 30.57 | 30.61 | **0.04** | 0.8967 | 0.9018 | **0.0051** |
| News | 32 | 32.68 | 32.45 | 75 | 56 | 30.86 | 31.37 | **0.51** | 0.9156 | 0.9241 | **0.0085** |
| | 48 | 48.57 | 48.41 | 88 | 57 | 32.55 | 32.89 | **0.34** | 0.9355 | 0.9395 | **0.0040** |
| Carpho | 32 | 32.18 | 32.10 | 78 | 76 | 30.94 | 31.43 | **0.49** | 0.9107 | 0.9174 | **0.0067** |

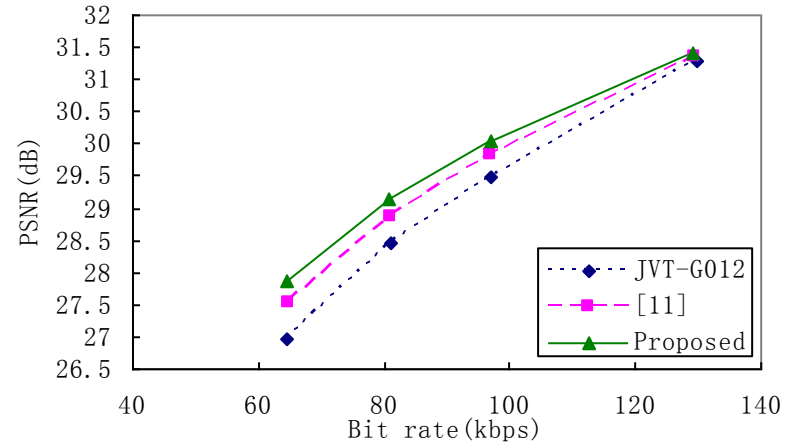
| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ne | 48 | 48.08 | 48.15 | 92 | 86 | 32.91 | 33.27 | **0.36** | 0.9358 | 0.9382 | **0.0024** |
| Salesman | 32 | 32.10 | 32.01 | 76 | 50 | 31.15 | 31.59 | **0.44** | 0.8793 | 0.8872 | **0.0079** |
| | 48 | 49.37 | 47.72 | 86 | 60 | 33.02 | 33.19 | **0.17** | 0.9094 | 0.9160 | **0.0066** |

