Research Article
A Rate-Distortion Optimized Coding Method for Region of Interest in Scalable Video Coding

Hongtao Wang, Dong Zhang, and Houqiang Li

Department of Electronic Engineering and Information Science, University of Science and Technology of China,
Hefei, Anhui 230027, China

Correspondence should be addressed to Houqiang Li; lihq@ustc.edu.cn

The support for region of interest (ROI) browsing, which allows dropping background part of video bitstreams, is a desirable feature for video applications. With the help of the slice group technique provided by H.264/SVC, rectangular ROI areas can be encoded into separate ROI slices. Additionally, by imposing certain constraints on motion estimation, ROI part of the bitstream can be decoded without background slices of the same layer. However, due to the additional spatial and temporal constraints applied to the encoder, overall coding efficiency would be significantly decreased. In this paper, a rate-distortion optimized (RDO) encoding scheme is proposed to improve the coding efficiency of ROI slices. When background slices are discarded, the proposed method uses base layer information to generate the prediction signal of the enhancement layer. Thus, the temporal constraints can be loosened during the encoding process. To do it in this way, the possible mismatch between generated reference frames and original ones is also considered during rate-distortion optimization so that a reasonable trade-off between coding efficiency and decoding drift can be made. Besides, a new Lagrange multiplier derivation method is developed for further coding performance improvement. Experimental results demonstrate that the proposed method achieves significant bitrate saving compared to existing methods.

1. Introduction

With the rapid development and continuous expansion of mobile communications, mobile internet services are becoming more and more popular. As a result, mobile video applications, such as mobile video broadcasting [1], mobile video conference [2], and mobile video surveillance [3], have become an active research area in recent years. However, due to the fact that mobile devices typically have limited communication bandwidth, constrained power capacity, and various display capabilities, there are several fundamental difficulties in deploying high-quality video service for mobile devices over wireless networks. Among them, one crucial problem for mobile video application is how to browse high-resolution (HR) videos on mobile devices with small screens. Traditional approaches typically downsize the video to achieve the required resolution, which will inevitably cause a loss of perceptual information and a waste of network bandwidth. Visual perception experiments show that small display screen is always a critical factor affecting the browsing

experiences. In fact, different parts of a picture do not equally attract people's attention. People are likely to pay more attention to a certain area, which is called region of interest (ROI), than to other areas of a picture. Thus, it is beneficial to optimize multimedia systems according to ROIs of video content, for example, to make ROI areas have better video quality [4] and to make background areas droppable when needed [5].

Scalable video coding extension of the H.264/MPEG-4 AVC video compression standard (H.264/SVC or SVC for short) [6] enables various functionalities to make encoded bitstreams more adaptive to dramatic variation of resource constraints, such as bandwidth, display capability, and power consumption. The base layer of SVC is compatible with H.264/AVC, so that it is easy to meet the requirement for the compatibility when upgrading video broadcasting infrastructures [7]. SVC is capable of encoding original video into different layers. Typically, it generates a high-quality video bitstream that contains one or more sub-streams, each of which corresponds to a degraded version of
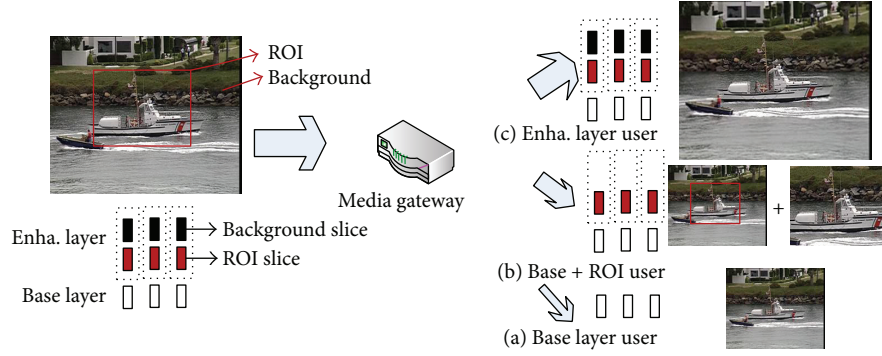
FIGURE 1: Application scenarios for SVC with ROI.

the original video signal with lower spatial resolution or lower temporal resolution or lower picture fidelity. Substreams can be extracted at media gateways according to network bandwidth and end device capability.

In addition to spatial, temporal, and quality scalabilities, SVC also supports ROI scalability [5]. With the help of the slice group technique (also known as Flexible Macroblock Ordering (FMO) [8]), macroblocks (MBs) in ROI and background areas can be coded into ROI slices and background slices, respectively. A low bitrate substream, which contains ROI slices, can be extracted from the high-quality bitstream without any transcoding operation [6]; therefore SVC can provide ROI browsing functionality for multimedia communication systems. As illustrated in Figure 1, the base layer of a bitstream can be coded with relatively low spatial resolution or modest fidelity to provide basic video quality for devices with small screen or low bandwidth. The enhancement layer can be coded with high resolution or high fidelity. To provide higher video quality, the enhancement layers may be composed of ROI slices and background slices to provide ROI scalability with FMO, which is allowed in the base baseline line profile of H.264/SVC. With the help of spatial scalability, one may choose either the QCIF (Quarter Common Intermediate Format) base layer or the CIF (Common Intermediate Format) enhancement layer according to resource constraints such as network bandwidth, screen size, and power consumption, as illustrated in scenarios (a) and (c). Additionally, ROI scalability introduces a new scenario (scenario (b)), in which both base layer and ROI of enhancement layer are delivered and one may choose between whole scene with low resolution and zoomed ROI area with high resolution for a better experience.

To develop a video application system with ROI browsing functionality based on SVC (like the application in Figure 1), ROI areas should be detected firstly in each frame, and then they can be coded into ROI slices using FMO. The detection of ROI of a picture has been widely studied with various attention models [9–11]. Given the fact that the ROI of a certain picture may be quite different for different people, it is practical and reasonable for a video application system to allow users to freely select some initial interested objects and then track these objects in subsequent video frames to locate ROI areas. Existing tracking methods can

be used to perform the tracking operation. These methods can be categorized into pixel domain (Pel-domain) [12–14] and compressed domain (Com-domain) [15–17] approaches. Generally speaking, the Pel-domain approaches can achieve better tracking accuracy than the Com-domain ones yet with higher complexity. For the simulations in this paper, a color based Monte Carlo tracking technique introduced by Prez et al. [14] is applied for ROI tracking.

To support ROI scalability, ROI slices should be self-contained, in other words, decodable in the absence of other slices of the same picture [18]. Thus, during the encoding process, dependencies between different slices, such as dependencies introduced by intraprediction and motion vector prediction, should be prohibited [8]. Besides, in order to provide acceptable visual quality in case background slices are all discarded, several additional temporal constraints are suggested to avoid using background slices in reference pictures to predict current ROI slice. The constrained motion vector method [19] presents an H.264/SVC compatible way where the MBs belonging to current ROI slice should use corresponding ROI slices in reference pictures only as reference. Besides, since the reference frames should be upsampled with a 6-tap interpolation filter for quarter-pel motion estimation (ME), the fractional pixel located within two pixels of the ROI slice boundary must also be ignored during motion estimation. However, this may significantly decrease the coding efficiency. Thus, Bae et al. [20] suggest using half-sample interpolation method for fractional pixel interpolation, where the slice boundary is treated as picture boundary. However, this method has little improvement compared with the constrained motion vector method; therefore, it has not been adopted into the H.264/SVC standard. Generally speaking, the above approaches all make strict truncation of temporal prediction, which leads to significant degradation of coding performance. Fortunately, for SVC application in Figure 1, since the base layer information is available when ROI of enhancement layer is being encoded, it is better to adopt a more flexible method to improve the coding performance of the enhancement layer ROI slice.

