Adaptive Scalable Video Coding: An HEVC-Based Framework Combining the Predictive and Distributed Paradigms

Xiem HoangVan, João Ascenso, and Fernando Pereira, *Fellow, IEEE*

***Abstract*— The emerging scalable High Efficiency Video Coding (SHVC) video coding standard provides an efficient solution for transmission of video over heterogeneous and time dynamic networks, terminals, and usage environments. The encoding complexity and the error sensitivity associated with the efficient HEVC coding tools adopted in SHVC make this scalable codec less attractive to some emerging applications such as video surveillance, visual sensor network, and remote space transmission where these requirements are critical. To address the requirements of these application scenarios including scalability, this paper proposes a novel HEVC-based framework offering quality scalability on top of an HEVC compliant base layer while appropriately combining the predictive and distributed coding paradigms. To achieve the best enhancement layer compression efficiency, two novel coding tools are proposed, notably a machine learning-based side information creation mechanism and an adaptive correlation modeling process. The experimental results reveal that the rate-distortion performance of the proposed distributed scalable video coding-HEVC solution outperforms the relevant alternative coding solutions, notably by up to 52.9% and 23.7% BD-rate gains regarding the HEVC-Simulcast and SHVC standard solutions, respectively, for an equivalent prediction configuration, while achieving a lower encoding complexity.**

***Index Terms*— Correlation modeling (CM), distributed video coding (DVC), High Efficiency Video Coding (HEVC) standard, predictive video coding, quality scalability, scalable HEVC (SHVC) standard, scalable video coding (SVC), side information (SI) creation.**

I. Introduction

THE compression efficiency benefits associated with the recent High Efficiency Video Coding (HEVC) [1] standard, notably 50% compression gains for the same perceptual quality when compared with the previous H.264/Advanced Video Coding (AVC) standard [2], and the increasing market relevance of heterogeneous and dynamic transmission environments, have naturally stimulated the development of an efficient HEVC scalable extension with significant compression performance regarding the already available scalable video coding (SVC) standard [3] which previously extended the H.264/AVC standard. Toward this target, the International Organization for Standardization/International Electrotechnical Commission Moving Picture Experts Group and International Telegraph Union – Telecommunication Standardization Sector Video Coding Experts Group groups have launched in 2012 a joint call for proposals targeting an efficient scalable extension of the HEVC standard [4], well known as scalable HEVC (SHVC), which should also provide base layer (BL) HEVC backward compatibility.

The novel SHVC standard [5] does not use the same conceptual approach adopted in SVC [3] where new macroblock-level signaling capabilities were defined to indicate whether the enhancement layer (EL) macroblock is predicted from the BL or from the current EL. In SHVC, the BL-reconstructed picture is taken as an inter-layer (IL) reference picture to be included in the EL prediction buffer, eventually after some IL processing. This scalable coding approach requires changes only at the HEVC high-level syntax, thus significantly increasing the HEVC and SHVC compatibility and easing its implementation and deployment.

The SHVC standard is able to provide high compression efficiency, notably regarding the prior SVC standard [5]. However, to achieve this higher compression efficiency, additional computational complexity had to be invested, notably at the encoder, as it happens for HEVC regarding H.264/AVC [6]. Moreover, its high dependency on the quality of the EL predicted pictures makes this solution even less robust to channel errors and drift, which may be a problem for some application scenarios, especially when wireless networks are involved. These drawbacks make the SHVC standard less appropriate for many emerging applications such as video surveillance, visual sensor networks, and remote space transmission, which require scalable coding solutions having not only acceptable compression performance but also low encoding complexity and high error resilience. However, even for these applications, HEVC backward compatibility in the BL is still a critical requirement considering its expected widespread deployment. Considering the need for an efficient low encoding complexity error-resilient and yet HEVC-backward-compatible SVC solution, this paper proposes a powerful SHVC-based framework combining the strengths of the predictive and distributed video coding (DVC) paradigms, labeled DSVC-HEVC from distributed SVC based on HEVC. In the

Manuscript received July 22, 2015; revised October 28, 2015; accepted December 21, 2015. Date of publication March 16, 2016; date of current version August 2, 2017. This paper was recommended by Associate Editor F. Wu.

X. HoangVan is with the Faculty of Electronics and Telecommunication, Vietnam National University–University of Engineering and Technology, Hanoi 10000, Vietnam (e-mail: hoang.xiem@lx.it.pt).

J. Ascenso is with the Multimedia Signal Processing Group, Instituto Superior Técnico and the Instituto de Telecomunicações, Lisbon 1049-001, Portugal.

