electronics

MDPI

Article

Adaptive Quantization Parameter Estimation for HEVC Based Surveillance Scalable Video Coding

Xiem HoangVan

Faculty of Electronics and Telecommunications, University of Engineering and Technology, Vietnam National University, Hanoi 123106, Vietnam; xiemhoang@vnu.edu.vn

Received: 6 May 2020; Accepted: 27 May 2020; Published: 30 May 2020

Abstract: Visual surveillance systems have been playing a vital role in human modern life with a large number of applications, ranging from remote home management, public security to traffic monitoring. The recent High Efficiency Video Coding (HEVC) scalable extension, namely SHVC, provides not only the compression efficiency but also the adaptive streaming capability. However, SHVC is originally designed for videos captured from generic scenes rather than from visual surveillance systems. In this paper, we propose a novel HEVC based surveillance scalable video coding (SSVC) framework. First, to achieve high quality inter prediction, we propose a long-term reference coding method, which adaptively exploits the temporal correlation among frames in surveillance video. Second, to optimize the SSVC compression performance, we design a quantization parameter adaptation mechanism in which the relationship between SSVC rate-distortion (RD) performance and the quantization parameter is statistically modeled by a fourth-order polynomial function. Afterwards, an appropriate quantization parameter is derived for frames at long-term reference position. Experiments conducted for a common set of surveillance videos have shown that the proposed SSVC significantly outperforms the relevant SHVC standard, notably by around 6.9% and 12.6% bitrate saving for the low delay (LD) and random access (RA) coding configurations, respectively while still providing a similar perceptual decoded frame quality.

Keywords: HEVC; surveillance scalable video coding; long-term reference coding; quantization parameter adaptation

1. Introduction

In recent years, there has been an accelerated expansion of surveillance systems to cope with security and safety's threats. Considerable numbers of surveillance cameras have been mounted in public and private areas [1]. The emergence of large video surveillance infrastructures leads to a massive amount of content that must be stored, analyzed and managed by security teams with limited resources. Furthermore, the heterogeneity of networks, display devices and transmission environments has been rising as a critical issue in the modern video communication. To fulfil this challenge, it is necessary to have a dynamic and adaptable surveillance video compression system, which not only improves the compression efficiency but also adapts to the variation of networks and transmissions as depicted in Figure 1.

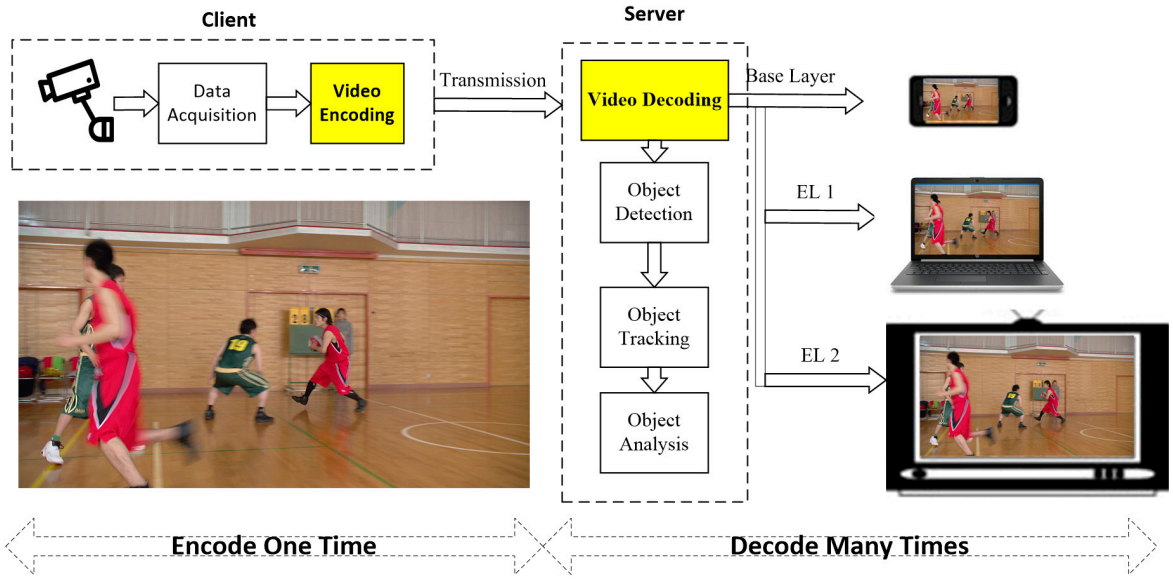

Figure 1. An illustration of a visual surveillance system.

High Efficiency Video Coding (HEVC) [2] is the most recent video coding standard that provides 50% bitrate saving when compared to the prior H.264/AVC standard [3]. Many efficient coding tools have been proposed for HEVC, including a quad-tree partitioning structure, angular intra prediction, merge modes, sample adaptive offset (SAO) filter and variable transform block sizes. In addition, rate-distortion optimization (RDO) has also been adopted to select the optimal coding modes [4]. However, these coding tools are originally designed for ordinary video contents. The surveillance videos are special and different from the generic videos in several aspects: (1) the cameras in the video surveillance are usually stationary. Therefore, the acquired videos are generally with stationary backgrounds and have stronger long-term redundancy than the generic videos; (2) the motion patterns are generally simpler than that in the generic videos. Hence, conventional inter-coding prediction may not be efficient.

To take advantage of the surveillance video characteristics, some coding improvement methods have been introduced for HEVC based surveillance video coding structure [5–7]. The improvement methods can be classified into: (1) background modeling based long-term reference (LTR) [6] and key frame based long-term reference [7]. Although previous works have shown some promising results for surveillance video coding, they are unable to provide the adaptive streaming capability due to the single-layer compression nature of the HEVC. As shown in Figure 1, users may want to decode the bitstreams from different devices or network capacities. In these cases, scalable video coding based approach becomes essential.

Scalable High Efficiency Video Coding (SHVC) is the latest scalable video coding solution that provides adaptive video compression capability for a large number of video transmission environments and displaying devices. As an extension, SHVC inherited most of the coding tools provided for HEVC standard [2]. Similar to the prior scalable video coding (SVC) solution [8], SHVC follows a layered coding structure with one base layer (BL) and one or several enhancement layers (EL). However, SHVC was also not originally designed for videos captured from the visual surveillance system.

In this paper, we propose an efficient HEVC based surveillance scalable video coding (SSVC) framework that is built on the top of the HEVC architecture, to provide the compression efficiency and adaptive streaming capability by exploiting the LTR coding concept and a novel quantization parameter adaptation model. The main contributions of this paper can be summarized as:

1. HEVC based surveillance scalable video coding scheme: to meet the dynamic changes of transmission environment and the diversity of displaying devices, we extend the HEVC with a

layered coding structure in which the HEVC architecture is modified to compress the surveillance videos at several quality/fidelity levels;

2. Adaptive long-term reference solution: to reduce the temporal redundancy that exists in videos captured from surveillance system, we propose to adaptively employ the LTR coding concept. In the proposed LTR solution, a low complexity LTR frame is created and adaptively updated based on the video content;
3. Quantization parameter model: to optimize the SSVC compression efficiency, we propose to adaptively adjust the quantization parameters for frame at the LTR positions. This is performed based on an extensive statistical analysis of the RD performance and the quantization parameter optimization.

Experiments conducted for a common set of surveillance videos have shown that the proposed SSVC significantly outperforms the relevant SHVC standard, notably by around 6.9% and 12.6% bitrate saving for the low delay (LD) and random access (RA) coding configurations, respectively while still providing a similar perceptual decoded frame quality.

The rest of this paper is organized as follows. Section 2 briefly summarizes the background and related works on scalable and surveillance video coding. Afterwards, Section 3 describes the proposed SSVC solution while Section 4 presents the QP adaptation model. Section 5 evaluates the compression performance of the proposed SSVC. Finally, Section 6 gives some main conclusions and ideas for future works.

2. Related Works

To provide the scalability features, scalable video coding is a promising solution while to cope with the surveillance video characteristics. Several improvement methods have been introduced for surveillance video coding. Therefore, this section will review the related works on scalable and surveillance video coding.

2.1. Scalaable Video Coding

Scalable video coding (SVC) facilitates the encoding of a bitstream containing various representations such as spatial resolutions, frame rates and quality fidelities, which are designed to meet the requirements of the heterogeneous display and computational capabilities of the target devices. A client with restricted resources such as display resolution, processing power or bandwidth can only decode a part of the delivered bitstream.

