VNU Journal of Science: Comp. Science & Com. Eng, Vol. 34, No. 1 (2018) 38-51

VNU Journal of Science:
Computer Science and Communication Engineering

Journal homepage: http://www.jcsce.vnu.edu.vn/index.php/jcsce

Efficient and Low Complexity Surveillance Video Compression using Distributed Scalable Video Coding

Le Dao Thi Hue[1], Luong Pham Van[2], Duong Dinh Trieu[1], Xiem HoangVan[1,*]

[1]*VNU University of Engineering and Technology, 144 Xuan Thuy, Cau Giay, Hanoi, Vietnam*
[2]*Department of Electronics and Information Systems, Ghent University*

Abstract

Video surveillance has been playing an important role in public safety and privacy protection in recent years thanks to its capability of providing the activity monitoring and content analyzing. However, the data associated with long hours surveillance video is huge, making it less attractive to practical applications. In this paper, we propose a low complexity, yet efficient scalable video coding solution for video surveillance system. The proposed surveillance video compression scheme is able to provide the quality scalability feature by following a layered coding structure that consists of one or several enhancement layers on the top of a base layer. In addition, to maintain the backward compatibility with the current video coding standards, the state-of-the-art video coding standard, i.e., High Efficiency Video Coding (HEVC), is employed in the proposed coding solution to compress the base layer. To satisfy the low complexity requirement of the encoder for the video surveillance systems, the distributed coding concept is employed at the enhancement layers. Experiments conducted for a rich set of surveillance video data shown that the proposed surveillance - distributed scalable video coding (S-DSVC) solution significantly outperforms relevant video coding benchmarks, notably the SHVC standard and the HEVC-simulcasting while requiring much lower computational complexity at the encoder which is essential for practical video surveillance applications.

Received 15 March 2018, Accepted 22 September 2018

Keywords: Surveillance video coding, HEVC standard, distributed source coding, joint layer prediction, scalable video coding.

1. Introduction

Video surveillance systems have been gaining its important role in many areas of human life, including public safety and private protection [1]. Such a system provides real-time monitoring and analysis of the observed environment. Real-world video surveillance applications typically require storing videos without neglecting any part of scenarios for weeks or months. This process generates a huge amount of data. Moreover, the heterogeneity of devices, networks and environments is also gaining a request of adaptation solutions. In this scenario, there is a critical need of a powerful video coding scheme that is featured by high coding efficiency, scalability and low encoding complexity capabilities.

* Corresponding author. Email.: xiemhoang@vnu.edu.vn
https://doi.org/10.25073/2588-1086/vnucsce.198

Figure 1 shows a basic diagram of a video surveillance system (VSS) using scalable video coding [2]. A VSS typically includes two main parts, the provider and users. The video is firstly captured and processed at the provider by a surveillance camera. Such camera can be either analog or digital type. The captured video is then compressed and sent to the users. At the user side, video data is decompressed before using for object detection, activity tracking, and/or event analysis.

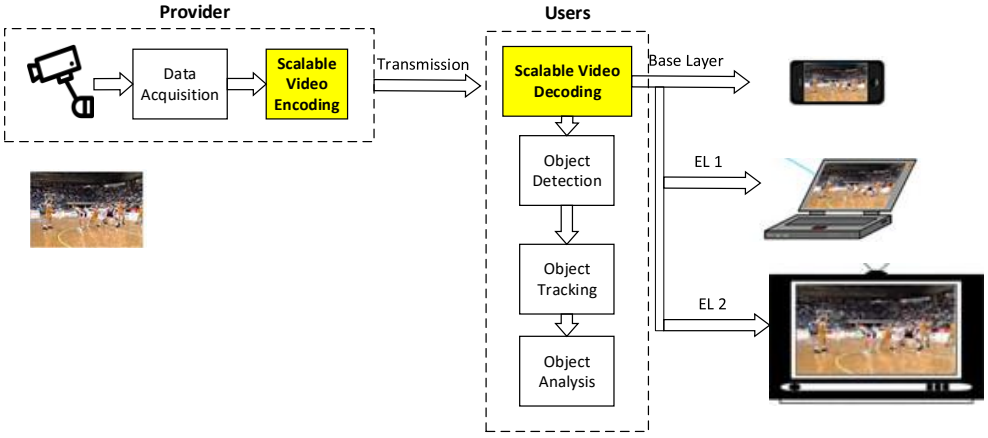


Figure 1. A video surveillance system with scalable video coding.

These surveillance applications usually require the storage of video data over a period for automatic analysis and future use. However, the storage of raw video data captured directly from cameras can be very expensive. Therefore, a video compression solution to reduce the storage space of raw surveillance video data is very essential. Besides, due to the heterogeneity of the user devices, networks and environments, e.g., smartphone, laptop or television, it is reasonable to compress the surveillance video data in a layered coding structure with one base layer and one or several enhancement layers. The layered coding structure is usually adopted in a scalable video coding scheme, such as SVC standard [2]. In this solution, the scalable bitstream makes the surveillance camera system more adaptive to the variation of network conditions and user devices.

The current video coding standards such as the High Efficiency Video Coding (HEVC) [3] and its extension, the Scalable High Efficiency Video Coding (SHVC) [4] are mainly designed for generic video content. Considering the relative static background characteristic of surveillance video data, the authors in [5, 6] proposed a background modeling based adaptive prediction for surveillance video coding. Afterwards, a large number of surveillance video coding improvements have been presented in [7-9]. However, since the surveillance video coding scheme is usually developed based on the conventional predictive video coding standards, e.g., H.264/AVC [5-7] or HEVC [9], its compression performance usually come along with the high computational complexity, hence making the encoder extremely heavy. In this case, the low encoding complexity requirement for a video surveillance system may not be satisfied. In addition, the prior surveillance video coding solutions [5-9] are unable to achieve the scalability capability as only one compression layer is used.