In this paper, an efficient ROI coding algorithm under the SVC scalable baseline profile is proposed. When the enhancement layer contains ROI slices, information of the base layer is adopted to improve the coding efficiency of them.

The framework of the proposed algorithm is illustrated in Figure 2. The ROI area of an input picture is coded as a ROI slice by the proposed enhancement layer encoder. It improves the coding efficiency by using rate-distortion optimized (RDO) mode decision, which takes into consideration the error propagation due to the loss of the background slices, instead of directly restricting the motion vectors. A new Lagrange multiplier derivation method, which is associated with the proportion of ROI area, is also derived and used in the RDO model for further improving the ROI slice encoding performance.

The remainder of this paper is organized as follows. The proposed ROI coding method is introduced in Section 2. Then Section 3 shows some experimental results to verify the benefits of the proposed RDO method and is followed by the conclusion in Section 4.

2. ROI Coding for H.264/SVC

In H.264/SVC, a rectangular ROI area of a picture can be coded into a separate slice using FMO technique. However, to support ROI browsing functionality, additional effort should be made to ensure that ROI slices are independent of background data. Several approaches can be used to encode ROI slices. Most of them apply constraints to temporal prediction to enable fully independent decoding of ROI slices, for example, the constrained motion vector method [10, 19] and the half-sample interpolation method [20]. However, in H.264/SVC, such strict constraints may severely degrade the coding efficiency for enhancement layer users. Fortunately, since the corresponding base layer data is always contained in an enhancement layer bitstream, the base layer information is available to be adopted to further improve the coding efficiency of the ROI area of the enhancement layer. In this section, the existing ROI coding approaches are introduced first, and then the proposed RDO based framework for coding ROI slices is presented.

2.1. Existing ROI Coding Methods

2.1.1. The Constrained Motion Vector Method. In H.264/SVC, motion estimation (ME) and motion compensation (MC) are performed using motion vectors with the accuracy of quarter-pixel luma samples. If the motion vector represents a fractional pixel position, interpolation is performed to generate predicted signal value.

As shown in Figure 3, the half-pixel samples are interpolated first from neighboring integer-pixel samples using a 6-tap finite impulse response (FIR) filter [21]. This means that each half-pixel sample is a weighted sum of 6 neighboring integer samples. Once all the half-pixel samples are available, the quarter-pixel samples are interpolated with neighboring half- and/or integer-pixel samples using bilinear interpolation.

The half-pixel sample "b" is interpolated by

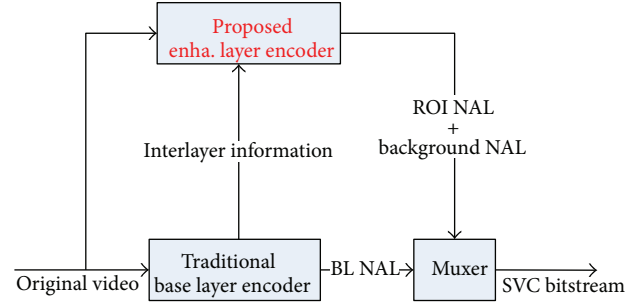$$b = (20(G + H) - 5(F + I) + (E + J) + 16) \gg 5. \quad (1)$$

FIGURE 2: Architecture of ROI enabled SVC coding system.

And the quarter-pixel sample a is interpolated by

$$a = (b + G + 1) \gg 1. \quad (2)$$

Since the derivation process of some fractional pixels within two pixels inside the ROI boundary (labeled as "unavailable fractional pixel" in Figure 3) depends on pixels out of the ROI region, Hannuksela et al. [19] proposed not to use them as reference during ME/MC; therefore, the dependency between ROI and background can be removed.

2.1.2. The Half-Sample Interpolation Method. To loosen the restriction imposed on the ME process, the half-sample interpolation method [20] modifies the fractional pixel interpolation process to extend the pixels on slice boundaries using the 6-tap interpolation; for example, the half-sample "b" in Figure 3 is generated by

$$b = (20(G + H) - 10I + 2J + 16) \gg 5. \quad (3)$$

In the actual implementation, half-sample interpolation method is only adopted when generating reference signal for ROI areas. Original interpolation method in H.264/SVC is still used to generate reference frames of background slices so that better coding efficiency can be achieved.

The above two methods aim to prohibit using samples in the background area of the reference frame during ME/MC, which also prevent a lot of MBs from being coded with proper prediction block located in or overlapped with background area. Therefore the coding efficiency is severely reduced compared with the original coding method.

2.2. Proposed RDO Based ROI Coding Framework. In application scenario (b) illustrated in Figure 1, when background slices of the enhancement layer are dropped, pixels in them cannot be used as reference to decode the ROI area of enhancement layer, but the base layer information is still available. So the background pixels of the base layer can be reconstructed and used to generate reference frames for ROI areas of enhancement layer using error concealment techniques. However, the mismatch between original and error-concealed reference frames may probably cause severe error propagation; therefore, the coding modes of MBs (which may use the error-concealed blocks as reference) should be selected carefully.
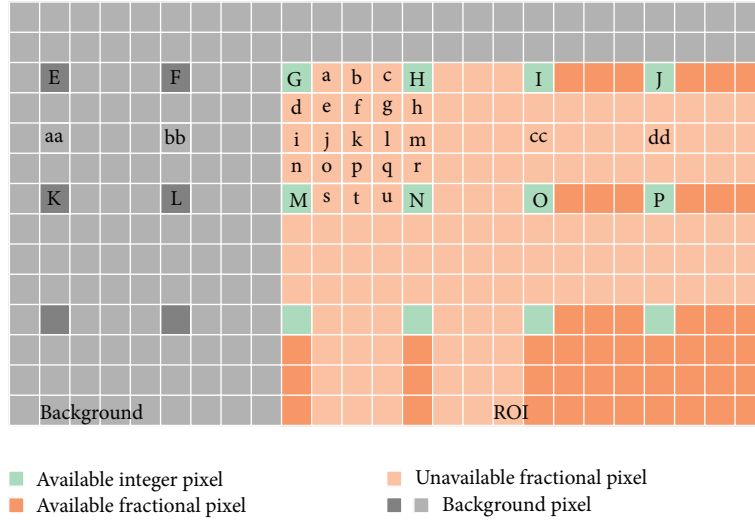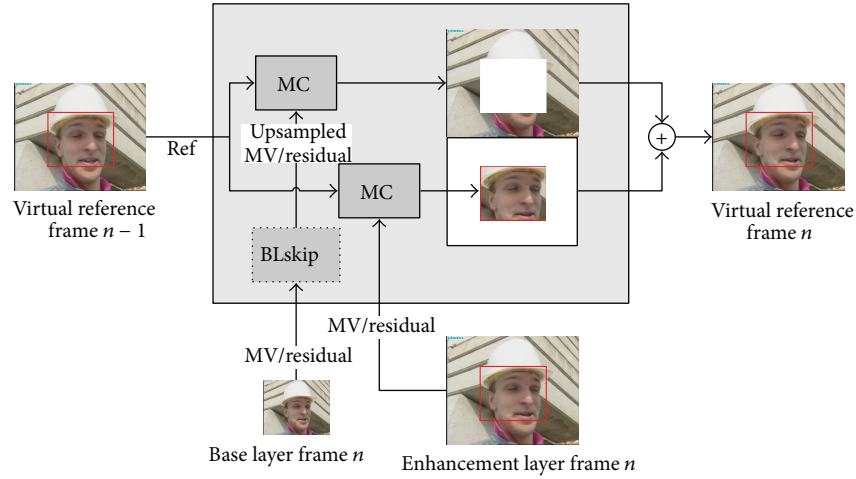
FIGURE 3: Interpolation of fractional pixels.