F. Pereira is with the Electrical and Computer Engineering Department, Instituto Superior Técnico and the Instituto de Telecomunicações, Lisbon 1049-001, Portugal.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2016.2543120

proposed DSVC-HEVC solution, the BL frames are compliantly coded with HEVC Inter while the EL frames are coded with a novel DVC-based solution [7], [8]. As the EL is Intra coded following DVC principles, the EL residue in the proposed DSVC-HEVC is just the difference between the original and the BL-decoded frame, which means that no drift at the EL occurs when there are errors or packet losses.

To efficiently exploit the BL motion information available at both the encoder and decoder, a novel DVC side information (SI) creation solution is proposed in this paper, which has a direct impact on the EL compression efficiency. The final SI is created by adaptively selecting one out of several SI candidates using a machine learning approach based on a multiclass support vector machine (SVM) [9], [10]. Moreover, an adaptive correlation modeling (ACM) solution is proposed with improved accuracy based on several correlation compensation modes. These correlation compensation modes are dynamically selected using a rate-distortion optimization (RDO) approach based on a low complexity (LC) SI creation process. The proposed scalable DSVC-HEVC framework outperforms the most relevant alternative simulcast and scalable coding solutions, notably by up to 52.9% and 23.7% BD-rate gains with respect to the HEVC-Simulcast and SHVC-InterBL_IntraEL (to be defined later) benchmarks, while offering a lower encoding complexity.

To achieve its objectives, this paper has been organized as follows. Section II reviews the relevant background work, while Section III presents the proposed DSVC-HEVC framework, notably the encoder and decoder architectures. After, Section IV presents the proposed SI creation solution, while Section V presents the ACM. Section VI analyzes the performance of each novel coding tool as well as the overall DSVC-HEVC framework in terms of Rate Distortion (RD) performance and encoding complexity. Finally, Section VII presents the main conclusions and ideas for future work.

II. Related Work

This section includes three parts: review of the related work on DSVC; discussion of the relevant work on SI creation; and related work on CM.

A. Distributed Scalable Video Coding

DSVC regards all coding solutions where DVC principles [11], [12] are used to code the video data while providing different types of scalability. Although it is possible to code both the BL and ELs with a DVC approach, a backward compatibility requirement with some other coding solution motivates the adoption of a coding architecture where the BL is coded with a predictive video coding standard, while the ELs are coded with a DVC-based codec. In [13], a scalable version of the Power-efficient, Robust, hIgh-compression, Syndrome-based Multimedia coding (PRISM) codec [8] is proposed to address both spatial and temporal scalabilities where the BL is backward compatible with H.264/AVC and the EL is coded with the PRISM tools. In [14], a scalable video codec is proposed where a prediction is created using H.264/AVC motion compensation (MC) techniques; also, a bank of Low Density Parity Check (LDPC) codes is applied to the quantized discrete cosine transform (DCT) coefficients to generate parity information. At the decoder, the same prediction is created as SI. Later, Wang *et al.* [15] proposed to improve the MPEG-4 fine grain scalability (FGS) [16] RD performance by exploiting the EL temporal redundancy at the decoder. Refinement bitplanes are encoded with a hybrid approach using either LPDC codes and IL prediction or conventional FGS variable-length coding tools. To achieve both scalability and error resilience, a layered DVC codec using H.264/AVC in the BL and DCT in the EL followed by nested scalar quantization (NSQ) and irregular LDPC codes has been proposed in [17]. In [18], a coding solution where two SI creation methods are adopted is proposed: 1) a simple SI creation method at the encoder to estimate the coding rate and 2) a two-stage SI creation method at the decoder to reconstruct the EL information. In [19], an RD performance analysis of temporal scalable DVC regarding the classical (predictive) SVC approach is proposed. As most of the DSVC solutions [13]–[19] available in the literature are still H.264/AVC backward compatible in the BL and the EL is not efficient enough, their compression efficiency is limited, notably when compared with the recent SHVC standard.

Recently, HoangVan *et al.* [20] proposed a DSVC solution with BL HEVC backward compatibility where the BL frames are purely Intra coded while the EL frames are Intra encoded but Inter decoded. As reported, this DSVC solution is able to achieve better compression efficiency with lower encoding complexity when compared with the SHVC Intra standard. However, the performance of this DSVC solution [20] is limited since a BL Intra-only coding configuration is employed which significantly penalizes the overall RD performance. On the contrary, the solution proposed in this paper generalizes and significantly advances the previous DSVC solution in technology and performance by making the BL backward compatible with an HEVC Inter coding solution. Moreover, as BL motion information becomes available, the EL DVC tools may take benefit of motion information without additional complexity, thus opening the doors to a novel scalable coding architecture and associated tools, targeting a harmonious integration of the BL and EL in terms of motion information.

