REAL-TIME REGION-OF-INTEREST VIDEO CODING USING CONTENT-ADAPTIVE BACKGROUND SKIPPING WITH DYNAMIC BIT REALLOCATION

Haohong Wang, Yi Liang, and Khaled El-Maleh
Qualcomm Inc., San Diego, USA

ABSTRACT

In this paper, we propose a low-complexity content-adaptive background skipping scheme for region-of-interest (ROI) video coding in real-time wireless video telephony applications. Based on the real-time content information of the current and previous frames, such as foreground shape deformation, foreground and background motion, and background texture complexity, the proposed algorithm dynamically decides whether to skip Non-ROI macroblocks of current frame and reallocate saved bits to ROI and coded non-ROI. In this work, we proposed and compared three bit reallocation strategies, and the one predicting the future events using a Bayesian model is found the most efficient. Experimental results indicate that the proposed scheme significantly outperforms other methods by up to 2.5dB.

1. INTRODUCTION

The rapid growth in interactive multimedia applications has resulted in spectacular strides in wireless signal processing and communications systems. As a very useful technique for video telephony, Region-of-Interest (ROI) video coding [1-5] provides users more flexibility and interactivity in specifying their desires and enable encoders more efficiency in controlling the visual quality of coded video sequences. This way, the perceptually-important region, for example human faces, can be coded at higher quality to effectively improve the subjective quality of the coded video sequence. In [1], a face model is used to assist encoding facial feature related areas with different quantization parameters and temporal resolution. [2] considers human visual sensitivity variations with eccentricity in macroblock-level quantizer assignment. In [3-4], finer quantizers are assigned to the foreground and coarser quantizers are assigned to the background. [4] uses filters on background to decrease its bitrate usage to save bits for the foreground. We proposed in [5] the first effort to develop an optimized ρ-domain (ρ represents the number or percentage of non-zero quantized AC coefficients in a macroblock in video coding [6]) bit allocation scheme for ROI video coding. The major difference between ρ-domain and QP-domain rate control model used in [1-4] is that the former model is more accurate and thus effectively reduces the rate fluctuations.

In this work, we propose a generalized adaptive background skipping scheme that takes into account the frame content variation in skip mode decision and rate control. An initial effort of this exploration [5] uses a prototyped unit-based background skipping approach where every two consecutive frames are grouped into a unit in which the non-ROI of the second frame is skipped (not coded but replaced by the macroblocks of the first frame in the same locations) if the distortion caused by the skipping is smaller than a pre-defined threshold. In this paper, a much more flexible background skipping scheme is proposed which considers the real-time frame content statistics, such as foreground (or ROI) shape deformation, foreground and background (or non-ROI) movement, and the accumulated skipped distortion due to skipped background, in a jointly framework to make runtime skip mode decisions. In addition, the algorithm dynamically reallocates the saved bits due to skipping to other regions, and adjusts the bit allocation in both frame and macroblock levels. In this work, we propose and compare three different strategies for bit reallocate. An aggressive strategy that predicts the future possible events with Bayesian model and allocates resources based on the prediction is shown the most favorable.

The rest of the paper is organized as follows. Section 2 presents design philosophy; Section 3 proposes the adaptive background skipping approach; Section 4 proposes the dynamic bit reallocation strategies, and Section 5 demonstrates the experimental results. We draw conclusions in the last section.

2. BACKGROUND AND DESIGN PHILOSOPHY

In this work, we assume the ROI is known at the encoder, which is possible if the ROI is either automatically detected or specified by the end-user. A general problem is to find out the number and locations of the frames in a video sequence whose non-ROIs are to be skipped and the number of bits to be allocated to each frame that ensures the best visual quality of the video sequence. However, to obtain an ideal optimal solution is almost impossible, because (1) the visual quality metric that balances the spatial and temporal quality for a video sequence is still an open issue; (2) in real-time communication systems, the future frames normally are not available when the encoder processes the current frame, therefore the optimality is not achievable.

Our goal is to find a low-complexity practical solution for real-time applications that achieve good perceptual video quality. Some perceptual rationales, as used in [7-8], need

to be considered in our algorithm design, for example, the human visual system (HVS) is more sensitive to the temporal changes than to the spatial details when the frame contains high motion activities.