Distributed video coding (DVC) is another coding approach, targeting the low complexity requirement at the encoder and the robustness to error propagation at the decoder [10]. DVC was developed from two well-known information theorems, Slepian-Wolf [11] and Wyner-Ziv [12]. There have been great

attentions on DVC in recent decades with many significant contributions, notably on both practical coding architectures and improving coding tools [13, 14]. In DVC, the temporal correlation is mainly exploited at the decoder side by a so-called side information creation [15] while the encoder side is designed in a very light way. Hence, this coding solution is very attractive to emerging video coding applications, e.g., visual sensor networks, surveillance systems, and remote sensing. Recent researches have also shown that DVC is generally suitable for encoding videos featured by low and static motion contents [10, 16]. As assessed in [17], the DVC practical coding solution requires much lower encoding complexity than the traditional predictive video coding standards, e.g., H.264/AVC or HEVC while providing a more robust error resilience, yet compression efficient video coding scheme.

In this context, considering for the need of a powerful video coding solution that typically requires the high compression efficiency, scalability and low complexity capability, we proposed in this paper a novel scalable video coding solution, specially designed for surveillance video data. The proposed surveillance scalable video coding scheme is developed based on a combination of the traditional predictive video coding standards, HEVC and SHVC with the emerging distributed video coding paradigm [10]. As the layered coding structure is adopted, the proposed surveillance - distributed scalable video coding solution, namely S-DSVC, is able to provide the quality and temporal scalability features. In addition, several coding tools are also introduced to further increase the compression performance of the proposed S-DSVC solution. Experimental results revealed that the proposed S-DSVC solution significantly outperforms other relevant video coding benchmarks, notably the HEVC-simulcasting and the SHVC standards.

The rest of the paper is organized as follows. Section 2 reviews the relevant background work, while Section 3 describes the proposed S-DSVC architecture and its advanced coding tools. Afterwards, Section 4 analyses the S-DSVC performance in comparison with the HEVC-simulcasting and SHVC standard. Finally, Section 5 presents the main conclusions and ideas for future work.

2. Relevant background works

Since the proposed surveillance video coding solution is mainly developed based on the combination of the distributed and predictive coding paradigms while also providing the scalability capability, this Section describes the two most relevant background works, the distributed video coding and the scalable video coding.

2.1. Distributed video coding

The distributed video coding theoretical foundations go back to the 70's when Slepian and Wolf [11] established the achievable rates for lossless coding of two correlated sources. The Slepian and Wolf theorem (1973) states that the minimum rate to encode two correlated sources, X and Y is the same as the minimum rate for joint encoding, this means the join entropy $H(X,Y)$, with an arbitrarily small error probability for long sequences, provided that their correlation is known at the encoder and decoder. This theorem is important since it was the first establishing the rate boundary for a separate encoding but joint decoding of two correlated sources as presented in the following inequalities.

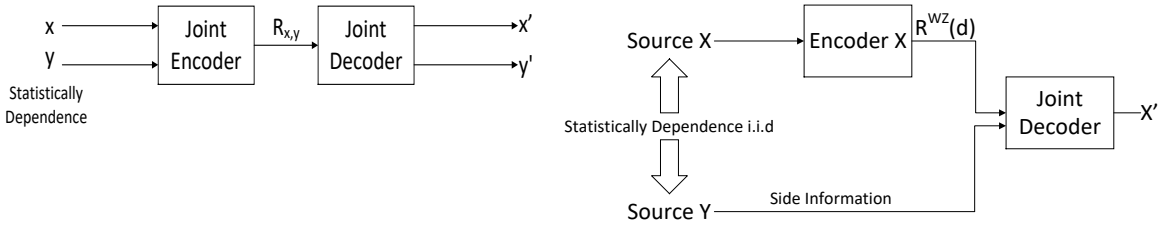
$$R_X \geq H(X \mid Y), R_Y \geq H(Y \mid X)$$
$$R_X + R_Y \geq H(X,Y) \quad (1)$$

where $H(X,Y)$and $H(Y \mid X)$denote the conditional entropy and $H(X,Y)$ denotes the joint entropy of source X and Y, respectively.

However, the Slepian and Wolf theorem refers only to the lossless coding scenario that is not the most exciting for practical video coding solutions due to the associated low

compression ratios. In 1976, Wyner and Ziv [12] extended the Slepian and Wolf theorem to the lossy compression case. The Wyner and Ziv theorem states that, for a source X with side information Y available at the decoder, the rate required to achieve a certain distortion when some side information is available at the decoder only obeys to $R_{(X|Y)}(D) \leq R^{WZ}_{(X|Y)}(D)$ where $R_{(X|Y)}(D)$ is the rate obtained when the SI is available at both the encoder and decoder. Therefore, when the statistical dependency is exploited only at the decoder, the minimum rate to transmit X at the same distortion D may increase or be the same compared to the case where the statistical dependency is exploited at both the encoder and decoder (commonly adopted in the video coding standards, e.g., H.264/AVC and HEVC).

In general, the Slepian-Wolf and Wyner-Ziv theorems proved that it is possible to achieve the same rate for the coding systems exploiting the statistical dependency only at the decoder as for the systems where the dependency is exploited at both the encoder and decoder as specified in the following conceptual coding diagrams:



(a) (b)

Figure 2. Conceptual illustration of the predictive and distributed video coding paradigms. a) Predictive video coding; b) Distributed video coding

Based on the Slepian-Wolf and Wyne-Ziv theorems, distributed video coding (DVC) provides a statistical framework where the correlation noises statistics are exploited at the decoder only; this correlation noise regards the difference between the original data only available at the encoder and the side information available at the decoder. DVC is a promising coding solution for many emerging applications such as wireless video surveillance systems, multimedia sensor networks, mobile camera phones, and remote space transmission since DVC is able to provide the following functional benefits: i) flexible allocation of the overall video codec complexity; ii) improved error resilience; iii) codec independent scalability; and iv) exploitation of multiview correlation without camera/single encoder communication [10].