In the recent decades, several specific scalable video profiles have been included in video codecs such as MPEG-2 [9], MPEG-4/FGS (Fine grain scalability) [10]. However, the scalability features resulted in the significant reduction of compression performance, notably when compared with non-scalable video profiles. Consequently, scalable profiles have been scarcely utilized in real applications. In October 2007, a novel scalable video coding solution (SVC), extended from the H.264/AVC standard, had been introduced [8]. As reported, the H.264/SVC significantly outperforms the relevant H.264/AVC simulcasting benchmark in terms of compression performance. H.264/SVC was widely recognized as one of the state-of-the-art SVC codecs that provides spatial, temporal and quality scalability features. The H.264/SVC standard follows a layered coding structure where one BL and one or several ELs are created on the top of the H.264/AVC standard. Three inter-layer redundancies, inter-layer intra, inter-layer motion and inter-layer residual coding modes were introduced.

The achievement of the HEVC standard with respect to the prior H.264/AVC has encouraged the development of a newly HEVC scalable extension (namely SHVC) [11]. SHVC also follows the layered coding structure as the prior H.264/SVC. However, inter-layer coding modes were no longer used in SHVC architecture. Instead, the high level syntax (HLS) was introduced in which the BL reconstructed information, i.e., texture and motion were used at the EL as new reference. This coding approach makes SHVC easy in development and deployment. Figure 2 shows a conceptual architecture of

the SHVC encoder standard where HLS approach was employed, the inter layer reference picture is created based on the BL information.

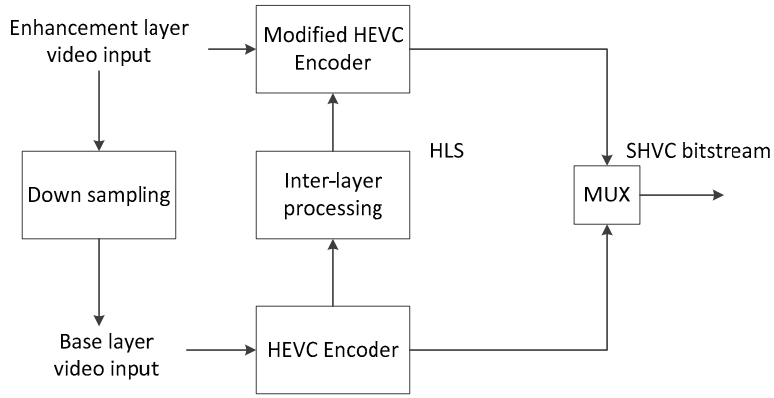

Figure 2. Conceptual architecture of the Scalable High Efficiency Video Coding (SHVC) encoder.

Since SHVC adopted the HLS approach mentioned above, recent research to improve its compression efficiency mainly focused on the inter-layer processing. The main target is to provide better predictions for the EL coding unit based on the available BL decoded data. The generalized inter-layer residual prediction method in [12] creates a new prediction of EL by adding its temporal prediction with a collocated BL residue using several pre-defined weighting values. However, it requires more computation and memory usage compared to SHVC due to its increased number of additional coding modes. Later, a solution combining temporal and inter-layer prediction is proposed in [13]. It can be implemented by adding a new generalized combined prediction mode to SHVC. While the work in [14] disclosed an improved EL merge mode solution, which adaptively refines the merge motion vector and linearly combines the reconstructed BL picture with a refined EL merge prediction to achieve a better merge prediction quality, thus improving the SHVC RD performance. Extending this work, the authors in [15,16] proposed a joint layer prediction for SHVC which provided around 8% of bitrate saving when compared to the standard SHVC. However, none of these improvement methods were designed for surveillance video content. Finally, to address the error propagation problem happening in SHVC transmission, the work in [17,18] proposed several error concealment solutions for quality and spatial scalability SHVC, respectively.

2.2. Surveillance Video Coding

Since HEVC is originally designed for video content captured from generic cameras, HEVC may not be suitable to compress videos captured from surveillance cameras that usually contain few activities and large background areas [7]. Considering this fact, Zhang et al. proposed in [6] an efficient coding solution for videos captured from stationary cameras. In this proposal, a high quality background frame was generated and employed to efficiently compress the surveillance video. Following this direction, several background frame creation models had been presented in [19–22]. Paul et al. in [23] introduced a Gaussian Mixture Modeling (GMM) method to generate a background picture as a long-term reference. Since GMM brings more float computation, this would greatly increase the encoding time. To address this problem, Zhang et al. proposed in [24] a simple averaging method to generate the background picture. In Zhang's work, the whole background picture was coded into the stream as a special I-picture with a smaller quantization parameter. However, this may cost a large number of bits to transmit the background picture within a short time, which probably results in a traffic burst and thus packet losses as well as severe video quality degradation.

To address the traffic burst for the whole background picture, Chen et al. proposed in [25] a block-composed background reference video coding scheme. In this work, the authors split the whole background picture into background coding tree blocks to achieve smooth bitrate output. However,

these methods were mainly designed for video captured by stationary cameras and thus only suitable for such systems.

To extend for surveillance videos captured from moving cameras, Wang et al. proposed in [26] a new background modeling and referencing solution. In this work, the global motion compensation method was used to generate the background reference. Afterwards, the motion background coding tree units were adaptively selected by anchoring the input video frame on the modeling background frame. In addition, a quantization parameter adjustment method was proposed to further improve the HEVC compression performance. Although a significant compression improvement can be achieved with the prior surveillance video coding approaches [19–26], none of these works is able to cope with the dynamic changing of transmission environment and the variety of displaying devices.

To achieve scalability and compression efficiency simultaneously, the work in [27] proposed an HEVC based surveillance scalable video coding (SSVC) solution. In this work, the LTR coding approach was employed for both base and enhancement layers. The first frame appeared in video is simply cast as the LTR. However, since the importance of the LTR picture and the complexity associated to LTR creation were not well investigated, there are rooms for further improving the SSVC coding performance.

3. Proposed Surveillance Scalable Video Coding

In this section, we describe the proposed HEVC based SSVC solution. Firstly, we discuss the characteristics of surveillance videos. Afterwards, we present the overall SSVC framework. Finally, we describe an adaptive LTR (ALTR) mechanism.

3.1. Characteristics of Surveillance Videos

Generally, surveillance videos are captured by stationary cameras or moved with a narrow angle. It is crucial to exploit this special characteristic for coding surveillance video content. To reveal this fact, we present in this section both subjective and objective observations.

For subjective testing, we compute and illustrate the difference between two frames having large temporal distance obtained from three surveillance videos, namely Crossover, Mainroad and Intersection [28]. The subjective results are shown in Figure 3. The black area shown in Figure 3c,f,i is the background that consistently maintains along frames on video while the white area represents the foreground with movement objects.

This observation confirms that the video captured from surveillance systems largely contains high temporal correlation and hence results in a large background area. For this reason, if the background area is well detected, only difference between the foreground and background information should be coded and sent to the decoder to fully reconstruct the video. This can greatly increase the compression efficiency.

For objective testing, we compute the average pixel difference (APD) between a pair of frames X_t, X_{t+i} as

$$APD = \frac{1}{N}\sum_{i=0}^{N-1}\left|X_t(i) - X_{t+k}(i)\right|, \tag{1}$$

Here, kth is the temporal distance between two frames; ith is the pixel position and N is the number of pixels in a frame. It should be noted that if $k = 1$, two frames are consecutive, i.e., X_t, X_{t+1}.

This metric is computed and compared between surveillance videos, Crossover, Mainroad, Intersection and a generic video, BasketballDrill.

As shown in Figure 4, the APD computed for surveillance videos is much smaller than that of the generic video. For generic video, i.e., BasketballDrill, the APD is larger when the temporal distance increases. In generic video, not only the foreground, but also the background information changes with time. In surveillance videos, the background information rarely changes. Instead, the foreground

information slowly changes; resulting a small APD value. This again confirms that the high temporal correlation between frames is usually obtained in surveillance videos.