**Table 2.** Performance comparisons for CIF sequences at various target bit rate

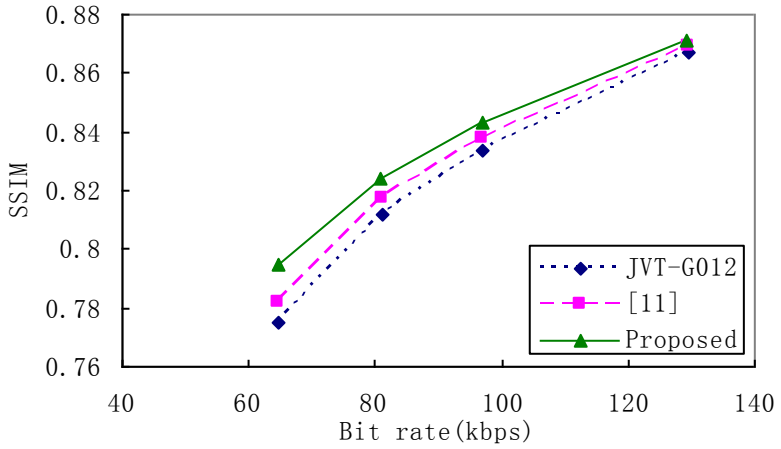
| **Sequence** | **Target bit rate (kbps)** | **Actual bit rate (kbps)** | | **Actual encoding frame number** | | **PSNR(dB)** | | | **SSIM** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **JVT-G012** | **Proposed** | **JVT-G012** | **Proposed** | **JVT-G012** | **Proposed** | **Gain** | **JVT-G012** | **Proposed** | **Gain** |
| Monitor | 64 | 64.46 | 64.26 | 76 | 57 | 32.24 | 32.60 | **0.36** | 0.9264 | 0.9297 | **0.0033** |
| | 80 | 80.25 | 80.34 | 81 | 57 | 33.13 | 33.20 | **0.07** | 0.9317 | 0.9327 | **0.0010** |
| Container | 96 | 96.23 | 96.69 | 79 | 55 | 31.94 | 33.12 | **1.18** | 0.8748 | 0.8862 | **0.0114** |
| | 128 | 128.16 | 129.45 | 84 | 65 | 32.94 | 33.87 | **0.93** | 0.8859 | 0.8950 | **0.0091** |
| Foreman | 96 | 96.97 | 96.87 | 72 | 78 | 29.49 | 30.04 | **0.55** | 0.8334 | 0.8433 | **0.0099** |
| | 128 | 129.67 | 129.20 | 87 | 80 | 31.30 | 31.41 | **0.11** | 0.8670 | 0.8708 | **0.0038** |
| News | 64 | 65.32 | 64.58 | 67 | 60 | 30.62 | 31.27 | **0.65** | 0.9047 | 0.9108 | **0.0061** |
| | 80 | 81.16 | 80.24 | 73 | 59 | 31.70 | 32.09 | **0.39** | 0.9162 | 0.9220 | **0.0058** |
| Coast guard | 96 | 95.73 | 96.02 | 60 | 66 | 24.53 | 25.70 | **1.17** | 0.5904 | 0.6334 | **0.0430** |
| | 128 | 128.52 | 127.83 | 73 | 82 | 25.74 | 26.57 | **0.83** | 0.6465 | 0.6740 | **0.0275** |
| Harbour | 128 | 128.22 | 128.07 | 63 | 73 | 23.53 | 24.25 | **0.72** | 0.6967 | 0.7240 | **0.0273** |
| | 192 | 192.68 | 192.73 | 82 | 83 | 25.51 | 25.77 | **0.26** | 0.7883 | 0.7995 | **0.0112** |

More experimental results between our proposed scheme and JVT-G012 for CIF test sequences with different motion activity (low: monitor, container; moderate: foreman, news; high: coastguard, harbour) at various target bit rate are shown in **Table 2**. We can see that the PSNR gain for container sequence with low motion at 96kbps is 1.18 dB and the SSIM gain is 0.0114. And the PSNR gain for foreman sequence with moderate motion is 0.55 dB and the SSIM gain is 0.0099. While for coastguard sequence with high motion the PSNR gain is 1.17 dB and the SSIM gain is 0.0430. And the PSNR and SSIM gain at low bit rate are usually more remarkable than at high bit rate for most test sequences, which further proves the application situations of the proposed frameskip scheme.

**Fig. 8** and **Fig. 9** show the coding subjective and objective rate distortion performance for foreman and coastguard CIF sequences at various bit rate by JVT-G012, objective metric based frameskip scheme [11] and our proposed scheme. We can see that objective metric based frameskip scheme improves PSNR to some degree compared to JVT-G012, but the SSIM gain is relatively slight. While our scheme simultaneously improves the PSNR and SSIM index more significantly, especially the coding gain at low bit rate is more obvious than the gain at high bit rate.