The proposed coding algorithm is briefly described as follows: For each encoded frame, an initial frame-level bit allocation is done by allocating available bits uniformly among the remaining frames in the rate control window. Then, based on a number of content cues, for example, ROI shape deformation and motion activities, and a set of pre-defined rules, the decision whether to skip the current non-ROI is made, and then the ρ budget for current frame is adjusted to favor the bit reallocation on ROI macroblocks. Next, an optimized macroblock-level bit allocation is conducted, and the frame is coded with the assigned quantizeation parameters.

Clearly, the proposed design favors both spatial and temporal quality. By background skipping and reallocating bits from non-ROIs to ROIs, the spatial visual quality of the frames are improved. On the other hand, the solution is content-adaptive in the sense that it follows some human perceptual rationales for improving temporal video quality. More details are presented in the following sections.

3. CONTEXT-ADAPTIVE BACKGROUND SKIPPING

In this work, the mode of background skipping is dynamically determined based on the content context such as background and foreground activities. Let us first define two filters $F(\{x_n\}, M, Th)$ and $G(\{x_n\}, M, Th)$, where $\{x_n\}$ is a set of real numbers in which x_n is the nth item, M is an integer number, Th is a percentage threshed, and

$$F(\{x_n\}, M, Th) = \begin{cases} 1 & x_n > (Th\,\%\, of\ items\ in\ x_{n-M}, ..., x_{n-1}) \\ 0 & otherwise \end{cases}, \quad (1)$$

and

$$G(\{x_n\}, M, Th) = \begin{cases} 1 & if\ \dfrac{x_n - x_{n-M}}{x_{n-M}} \geq Th\%, \\ 0 & otherwise \end{cases} \quad (2)$$

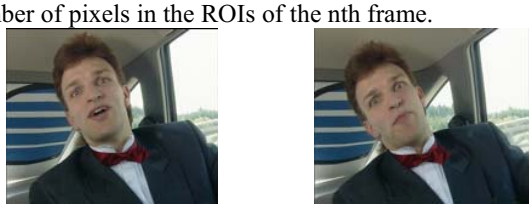
The filter (1) detects within a local window (fixed length of M) if the current value x_n is in the top position (above more than $Th\%$ of items), and the filter (2) detects if there is an increase from x_{n-M} to x_n by more than $Th\%$. These filters will be used in detecting the content status or status change, which indirectly affects the skip mode decision.

We consider both foreground and background activities in the framework. When large amount of motion occurs in background regions, the frequency of background skipping should be reduced. On the other hand, when the foreground contains large amount of activities, skipping background might be helpful to reallocate more bits to code the foreground. Let us denote by $\{\zeta_n\}$ the amount of foreground activities, which represents both global activities, such as object movement/rotation and shape deformation, and local activities such as change of facial expression (as shown in Fig. 1), and $\{\chi_n\}$ the amount of background activity, then we define these content cues as

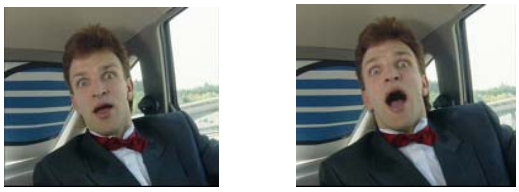
$$\chi_n = \sum_{i \in non-ROI} \left| MV_i \right|, \quad (3)$$

$$\zeta_n = \frac{C_n}{T_n} \sum_{i \in ROI} \left| MV_i \right|, \quad (4)$$

where MV_i is the motion vector of the ith macroblock in the nth frame, C_n is the total number of non-overlapped pixels in the ROIs of the (n-1)th and nth frames, and T_n is the total number of pixels in the ROIs of the nth frame.