2.2. Scalable video coding

Scalable Video Coding (SVC) is a highly attractive solution to the problems posed by the characteristics of modern video transmission systems. The term “scalability” in this paper refers to the coding capability of video compression solution to adapt it to the various needs or preferences of end users as well as to varying terminal capabilities or network conditions. In SVC, the video bitstream contains a base layer (BL) and or several enhancement layers (ELs) [2]. ELs are added to the BL to further enhance the quality or resolution fidelities of the BL coded video. The improvement can be made by increasing the spatial resolution, video frame-rate or video quality, corresponding to spatial, temporal and quality/SNR scalability, respectively.

Figure 3 shows an example of a SVC scheme with two layers, one base and one enhancement layers, providing the quality scalability feature. In this coding structure, the

Inter-layer processing aims to exploit the correlation between layers.

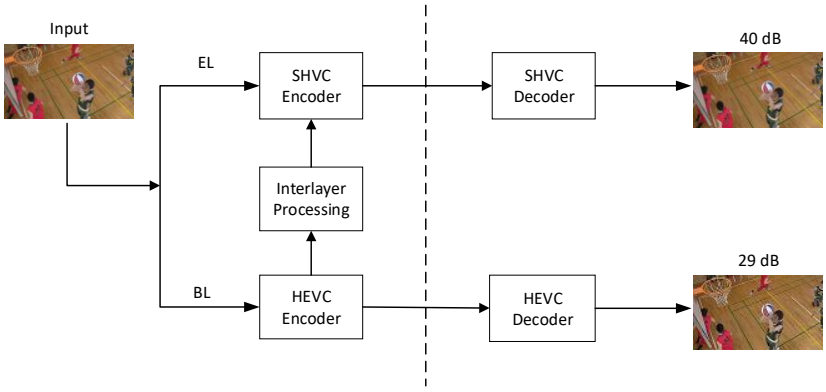


Figure 3. A conceptual structure of the SVC.

With its capabilities, SVC is generally suitable for video streaming over heterogeneous networks, devices or coding environments. Therefore, the scalability is a desirable feature for video transmission over most practical networks, especially for the case of video surveillance network as illustrated in Figure 1.

3. Proposed surveillance - distributed scalable video coding

Considering the need for a powerful surveillance video compression solution which contains the high compression performance, yet low complexity while able to provide the scalability function, we present in this Section a novel surveillance distributed scalable video coding solution, which combines the predictive and distributed coding paradigms. Before describing the proposed video coding solution, it is desired to have a brief analysis of the surveillance video content.

3.1. Surveillance video data: An analysis

In a video surveillance system, the camera is usually set at a certain position or moved with a very small motion and angle. Consider this fact, several experiments have been performed on various training video samples. For surveillance video, three training sequences obtained from the PKU-SVD-A dataset [18, 19], namely Mainroad, Classover, and Intersection while for generic video, the BasketballDrill sequence obtained from [20] are used.

First, to assess the temporal correlation and the motion activity between consecutive frames of surveillance video, a frame difference (FD) metric is computed as below:

$$FD_t = \sum_{i=1}^{N} |F_t(i) - F_{t+1}(i)| \quad (2)$$

Where t^{th} and i^{th} are the frame index and the pixel position in each frame F_t, respectively, and N denote the total number pixels of each video frame.

Since the training videos may have different spatial resolution, it is proposed to use the pixel-averaged difference *(PAD)* as computed in below to assess the motion characteristics along sequence:

$$PAD_t = \frac{FD_t}{N} \quad (3)$$

Figure 4 illustrates the *PAD* statics along consecutive frame pair obtained for the mentioned surveillance and standard videos. As shown, the *PAD* between frames in surveillance videos, notably *Mainroad, Classover, and Intersection* is greatly smaller than that of the standard video, *BasketballDrill*. In this context, the small *PAD* implies the high temporal correlation between

consecutive frames. Therefore, it is noted that the surveillance videos usually contain the low motion activity statistic.

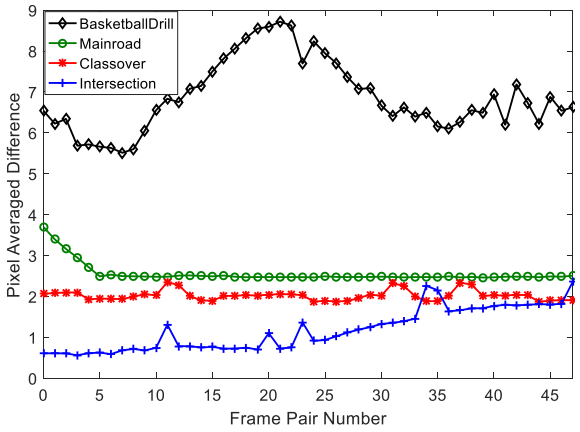


Figure 4. Pixel averaged difference between consecutive frames.

In the second experiment, we examine the background area inside each surveillance video frame by assessing the motion vector field associated to each video frame. Figure 5 illustrates the three frames captured from surveillance videos (a, b, c) and their corresponding motion vector field (d, e, f).

As shown in Figure 5, the size of motion area in surveillance videos is smaller than that of background area. Therefore, it can be concluded that in a surveillance video, the static scenes usually take a high percentage.

This important characteristic is employed in this work to build an effective video compression architecture, especially for the video surveillance system. In the next subsection, we describe more details on the coding solution proposed for the video surveillance system.