FIGURE 4: Generation of virtual reference frame n.

Let P and P' denote the original reference frame (ORF) and the error-concealed reference frame (named as the virtual reference frame, VRF), respectively, where P' may be generated using base layer information through error concealment techniques. In the proposed RDO framework, the mismatch between P and P', together with source error introduced by quantization in the encoding loop, is considered as total distortion. The RDO evaluation for mode decision is based on this total distortion. Furthermore, the Lagrange multiplier is modified to take account of the proportion of ROI area for better performance.

2.2.1. Generation of Virtual Reference Frame (VRF). Figure 4 shows the generation of intercoded VRF n at encoder side. Since the background slice is assumed to be discarded, the pixels belonging to the background slice are estimated using the base layer information with the same error concealment method that the decoder uses (in this paper, the well-known

BL-skip method [22] is adopted); the pixels belonging to the ROI slice are generated by motion compensation using their own motion vectors and residuals while taking the former VRF $n-1$ as reference. Then, the VRF n also serves as the reference frame of the following VRFs. The generation of intracoded VRF is similar to that of intercoded VRF except that upsampled textures are directly used to imitate the background slice. Notice that, in actual implementation, the upsampled motion vectors (MVs), residuals, and textures can be easily obtained when calculating the cost of "base layer mode." Thus, only one additional MC operation for each MB is needed to generate VRF.

2.2.2. Proposed RDO Mode Decision. In the mode decision process of a macroblock, the coding mode with the minimum RD cost is selected:

$$J = D + \lambda R, \qquad (4)$$

where D and R are distortion introduced and bits consumed by the coding mode under consideration, respectively. λ is the Lagrange multiplier.

For an MB in a ROI slice, the proposed mode decision scheme considers both the distortion introduced by the difference between the reconstructed MB and the original MB (termed source distortion) and the mismatch between reference MB in P and P' (termed mismatch distortion). So the RD cost function J for mode decision becomes

$$J = D_s + D_c + \lambda R, \tag{5}$$

where D_s stands for the source distortion and D_c is the mismatch distortion.

For an MB in a background slice, a basic assumption is that users who receive enhancement layer background slices should also receive ROI slices. Such assumption is reasonable considering ROI slices are more important and thereby are more protected than background slices. Thus, D_c becomes zero, and the cost function for mode decision is now degraded to its original form:

$$J = D_s + \lambda R. \tag{6}$$

Figure 5 depicts in detail the implementation of the proposed RDO mode decision of an MB.

(1) Firstly, given a mode m, the best motion vector mv_{best} for each partition is selected using the original reference frame. And let the corresponding predictor be P_m.

(2) Then, calculate the source distortion D_s and the cost bits R through the encoding process of mode m.

(3) If current MB belongs to ROI area, then find the new predictor P'_m from previous VRF with motion vector mv_{best} and calculate the mismatch error $D_c = (P_m - P'_m)^2$. Note that the following selection should be made: for distortion calculation in mode decision, the previous VRF is used as reference, while for distortion calculation in ME, the previous ORF is still used.

(4) Calculate the RD cost using (5) or (6), and turn to step (1) for the next mode.

(5) Finally, the mode with the minimum J among all candidate modes is selected as the best mode.

The benefit of the proposed RDO method is illustrated through the rate-distortion (RD) performance comparison in Figures 6 and 7 for both spatial and quality SVC. In Figure 6, the spatial SVC bitstream contains a QCIF base layer and a CIF enhancement layer. In Figure 7, the quality SVC bitstream contains a CIF base layer and a CIF enhancement layer. The intraperiod is set to 30. Four pairs of quantization parameter (QP) are chosen for the test: for spatial SVC, QP pairs for QCIF base layer and CIF enhancement layer are (22, 26), (26, 30), (30, 34), and (34, 38), respectively, and, for quality SVC, the QP pairs are (30, 26), (34, 30), (38, 34), and (42, 38), respectively. The original method in H.264/SVC, which uses the ORFs without any constraints on temporal prediction, is simulated as the anchor. Three collections of
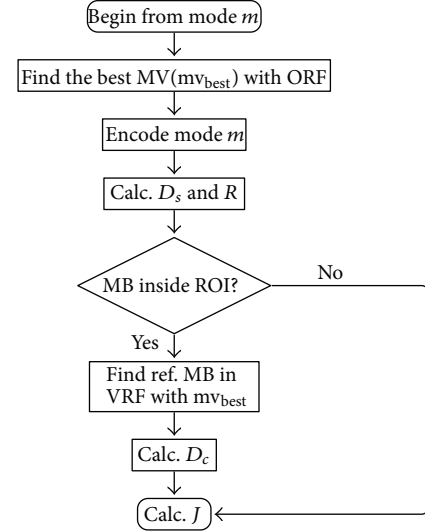

Figure 5: The proposed RD cost calculation process for a macroblock.

data are presented in each figure, where "mdrdo" stands for the proposed RDO based mode decision method, and "mv_constrain" and "half-interpolation" stand for the ROI coding methods mentioned in Section 2.1.

Coding efficiency in the following two scenarios is considered. The first is that enhancement layer slices are received completely, which means that the quality of full resolution enhancement layer (labeled "Enc full") should be considered. The second scenario is that background slices are all discarded, which means only the quality of ROI areas (labeled "Dec ROI") affects the user experience. Average bitrate savings, which are calculated via the excel add-in proposed in VCEG-AE07 [23] (lager Δ-bitrate value means worse performance), of the above three coding methods compared with "orig" method are presented.

From Figures 6(a) and 7(a), we can see that the coding efficiency of those three methods (for the whole SVC bitstream) is all lower compared with the "orig." method, which uses the perfectly decoded reference frames as prediction. However, the proposed method achieves significant improvement compared with the other two methods because of the use of a better reference for coding ROI slices. The average performance gain compared with the one termed "half-interpolation" over the tested sequences is about 5% (Figure 6(a)) and 7% (Figure 7(a)) for spatial and quality SVC, respectively.