B. Side Information Creation

SI plays a key role in a DVC framework as it has a direct impact on the quality of the final reconstruction information [7]. Similarly, SI also plays a critical role on the final DSVC RD performance. In DSVC, SI creation is usually performed at the decoder to reconstruct the EL frames after knowing the correlation between the encoder EL and the decoder SI [20]. Many SI creation solutions have been proposed in the literature so far, notably based on frame interpolation [21] and extrapolation [22]. Other solutions create the SI at the decoder in a tentative way [23] or rely on auxiliary data received from the encoder for the parts of the image that are more difficult to estimate [24]. Moreover, the initial SI estimation can be refined based on the data already decoded for each frame, e.g., the successive DCT coefficients [25]. Moreover, hybrid approaches combining these elementary approaches can also be used. In [26], two SI candidates are refined and an

SVM solution is adopted to combine them. A detailed review of DVC SI creation solutions is available in [27]. While the above SI creation solutions can be directly employed in a DSVC framework, this is not an efficient solution. In fact, in SVC, not only the EL-decoded forward and backward frames but also the decoded BL frame and its motion information can be exploited.

C. DSVC Correlation Modeling

Similar to the SI creation, CM also plays a critical role in the final DSVC RD performance as it has a direct impact on both the bitrate and the EL-reconstructed quality. CM aims to define the correlation between the EL residue and SI residue (SIR) expressed through the number of least significant bitplanes n_{LSB} that need to be coded to fully recover the EL residue at the decoder. In [28] and [29], an encoder-based CM (ECM) solution is proposed which only computes the correlation at the encoder by exploiting the original information; then, the computed n_{LSB} are sent to the decoder, notably at block level. This solution has two drawbacks.

1) The correlation information expressed at block level may not be appropriate for every coefficient in a block.
2) Extra bitrate is necessary to code the estimated correlation information to the decoder.

In [30], an asymmetric CM solution has been proposed where n_{LSB} is independently determined at the encoder and the decoder using a LC-SI creation solution at the encoder and a higher complexity SI creation solution at the decoder. However, this approach may lead to n_{LSB} mismatches between the encoder and the decoder due to the use of different SI creation solutions, which reduces the DSVC RD performance. Considering these weaknesses, an encoder–decoder-based CM (EDCM) solution is proposed in [20] where the correlation information is computed at both the encoder and decoder in the same way. This CM solution is able to avoid the correlation information mismatch while working at a very fine granularity level, notably the coefficient level, and without extra bitrate. However, this CM solution requires motion estimation (ME) and compensation techniques to create the SI at the encoder, which should be avoided when low encoding complexity is a critical requirement. Moreover, as only decoded information is employed, inaccurate correlation information estimation may occur, thus reducing the final DSVC RD performance. To overcome these problems, this paper proposes some significant steps forward by adopting a novel ACM solution.

III. Distributed Scalable Video Coding Solution

This section presents the overall architecture of the proposed DSVC-HEVC solution, combining the predictive and distributed coding paradigms.

A. Prediction Structure

Considering its impact, Fig. 1 illustrates the prediction structure adopted for the proposed DSVC-HEVC solution [Fig. 1(a)] as well as for the DSVC (Intra BL) solution in [20] [Fig. 1(b)] and the two most relevant benchmarks, SHVC-InterBL_IntraEL [Fig. 1(c)] where the BL frames are Inter coded and the EL frames Intra coded (to limit the encoding complexity, error sensitivity, and drift) and SHVC-InterBL_InterEL [Fig. 1(d)] where both BL and EL frames are Inter coded. In Fig. 1, WZ refers to Wyner-Zv frames which are DVC coded and K to key frames which are conventionally coded.

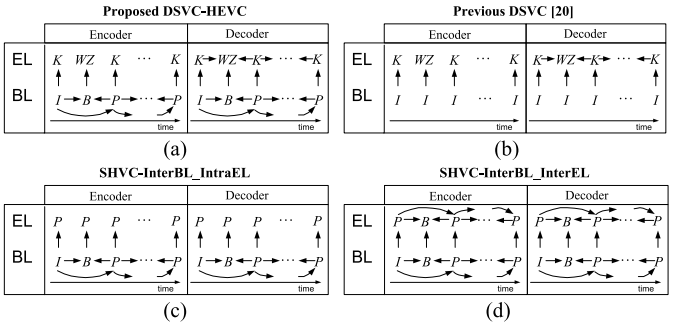

Fig. 1. Prediction structures. (a) Proposed DSVC-HEVC. (b) DSVC (Intra BL) from [20]. (c) SHVC-InterBL_IntraEL. (d) SHVC-InterBL_InterEL (for simplicity, only two scalable layers and GOP size of 2 are shown).