3.2. Distributed scalable surveillance video coding architecture

Figure 6 illustrates the architecture of the proposed surveillance video coding solution, in which the novel distributed coding elements are highlighted. The proposed approach also follows a layered coding approach to provide the scalability feature. The distributed coding concept is used at the enhancement layers while the predictive video coding paradigm, notably the HEVC is used at the base layer. To achieve the low computation complexity requirement, both base and enhancement layers are Intra coded; thus, resulting a low computational complexity at the encoder side.

The basic idea of the proposed solution is that the EL residue is coded exploiting some temporal correlation in a distributed way [10], and thus only a part of the EL residue, which cannot be estimated with the decoder side information (SI) creation, is coded and sent to the decoder. To avoid sending information that can be inferred at the decoder, a correlation model (CM) determines the number of least significant bitplanes, that should be different between the EL and the SI residues, thus, must be coded and transmitted.

For the EL coding, the DVC approach has been employed in our proposed method where the input video frames are split into two parts: the key and WZ frames as shown in Figure 6. In this approach, the key frames are coded with the conventional SHVC encoder [4] while the WZ frames are coded using the syndrome creation, syndrome encoding, and correlation modelling. At the decoder, the received bitstream is processed to obtain the original video data using syndrome decoding, syndrome reconstruction, correlation modelling, and side information (SI) residue creation. In such coding scheme, the low complexity features of DVC are again effectively exploited in this approach where both key and WZ frames are coded using a simple Intra and transform coding approaches; thus, no complex motion estimation is performed at the proposed S-DSVC encoder [14].

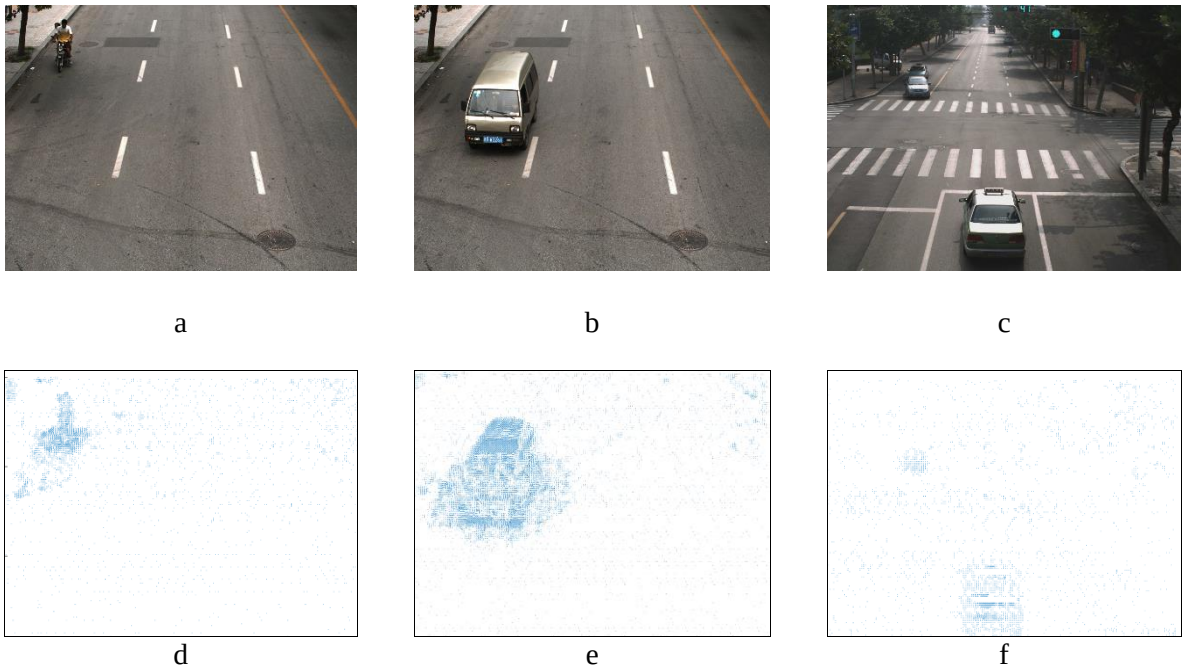


Figure 5. Example of surveillance video frames and their motion vector fields.

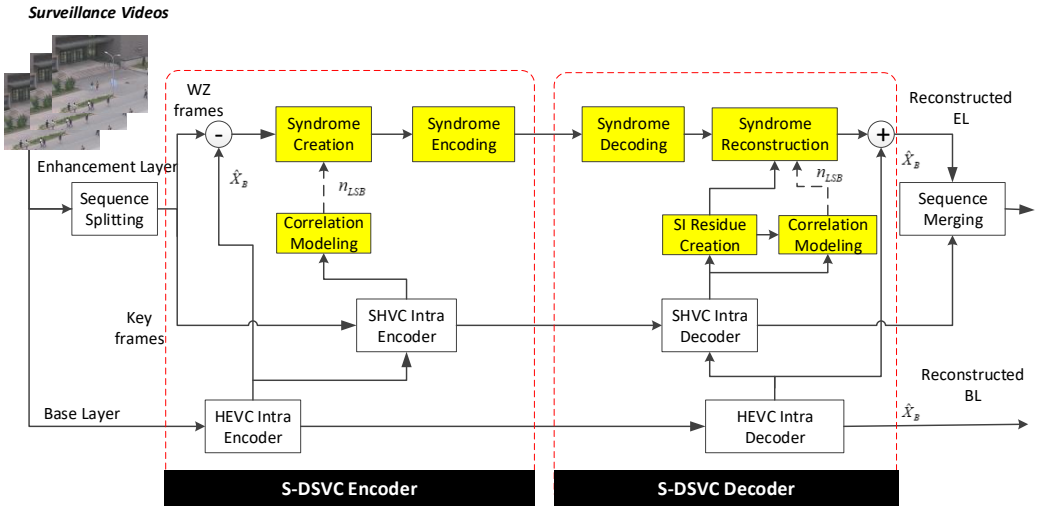


Figure 6. Proposed Surveillance - Distributed Scalable Video Coding architecture.