Though the performances of those methods are inferior compared with original method when enhancement layer slices are all received, however, considering the most important ROI browsing scenario in which all the background slices may be discarded, the RD performance of those methods is much better than that of "orig." method. As illustrated in Figures 6(b) and 7(b), the bitrate saving of the proposed method is up to 50% (about 30% and 4% for spatial and quality SVC on average, resp.) compared with "orig." method. The lower gain for quality SVC compared with spatial SVC is consistent with the common knowledge that a better concealment
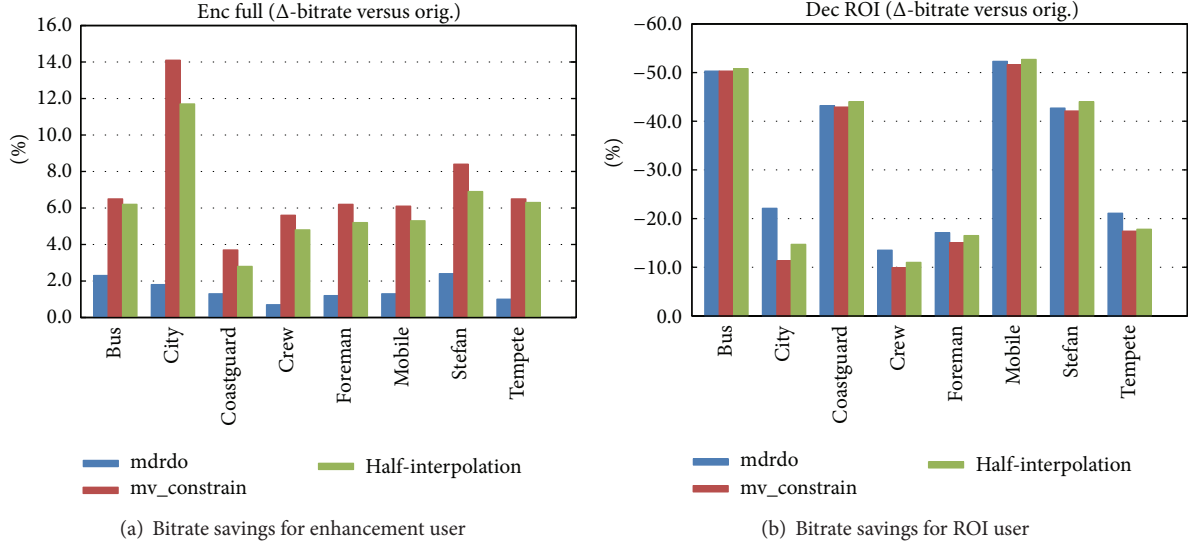
(a) Bitrate savings for enhancement user

(b) Bitrate savings for ROI user

FIGURE 6: RD performance for spatial SVC.

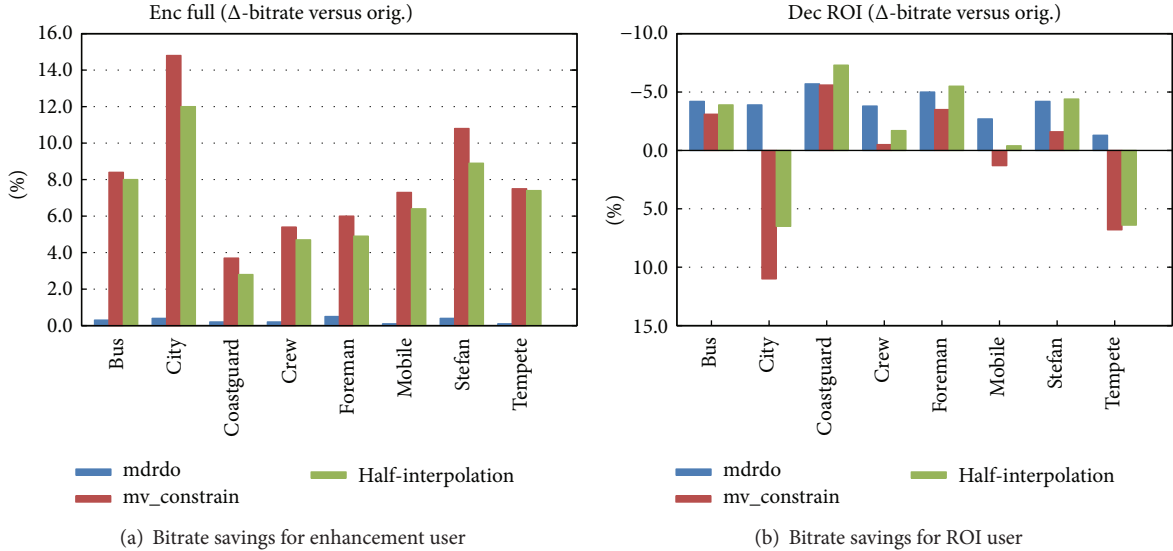(a) Bitrate savings for enhancement user

(b) Bitrate savings for ROI user

FIGURE 7: RD performance for quality SVC.

quality will be obtained when the base layer and enhancement layer have the same resolution; thus, acceptable quality may be obtained even with "orig." method for the quality coding configuration (CIF-CIF). Still, the proposed method outperforms "mv_constrain" and "half-interpolation" methods, and the average gain compared with "half-interpolation" is 3% (Figure 6(b)) and 2.5% (Figure 7(b)) for spatial and quality SVC, respectively.

2.2.3. The Selection of Lagrange Multiplier. In RDO optimization, the Lagrange multiplier should be carefully selected to ensure that the most suitable modes are chosen. In this paper, a refined Lagrange multiplier selection method for background and ROI slices is proposed to further improve the RD performance of ROI slices. In H.264/AVC, the Lagrange multiplier λ can be calculated as follows.
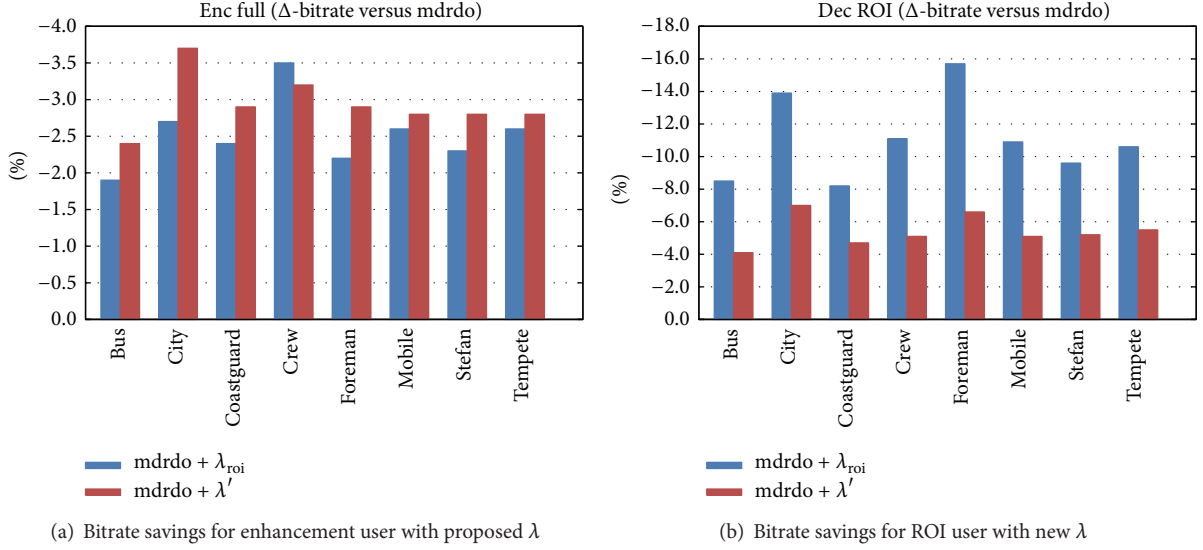
Supposing R and D in (4) to be differentiable everywhere, the minimum cost J is given by setting its derivative to zero, thus, leading to