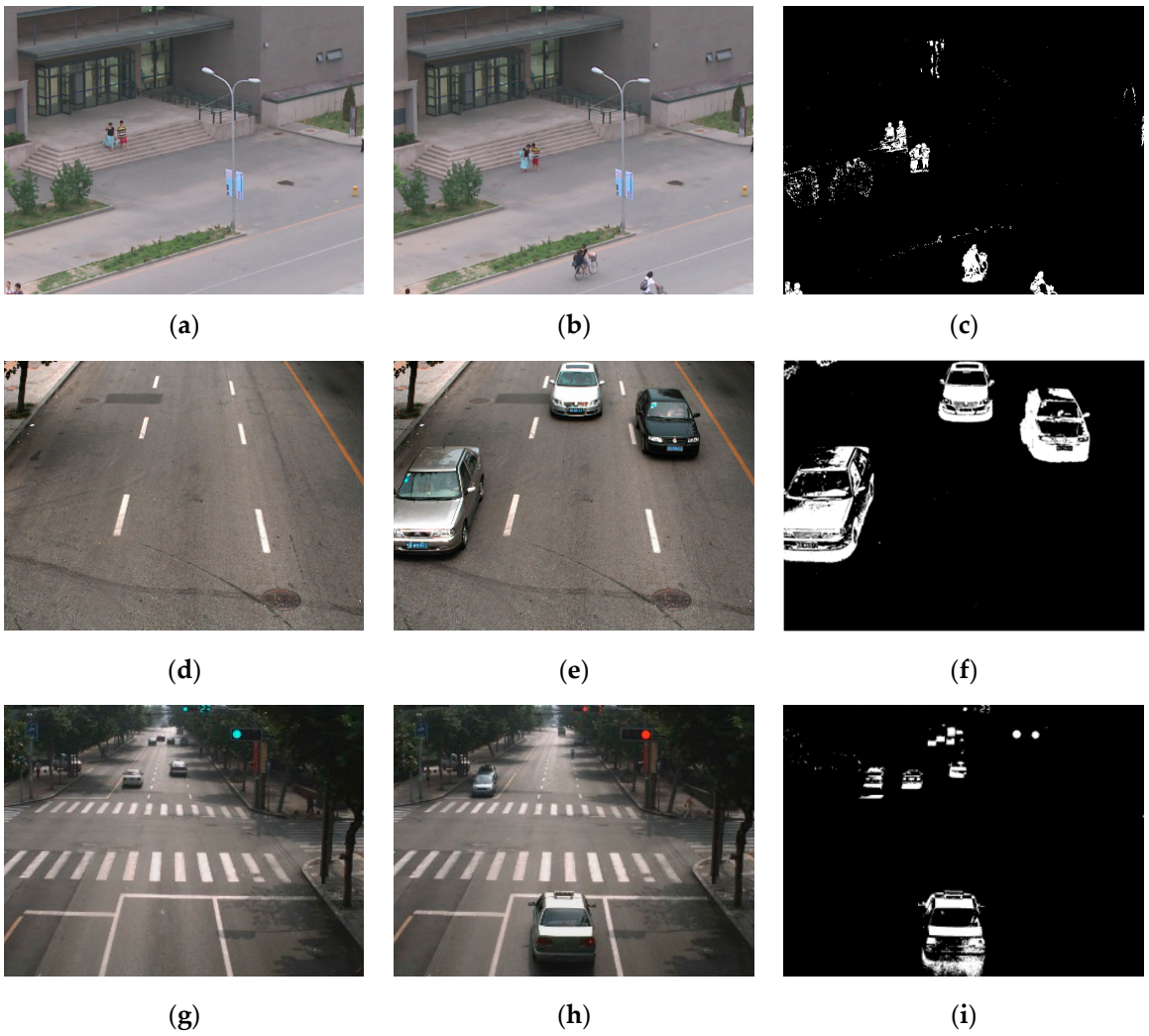

Figure 3. An illustration of frame difference in surveillance videos: (**a**–**c**) Classover #1st frame, 200th frame and the associated difference; (**d**–**f**) Mainroad #380th frame, 512th frame and the associated difference; (**g**–**i**) Intersection #1st frame, 280th frame and the associated difference.

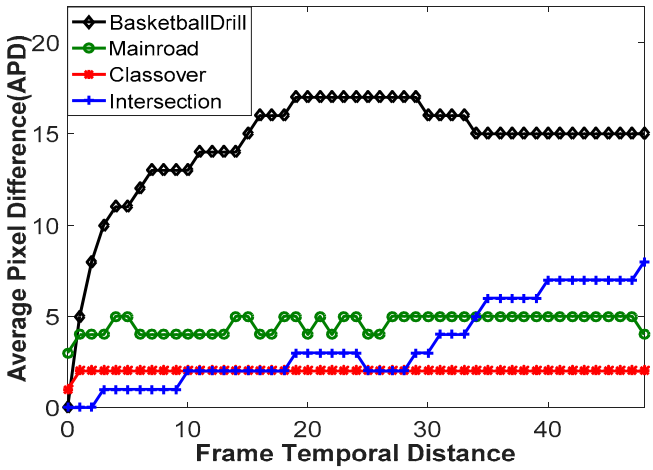

Figure 4. Average pixel difference (APD) computed for a pair of frames in surveillance and generic video.

Following these observations, we propose a novel HEVC based SSVC solution, which consists of an adaptive LTR and a QP adaptation mechanism.

3.2. Overall SSVC Framework

Figure 5 illustrates the proposed SSVC architecture. The proposed SSVC framework compresses the surveillance video with one BL and one or several ELs to provide the quality scalability. To exploit the high temporal correlation characteristic observed in surveillance videos, we employ an adaptive LTR coding approach. As highlighted, content of the surveillance video is analyzed to create and adjust the LTR picture, which is then put back to the decoded picture buffer (DPB) of either BL or EL depending on which coding layer is considered.

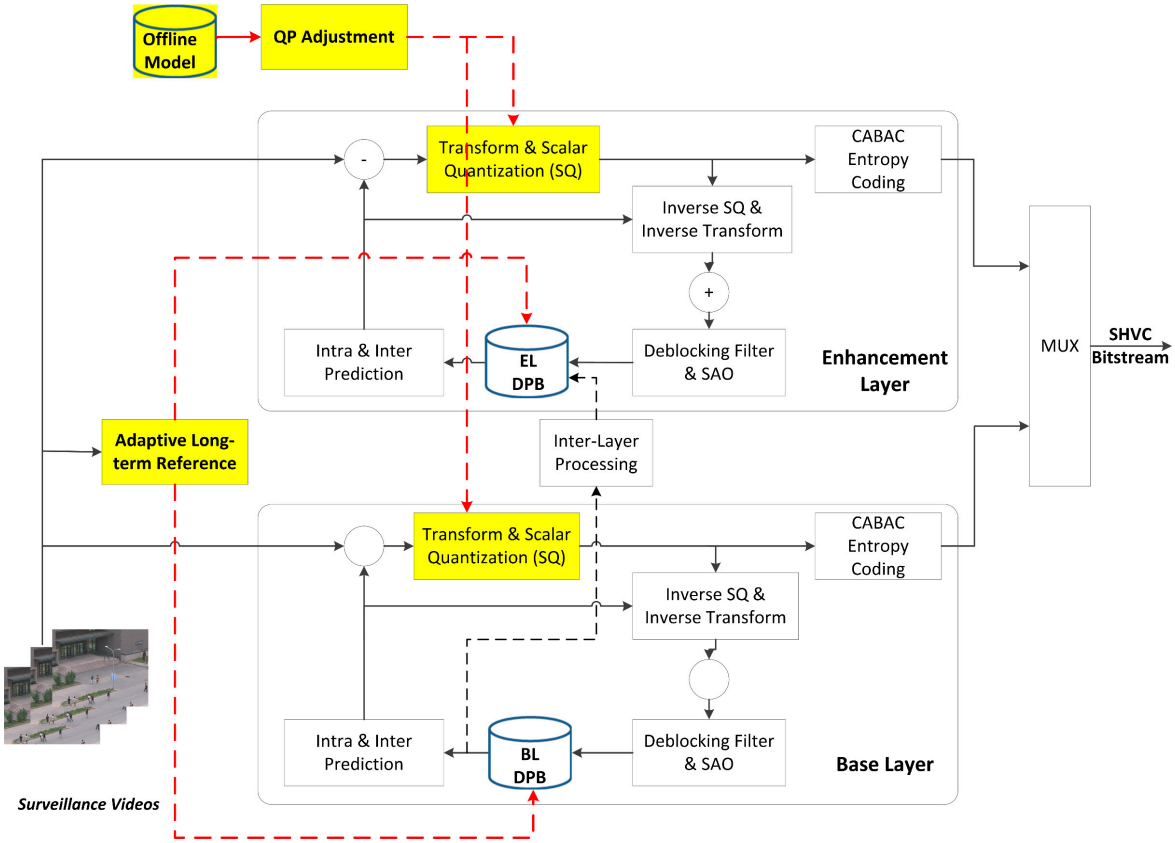

Figure 5. Proposed surveillance scalable video coding (SSVC) architecture (novel modules are highlighted).