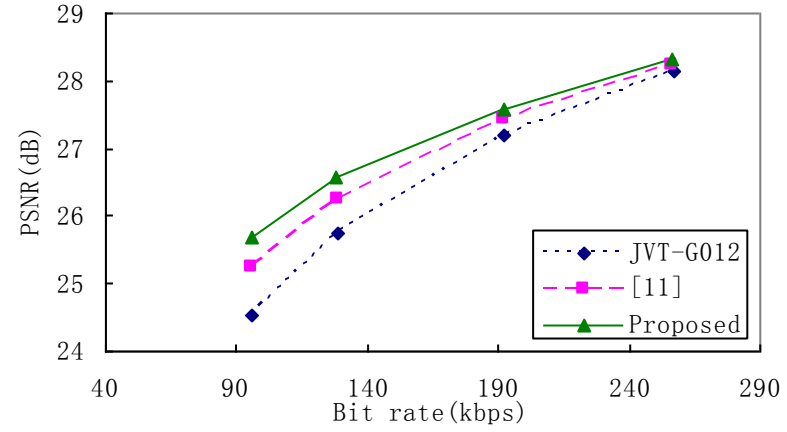



**(a)**

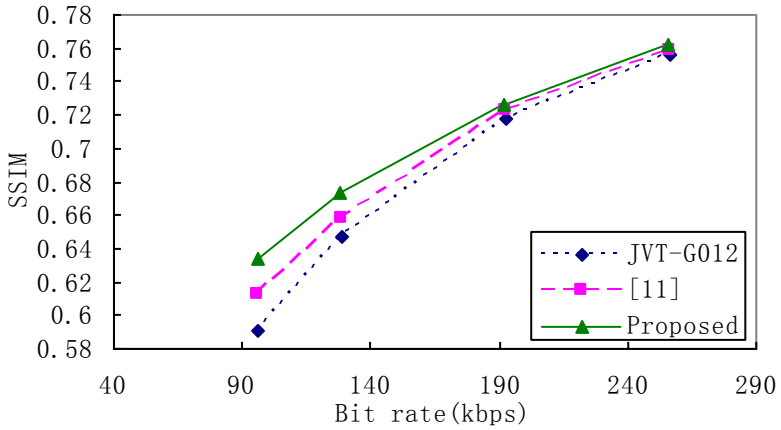



**(b)**

**Fig. 8.** Performance comparison for foreman_cif, **(a)** PSNR-rate curve and **(b)** SSIM-rate curve



**(a)**




**(b)**

**Fig. 9.** Performance comparison for coastguard_cif, **(a)** PSNR-rate curve and **(b)** SSIM-rate curve

The computation complexity of our proposed scheme combined with frame layer RC is comparative to JVT-G012 frame layer RC with buffer-based frameskip, since the overhead mainly come from the 4×4 block based motion compensated frame interpolation (MCFI) and non-lapped 8×8 square block based SSIM calculation when judging if a frame should be skipped, which consumes a few computation but has relative low complexity compared with computation intensive rate distortion optimization (RDO) based subpixel motion estimation (ME) and mode decision processes during encoding. The total encoding time comparisons for the first 100 frames of QCIF test sequences at various target bit rate between JVT-G012 and our porposed scheme are shown in **Table 3**. From which we can see that, our scheme usually has less encoding time than JVT-G012 for most sequences, that is because our scheme actually encodes fewer frames and skips more frames. For example, for container sequence at 48kbps, JVT-G012 actually encodes 88 frames using 124.06s while our scheme actually encodes 63 frames using 102.49s. While for froeman sequence at 32kbps, JVT-G012 actually encodes 69 frames using 77.02s and our scheme encodes 68 frames using 96.64s, which proves our scheme will consume a litter more time due to MCFI and SSIM computation if the two schemes actually enocde almost the same number of frames. So our proposed scheme has low complexity and is completely appropriate for low delay real-time applications.

**Table 3..** Total encoding comparisons for QCIF sequences at various target bit rate

| **Sequence** | **Target bit rate (kbps)** | **Actual encoding frame number** | | **Total encoding time (s)** | | |
|---|---|---|---|---|---|---|
| | | **JVT-G012** | **Proposed** | **JVT-G012** | **Proposed** | **Δ Time (%)** |
| Container | 32 | 78 | 57 | 94.91 | 63.18 | **-33.43** |
| | 48 | 88 | 63 | 124.06 | 102.49 | **-17.39** |
| Grandma | 32 | 82 | 56 | 104.94 | 78.28 | **-25.41** |
| | 48 | 91 | 66 | 128.58 | 93.55 | **-27.24** |
| Foreman | 32 | 69 | 68 | 77.02 | 96.64 | **+25.47** |

| | | | | | | |
|---|---|---|---|---|---|---|
| | 48 | 89 | 78 | 112.03 | 107.48 | **-4.06** |
| News | 32 | 75 | 56 | 97.50 | 61.92 | **-36.49** |
| | 48 | 88 | 57 | 115.36 | 90.81 | **-21.28** |
| Carphone | 32 | 78 | 76 | 90.95 | 92.11 | **+1.28** |
| | 48 | 92 | 86 | 109.43 | 110.12 | **+0.63** |
| Salesman | 32 | 76 | 50 | 105.52 | 78.86 | **-25.27** |
| | 48 | 86 | 60 | 112.41 | 87.82 | **-21.88** |