$$\lambda = -\frac{\mathrm{d}D}{\mathrm{d}R}. \tag{7}$$

Then, the Lagrange multiplier λ for single layer video coding can be solved through the rate model R_{single} (8) and distortion model D_{single} (9) [24]:

$$R_{\text{single}} = a \log_2\left(\frac{b}{D_{\text{single}}}\right), \tag{8}$$

$$D_{\text{single}} = \frac{Q^2}{12}, \tag{9}$$

(a) Bitrate savings for enhancement user with proposed λ

(b) Bitrate savings for ROI user with new λ

FIGURE 8: RD performance for spatial SVC with new λ.

where a and b are two constants and Q is the quantization step. According to (8) and (9), the derivative of D and R can be calculated by

$$
\begin{aligned}
dD_{\text{single}} &= \frac{Q}{6} dQ, \\
dR_{\text{single}} &= -\frac{2a}{\ln 2 \cdot Q} dQ.
\end{aligned}
\tag{10}
$$

Putting (10) into (7) and letting $\ln 2 = 12ac$, the λ for single layer is finally derived as

$$
\lambda(Q) = \frac{\ln 2}{12a} \cdot Q^2 = c \cdot Q^2,
\tag{11}
$$

where c is a constant, which is experimentally suggested to be 0.85 [24], though others proposed 0.68 [25].

During the development of H.264/SVC, such λ is directly used in H.264/SVC reference software Joint Scalable Video Model (JSVM) [26]. However, applying such λ, which is derived for single layer Lagrange multiplier, into multilayer scenario is inappropriate, since the correlation between layers is not considered in the Lagrange multiplier selection. To improve the overall coding performance, an encoder-only optimization contribution for RDO in SVC is presented by Li et al. [27], which is adopted into later JSVM in an optional way. According to this method, the Lagrange multiplier is derived as

$$
\lambda' = \frac{\beta \cdot (Q + \Delta)}{\beta \cdot (Q + \Delta) + Q} \cdot \lambda(Q),
\tag{12}
$$

where β denotes the resolution ratio of the two layers and $(Q + \Delta)$ and Q are the quantization steps for the base layer and enhancement layer, respectively.

Similarly, for a specific user who requires base layer slices together with ROI slices, the best Lagrange multiplier for ROI slices can be derived as follows.

Let the joint cost be

$$
\begin{aligned}
J &= \omega_b \cdot J_b + \omega_{\text{roi}} \cdot J_{\text{roi}} \\
&= \omega_b \cdot (D(Q + \Delta) + \lambda_b \cdot R(Q + \Delta)) \\
&\quad + \omega_{\text{roi}} \cdot (\alpha \cdot D(Q) + D_c + \lambda_{\text{roi}} \cdot (R(Q + \Delta) + \alpha \cdot R(Q))),
\end{aligned}
\tag{13}
$$

where J_b and J_{roi} are the RD cost functions for base layer and enhancement layer with ROI area, respectively, ω_b and ω_{roi} are the contribution weights of base layer cost and ROI area cost, respectively, and $\omega_b + \omega_{\text{roi}} = 1$. Similar to [27], the term α denotes the resolution ratio between enhancement layer ROI area and base layer and is introduced as an approximation for bitrate ratio between the ROI slice and base layer slice.

The term D_c is the mismatch distortion (as described in (5)) introduced by difference between ORF and VRF. D_c is much larger than $D(Q)$ and thus can be regarded as independent with Q. Therefore, put (8) and (9) into (13) and then set the derivative of J to zero; the Lagrange multiplier λ_{roi} can be solved as

$$
\begin{aligned}
\lambda_{\text{roi}} = Q\Bigg(&\left(\frac{\omega_b}{\omega_{\text{roi}}}\right) \cdot \left(\ln 2 \cdot (Q + \Delta)^2 - 12a\lambda_b\right) \\
&+ \ln 2 \cdot \alpha Q(Q + \Delta)\Bigg) \\
&\times (12a(\alpha \cdot (Q + \Delta) + Q))^{-1}.
\end{aligned}
\tag{14}
$$

Considering the base layer Lagrange multiplier λ_b is determined by single layer λ selection method (11); put $\lambda_b = c \cdot (Q + \Delta)^2$ into (14) to derive λ_{roi} as