In addition, we propose a QP adaptation mechanism, which adaptively adjusts the QP for coding pictures at LTR and non-LTR positions to achieve higher compression performance for the proposed SSVC. Other coding tools like quad-tree partitioning structure, intra, inter predictions; variable discrete cosine transforms and context adaptive binary arithmetic entropy coding are kept as in HEVC standard [2].

Finally, both BL and EL bitstreams are combined to form the scalable bitstream.

3.3. Adaptive Long-Term Reference Coding

The SHVC reference picture management is basically similar to that of the HEVC [2] in which a DPB is used for reference. Pictures in DBP can be marked as "short-term" or "long-term" reference. For SHVC, the EL DPB can refer to the BL decoded picture. This high-level syntax approach makes SHVC easy in deployment, especially when compared to the prior SVC standard [8].

LTR coding has shown that significant compression improvement can be achieved for coding videos with slow movement objects and having large background areas, e.g., surveillance and conference videos [17–26]. The LTR coding consists of two important steps: (i) the LTR creation and (ii) the LTR exploitation.

Generally, the better quality of LTR picture, the fewer bitrate is needed to code the video. However, the quality of LTR picture is usually proportional with the complexity associated to the LTR creation. Since the encoder based LTR creation approach presented in [6,19,20] greatly increases the complexity associated to the background reference creation and introduces the overhead bit-rate, we propose in this paper a decoder based LTR creation. In this method, the LTR is initialized with a decoded picture obtained from the first frame of the input video and adaptively updates it along the video. Our observation shown in Section 2 indicates that the pictures of surveillance video usually have high temporal correlation even with a large temporal distance. Therefore, the decoded picture of the first frame in surveillance video can be used to initialize the LTR. Figure 6 shows the prediction structure comparison between the standard SHVC and the proposed SSVC for the RA configuration with a group of pictures (GOP) of 8 [11]. The temporal layer index (TId) is therefore ranging from zero to three.

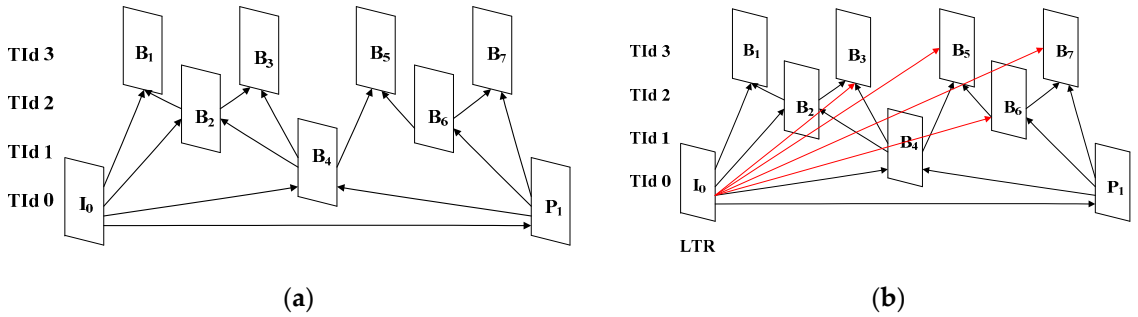

Figure 6. The prediction structure of: (**a**) the standard SHVC; (**b**) the proposed SSVC (highlighted the new referencing with red arrow).

In the decoder based LTR coding approach, only decoded frames at base and enhancement layers are employed to create the LTR. Therefore, no overhead bitrate is needed to achieve a similar LTR at the decoder side. This solution naturally requires low complexity. However, the initial LTR may not work well for coding the remaining pictures, especially when movement objects appear. To overcome this problem, we propose to adaptively update the LTR picture using a content analysis as shown in Figure 7.

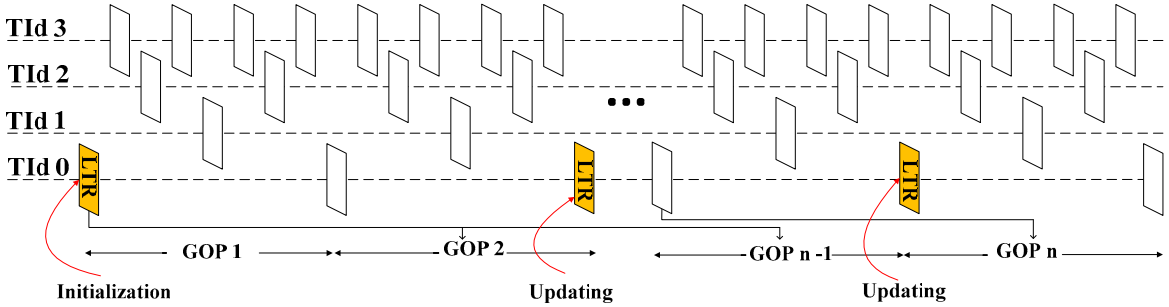

Figure 7. Adaptive long-term reference (LTR) selection.

In practice, whenever new objects appear on video, the difference between the current and its LTR pictures will greatly increase. In this case, the compression for the following pictures using the LTR concept will not be efficient. Therefore, we propose to use a sum of absolute difference (SAD) metric computed between the currently decoded picture and its LTR ones to decide whether or not the LTR should be updated. The SAD metric is measured as:

$$SAD_t = \sum_{i=0}^{N-1} \left| \hat{X}_t(i) - LTR_t(i) \right|, \tag{2}$$

where ith is the pixel index of a frame having N pixels and tth denotes the frame index in a surveillance video.

Generally, the background information should be updated if the SAD is large and vice versa. To identify this case, an adaptive thresholding method is used. An initial threshold (ThD) can be simply computed as an averaged SAD from the first LTR picture to the current picture, this means:

$$ThD_t = \frac{1}{t} \sum_{j=0}^{t-1} SAD_j. \tag{3}$$

Then, if $(SAD_t < ThD_t)$, the current LTR is continue used; otherwise, the most recent reference frame will be marked as LTR. The ALTR solution can be performed as in Table 1.

Table 1. Pseudo-code of the ALTR algorithm.

Input: Sequence to compress Output: LTR indexing
Initialize LTR by the first decoded frame
1. $LTR = \hat{X}_0$
Initializing ***ThD****:*
2. $ThD_0 = 0$;
3. for t = 0, 1, . . . , (GOP_size -1) do:
4. $SAD = \sum_{i=0}^{N-1} \left| \hat{X}_{t+1}(i) - \hat{X}_t(i) \right|$
5. $ThD_1 = ThD_0 + SAD$
6. $ThD_1 = ThD_1$/GOP_size
end for
Checking the Update:
7. for k = GOP_size, 2 × GOP_size, . . . to the last GOP do:
8. $SAD_k = \sum_{i=0}^{N-1} \left| \hat{X}_k(i) - LTR_k(i) \right|$
9. if $(SAD_k > ThD_{k-1})$:
10. $ThD_k = SAD_k$
11. $LTR = \hat{X}_k(i)$
12. end if
13. end for

4. LTR Quantization Parameter Adaptation

Conventional rate-distortion optimization (RDO) theory applied to video coding [4,29] aims to find the optimal coding mode, CU size or coding parameters by searching for the ones which minimizes the RD cost, J_{cost} as:

$$\begin{gathered} \arg\min J_{cost} \\ \text{w.r.t. } J_{cost} = D + \lambda \times R, \end{gathered} \tag{4}$$

where D is the distortion computed between the reconstructed and the original pictures, R is the coding rate and λ is the Lagrange multiplier.

In video coding, the quality of coded picture can be adjusted by the QP value, which ranges from 0 to 51 [2]. The larger QP, the smaller bitrate is obtained for the compressed video. Consequently, the lower quality of decoded picture can be achieved. The typical RD curve can be demonstrated as in Figure 8.

In the proposed SSVC scheme, the LTR picture is referred by not only the next frame but also the consecutive ones. Therefore, the quality of LTR picture will directly affect to the overall SSVC

compression performance. In this regard, we propose to adaptively adjust the QP for pictures at LTR position. Given that the initial QP of sequence is QP_{Seq}, the QP set for pictures at the LTR position, QP_{LTR}, can be adjusted as:

$$QP_{LTR} = QP_{Seq} - \Delta QP \tag{5}$$

where, ΔQP is the quantization parameter offset between the pictures at the LTR and remaining positions; $\Delta QP = \{0, 1, 3, \ldots\}$. $\Delta QP = 0$ refers to the case a similar QP can be used for pictures at both LTR and other positions.