(a) Global activity (face movement)



(b) Local activity (change of facial expression)

Figure 1. Examples of foreground activity

Let us denote by $\{\sigma_{B_n}^2\}$ the total energy of the background residue per frame for the video sequence, which is also the distortion due to skipped background. In this work we determine the skip mode by

$$\begin{aligned} S_n = {} & F(\{\zeta_n\}, M_2, Th_{\zeta 1}) G(\{\zeta_n\}, 1, Th_{\zeta 2}) + [1 - F(\{\zeta_n\}, M_2, Th_{\zeta 1}) \\ & G(\{\zeta_n\}, 1, Th_{\zeta 2})][1 - G(\{\sigma_{B_n}^2\}, p, Th_\sigma)][1 - F(\{\chi_n\}, M_1, Th_{\chi 1})] \\ & [1 - G(\{\chi_n\}, 1, Th_{\chi 2})], \end{aligned} \quad (5)$$

where $Th_\sigma, M_1, Th_{\chi 1}, Th_{\chi 2}, M_2$ and $Th_{\zeta 1}$ are thresholds and local window sizes defined by users, and p is an integer number that the $(n-p)$th frame codes background while the $(n-p+1)$th, $(n-p+2)$th, … and $(n-1)$th frames skip background. In Eq. (5), $S_n > 0$ means the background of the current frame is skipped, otherwise, it is coded. Clearly it can be observed that the algorithm chooses to skip background if the foreground contains large amount of activity and the amount is quickly increasing. In addition, the background will be coded if background contains large motion or the accumulated distortion due to skipped background is high.

4. DYNAMIC BIT REALLOCATION

4.1 Optimized macroblock-level ρ allocation

In this work, we use the same optimized macroblock-level bit allocation scheme in [5]. Let us denote by N the number of macroblocks in the frame, $\{\rho_i\}$ and $\{\sigma_i\}$ the set of ρ's

and standard deviation of the *i*th macroblock, and ρ_{budget} the total ρ budget for the frame, then the optimized ρ allocation on the *i*th macroblock can be represented by

$$\rho_i = \frac{\sqrt{w_i\sigma_i}}{\sum_{j=1}^{N}\sqrt{w_j\sigma_j}}\rho_{budget}, \quad (6)$$

where $$w_i = \begin{cases} \frac{\alpha}{K} & \text{if it belongs to ROI} \\ \frac{1-\alpha}{(N-K)} & \text{if it belongs to Non - ROI} \end{cases}, \quad (7)$$

where α is the ROI perceptual importance factor, and *K* is the number of macroblocks within the ROI.

4.2 Frame-level ρ reallocation

Let us denote by $\{\rho_n^{budget}\}$ the ρ budget obtained from the frame-level rate controller, as we mentioned in section 2, the budget is initially assigned as the average remaining ρ for the rest of the frames in the rate control window. Due to the skipping of the background, dynamic adjustment on the frame ρ budget is necessary.

We consider three types of bit reallocation strategies: (1) *Normal strategy*, which proportionally reduces the ρ budget based on the texture complexity of ROI and non-ROI when the background is skipped; (2) *Conservative strategy*, which proportionally reduces the ρ budget when the background is skipped, and saves these ρ 's for the latest future frame coding its background; (3) *Aggressive strategy*, which estimates the future skipping events based on the statistics and patterns of the previous background skipping history, and then determines the ρ budget based on the estimation. Mathematically, these strategies are represented by equations (8), (9) and (10) below respectively.

Let us denote by $\{\rho_n^{adjusted}\}$ the adjusted ρ budget, so for the *normal strategy*,

$$\rho_n^{adjusted} = \begin{cases} \rho_n^{budget} & \text{if } S_n = 0 \\ \frac{\sum_{i\in ROI}\sqrt{w_i\sigma_i}}{\sum_{i\in ROI}\sqrt{w_i\sigma_i} + \sum_{i\in NON-ROI}\sqrt{w_i\sigma_i}}\rho_n^{budget} & \text{otherwise} \end{cases}. \quad (8)$$

The *conservative strategy* is conceptually similar to the traditional banking operation, where a customer can cash the maximum of the total deposit of his/her account. In this case, the saving of ρ 's in frames with background skipping is deposited for the nearest future frame, whose background is coded. Hence,

$$\rho_n^{adjusted} = \begin{cases} p\rho_{n-p+1}^{budget} - \sum_{i=1}^{p-1}\rho_{n-i}^{adjusted} & \text{if } S_n = 0 \\ \frac{\sum_{i\in ROI}\sqrt{w_i\sigma_i}}{\sum_{i\in ROI}\sqrt{w_i\sigma_i} + \sum_{i\in NON-ROI}\sqrt{w_i\sigma_i}}\rho_n^{budget} & \text{otherwise} \end{cases}. \quad (9)$$