As illustrated in Fig. 1, the proposed DSVC-HEVC solution codes the BL frames with an HEVC Inter-coding solution instead of the HEVC Intra-coding solution adopted in the DSVC codec [20]. In this way, the proposed DSVC-HEVC solution can exploit at the ELs the temporal motion information created at the BL without any additional complexity cost. In DSVC-HEVC, the key frames are IL predictively coded from the corresponding BL frame, this means without exploiting any temporal redundancy at the EL encoder to avoid drift. With this type of prediction structure, several benefits can be achieved.

1) Error robustness is improved as temporal error propagation is avoided at the ELs since the key frames and WZ frames are independently coded.
2) As BL Inter coding is more general than BL Intra coding, it is always possible to constrain Inter to become Intra while the opposite is not true.
3) Compression efficiency is improved as better EL SI may be created by exploiting the available BL motion information.

B. Encoder Architecture and Walkthrough

Although the proposed DSVC-HEVC architecture only considers currently temporal and quality scalabilities, it can be extended in future work to also support spatial scalability by integrating appropriate down and up-sampling filters as done in the SHVC standard [5]. In the proposed DSVC-HEVC solution, temporal scalability directly results from the adopted prediction structure while quality scalability is achieved with the layered coding approach. Fig. 2 illustrates the DSVC-HEVC encoder architecture where the shaded blocks correspond to the novel coding tools proposed in this paper. Although only two quality layers are shown, the proposed SVC framework can be extended for any number of scalable layers. In such cases, each EL should exploit the decoded information from both the previous EL and the BL, notably the previous EL-decoded frame and the BL motion vector field (MVF).

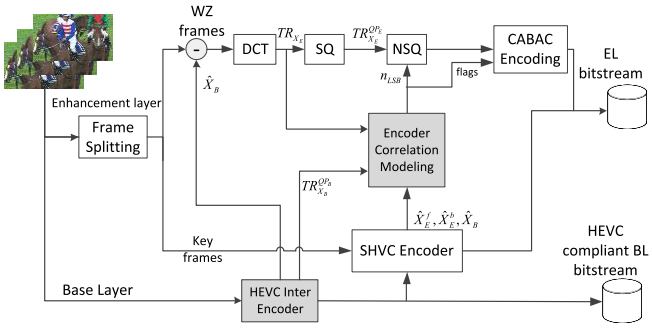

Fig. 2. DSVC-HEVC encoder architecture (highlighting the novel tools).

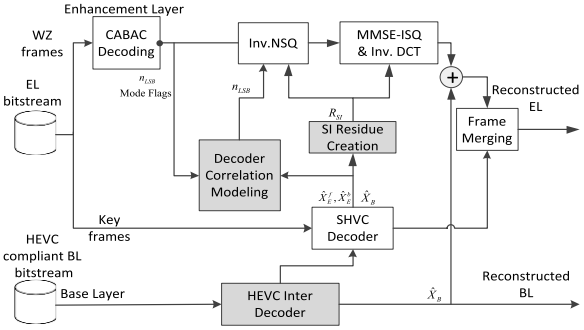

Fig. 3. DSVC-HEVC decoder architecture (highlighting the novel tools).

In the DSVC-HEVC solution, the BL frames are coded with a compliant HEVC Inter encoder. Therefore, together with the BL-decoded frame ($\hat{X}_B$) and the BL quantized transformed residue ($\mathrm{TR}_{X_B}^{\mathrm{QP}_B}$), the BL motion information (mv_B) can be exploited to better code the EL frames. In DSVC-HEVC, the EL residue (R_{X_E}) corresponds to the difference between the original (X) and the BL-decoded frames. However, only the part of the EL residue that cannot be estimated at the decoder is coded. The amount of coded EL residue is indicated by the number of least significant bitplanes to code, n_{LSB}, which cannot be perfectly inferred from the decoder SI, R_{Y_E}, and associated correlation model. The sequence of EL coding steps is as follows.

1) *Frame 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 group of picture (GOP) size.
2) *DCT and SQ:* For the WZ frames, the EL residue is created by subtracting the BL-decoded frame from the original frame. This residue is transformed with the integer DCT and scalar quantized with an EL quantization step size to create the EL quantized residue ($\mathrm{TR}_{X_E}^{\mathrm{QP}_E}$). In this case, the DCT sizes are from 4×4 to 32×32.
3) *ECM:* In this step, the correlation information between the encoder EL and decoder SI quantized residues is estimated for each coefficient. This paper proposes an ACM solution with two key novelties.
   a) Instead of using a high complexity SI creation for estimating n_{LSB} as in [20], the proposed correlation model uses an LC-SI creation solution, notably without performing any EL ME.
   b) A set of novel n_{LSB} compensation modes are used to refine an initial n_{LSB} estimation using an RDO approach at block level.