$$
\begin{aligned}
\lambda_{\text{roi}} &= \frac{\alpha \cdot (Q + \Delta)}{\alpha \cdot (Q + \Delta) + Q} \cdot \left(c \cdot Q^2\right) \\
&= \frac{\alpha \cdot (Q + \Delta)}{\alpha \cdot (Q + \Delta) + Q} \cdot \lambda(Q).
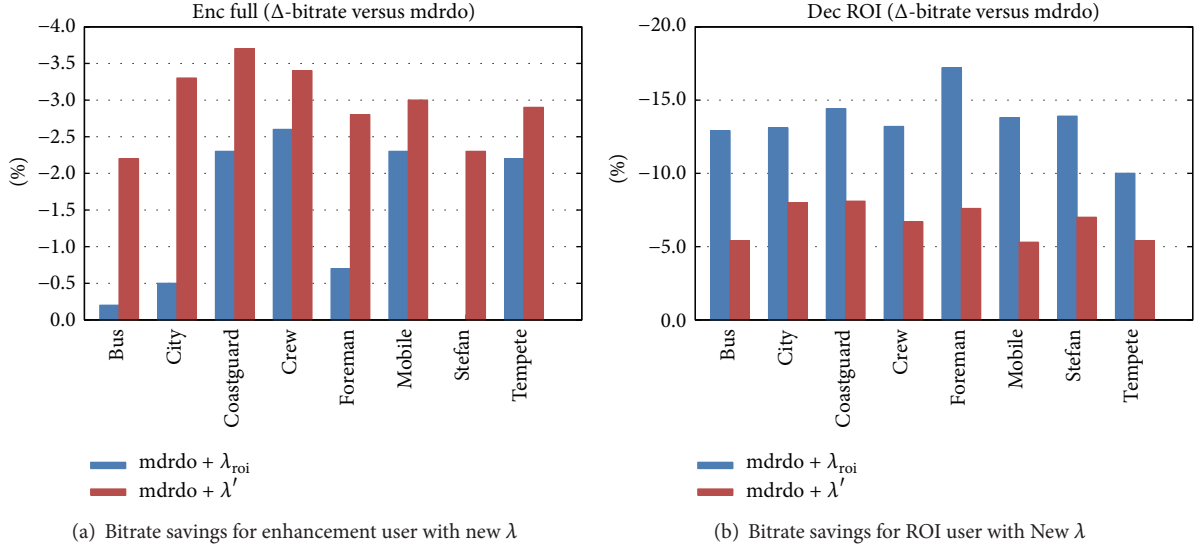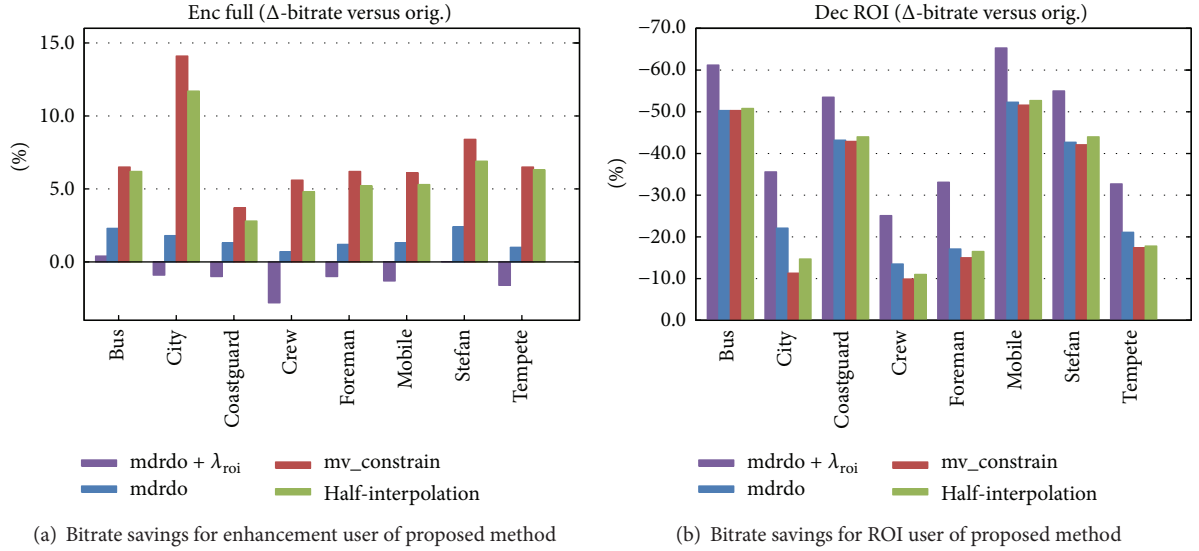\end{aligned}
\tag{15}
$$

(a) Bitrate savings for enhancement user with new λ

(b) Bitrate savings for ROI user with New λ

FIGURE 9: RD performance for quality SVC with new λ.

(a) Bitrate savings for enhancement user of proposed method

(b) Bitrate savings for ROI user of proposed method

FIGURE 10: RD performance for spatial SVC with proposed RDO framework.

Note that the derived λ_{roi} is similar to (13). However, the rate ratio has a different form since the area of ROI has been taken into consideration.

Let QP_b and QP_e denote the quantization parameters for base layer and enhancement layer, respectively. Then, according to the relationship between quantization steps and quantization parameters, which has been defined in H.264/AVC standard [28], QP_b, QP_e, and Q conform to

$$Q + \Delta = 2^{(QP_b-12)/6},$$
$$Q = 2^{(QP_e-12)/6}. \tag{16}$$

Equation (14) can be simplified according to the quantization parameter difference between base and enhancement layers; namely, $\Delta QP = QP_b - QP_e$. And finally, the modified Lagrange multiplier λ_{roi} can be obtained through the following equation for SVC enhancement layer when ROI is enabled:

$$\lambda_{roi} = \frac{\alpha \cdot 2^{\Delta QP/6}}{\alpha \cdot 2^{\Delta QP/6} + 1}. \tag{17}$$

Similar to the RDO mode decision part, the performance of the proposed RDO method with modified Lagrange multiplier is shown in Figures 8 and 9 for spatial and quality SVC, respectively. The same coding parameters are used. "Enc full" and "Dec ROI" denote the scenarios when background slices are received and discarded, respectively. The proposed "mdrdo" method with Lagrange multiplier λ' [27] and with the proposed λ_{roi} (17) is simulated separately.
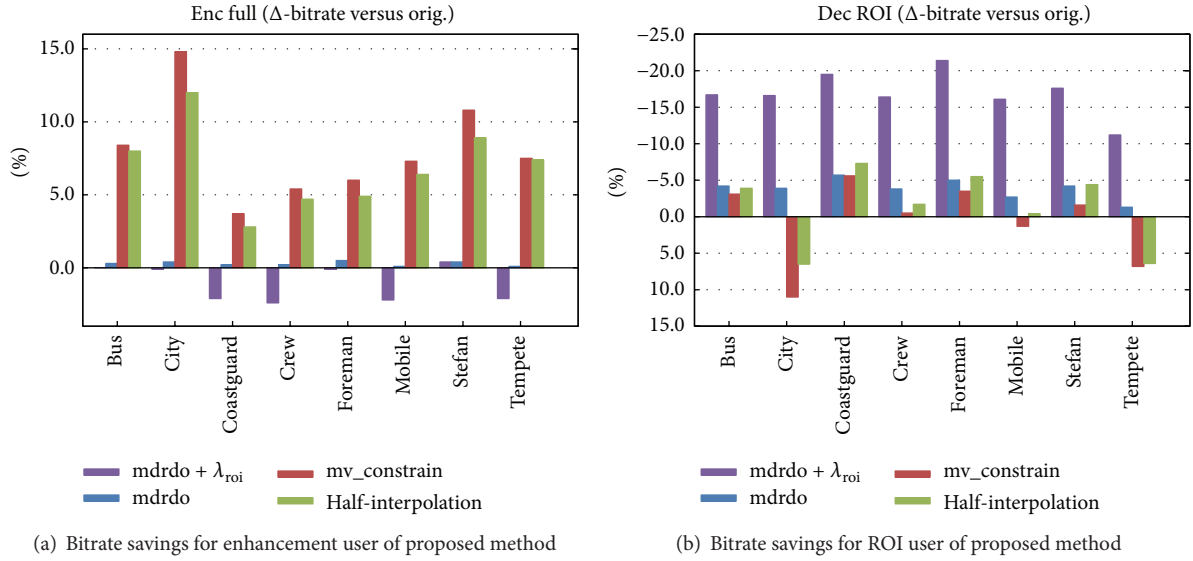
(a) Bitrate savings for enhancement user of proposed method

(b) Bitrate savings for ROI user of proposed method

FIGURE 11: RD performance for quality SVC with proposed RDO framework.