The *aggressive strategy* predicts the future possible events and allocates resources based on the prediction. Here, we assume that the future frames with skipped backgrounds have similar complexity in foreground as the current frame; therefore, once we estimate that there will be *q* frames with skipped background following the current frame, we can calculate the adjusted ρ budget by

$$\rho_n^{adjusted} = \begin{cases} p\rho_{n-p+1}^{budget} - \sum_{i=1}^{p-1}\rho_{n-i}^{adjusted} & \text{if } S_n = 0, n \le M \\ \frac{\sum_{i\in ROI}\sqrt{w_i\sigma_i} + \sum_{i\in NON-ROI}\sqrt{w_i\sigma_i}}{2(\sum_{i\in ROI}\sqrt{w_i\sigma_i} + \frac{1}{q+1}\sum_{i\in NON-ROI}\sqrt{w_i\sigma_i})}\rho_n^{budget} + & \\ \frac{p\rho_{n-p+1}^{budget} - \sum_{i=1}^{p-1}\rho_{n-i}^{adjusted}}{2} & \text{if } S_n = 0, n > M \\ \frac{\sum_{i\in ROI}\sqrt{w_i\sigma_i}}{\sum_{i\in ROI}\sqrt{w_i\sigma_i} + \sum_{i\in NON-ROI}\sqrt{w_i\sigma_i}}\rho_n^{budget} & \text{otherwise} \end{cases}, \quad (10)$$

where it acts exactly the same as the *conservative strategy* for the first *M* frames (*M* is an adjusted constant value). In this period, the statistics are collected for future *q* estimation. When $n>M$ and $S_n=0$, ρ is assigned an average value considering the previous saving and the predicted future saving due to background skipping. We estimate *q* by using a Bayesian model and convert the problem into a multi-class classification problem, where the classes are represented by all possibilities of *q*, and the feature vector used in making classification decision is $x_n=(\chi_n, \zeta_n, \sigma_{B_n}^2)$. By defining thresholds for χ_n, ζ_n and $\sigma_{B_n}^2$, we can map the space of $\{x_n\}$ into eight classes $\{y_n\}$ ($y_n=0, 1, \ldots,$ or 7). Therefore, for the current frame, the best selection for *q* is the one maximizing the probability

$$P(q \mid y_n) = \frac{P(y_n \mid q)P(q)}{P(y_n)}, \quad (11)$$

Thus, it is the *q* that maximizes $P(y_n \mid q)P(q)$. The probabilities of $P(y_n \mid q)$ and $P(q)$ can be obtained using histogram analysis based on the statistics of the previously processed frames. Let us denote by $H_q(y)$ the counts of frames with coded background, and they follow *q* frames with skipped background with feature vector *y*, then

$$P(y_n \mid q) = \frac{H_q(y_n)}{\sum_y H_q(y)}, \quad (12)$$

and $P(q)$ is obtained in the similar fashion.

5. EXPERIMENTAL RESULTS

We conducted test experiments using the H.263 Profile 3 codec, and we tested a number of QCIF sequences at bitrates from 32kbps to 64kbps. The perceptual PSNR defined in [5] is used for video quality evaluation. In the first experiment, three reallocation strategies are compared as shown in Fig. 2. Clearly, both of the *conservative* and *aggressive strategies* outperform the *normal strategy*. The aggressive *strategy* slightly outperformed the *conservative strategy* at higher bit rate end. Although it requires extra computational complexity, the *aggressive strategy* has a good potential for video sequences with repeated patterns or have self-similarity characteristics.

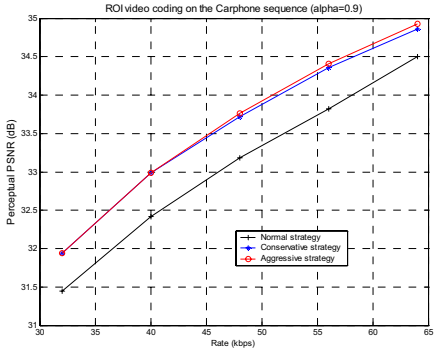


Figure 2 Comparison of three bit reallocation strategies

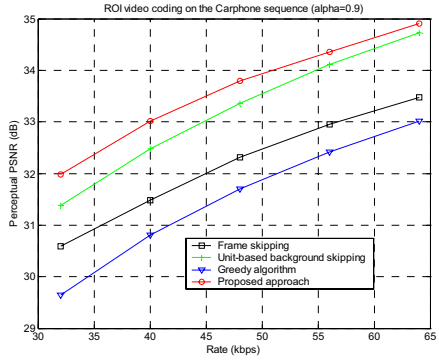


Figure 3. Performance comparison on "Carphone"