   Since original information is used to select one out of four n_{LSB} compensation modes, one 2-b flag is needed to signal the selected n_{LSB} compensation mode. The novel encoder ACM is described in Section V-A.
4) *NSQ:* Based on the computed n_{LSB} from the previous step and the quantized EL residue (with EL quantization step) DCT coefficients $\mathrm{TR}_{X_E}^{\mathrm{QP}_E}$, an NSQ technique [28] is used to create the so-called *syndrome* (S), corresponding to the n_{LSB} least significant bitplanes of $\mathrm{TR}_{X_E}^{\mathrm{QP}_E}$.
5) *CABAC Encoding:* The syndrome generated from the previous step and the compensation mode flag obtained in the encoder correlation modeling are coded with HEVC Context Adaptive Binary Arithmetic Coding (CABAC) entropy coding.

Finally, the two bitstreams associated with the key and WZ frames are independently and appropriately packetized to offer temporal and quality scalability.

C. Decoder Architecture and Walkthrough

Fig. 3 illustrates the proposed DSVC-HEVC decoder architecture. First, the BL frames are decoded with a certain quality level (depending on the BL quantization step) using an HEVC Inter decoder. Then, the EL key frames are decoded with a (normative) SHVC decoder.

The EL WZ frames are decoded with the novel DSVC-HEVC decoder and the novel tools proposed as follows.

1) *CABAC Decoding:* First, the received EL WZ bitstream is entropy decoded with an HEVC CABAC decoder. While the main information is the decoded syndrome, some auxiliary information is also decoded, notably the n_{LSB} compensation mode flag.
2) *SIR Creation:* The SIR creation module plays a key role in the DSVC-HEVC EL codec as it is used to conditionally decode the source, notably to reconstruct the EL residue that is added to the decoded BL frame to obtain the reconstructed EL frame. When the SI quality is high, a smaller n_{LSB} needs to be coded and sent to the decoder to reconstruct the EL coded residue, thus increasing the DSVC-HEVC RD performance. In this paper, a novel SI creation solution (presented in Section IV) is proposed considering the availability of BL motion information.
3) *Decoder CM:* The decoder CM also plays a critical role in the DSVC-HEVC EL codec as the number of Least Significant Bitplanes (LSBs) to be coded is calculated in this step and ideally there should be no mismatches between the encoder and the decoder; this means that the decoder estimated n_{LSB} should be always similar to the encoder estimated n_{LSB}. Since the mismatch-free CM associated with the initial n_{LSB} estimation may not be accurate enough, the final n_{LSB} accuracy is improved by using a n_{LSB} compensation mode flag; this flag is used by the decoder CM (presented in Section V-B) to decide how the initial n_{LSB} should be compensated to improve its accuracy for each quantized coefficient.

4) *Inverse NSQ:* The inverse NSQ is used to reconstruct the scalar quantized coefficients from which the most significant bitplanes were removed by the NSQ at the EL encoder. The proposed DSVC-HEVC solution employs the same inverse NSQ module as in [20] and [28].
5) *Minimum Mean Squared Error (MMSE)-Based Inverse Quantization and Inverse Discrete Cosine Transform:* In this step, the DCT coefficients are statistically reconstructed using a well-known MMSE-based reconstruction solution [31]. After, the DCT coefficients are inversely transformed to create the decoded EL residue.

Finally, the EL-decoded residue is combined with the BL-decoded frame to reconstruct the WZ frame. These decoded WZ frames and the EL-decoded key frames obtained from the SHVC Intra decoder are combined to create the EL-decoded video sequence. In summary, the proposed DSVC-HEVC architecture adopts an LC pure Intra-EL coding solution (without temporal predictions), which allows to stop the error propagation and drift when the bitstream is corrupted. Moreover, since the rate control is performed at both the encoder and decoder through a mismatch-free correlation model defining the size of the syndromes to transmit, no feedback channel is needed as it is common in DVC solutions inspired in the Stanford DVC architecture [7].

IV. Machine Learning-Based Side Information Creation

In this scalable coding context, better SI can be created by taking benefit of all the available information, notably the BL motion field and the previously decoded BL and EL frames. Acknowledging this fact, this paper proposes a novel SI creation method that generates three SI candidates, which mutually mitigate their weaknesses in different contexts. Then, the final SI is created by adaptively selecting one out of three SI candidates using a well-known machine learning-based classification technique, an SVM [9], [10]. To better understand the proposed SI creation solution, this section will start by describing the overall SI creation architecture and each SI candidate creation approach before going into the machine learning-based SI selection.