Finally, the optimal ΔQP value adjusted for pictures at LTR positions can be achieved by:

$$\begin{gathered} \arg\min J_{cost}(\Delta QP) \\ \text{w.r.t. } J_{cost}(\Delta QP) = D(\Delta QP) + \lambda \times R(\Delta QP) \end{gathered} \tag{6}$$

There have already been several researches on modeling the RD cost and associated optimal parameters, such as the quadratic RD function proposed in [30] to reveal the relationship between rate and quantization parameter. In other work, the authors in [31] proposed a rhρ-domain RD analysis while the authors in [32] proposed a λ-domain rate control algorithm. Recently, the authors in [33] proposed a frame bitrate allocation, which models the relationship between the inter-frame dependency and the target bitrate while the authors in [34] introduced a frame based QP adjustment mechanism for HEVC. However, those RD models were mainly designed for rate control problem in either H.264/AVC or HEVC standards.

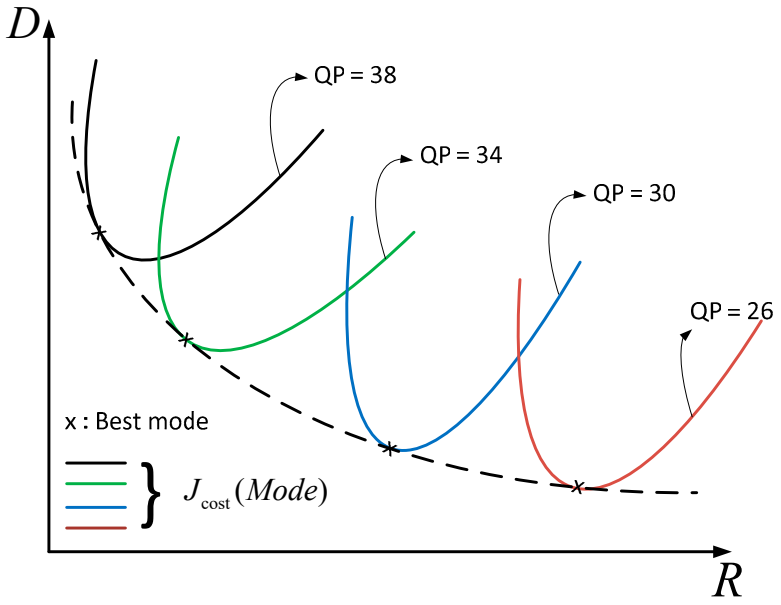

Figure 8. A typical operational rate-distortion (RD) curve.

In this paper, to reveal the relationship between ΔQP and the SSVC RD performance, we conducted a statistical learning experiment for several surveillance sequences, i.e., Overbridge and Crossroad obtained from [28]. In this experiment, the J_{cost} are measured for several ΔQPs, i.e., $\Delta QP = \{0, 1, 2, \ldots, 10\}$. The J_{cost} is computed as in Equation (6) in which the $D(\Delta QP)$ is measured by the mean square error (MSE) between the original and reconstructed pictures while the Lagrange multiplier λ is set proportionally with QP, i.e., $\lambda = 0.85 \times 2^{(QP-12)/3}$ [35].

To find the relationship between ΔQP and J_{cost}, we examined several linear and non-linear functions, including linear, exponential, second-order, third-order and fourth-order polynomials [36] as the following expressions:

- Linear model:

$$J_{cost}(\Delta QP) = a \times \Delta QP \tag{7}$$

- Exponential model:

$$J_{cost}(\Delta QP) = a \times e^{\Delta QP} \tag{8}$$

- Second-order polynomial model (denoted as 2nd Poly):

$$J_{cost}(\Delta QP) = a \times \Delta QP^2 + b \times \Delta QP + c \tag{9}$$

- Third-order polynomial model (denoted as 3rd Poly):

$$J_{cost}(\Delta QP) = a \times \Delta QP^3 + b \times \Delta QP^2 + c \times \Delta QP + d \tag{10}$$

- Fourth-order polynomial model (denoted as 4th Poly):

$$J_{cost}(\Delta QP) = a \times \Delta QP^4 + b \times \Delta QP^3 + c \times \Delta QP^2 + d \times \Delta QP + f \tag{11}$$

here, (a, b, c, d, f) are model parameters which can be experimentally determined by an offline training process.

To identify the most suitable modeling function, we conducted several experiments and illustrated in Figure 9 the results obtained for $J_{cost}(\Delta QP)$ and several fitted models. In addition, we compute and compare the coefficient of determination R-squared (R^2) [37] of the fitting model. Intuitively, R^2 expresses the "goodness of fit" of a model. ($R^2 = 1$) indicates that the model perfectly fits between the mentioned models.

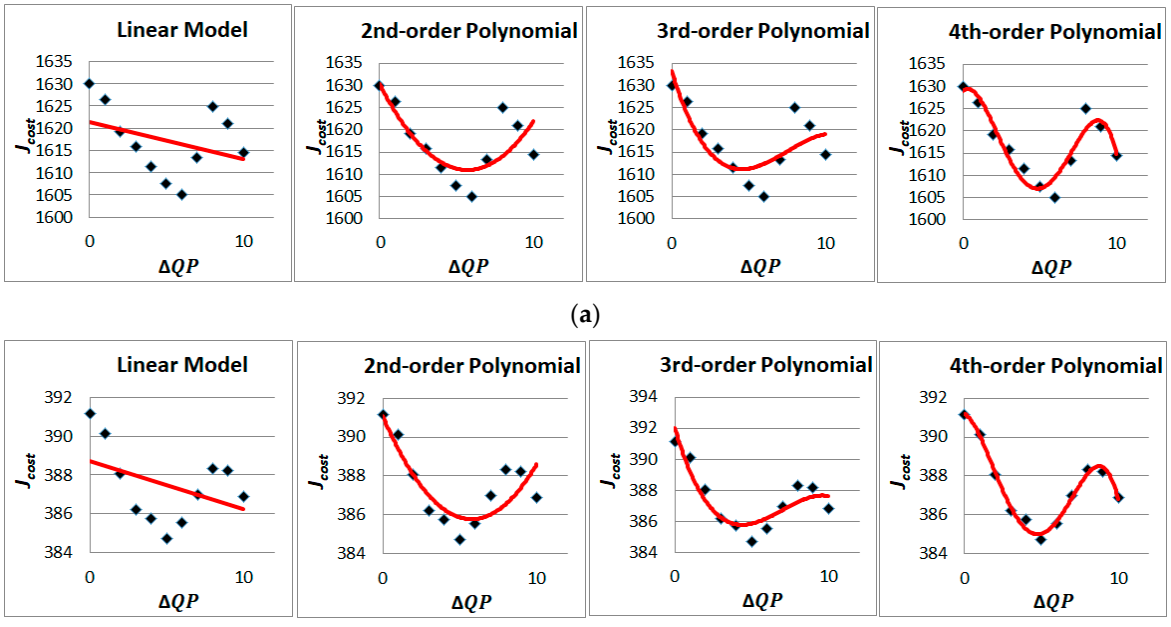

Figure 9. $J_{cost}(\Delta QP)$ vs. ΔQP with fitting models for (**a**) Overbridge; (**b**) Crossroad.

From the results obtained in Table 2 and Figure 9, it is realized that among the studied models, the fourth-order polynomial model achieves the highest correlation between (J_{cost}) and ΔQP. Therefore, we adopted in this paper a fourth-order polynomial model based QP adaptation solution.

Table 2. R-squared computation for various fitting models.

Sequences	QP_{Seq}	Fitting Model				
		Linear	**Exponential**	**2nd Poly**	**3rd Poly**	**4th Poly**
Overbridge	38	0.1855	0.1849	0.8238	0.9633	0.9831
	34	0.1420	0.1416	0.7762	0.9137	0.9696
	30	0.0203	0.0201	0.6504	0.7158	0.8470
	26	0.1247	0.1241	0.6073	0.6700	0.8966
Crossroad	38	0.0493	0.0497	0.6593	0.9517	0.9764
	34	0.0007	0.0007	0.6069	0.7948	0.9345
	30	0.1722	0.1710	0.7279	0.8414	0.9856
	26	0.0936	0.0935	0.4525	0.6520	0.8991
Average		0.0985	0.0982	0.6630	0.8128	0.9365

From the adopted model, an optimal delta QP can be derived as:

$$\Delta QP \triangleq \left(\frac{\partial (J_{cost})}{\partial (\Delta QP)} = 0 \right) \tag{12}$$

By using a QP adjustment mechanism for LTR coding, a higher quality LTR can be achieved while keeping a reasonable consumed bitrate. The model parameters $(a,\ b, c, d, f)$ are experimentally computed using an offline learning process as specified in Section 5.