In summary, the sequence of EL encoding steps can be summarized as:

E1. Sequence splitting: First, the EL frames are split into the key and WZ frames. The number of WZ frames between two consecutive key frames is defined by the GOP size. Naturally, the GOP size of 2 is commonly used due to its balance between the compression efficiency and the decoding delay requirement of video.

E2. "Syndrome creation: For the WZ frames, the EL residue is created by subtracting

the BL decoded frame from the original frame. This residue is then transformed with the integer discrete cosine transform (DCT) and scalar quantized with an EL quantization step size to create the EL quantized residue. In the proposed S-DSVC solution, only a part of EL quantized residue, called syndrome, is coded and sent to the decoder. The syndrome size is mainly characterized by the correlation between the original residue and the side information residue created at the decoder.

E3. Correlation modeling (CM): In order to efficiently compress the EL residue, the correlation between the original EL residue and the decoder side information residue is estimated at this step. Here, the correlation degree is determined through a number of least significant bits, n_LSB, which needs to be transmitted to the receiver. In this paper, n_LSB can be computed as similar to our previous work [21].

E4. Syndrome encoding: The syndrome created from the previous step is finally compressed using a common context adaptive binary arithmetic coding (CABAC) solution as common in predictive video coding standards such as H.264/AVC and HEVC.

At the receiver, the sequence of EL decoding steps includes:

D1. Syndrome decoding: Firstly, the EL received syndrome is decoded using the context adaptive binary arithmetic decoding (CABAD) solution. The syndrome is important part of the original information which cannot be estimated at the decoder using the side information (SI) creation solution presented in the next step.

D2. SI residue creation: Side information is a noisy version of the original information which can be created at the decoder side. Naturally, the higher quality of SI, the lower bitrates needed to send to the decoder. Therefore, the quality of SI plays an utmost important role in the proposed S-DSVC solution. Considering the high temporal correlation between consecutive frames in a surveillance video sequence, it is proposed in this paper an efficient SI creation solution as described in the next sub-section.

D3. Correlation modeling: Similar to the encoder, the correlation modeling proceeded in the decoder also aims to estimate the correlation between the encoder original and the decoder SI residues. This correlation is also represented through a number of significant bitplanes and computed as in the encoder side.

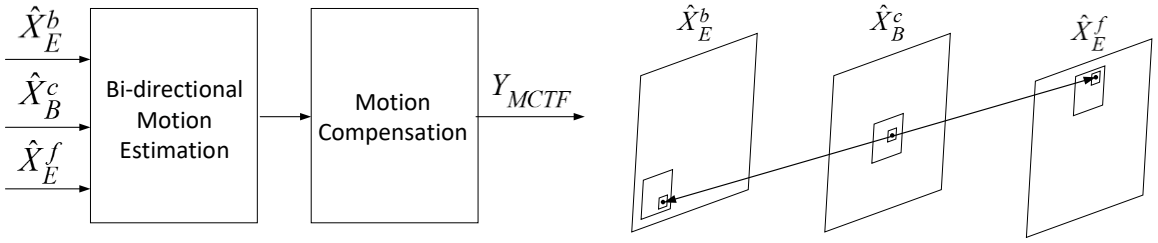
D4. Syndrome reconstruction: Finally, the EL information is reconstructed using the syndrome sent from the receiver and the SI residue computed at the decoder. To achieve the highest EL frame quality, a statistical reconstruction solution as presented in [22] is adopted.

3.3. Proposed SI frame creation

In order to create the SI frame, we propose a novel scheme, namely, Motion compensated temporal filtering (MCTF), which can effectively exploit the high temporal correlation features (between two consecutive EL key frames characterized for the surveillance video. Figure 7 shows the proposed MCTF scheme where the input frames include the BL current, the EL forward and backward decoded frame, $\hat{X}_B^c$, $\hat{X}_E^f$, $\hat{X}_E^b$, respectively.

As presented in Figure 7, the temporal correlation is exploited to improve the BL frame quality by finding the displacement of each lower quality BL block in the two (higher quality) EL frames, and then averaging the EL displaced and BL blocks to obtain the final SI frame. Therefore, the MCTF can be performed as follows:

Bi-directional motion estimation (BiME): This step aims to find a set of MVs representing well the motion of each decoded BL frame $\hat{X}_B^c$ block with respect to the EL decoded backward frame, $\hat{X}_E^b$, and EL decoded forward frame, $\hat{X}_E^f$. The BiME will result in a pair of symmetric MVs, one pointing to $\hat{X}_E^b$, and another pointing to $\hat{X}_E^f$.



a) MCTF scheme b) Bi-directional motion estimation

Figure 7. Proposed MCTF scheme.

Motion compensation (MC): Using the MVs obtained from the previous step, two SI candidate frame estimations ($\hat{X}_E^f(i+\overrightarrow{mv})$, $\hat{X}_E^b(i-\overrightarrow{mv})$) are obtained by performing motion compensation based on the two EL reference backward and forward frames. Next, these motion compensated estimations and the decoded BL frame are averaged to obtain the MCTF SI frame, Y_{MCTF}, as follows:

$$Y_{MCTF}(i) = \frac{1}{3}[\hat{X}_E^b(i-\overrightarrow{mv}) + \hat{X}_B^c(i) + \hat{X}_E^f(i+\overrightarrow{mv})] \quad (4)$$

where i is the i^{th} block in frame.

As determined in (4), the MCTF SI frame, Y_{MCTF}, is created not only from the decoded BL but also from two motion compensated frames derived from the previous and next EL decoded frames to consider both the spatial and temporal correlations. This can guarantee a good SI quality even when the BL decoded frame has lower quality.