(a) Input

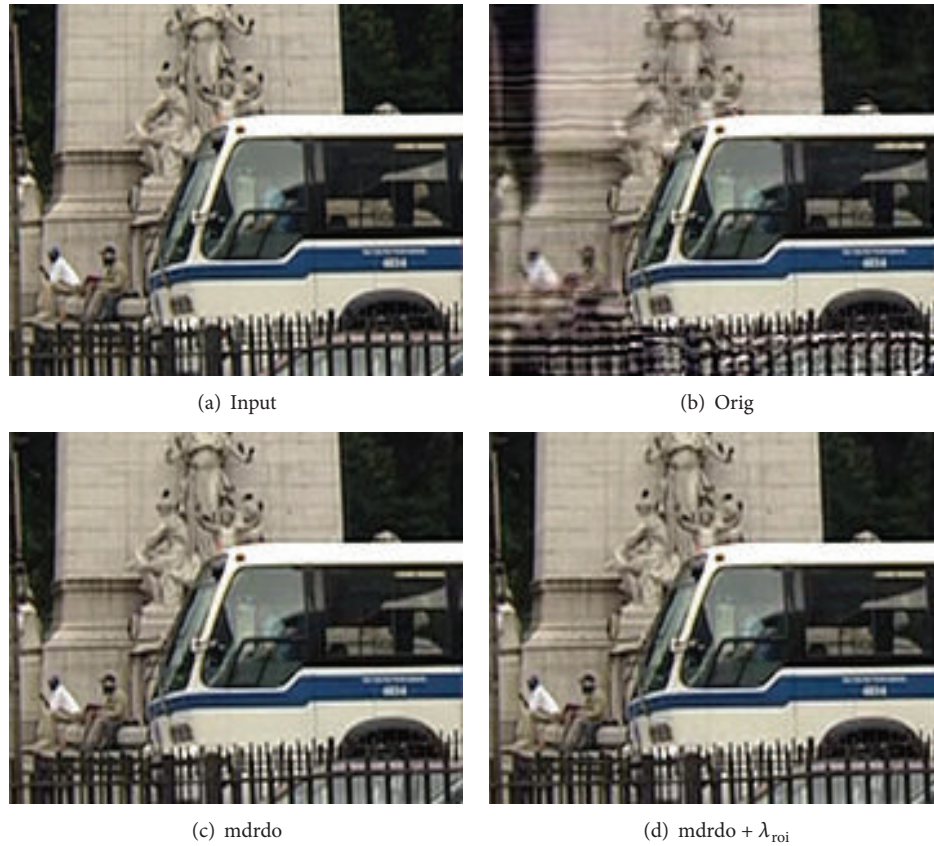(b) Orig

(c) mdrdo

(d) mdrdo + λ_{roi}

FIGURE 12: Visual quality comparison under "Dec ROI" scenario for the "bus" sequence.

And the "mdrdo" method with original single layer Lagrange multiplier (11) is performed as the anchor.

All the values in Figures 8 and 9 are negative, which means that both Lagrange multiplier modification methods have brought benefits to ROI and enhancement layer coding compared with the original Lagrange multiplier method. Compared with λ', the proposed λ_{roi} achieves a better performance for ROI, and the average gain is about 6% for spatial SVC (Figure 8(b)) and 7% for quality SVC (Figure 9(b)). Since λ_{roi} is always smaller than λ', the proposed λ_{roi} shifts

(a) Input

(b) Orig
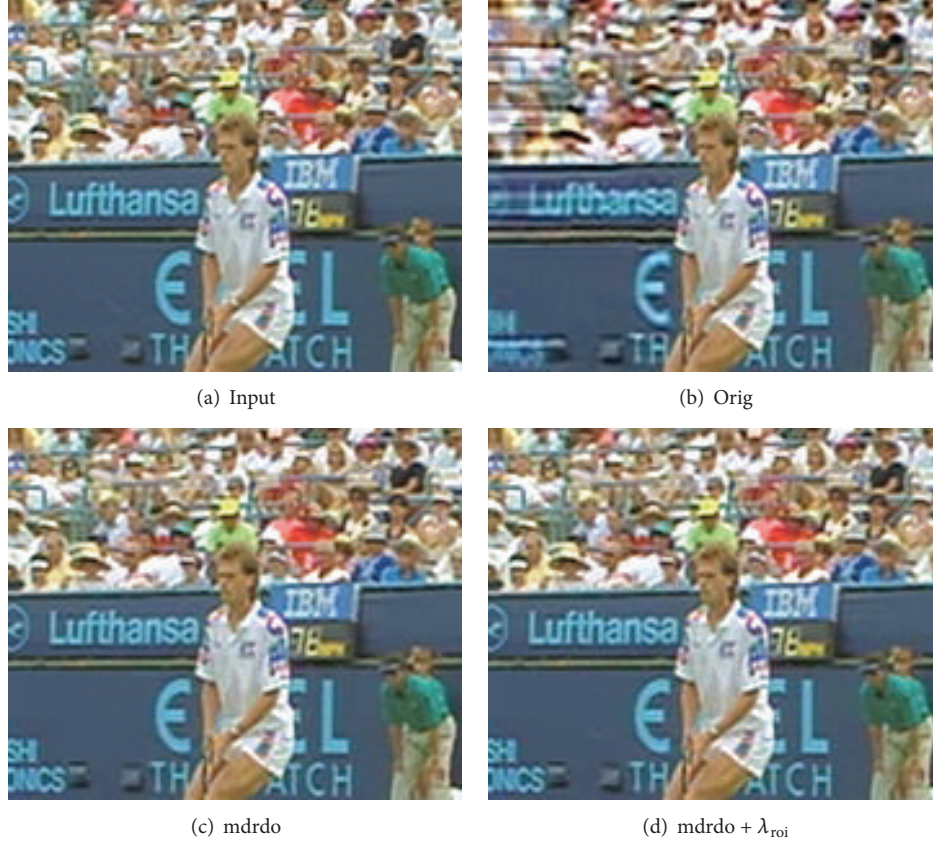

(c) mdrdo

(d) mdrdo + λ_{roi}

FIGURE 13: Visual quality comparison under "Dec ROI" scenario for the "stefan" sequence.

bits from a background slice to an ROI slice. Thus, for the whole enhancement layer, the proposed λ_{roi} leads to trifling loss of coding efficiency compared with λ'. The average loss of "mdrdo + λ_{roi}" compared with "mdrdo + λ'" is about 1% for spatial SVC (Figure 8(a)) and 1.5% for quality SVC (Figure 9(a)), respectively. Although the performance of the background area is sacrificed, it is worthwhile given that the performance of semantic-important ROI area is improved.

3. Simulation Results

To illustrate the overall improvement of the proposed framework, the benefits of the proposed RDO method together with modified Lagrange multiplier are shown in Figures 10 and 11 for spatial and quality SVC, respectively. The same coding parameters, as described in Section 2.2, are still used. Now, thanks to the modification of Lagrange multiplier, the proposed "mdrdo + λ_{roi}" method achieves significant bitrate saving compared with the coding scheme using the original Lagrange multiplier method (denoted as "orig."), regardless of the background slices being lost or not. Accordingly, the gain of "mdrdo + λ_{roi}" method, which uses the modified Lagrange multiplier, is also increased compared with "mv_constrain" and "half-interpolation" methods which use the original Lagrange multiplier method. For example, the average gain compared with "half-interpolation" method

under "Dec ROI" scenario is now 15% (Figure 10(b)) and 12% (Figure 11(b)) for spatial and quality SVC, respectively.