5. Performance Evaluation and Discussions

5.1. Test Conditions

To assess the proposed SSVC solution, six common surveillance videos obtained from PKU-SVD-A dataset [28] were used in the experiments in which two sequences, Intersection and Mainroad have size of 1600 × 1200 while four sequences, Bank, Campus, Classover and Office have size of 720 × 576. The name and characteristics of those sequences are specified in Table 3 while Figure 10 illustrates the first frame of each sequence.

Table 3. Summarization of test conditions.

Test sequences, spatial resolution, frame rate and number of frames	1. Bank, 720 × 576, @30Hz, 297 frames 2. Campus, 720 × 576, @30Hz, 297 frames 3. Classover, 720 × 576, @30Hz, 297 frames 4. Intersection, 1600 × 1200, @30Hz, 297 frames 5. Mainroad, 1600 × 1200, @30Hz, 297 frames 6. Office, 720 × 576, @30Hz, 297 frame
QP setting (BL, EL)	{(38; 34), (34; 30), (30; 26), (26; 22)}
Coding configurations	Low Delay (LD) and Random Access (RA)
Other coding environment	- Microsoft visual studio 2017, C/C++ programming - Processor: Intel® Core i7—3.2 GHz - RAM: 8.00 GB - System: Win 10, 64-bit
Codec comparison	- SHVC: Standard SHVC obtained in [38] - Ref [20]: SSVC with LTR proposed in [27] - Ref [27]: SSVC with QPA proposed in [34] - ALTR: SSVC with only ALTR described in Section 3 - QPA: SSVC with only QPA described in Section 3 - Proposal: Proposed SSVC with both ALTR and QPA

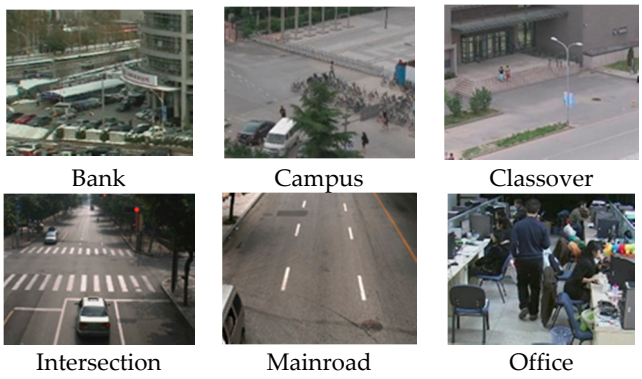

Figure 10. Illustration of the first frame for test surveillance videos.

Video compression benchmarks include the state-of-the-art SHVC standard, obtained with the SHM reference software version 12.3 [38], the prior SSVC solution [27] and the SSVC with a QP-lambda model proposed in [34]. The QPA (quantization parameter adaptive) model parameters (a, b, c, d, f) are experimentally learn by using a set of training sequences including Overbridge, Crossroad, BQSquare, Vidyo1, Vidyo3 obtained in [28,39].

5.2. SSVC Compression Performance Assessment

To assess the RD performance improvement, the common BD-rate saving [40] computed between the previous SSVC [27], the SSVC with the QPA proposed in [34] and the proposed SSVC with respect to the standard SHVC are performed. Table 4 shows the BD-rate saving comparison between the proposed SSVC and the reference benchmarks. Here the overall YUV-PSNR [41] is generally computed as the follows:

$$PSNR_{YUV} = \frac{6 \times PSNR_Y + PSNR_U + PSNR_V}{8} \tag{13}$$

Table 4. BD-rate saving with the proposed SSVC, ALTR, QPA, Ref [27] and Ref [34] when compared to the standard SHVC.

Sequences	Low Delay					Random Access				
	Ref [27]	**Ref [34]**	**ALTR**	**QPA**	**Proposal**	**Ref [27]**	**Ref [34]**	**ALTR**	**QPA**	**Proposal**
Bank	−2.97	−6.92	−3.05	−3.83	−8.20	−2.51	−11.60	−2.77	−9.89	−16.16
Campus	−2.52	−5.80	−2.36	−4.79	−8.63	−2.18	−9.49	−2.80	−9.30	−14.23
Classover	−1.40	−2.58	−1.29	−4.04	−6.09	−2.34	−5.68	−2.43	−7.46	−11.81
Intersection	−1.70	N/A	−1.71	−1.25	−3.00	−1.56	N/A	−1.66	−3.35	−6.49
Mainroad	−5.08	N/A	−5.62	−2.15	−9.05	−4.74	N/A	−4.21	−8.14	−13.70
Office	−1.65	−3.22	−1.75	−3.87	−6.45	−1.85	−5.21	−1.81	−7.68	−13.45
Averages	−2.55	−4.63	−2.63	−3.32	−6.90	−2.53	−8.00	−2.61	−7.64	−12.64

As shown in Table 4, some conclusions and discussions can be obtained as:

1. The proposed SSVC significantly outperforms the standard SHVC for all test sequences and coding configurations. The BD-rate saving can achieve up to 9% and 16% for LD and RA configurations, respectively.
2. In average, the BD-rate saving achieved with the proposed SSVC is 6.9% for LD coding configuration and 12.6% for RA case.
3. For sequences containing few movement objects like Mainroad, the LTR method shows an important BD-rate improvement.

4. For sequences having multiple movement objects like Intersection, it is difficult to obtain a LTR with large background area. Hence, the coding achievement with the proposed ALTR and QPA is just about 3% and 6.5% for LD and RA configurations, respectively.
5. Compared to the prior SSVC with only LTR coding [27], the proposed SSVC achieved better BD-rate saving for all test sequences. This comes from the QPA mechanism as described in Section 4.
6. Compared to the SSVC with QPA method proposed in [34], the SSVC with proposed QPA also achieved better BD-rate saving for all test sequences. This comes from the soft-adaptive QP mechanism proposed in Section 4. It should be noted that, the previous QPA proposed in [34] has limitation in obtaining a set of pre-defined lambda values.

Finally, the BD-rate saving can indirectly be assessed through the QPA model accuracy. In this case, to assess the QPA model accuracy, we computed and shown in Figure 11 the $J_{cost}(\Delta QP)$ for $\Delta QP = \{1, 2, 3, \ldots, 10\}$ and associated fitting model using the proposed fourth order polynomial model described in Equation (11). From the obtained results, it again confirms that the assumed QPA model highly represents the correlation between the SSVC RD performance and the ΔQP. This leads to the high compression gain with the proposed QPA mechanism.

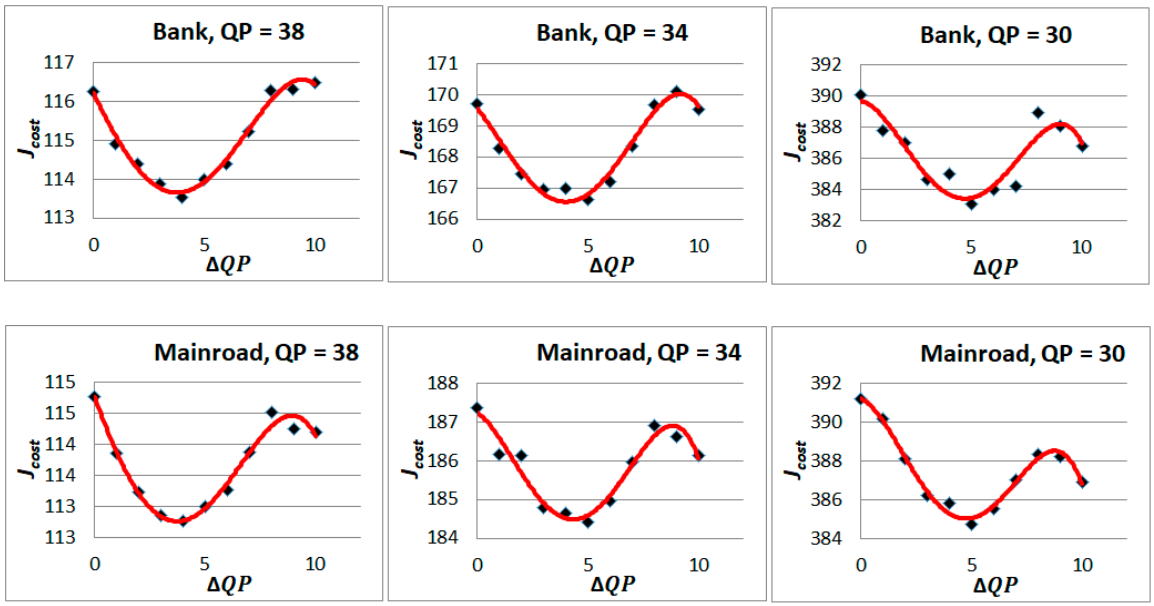

Figure 11. QPA model accuracy assessment where red curve is the fitting model.