## 5. Conclusion

In this work, a novel SSIM-based adaptive frameskip scheme with low complexity is proposed. There are four major contributions in our work. First, active content adaptive frame skipping is used instead of passive or forced frame skipping to improve coding performance and spatial-temporal quality especially for low bit rate H.264 video coding. Second, the recovery quality of skipped frames at the decoder side is considered when designing the frameskip strategy during encoding. Third, SSIM-based subjective quality metric is used to determine the number of frame skipped by assessing recovery quality if the frame is skipped and the similarity between two successive frames instead of objective quality difference such as MAD. Finally, an integrated frame layer RC scheme with our adaptive frameskip strategy is proposed and compared with classic JVT-G012 and objective metric based frameskip scheme. Experimental results show that our scheme acquires improved overall coding performance and better balance of spatial-temporal quality especially for low bit rate video coding. The average objective quality (PSNR) and subjective quality (SSIM) are both improved a lot, and the significant improvement in perceptual quality can be observed clearly compared with passive buffer-based frameskip scheme JVT-G012 and prior objective metric based frameskip.

## References

[1] Z. G. Li, F. Pan, K. P. Lim, G. N. Feng, X. Lin and S Rahardja, "Adaptive basic unit layer rate control for JVT," *Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG. JVT-G012, 7th Meeting*, Mar. 2003. Article (CrossRef Link)

[2] W. Yuan, S. Lin, Y. Zhang, H. Luo and W. Yuan, "Optimum bit allocation and rate control for H.264/AVC," *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6) 15th Meeting, JVT-O016*, Apr. 2005. Article (CrossRef Link)

[3] D. K. Kwon, M. Y. Shen and C. C. J. Kuo, "Rate control for H.264 video with enhanced rate and distortion models," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 5, pp. 517-529, May. 2007. Article (CrossRef Link)

[4] S. S. Rodriguez, O. Esteban, M. Lopez and F. Maria, "Cauchy-density-based basic unit layer rate controller for H.264/AVC," of *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 8, pp. 1139-1143, Aug. 2010. Article (CrossRef Link)

[5] L. Xu, D. B. Zhao, X. Y. Ji, S. Kwong and W. Gao, "Window-level rate control for smooth picture quality and smooth buffer occupancy," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 723-723, Mar. 2011. Article (CrossRef Link)

[6] T. S. Ou, Y. H. Huang and H. H. Chen, "SSIM-based perceptual rate control for video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 5, pp. 682-291, May. 2011. Article (CrossRef Link)