When all background slices of enhancement layer are discarded, decoding drift occurs in both "orig." method and the proposed method. To show the impact of such drift in both cases, visual quality comparison of "mdrdo," "mdrdo + λ_{roi}," and "orig." methods is given for decoded ROI. In Figures 12 and 13, the decoded QCIF sized ROI areas under "Dec ROI" scenarios for spatial SVC with a QICF base layer (QP 22) and a CIF enhancement layer (QP 26) are illustrated. The "input" shows the original QCIF ROI area of the input sequence. It can be seen that the "orig." method, which applies no constraints to temporal prediction, suffers a severe quality degradation, especially at the boundaries of ROI areas (Figures 12(b) and 13(b)). By taking into consideration the error propagation during the mode decision, the proposed "mdrdo" and "mdrdo + λ_{roi}" methods can effectively restrict the errors. The error propagation is unnoticeable in Figures 12(c) and 12(d) and 13(c) and 13(d).

4. Conlusions

In this paper, a ROI coding framework based on H.264/SVC is proposed. Firstly, a rate-distortion optimized mode decision method is proposed, which loosens the temporal constraints when coding ROI slices in SVC enhancement layer. The mismatch between reference frames is considered during

the RDO process, whenever the background slices are discarded or kept. Then, a new Lagrange multiplier estimation algorithm is derived to improve the coding efficiency of ROI slices. Compared with the existing constraint-based methods, such as the constrained motion vector method and the half-sample interpolation method, experimental results show that the proposed method achieves significant bitrate saving while maintaining both higher objective and subjective video quality.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] Ç. Atici and M. Sunay, "High data-rate video broadcasting over 3G wireless systems," *IEEE Transactions on Broadcasting*, vol. 53, no. 1, pp. 212–223, 2007.

[2] H. Cycon, T. Schmidt, G. Hege, M. Wahlisch, and M. Palkow, "An optimized h.264-based video conferencing software for mobile devices," in *Proceedings of the IEEE International Symposium on Consumer Electronics (ISCE '08)*, pp. 1–4, 2008.

[3] G. Gualdi, A. Prati, and R. Cucchiara, "Video streaming for mobile video surveillance," *IEEE Transactions on Multimedia*, vol. 10, no. 6, pp. 1142–1154, 2008.

[4] G. Muntean, G. Ghinea, and T. N. Sheehan, "Region of interest-based adaptive multimedia streaming scheme," *IEEE Transactions on Broadcasting*, vol. 54, no. 2, pp. 296–303, 2008.

[5] T. C. Thang, T. M. Bae, Y. J. Jung, Y. M. Ro, J.-G. Kim, and H. Choi, "Spatial scalability of multiple ROIs for surveillance video," ISO/IEC MPEG & ITUT VCEG JVT-O037, Busan, Republic of Korea, 2005.

[6] ISO/IEC JTC1/SC29/WG11, "Applications and requirements for scalable video coding," Document N6880, Hong Kong, 2005.

[7] H. Choi, I. H. Shin, J.-S. Lim, and J.-W. Hong, "SVC application in advanced T-DMB," *IEEE Transactions on Broadcasting*, vol. 55, no. 1, pp. 51–61, 2009.

[8] P. Lambert, W. D. Neve, Y. Dhondt, and R. V. de Walle, "Flexible macroblock ordering in H.264/AVC," *Journal of Visual Communication and Image Representation*, vol. 17, no. 2, pp. 358–375, 2006.

[9] R. Zhang, S. Zhang, and S. Yu, "Moving objects detection method based on brightness distortion and chromaticity distortion," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 3, pp. 1177–1185, 2007.

[10] X. Fan, X. Xie, H.-Q. Zhou, and W.-Y. Ma, "Looking into video frames on small displays," in *Proceedings of the 11th ACM International Conference on Multimedia (MULTIMEDIA '03)*, pp. 247–250, ACM, November 2003.

[11] Y. Wang, H. Li, X. Fan, and C. W. Chen, "An attention based spatial adaptation scheme for h.264 videos on mobiles," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 20, no. 4, pp. 565–584, 2006.

[12] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[13] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[14] P. Prez, C. Hue, J. Vermaak, and M. Gangnet, "Color-based probabilistic tracking," in *Computer Vision ECCV—2002*, A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, Eds., vol. 2350 of *Lecture Notes in Computer Science*, pp. 661–675, Springer, Berlin, Germany, 2002.

[15] P. Lambert, D. Schrijver, D. Deursen, W. Neve, Y. Dhondt, and R. Walle, "A real-time content adaptation framework for exploiting ROI scalability in H.264/AVC," in *Advanced Concepts for Intelligent Vision Systems*, J. Blanc-Talon, W. Philips, D. Popescu, and P. Scheunders, Eds., vol. 4179 of *Lecture Notes in Computer Science*, pp. 442–453, Springer, Berlin, Germany, 2006.

[16] F. Manerba, J. Benois-Pineau, R. Leonardi, and B. Mansencal, "Multiple moving object detection for fast video content description in compressed domain," *Eurasip Journal on Advances in Signal Processing*, vol. 2008, Article ID 231930, 2008.

[17] C. Käs and H. Nicolas, "Compressed domain indexing of scalable h.264/svc streams," *Signal Processing: Image Communication*, vol. 24, no. 6, pp. 484–498, 2009.

[18] ITU-T Rec, *Advanced Video Coding for Generic Audiovisual Services*, 2009.

[19] M. M. Hannuksela, Y.-K. Wang, and M. Gabbouj, "Isolated regions in video coding," *IEEE Transactions on Multimedia*, vol. 6, no. 2, pp. 259–267, 2004.

[20] T. M. Bae, T. C. Thang, D. Y. Kim, Y. M. Ro, J. W. Kang, and J. G. Kim, "Multiple region-of-interest support in scalable video coding," *ETRI Journal*, vol. 28, no. 2, pp. 239–242, 2006.

[21] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

[22] C. Ying, J. Boyce, and X. Kai, "Frame Loss Error Concealment for SVC," JVTQ046, Nice, France, 2005.

[23] S. Pateux and J. Jung, "An excel add-in for computing Bjntegaard metric and its evolution," VCEG-AE07, Marrakech, Morocco, 2007.

[24] T. Wiegand and B. Girod, "Lagrange multiplier selection in hybrid video coder control," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '01)*, vol. 3, pp. 542–545, Thessaloniki, Greece, October 2001.

[25] K. Takagi, "Lagrange multiplier and RD-characteristics (JVT-C084)," JVT-C084, Fairfax, Va, USA, 2002.

[26] JVT, *H.264/SVC Reference Software (JSVM 9.13.1) and Manual*, 2008.

[27] X. Li, P. Amon, and A. Hutter, "Lagrange multiplier selection for rate-distortion optimization in SVC, JVT-AD021," Geneva, Switzerland, 2009.

[28] *Advanced Video Coding (AVC)*, 3rd edition, 2004.