4. Performance evaluation

Generally, the compression efficiency of a video coding solution is assessed through the rate-distortion (RD) performance. This Section starts by describing the test conditions. Afterwards, the RD performance comparison between the proposed S-DSVC solution and relevant surveillance scalable coding benchmarks are presented.

4.1. Test conditions

The performance evaluation is carried out for six surveillance videos obtained from the PKU-SVD-A dataset [18, 19]. Figure 8 shows the first frames of the tested surveillance videos while TABLE I summarizes some of their main characteristics and the quantization parameters used for BL and EL compression. As usual, results are presented for the luminance component and the rate includes all frames (the BL frames, the EL key frames and EL WZ frames).

Bank Campus Classover

Crossroad Office Overbridge

Figure 8. Illustration of the first frame for the tested surveillance videos.

TABLE I. Summary of test conditions

Spatial resolution, temporal resolution, number of frames	720×576, @30Hz, 201 frames
GOP size	2 (Key-WZ-Key-…)
Quantization Parameters	QP_B = {38;34;30;26} $QP_E = QP_B - 4$

4.2. Overall rate distortion performance assessment

As mentioned above, in video coding research, the rate - distortion performance is usually used to assess a newly video coding solution. In this context, the two most relevant surveillance video coding benchmarks are compared with the proposed S-DSVC solution, notably the SHVC-intra [4] and the HEVC-simulcasting solution. It should be noted that, the SHVC-intra benchmark is carried out by compressing the surveillance video data with the SHVC reference software [23] and the Intra coding configuration while the HEVC-simulcasting is performed by compressing the surveillance video data with the HEVC reference software [24] and with two independent layers. The RD performance comparison is shown in Figure 9 while Table II presents the BD-Rate [25] saving when comparing the proposed S-DSVC with the relevant benchmarks.

Table II. BD-Rate saving

Sequences	SHVC-intra vs. HEVC-simulcasting	Proposed S-DSVC vs. HEVC-simulcasting	Proposed S-DSVC vs. SHVC-intra
Bank	-32.85	-39.19	-9.04
Campus	-28.93	-38.54	-9.41
Classover	-28.93	-36.83	-10.58
Crossroad	-34.58	-38.62	-5.93
Office	-32.41	-37.08	-6.55
Overbridge	-34.14	-40.56	-9.46
Averages	*-31.97*	*-38.47*	*-8.49*

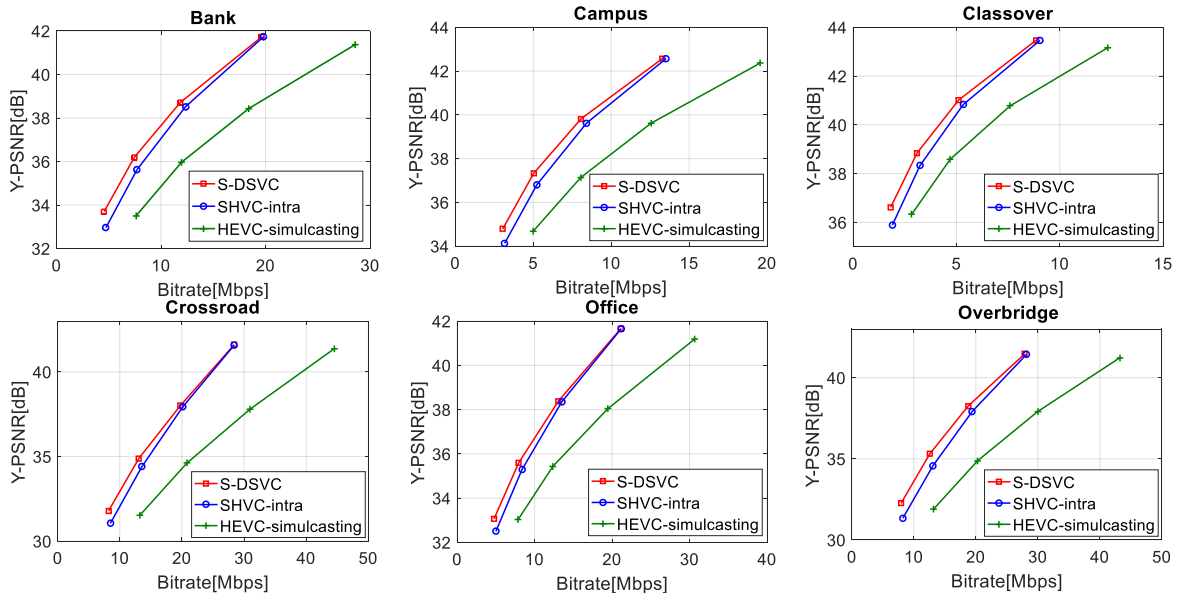


Figure 9. RD performance comparison for the test surveillance videos.

From the obtained results, it is able to derive some conclusions:

• As shown in Figure 9 and Table II, the proposed S-DSVC solution significantly outperforms the HEVC-simulcasting benchmark with around 38.5% bitrate saving while maintaining the similar perceptual quality.

• The higher compression gains are achieved for sequences containing fewer motion activities and objectives, e.g., *Overbridge*, and *Classover*. This because the side information creation in the proposed coding structure usually performs well for such low motion video content and thus, resulting a high compression efficiency for the proposed S-DSVC solution.

• The proposed S-DSVC also achieves a better compression gain when compared to the conventional SHVC standard, notably with around 8.5% bitrate saving.

4.3. S-DSVC complexity assessment

In this section, we assess the complexity associated with the proposed S-DSVC architecture as well as comparing with the well-known SHVC standard. To achieve this object, the processing time [second] is employed as a representative of the computational complexity of each coding solution. The configuration of the computer used for testing is specified in Table III.