[7] J. Chen and H. Hang, "Source model for transform video coder and its application-Part II: Variable frame rate coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 2, pp. 299-311, Feb. 1997. Article (CrossRef Link)
[8] E. C. Reed and F. Dufaux, "Constrained bit-rate control for very low bit-rate streaming-video applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 7, pp. 882-889, Jul. 2001. Article (CrossRef Link)
[9] E. Reed and J. Lim, "Optimal multidimensional bit-rate control for video communication," *IEEE Transactions on Image Processing*, vol. 11, no. 8, pp. 873-885, Oct. 2002. Article (CrossRef Link)
[10] F. Pan, X. Lin, S. Rahardja, K. P. Lim, Z. G. Li, D. J. Wu and S. Wu, "Proactive frame-skipping decision scheme for variable frame rate video coding," *IEEE International Conference on Multimedia and Expo(ICME)*, pp. 1903-1906, Jun. 2004. Article (CrossRef Link)
[11] F. Pan, Z. Lin, X. Lin, S. Rahardja, W. Juwono and F. Slamet, "Adaptive frame skipping based on spatio-temporal complexity for low bit-rate video coding," *Journal of Visual Communication and Image Representation*, vol. 17, no. 3, pp. 554-563, 2006. Article (CrossRef Link)
[12] A. Thammineni, A. Raman, S. C. Vadapalli and S. Sethuraman, "Dynamic frame-rate selection for live LBR video encoders using trial frames," *IEEE International Conference on Multimedia and Expo (ICME)*, pp. 817-820, Jun. 2008. Article (CrossRef Link)
[13] J. Jun, S. Lee, Z. He, M. Lee and E. S. Jang, "Adaptive key frame selection for efficient video coding," *Lecture Notes in Computer Science on Advances in Image and Video Technology*, vol. 4872, pp. 853-866, 2007. Article (CrossRef Link)
[14] P. Usach, J. Sastre and J. M. Lopez, "Variable frame rate and gop size H.264 rate control for mobile communications," *IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1772-1775, Jul. 2009. Article (CrossRef Link)
[15] S. Liu and C. C. J. Kuo, "Joint temporal-spatial bit allocation for video coding with dependency," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 1, pp. 15-26, Jan. 2005. Article (CrossRef Link)
[16] A. Vetro, Y. Wang and H. F. Sun, "Rate-distortion optimized video coding considering frameskip," *International Conference on Image Processing (ICIP)*, pp. 534-537, Oct. 2001. Article (CrossRef Link)
[17] J. Lee, A. Vetro, Y. Wang and Y. Ho, "Bit allocation for MPEG-4 video coding with spatio-temporal tradeoffs," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 6, pp. 488-502, 2003. Article (CrossRef Link)
[18] S. Hwangjun and C. C. J. Kuo, "Rate control for low-bit-rate video via variable-encoding frame rates," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 4, pp. 512-521, Apr. 2001. Article (CrossRef Link)
[19] Y. T. Yang, Y. S. Tung and J. L. Wu, "Quality enhancement of frame rate up-converted video by adaptive frame skip and reliable motion extraction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 12, pp. 1700-1713, Dec. 2007. Article (CrossRef Link)
[20] Y. Chung, "A novel selective frame discard method for 3D video over IP networks," *KSII Transactions on Internet and Information Systems*, vol. 4, no. 6, pp. 1209-1221, Dec. 2010. Article (CrossRef Link)
[21] Y. S. Hong and H. Park, "A cost-effective rate control for streaming video for wireless portable devices," *KSII Transactions on Internet and Information Systems*, vol. 5, no. 6, pp. 1147-1165, Jun. 2011. Article (CrossRef Link)
[22] Z. Wang, A. C. Bovik, H. R. Sheikh E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, Apr. 2004. Article (CrossRef Link)
[23] http://iphome.hhi.de/suehring/tml.download/.
[24] Z. Gan, L. Qi and X. Zhu, "Motion compensated frame interpolation based on H.264 decoder," *Electronics Letters*, vol. 43, no. 2, pp. 96-98, Jan . 2007. Article (CrossRef Link)
[25] S. Bandyopadhyay, Z. Wu, P. Pandit and J. Boyce, "Frame loss error concealment for H.264/AVC," *ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6) 16th Meeting, JVT-P072*, Jul. 2005. Article (CrossRef Link)

**Ziguan Cui** received the M.S. degree from Nanjing University of Aeronautics and Astronautics, China, 2008. He is currently working toward the Ph.D. degree in signal and information processing at Nanjing University of Posts and Telecommunications, China. His research interests include video coding and transmission over wire and wireless channels, image processing.

**Zongliang Gan** received the Ph.D. degree in signal and information processing from Nanjing University of Posts and Telecommunications, China. He is currently a lecturer in the College of Communication and Information Engineering at Nanjing University of Posts and Telecommunications, China. His research interests include image processing, distributed video coding, and compressed sensing.

**Xiuchang Zhu** was born in 1947. He is currently a professor in the College of Communication and Information Engineering at Nanjing University of Posts and Telecommunications, China. His research interests include image processing, video analysis, and compressed sensing.