In subsequent experiments, we compared four different rate control approaches: (1) Macroblock-level greedy algorithm [6] where the bits are allocated to the macroblocks in a uniformly distributed manner. (2) Frame skipping algorithm that skips every other frame during encoding. (3) Unit-based background skipping algorithm [5] that groups every two frames into a unit and skips the background of the second frame within each unit; (4) The proposed approach which content-adaptively determines the frames with skipped background, and uses *aggressive strategy* for bit allocation. We ran experiments on "Carphone" and "Foreman" QCIF sequences, and the results are shown in Figs. 3 and 4. Clearly, the proposed approach significantly outperforms all other approaches in the whole bitrate range and the gain is up to *2.5*dB.

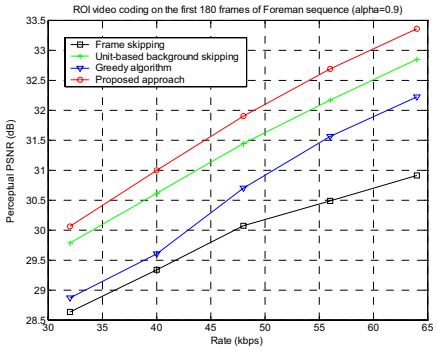


Figure 4. Performance comparison on "Foreman"

6. CONCLUSION

In this paper, we presented a content-adaptive background skipping scheme for ROI video coding. The scheme dynamically determines the background skip mode based on the current content information and the statistics of previously processed frames. Three bit reallocation strategies were proposed to work together with the optimized weighted bit allocation model in frame-level and macroblock-level bit allocation. Experimental results indicate that the proposed scheme has significant gains of up to *2.5* dB over the other approaches.

7. REFERENCES

[1] A. Eleftheriadis and A. Jacquin, "Automatic face location detection and tracking for model-assisted coding of video teleconferencing sequences at low bit-rates", *Signal Processing: Image Communications,* Vol. 7, No. 4-6, pp. 231-248, Nov. 1995.

[2] S. Daly, K. Matthews, and J. Ribas-Corbera, "As plain as the noise on your face: adaptive video compression using face detection and visual eccentricity models", *Journal of Electronic Imaging*, 10(1), Jan. 2001, pp. 30-46.

[3] D. Chai, and K. N. Ngan, "Face segmentation using skin-color map in videophone applications", *IEEE Trans. Circuits Systems for Video Technology*, Vol. 9, No. 4, June 1999, pp. 551-564.

[4] M. Chen, M. Chi, C. Hsu and J. Chen, "ROI video coding based on H.263+ with robust skin-color detection technique", *IEEE Trans. Consumer Electronics*, Vol. 49, No. 3, Aug. 2003. pp. 724-730.

[5] H. Wang and K. El-Maleh, "Joint adaptive background skipping and weighted bit allocation for wireless video telephony", in *Proc. International Conference on Wireless Networks, Communications, and Mobile Computing*, Maui, Hawaii, USA, June 2005.

[6] Z. He and S. K. Mitra, "A linear source model and a unified rate control algorithm for DCT video coding", *IEEE Trans. Circuits and System for Video Technology*, Vol. 12, No. 11, Nov. 2002. pp. 970-982.

[7] F. C. M. Martins, W. Ding, and E. Feig, "Joint control of spatial quantization and temporal sampling for very low bit rate video", in *Proc. ICASSP*, May 1996, pp. 2072-2075.

[8] F. Pan, Z. P. Lin, X. Lin, S. Rahardja, W. Juwono, and F. Slamet, "Content adaptive frame skipping for low bit rate video coding", in *Proc. 2003 Joint Conference of the Fourth International Conference on Information, Communications and Signal Processing, and the Fourth Pacific Rim Conference on Multimedia*, Vol. 1, Dec. 2003, Singapore, pp.230 – 234.