A. SI Creation Architecture and SI Candidate Creation

The SI, an estimate of the current original frame, is created for each EL coding block using ME and compensation techniques to exploit the temporal redundancy. In the DSVC-HEVC framework, a high-quality SI frame is estimated at the EL decoder using all the available decoded data, notably the BL MVF, the BL-decoded frame, $\hat{X}_B$, and the next past and future EL-decoded key frames, $\hat{X}_E^f$, $\hat{X}_E^b$. Fig. 4 illustrates the proposed SI creation architecture composed by two main sets of modules: 1) *SI candidate creation* and 2) *SI candidate selection*.

As shown in Fig. 4, the proposed SI creation process includes three SI candidate creation branches and an SI candidate selection process. The three block-level SI candidates should provide high-quality SI for different situations/contexts

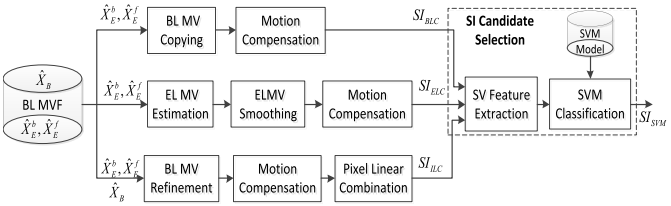

Fig. 4. Proposed SI creation architecture.

in terms of temporal correlation (TC) and IL correlation (ILC), thus providing a more robust solution assuming that the best SI candidate block may be selected. The procedures associated with the various alternative SI candidate creation branches are as follows.

1) BL Motion Field-Based SI Creation SI_{BLC}: The first branch creates the SI candidate by using the BL motion field and the next two past and future EL-decoded references $\hat{X}_E^f$ and $\hat{X}_E^b$, thus exploiting the motion field correlation between layers and the TC between EL frames. To create this SI candidate, the MVF is simply reused from the BL associated prediction unit (PU) and no ME is performed using the higher quality EL-decoded frames. The $\mathrm{SI_{BLC}}$ creation can be performed with the following two steps.

1) *BL MV Copying:* The BL frames are coded with HEVC Inter, and each BL coding unit may contain one or several PUs which may be either Inter- or Intra-mode coded. Therefore, depending on the prediction mode selected for the associated BL PU, two cases are relevant to create the SI MVF.
   a) If the associated BL PU is Inter coded, the motion information of the BL colocated PU, $\mathrm{mv}_B^f, \mathrm{mv}_B^b$, will be simply copied to the current EL SI block.
   b) If the associated BL PU is Intra coded, the corresponding BL-decoded PU will be used as $\mathrm{SI_{BLC}}$ and the BL motion field-based SI creation process for this block is ended.
2) *MC:* Using the motion information from the previous step, the SI candidate, $\mathrm{SI_{BLC}}$, is created by performing MC based on the forward and backward EL-decoded frames, $\hat{X}_E^f$, $\hat{X}_E^b$.

Since the motion information employed here is simply the BL motion field, this SI branch tends to be efficient when the BL motion field is highly correlated with the EL motion field.

2) EL Motion Field-Based SI Creation SI_{ELC}: The second SI branch creates the SI candidate by using only EL-decoded references, thus basically exploiting the TC between the neighboring EL-decoded frames. While the DVC popular motion-compensated temporal interpolation (MCTI) [21] framework could be used, another solution is proposed as the MCTI only uses forward ME to initialize the MVF between the two EL-decoded frames, and thus, the obtained MVF may not accurately capture the motion characteristics of all objects in the sequence. This SI candidate creation branch considers the following steps.

1) *EL MV Estimation:* First, two ME processes are performed to generate two MVFs.
   a) *Forward ME:* ME is applied to generate a *for-*

ward motion field (fwd_mv_E) defining the motion activity from the forward to the backward EL references ($\hat{X}_E^f \rightarrow \hat{X}_E^b$).

b) *Backward ME:* ME is used to generate a *backward motion field* (bwd_mv_E) defining the motion activity from the backward to the forward EL references ($\hat{X}_E^b \rightarrow \hat{X}_E^f$).

The motion vectors associated with each MVF will then serve as candidates for each nonoverlapping block in the interpolation frame in such a way that, for each block in the interpolation frame, from all the available candidate vectors, the motion vector intercepting the interpolated frame closer to the center of block under consideration will be selected; the blocks that are not crossed get a zero MV.