5.3. BD-Rate Improvement with ALTR and QPAs

To assess the contributions of ALTR and QPA method, we measure the BD-rate improvement of the proposed SSVC with ALTR and QPA, respectively. Figure 12 illustrates the BD-rate saving comparison between the proposed SSVC with ALTR, QPA and both of these methods, respectively. The standard SHVC is used as a benchmark. It should be noted that, for both cases, the QPA were adopted.

As shown in Figure 12, with ALTR, an important BD-rate saving can be achieved for both LD and RA coding configurations and for all six test sequences. The compression gain is highest for the Mainroad sequence that contains very few objects and low motion activities.

From the obtained results in Figure 12, it can also be concluded that the proposed QPA contributes a consistent coding improvement for both LD and RA configurations and for all test sequences.

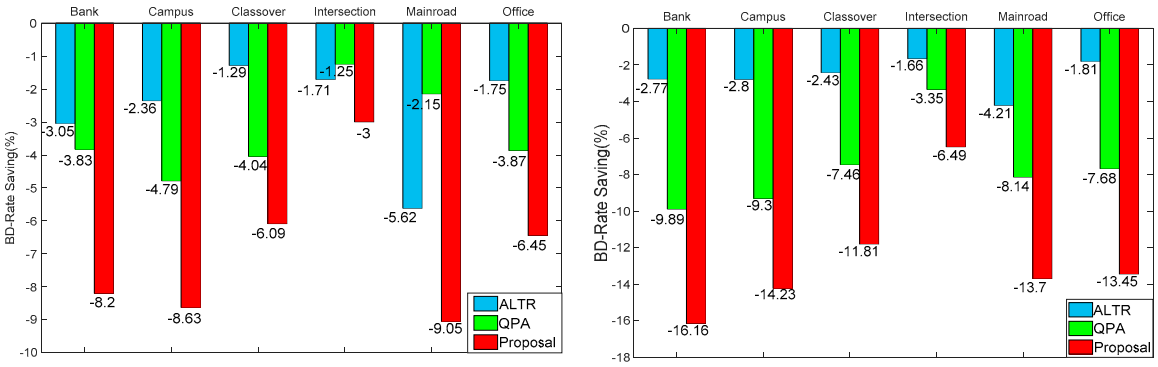

Figure 12. SSVC BD-rate saving with ALTR and QPA: (**a**) low delay; (**b**) random access.

5.4. Complexity Assessment

As the ALTR and QPA are additional coding tools when compared to the standard SHVC, they may increase the SSVC computational complexity. To assess this problem, we measure the encoding time (in second) for the proposed SSVC (ET_{SSVC}) and the standard SHVC (ET_{SHVC}). The encoding time increase (ETI) with the proposed coding solution is then computed as:

$$ETI = \frac{ET_{SSVC} - ET_{SHVC}}{ET_{SHVC}} \times 100 \tag{14}$$

Figure 13 illustrates the encoding time (s) comparison for various QPs while Table 5 presents the ETI (%) comparison for various surveillance videos.

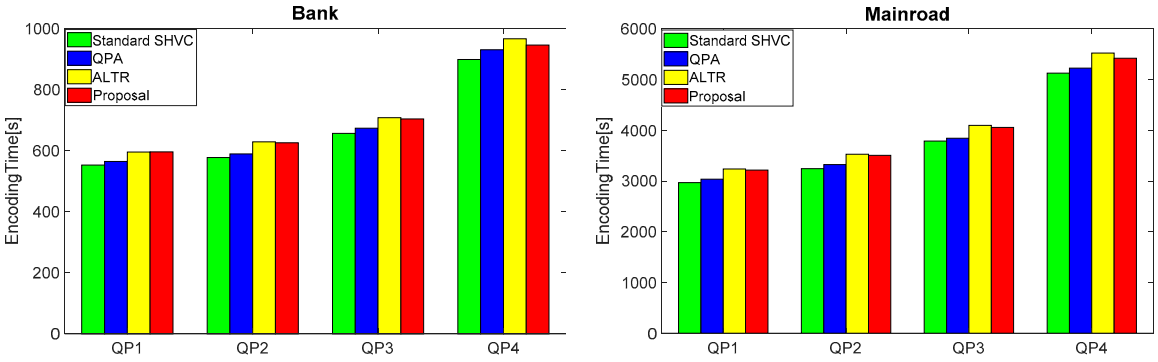

Figure 13. Encoding time comparison between various surveillance video coding solutions.

Table 5. Encoding time increase [%] comparison.

Sequence	QPA	LTR	ALTR	Proposal
Bank	2.55	2.56	7.97	7.17
Campus	2.66	2.85	8.10	7.93
Classover	2.33	1.78	8.55	7.71
Intersection	2.05	3.24	8.75	8.48
Mainroad	2.05	2.39	8.42	7.31
Office	1.79	1.09	8.19	7.56
Average	2.24	2.32	8.33	7.69

As shown in Figure 13 for all QPs, the encoding time increase with the proposed QPA and ALTR methods is negligible. Concretely, Table 5 reveals that compared to SHVC standard, only 7.7% encoding time increase can be recognized with the proposed SSVC. If the QPA process is employed together

with ALTR, the encoding time increase is slightly smaller than when only ALTR is employed. In this case, the good quality of LTR obtained by QPA mechanism may help better finding the optimal coding mode for SSVC. Finally, compared to the LTR based method, the ALTR asks more computations due to the checking and updating processes described in Table 1.

6. Conclusions

In this paper, we propose an efficient HEVC based surveillance scalable video coding solution for visual surveillance system. The proposed surveillance scalable video coding solution is developed on the top of the HEVC standard and exploits the motion characteristics observed in surveillance videos through an adaptive long-term reference coding. To optimize the LTR coding approach, we introduce a statistical QPA solution, which adaptively adjusts the QP for coding pictures at LTR and non-LTR positions. In the proposed QPA model, a fourth-order polynomial function is used to represent the relationship between the RD cost and the amount of QP adjustment for pictures at LTR positions. Finally, for compression performance, the proposed SSVC significantly outperforms the SHVC standard and the prior SSVC benchmark, notably with 6.9% and 12.6% BD-rate saving on average for LD and RA configurations, respectively. The future works can consider improving the accuracy of the LTR selection mechanism or performing an online model parameters estimation process.

Funding: This work has been supported by Vietnam National University, Hanoi (VNU), under Project No. QG.19.22.