Table III. Specification of the tested computer

Hardware configuration	Processor: Intel® Core™ i7-4800MQ @2.7 GHz RAM: 8.00 GB System: Win 10, 64-bit Environment: Microsoft Visual Studio 2017 Community

4.3.1. S-DSVC component analysis

As discussed, in contrast to the conventional predictive video coding standards [3, 4], the proposed S-DSVC shifts one of the most computational complexity parts, the motion estimation, to the decoder side. This results in a low encoding complexity video coding solution. To understand this effect, we measure and compare the complexity associated to each coding sides (encoder and decoder) for the proposed S-DSVC codec. Figure 10 shows the comparison between the encoding and decoding processes for six tested surveillance videos.

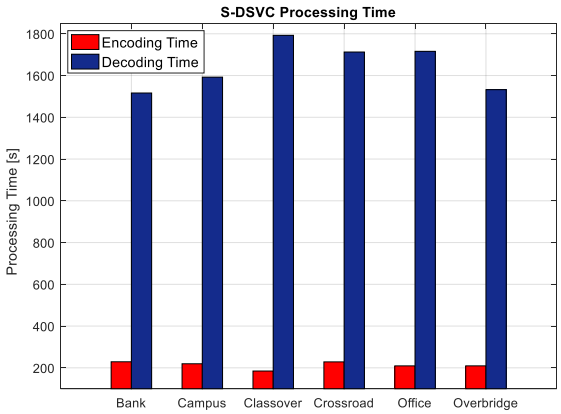


Figure 10. Encoding time vs. decoding time.

As shown, the computational complexity associated to the encoder side is much lower than that of the decoder side. To further understand the computational complexity associated each coding tools of the proposed S-DSVC solution, a complexity - component analysis is conducted for both the encoder and decoder and shown in Figure 11, and Figure 12, respectively.

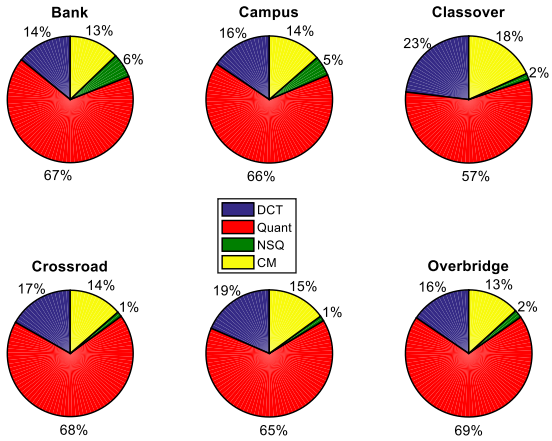


Figure 11. Encoding time – component.

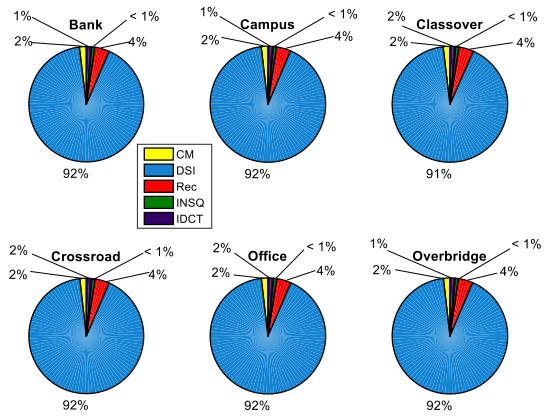


Figure 12. Decoding time - component.

It can be seen that during the encoding, the Quantization (Quant) consumes the highest percentage of processing time, about 60-70% and followed by the Discrete Cosine Transform (DCT), the Correlation Modeling and the Nested Scalar Quantization (NSQ).

At the decoder side, the Decoder Side Information (DSI) consumes the highest percentage of processing, with around 90%. This mainly comes from the high complexity - motion estimation process of DSI. Other components, correlation modeling, Inverse DCT (IDCT), Reconstruction (Rec), and Inverse Nested Scalar Quantization (INSQ) consumes less than 10% of overall decoding time.

4.3.2. S-DSVC versus SHVC

One of the main benefits with the proposed S-DSVC solution is the computational complexity associated to the encoder part. To demonstrate this advancement, we compare the encoding time [second] of the proposed S-DSVC with the SHVC benchmark. This experiment is conducted for six tested sequences and shown in Figure 13.

As it can be seen from Figure 13, the complexity associated to the proposed S-DSVC solution is much lower than that of the SHVC standard, notably about 60% encoding time reduction. This important feature makes the proposed S-DSVC solution suitable to a large number of video surveillance applications, which are usually constrained by the power and energy.

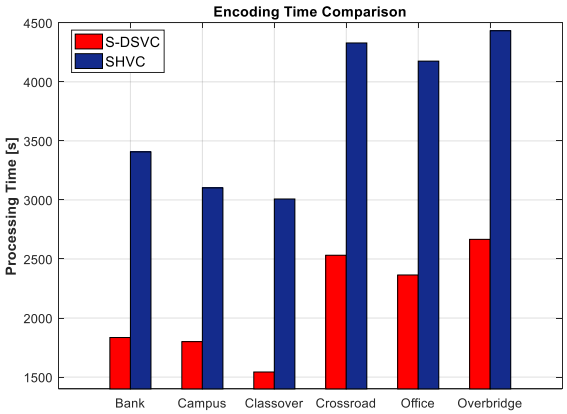


Figure 13. Encoding time comparison.