2) *EL MV Smoothing:* As it is sometimes observed that the initial MVFs have low spatial coherence, some spatial smoothing is beneficial toward better final SI. Therefore, weighted vector median filtering [21] is applied here to each initial SI MVF to create two smoothed MVFs, ($fwd\_mv_E^f$, $fwd\_mv_E^b$) and ($bwd\_mv_E^f$, $bwd\_mv_E^b$).

3) *MC:* Finally, using the two smoothed MVFs above, two SI interpolated blocks are created by applying bidirectional MC to the two EL-decoded key frames, $\hat{X}_E^f$, $\hat{X}_E^b$. Then, the two motion-compensated blocks are simply averaged to form the second SI candidate, $\mathrm{SI_{ELC}}$.

This solution uses only the EL-decoded information to create an SI candidate. Therefore, it should be selected when the EL quality is significantly better than the BL quality.

3) BL Motion Field Refinement-Based SI Creation SI_{ILC}: Finally, the third branch creates the SI candidate using all decoded information, notably the BL MVF and both the BL- and EL-decoded frames (notice that the first SI branch does not use the BL-decoded frame). As the previous SI candidate creation branches employ the MVF associated with either the BL- or EL-decoded information, they do not completely consider the ILC. Therefore, the third SI candidate branch includes ME associated with both the BL- and EL-decoded frames. Furthermore, as the quality of the motion-compensated frame may be low if the estimated MVF is inaccurate, a pixel linear combination of the compensated frame and the BL-decoded frame is proposed to create the final SI candidate. The $\mathrm{SI_{ILC}}$ creation branch works as follows.

1) *BL MVF Refinement:* First, the BL MVF is copied to the EL as described for the $\mathrm{SI_{BLC}}$ creation branch. These MVs are then refined using the BL and EL references. In this case, starting from the matching position associated with the BL motion vectors, $\mathrm{mv}_B^f, \mathrm{mv}_B^b$, in the EL reference pictures, the ME refinement process is applied to search for a better motion vector in terms of minimizing the sum of squared differences (SSD) between the BL collocated block and the EL motion-compensated blocks, thus obtaining $\mathrm{mv}_R^f, \mathrm{mv}_R^b$.

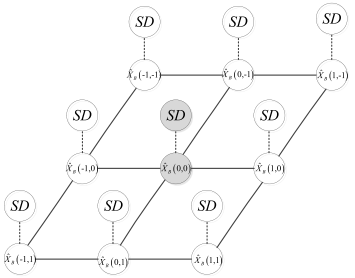

Fig. 5. Spatial neighborhood considered for SD regularization [the current pixel at position $(0, 0)$ is highlighted].

2) *MC:* Second, the obtained $\mathrm{mv}_R^f, \mathrm{mv}_R^b$ are used together with the EL references to create two corresponding motion-compensated blocks. These motion-compensated blocks are then averaged to create P_{mc}.

3) *Pixel Linear Combination:* Finally, the motion-compensated picture, P_{mc}, is linearly combined at pixel level with the BL-decoded block to form $\mathrm{SI_{ILC}}$ as follows:

$$\mathrm{SI_{ILC}}(x, y) = w(x, y) \times P_{\mathrm{mc}}(x, y) + (1 - w(x, y)) \times \hat{X}_B(x, y). \quad (1)$$

Here (x, y) is the pixel position while $w(x, y)$ is the weighting term defining the P_{mc} and $\hat{X}_B$ contributions to the $\mathrm{SI_{ILC}}$ candidate. As each pixel has specific features, it is proposed here to determine a different weight $w(x, y)$ for each pixel in the current block, which is computed based on the SD associated with the ILC and TC as follows.

a) *Pixel SD Calculation:* First, the SD associated with the ILC ($\mathrm{SD_{ILC}}$) and temporal layer correlation ($\mathrm{SD_{TC}}$) are computed for each pixel (x, y) using its BL forward and backward decoded frames coincident in time with the EL key frames, $\hat{X}_B^f$, $\hat{X}_B^b$, and EL motion-compensated key frames, $\hat{X}_E^f$, $\hat{X}_E^b$, as in (2) and (3), as shown at the bottom of this page.

b) *Pixel SD Regularization:* Since the above SDs are computed for each pixel without using original information, it may happen that they do not accurately express the real ILC and TC, thus leading to SD estimation errors and thus less appropriate weights. To obtain more robust weights, it is proposed to regularize these SDs using the spatial correlation, notably the available neighboring pixel information as illustrated in Fig. 5.