Acknowledgments: The authors would like to thanks the editors, reviewers, and MCT alumni Dao Thi Hue Le for their helpful comments to improve the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Desurmont, X.; Bastide, A.; Chaudy, C.; Parisot, C.; Delaigle, J.F.; Macq, B. Image analysis architectures and techniques for intelligent surveillance systems. *IET Image Process.* **2005**, *152*, 224–231. [CrossRef]
2. Sullivan, G.J.; Ohm, J.-R.; Han, W.-J.; Wiegand, T. Overview of the High Efficiency Video Coding (HEVC) standard. *IEEE Trans. Circuits Syst. Video Technol.* **2012**, *22*, 1649–1668. [CrossRef]
3. Wiegand, T.; Sullivan, G.J.; Bjontegaard, G.; Luthra, A. Overview of the H.264/AVC video coding standard. *IEEE Trans. Circuits Syst. Video Technol.* **2003**, *13*, 560–576. [CrossRef]
4. Sullivan, G.J.; Wiegand, T. Rate-distortion optimization for video compression. *IEEE Signal Process. Mag.* **1998**, *15*, 74–90. [CrossRef]
5. Zhang, X.; Liang, L.; Huang, Q.; Liu, Y.; Huang, T.; Gao, W. An efficient coding scheme for surveillance videos captured by stationary cameras. In Proceedings of the SPIE-VCIP, Huangshan, China, 11–14 July 2010.
6. Zhang, X.; Huang, T.; Tian, Y.; Gao, W. Background—Modelling—Based adaptive prediction for surveillance video coding. *IEEE Trans. Image Process.* **2014**, *23*, 769–784. [CrossRef]
7. Vetro, A.; Haga, T.; Sumi, K.; Sun, H. Object-based coding for long–term archive of surveillance video. In Proceedings of the IEEE International Conference on Multimedia and Expo, Baltimore, MD, USA, 6–9 July 2003.
8. Schwarz, H.; Marpe, D.; Wiegand, T. Overview of the Scalable Video Coding Extension of the H.264/AVC Standard IEEE Trans. *Circuits Syst. Video Technol.* **2007**, *17*, 1103–1120. [CrossRef]
9. ITU-T; ISO/IEC JTC 1. *Generic Coding of Moving Pictures and Associated Audio Information—Part 2: Video*; ITU-T Rec. H.262—ISO/IEC 13818-2; ISO: Geneva, Switzerland, 1994.
10. Olivier, A.; Alexandros, E.; Carsten, H.; Ganesh, R.; Liam, W. MPEG-4 Systems: Overview. *Signal Process. Image Commun.* **2000**, *15*, 281–298.
11. Boyce, J.M.; Ye, Y.; Chen, J.; Ramasubramonian, A.K. Overview of SHVC: Scalable Extensions of the High Efficiency Video Coding Standard. *IEEE Trans. Circuits Syst. Video Technol.* **2016**, *26*, 20–34. [CrossRef]
12. Li, X.; Chen, J.; Rapaka, K.; Karczewicz, M. Generalized inter-layer residual prediction for scalable extension of HEVC. In Proceedings of the IEEE International Conference on Image Processing, Melbourne, Australia, 15–18 September 2013.

13. Lai, P.; Liu, S.; Lei, S. Combined temporal and inter-layer prediction for scalable video coding using HEVC. In Proceedings of the Picture Coding Symposium, San Jose, CA, USA, 8–11 December 2013.
14. HoangVan, X.; Ascenso, J.; Pereira, F. Improving enhancement layer merge mode for HEVC scalable extension. In Proceedings of the Picture Coding Symposium, Cairns, Australia, 31 May–3 June 2015.
15. HoangVan, X.; Ascenso, J.; Pereir, F. Improving SHVC performance with a joint layer coding mode. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Shanghai, China, 20–25 March 2016.
16. HoangVan, X.; Jeon, B. Joint Layer Prediction for Improving SHVC Compression Performance and Error Concealment. *IEEE Trans. Broadcasting* **2018**, *65*, 504–520. [CrossRef]
17. Huu, T.N.; Trieu, D.D.; Jeon, B.; HoangVan, X. Performance evaluation of frame-loss error-concealment solutions for the SHVC standard. *IEIE Trans. Smart Process. Comput.* **2017**, *6*, 428–436. [CrossRef]
18. Huu, T.N.; Canh, T.N.; HoangVan, X.; Jeon, B. A Frame Loss Concealment Solution for Spatial Scalable HEVC using Base Layer Motion. In Proceedings of the IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, Jeju Island, Korea, 5–7 June 2019.
19. Chakraborty, S.; Paul, M.; Murshed, M.; Ali, M. An efficient video coding technique using a novel non-parametric background model. In Proceedings of the IEEE International Conference on Multimedia and Expo Workshops, Chengdu, China, 14–18 July 2014; pp. 1–7.
20. Huang, Z.; Hu, R.; Wang, Z. Background subtraction with video coding. *IEEE Signal Process. Lett.* **2013**, *20*, 1058–1061. [CrossRef]
21. Zhao, L.; Zhang, X.; Tian, Y.; Wang, R.; Huang, T. A background proportion adaptive Lagrange multiplier selection method for surveillance video on HEVC. In Proceedings of the IEEE International Conference on Multimedia and Expo, San Jose, CA, USA, 15–19 July 2013; pp. 1–6.
22. Sobral, A.; Vacavant, A. A comprehensive review of background subtraction algorithms evaluated with synthesis and real videos. *Comput. Vis. Image Underst.* **2014**, *122*, 4–21. [CrossRef]
23. Paul, M.; Lin, W.; Lau, C.; Lee, B. A long-term reference frame for hierarchical b-picture based video coding. *IEEE Trans. Circuits Syst. Video Technol.* **2014**, *24*, 1729–1742. [CrossRef]
24. Zhang, X.; Tian, Y.; Huang, T.; Dong, S.; Gao, W. Optimizing the hierarchical prediction and coding in HEVC for surveillance and conference videos with background modeling. *IEEE Trans. Image Process.* **2014**, *23*, 4511–4526. [CrossRef]
25. Chen, F.D.; Li, H.Q.; Liu, D.; Wu, F. Block-composed background reference for high efficiency video coding. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, *27*, 2639–2651. [CrossRef]
26. Wang, G.; Li, B.; Zhang, Y.; Yang, J. Background modelling and referencing for moving cameras-captured surveillance video coding in HEVC. *IEEE Trans. Image Process.* **2018**, *20*, 2912–2934.
27. Hoang Van, X.; Dao Thi Hue, L.; PhamVan, G. Adaptive long-term reference selection for efficient scalable surveillance video coding. In Proceedings of the IEEE International Symposium on Embedded Multicore/Many core Systems-on-Chip, Hanoi, Vietnam, 12–14 September 2018; Volume 1, pp. 69–73.
28. PKU-SVD-A. Available online: http://mlg.idm.pku.edu.cn/-resources/pku-svd-a.html (accessed on 2 January 2020).
29. Ortega, A.; Ramchadran, K. Rate—Distortion methods for image and video compression. *IEEE Signal Process. Mag.* **1998**, *15*, 23–50. [CrossRef]
30. Ma, S.; Gao, W.; Lu, Y. Rate-distortion analysis for H.264/AVC video coding and its application to rate control. *IEEE Trans. Circuits Syst. Video Technol.* **2005**, *15*, 1533–1544. [CrossRef]
31. He, Z.; Kim, Y.K.; Mitra, S.K. Low—Delay rate control for DCT video coding via ρ-domain source modelling. *IEEE Trans. Circuits Syst. Video Technol.* **2001**, *11*, 928–940.
32. Li, B.; Li, H.; Li, L.; Zhang, J. λ domain rate control algorithm for High Efficiency Video Coding. *IEEE Trans. Image Process.* **2014**, *24*, 3841–3854. [CrossRef]
33. Jing, H.; Fuzheng, Y. Efficient frame-level bit allocation algorithm for H.265/HEVC. *IET Image Process.* **2017**, *11*, 245–257.
34. Li, B.; Xu, J.; Zhang, D.; Li, H. QP refinement according to Lagrange multiplier for High Efficiency VIeo Coding. In Proceedings of the IEEE International Symposium on Circuits and Systems, Beijing, China, 19–21 May 2013.
35. Choi, J.; Park, D. A stable feedback control of the buffer state using the controlled Lagrange multiplier method. *IEEE Trans. Image Process.* **1994**, *3*, 546–557. [CrossRef] [PubMed]

36. Almira, J.M.; Szekelyhidi, L. Characterization of classes of polynomial functions. *Mediterr. J. Math.* **2016**, *13*, 301–307. [CrossRef]
37. Glantz Stanton, A.; Slinker, B.K. *Primer of Applied Regression and Analysis of Variance*; McGraw-Hill: New York, NY, USA, 1990; ISBN 978-0-07-023407-9.
38. SHVC Reference Software. Available online: https://hevc.hhi.fraunhofer.de/svn/svn_SHVCSoftware/ (accessed on 19 December 2019).
39. JVC Video Test Sequences. Available online: ftp://hevc@ftp.tnt.uni-hannover.de/testsequences/ (accessed on 2 January 2020).
40. Bjontegaard, G. Calculation of average PSNR differences between RD curves. Doc. VCEG-M33. In Proceedings of the 13th ITU-T VCEG Meeting, Austin, TX, USA, 2–4 April 2001.
41. Sullivan, G.J.; Ohm, J. Meeting report of the fourth meeting of the Joint Collaborative Team on Video Coding (JCT-VC). In Proceedings of the Joint Collaborative Team on Video Coding (JCT-VC) JCTVC-D500, Daegu, Korea, 20–28 January 2011.