5. Conclusions

In this paper, we presented a novel scalable video coding solution for compressing surveillance visual content. The proposed video coding solution follows a layered coding approach while efficiently combining the predictive and distributed video coding. The distributed coding approach is employed to compress the enhancement layer data while the HEVC standard is used to compress the base layer data. This selected solution is able to exploit the temporal correlation between surveillance video frames at the decoder while guaranteeing a backward compatibility with the well-known HEVC at the base layer. As assessed, the proposed scalable video coding solution significantly outperforms the relevant benchmarks. Moreover, with the adopted coding solution, the encoding complexity associated to the proposed S-DSVC is expected to be less than the traditional SHVC standard and the error robustness is improved. These issues can be addressed in the future researches.

Acknowledgements

This research is funded by Vietnam National Foundation for Science and

Technology Development (NAFOSTED) under grant number 102.01- 2016.15.

References

[1] M. Valera and S. Velastin, "Intelligent distributed surveillance systems: A review," IEE Proceedings - Vision, Image and Signal Processing, vol. 152, no. 2, pp. 192–204, Apr. 2005.

[2] H. Schwarz, D. Marpe, and T. Wiegand "Overview of the scalable video coding extension of the H.264/AVC standard," IEEE Circuits and Systems for Video Technology, vol. 17, no. 9, pp. 1103-1120, Sept. 2007.

[3] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1649-1668, Dec. 2012.

[4] J. M. Boyce, Y. Ye, J. Chen, and A. K. Ramasubramonian, "Overview of SHVC: Scalable Extensions of the High Efficiency Video Coding Standard", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 26, Issue 1, pp. 20-34, Sept. 2015.

[5] X. Zhang, L. Liang, Q. Huang, T. Huang, W. Gao, "A background model based method for transcoding surveillance videos captured by stationary camera," IEEE Picture Coding Symposium (PCS), Nagoya, Japan, pp. 78-81, 2010.

[6] X. Zhang, T. Huang, Y. Tian, and W. Gao, "Background-modeling-based adaptive prediction for surveillance video coding," IEEE Transactions on Image Processing, vol. 23, no. 2, pp. 769-784, 2014.

[7] X.G. Zhang, L.H. Liang, Q. Huang, Y.Z. Liu, T.J. Huang, and W. Gao, "An efficient coding scheme for surveillance videos captured by stationary cameras," IEEE International Conference on Visual Communication and Image Processing (VCIP), pp. 77442A1-10, 2010.

[8] S. Zhang, K. Wei, H. Jia, X. Xie, W. Gao, "An efficient foreground-based surveillance video coding scheme in low bit-rate compression," IEEE International Conference on Visual Communication and Image Processing (VCIP), San Jose, USA, Nov. 2012.

[9] X. Zhang, Y. Tian, T. Huang, S. Dong, W. Gao, "Optimizing the Hierarchical Prediction and Coding in HEVC for Surveillance and Conference Videos with Background Modeling," IEEE Transactions on Image Processing, vol. 23, no. 10, pp. 4511-4526, Oct. 2014.

[10] F. Pereira, L. Torres, C. Guillemot, T. Ebrahimi, R. Leonardi, and S. Klomp, "Distributed video coding: selecting the most promising application scenarios," Signal Processing: Image Communication, vol. 23, no. 5, pp. 339-352, June 2008.

[11] D. Slepian, J.K. Wolf, "Noiseless coding of correlated information sources," IEEE Information Theory Society, vol. 19, pp. 471-480, 1973.

[12] A.D. Wyner, J. Ziv, "The rate-distortion function for source coding with side information at the decoder," IEEE Information Theory Society, vol. 22, no. 1, pp. 1–10, 1976.

[13] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, M. Ouaret, "The DISCOVER Codec: Architecture, Techniques and Evaluation," IEEE Picture Coding Symposium (PCS), Lisboa, Portugal, Nov. 2007.

[14] R. Puri, A. Majumdar, and K. Ramchandran, "PRISM: A video coding paradigm with motion estimation at the decoder," IEEE Transactions on Image Processing, vol. 16, no. 10, pp. 2436-2448, Oct. 2007.

[15] F. Pereira, C. Brites, J. Ascenso, M. Tagliasacchi "Wyner–Ziv video coding: a review of the early architectures and further developments," IEEE International Conference on Multimedia and Expo (ICME), Hannover, Germany, June 2008.

[16] L. Liu, Z. Li, and E. J. Delp, "Efficient and Low-Complexity Surveillance Video Compression Using BackwardChannel Aware Compression," IEEE Circuits and Systems for Video Technology, vol. 19, no. 4, Apr. 2009.

[17] V. K. Kolavella and P. G. Krishna Mohan "Distributed video coding: codec architecture and implementation," Signal and Image Processing: An International Journal (SIPIJ), vol. 2, no. 1, pp. 151-163, Mar. 2011.

[18] W. Gao, Y. Tian, T. Huang, S. Ma, and X. Zhang, "IEEE 1857 standard empowering smart video surveillance systems," IEEE Intelligent Systems, 2013.

[19] PKU-SVD-A. [Online]. Available: http://mlg.idm.pku.edu.cn/-resources/pku-svd-a.html

[20] "Video test sequences," [Online]. Available: ftp://hevc@ftp.tnt.uni-hannover.de/testsequences/

[21] X. Hoang Van, J. Ascenso, F. Pereira, "HEVC backward compatible scalability: A low encoding complexity distributed video coding based approach," Signal Processing: Image Communication, vol. 33 pp. 51-70, Apr. 2015.

[22] X. HoangVan, J. Ascenso, and F. Pereira, "Optimal Reconstruction for a HEVC Backward Compatible Distributed Scalable Video Codec," IEEE Visual Communication and Image Processing (VCIP), Valletta, Malta, Dec. 2014.

[23] SHVC reference software, [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svnSHVCSoftware/.

[24] HEVC reference software, [Online]. Available:https://hevc.hhi.fraun hofer.de/svn/svn HEVCSoftware/.

[25] G. Bjontegaard, "Improvements of the BD-PSNR model," document ITU-T SC16/Q6, Doc. VCEG-AI11, 2008.