$$\mathrm{SD_{ILC}}(x, y) = \frac{\left(\hat{X}_E^f(x, y, \mathrm{mv}_R^f) - \hat{X}_B^f(x, y, \mathrm{mv}_R^f)\right)^2 + \left(\hat{X}_E^b(x, y, \mathrm{mv}_R^b) - \hat{X}_B^b(x, y, \mathrm{mv}_R^b)\right)^2}{2} \quad (2)$$

$$\mathrm{SD_{TC}}(x, y) = \frac{\left(\hat{X}_B(x, y) - \hat{X}_B^f(x, y, \mathrm{mv}_R^f)\right)^2 + \left(\hat{X}_B(x, y) - \hat{X}_B^b(x, y, \mathrm{mv}_R^b)\right)^2}{2} \quad (3)$$

As shown in Fig. 5, the SD regularization process for each pixel considers the SDs for a limited pixel neighborhood (i, j) weighted by a factor depending on the neighboring pixel intensity $(\hat{X}(i, j))$ of the BL-decoded information and their distance to the current pixel (x, y). The regularized SD $(\mathrm{RSD}(x, y))$ for the current pixel is computed as

$$\mathrm{RSD}(x, y) = \frac{\sum_{i=-W}^{W} \sum_{j=-W}^{W} (g(i, j) \times \mathrm{SD}(x + i, y + j))}{\sum_{i=-W}^{W} \sum_{j=-W}^{W} g(i, j)}. \tag{4}$$

Here, $g(i, j)$ is the weight associated with each neighborhood pixel (i, j) and W is the window size defining which neighboring pixels are used to regularize the current pixel SD. The neighboring pixel weight is computed as for bilateral filtering [32]

$$g(i, j) = e^{-\gamma_1 \times (\hat{X}(x,y) - \hat{X}(x+i,y+j))^2} \times e^{-\gamma_2 \times (i^2 + j^2)}. \tag{5}$$

In this paper, the control parameters γ_1 and γ_2 are set to 0.05 while W is set to 2.

c) *Weight Computation:* Finally, the weight $w(x, y)$ is computed using the regularized SDs associated with the ILC $(\mathrm{RSD}_{\mathrm{ILC}}(x, y))$ and TC $(\mathrm{RSD}_{\mathrm{TC}}(x, y))$

$$w(x, y) = \frac{\mathrm{RSD}_{\mathrm{ILC}}(x, y) + 1}{\mathrm{RSD}_{\mathrm{ILC}}(x, y) + \mathrm{RSD}_{\mathrm{TC}}(x, y) + 2}. \tag{6}$$

The term 1 is added to each Regularized Square Difference (RSD) to avoid dividing by zero when RSD is zero. The combination weight $w(x, y)$ obtained from (6) will be applied to (1) to create $\mathrm{SI}_{\mathrm{ILC}}$. As this SI candidate creation branch employs both BL- and EL-decoded information, this solution should be selected when the TC between frames is low or when the GOP size is large.

The quality of each SI candidate is highly dependent on the video content as well as on the techniques adopted in each SI candidate creation branch. Therefore, selecting the best SI at the block level is a rather challenging task as a wrong selection may kill the benefits of having complementary branches.

B. Machine Learning-Based SI Selection

To efficiently select one out of three SI candidates proposed above, the SI selection process is formulated as a classification problem. In this context, an SVM is adopted to solve this classification problem as it is able to provide high classification accuracy [33], [34]. Moreover, the SVM model can be statistically trained with the SI and original data, notably using an offline approach; it is able to learn the correlation between each SI and the original source, an approach that fits rather well the distributed coding paradigm. To design an efficient SVM-based SI selection, a set of SV features has to be carefully selected as it critically determines the classification accuracy as well as the training computational complexity [35].

1) SV Feature Definition: The SVM classification accuracy and the computational cost are highly associated with the quality of the SV features [35]. In this context, a set of SV features has been carefully selected considering what they can express in terms of showing the quality potential of each SI candidate branch. The following features were selected.

1) *MVF Features:* Since the MVF accuracy is directly associated with the quality of each SI candidate, it is important to select metrics expressing the accuracy of the MVF associated with each SI candidate branch. Thus, the SSD between the two EL-compensated reference blocks of each SI creation solution is used for MVF assessment. Therefore, three possible SV features are defined associated with each SI candidate

$$\mathrm{SV}_{\mathrm{mv}_B} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \left(\hat{X}_E^f(x, y, \mathrm{mv}_B^f) - \hat{X}_E^b(x, y, \mathrm{mv}_B^b)\right)^2 \tag{7}$$

$$\mathrm{SV}_{\mathrm{mv}_E} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \left(\hat{X}_E^f(x, y, fwd\_mv_E^f) - \hat{X}_E^b(x, y, fwd\_mv_E^b)\right)^2 \tag{8}$$

$$\mathrm{SV}_{\mathrm{mv}_R} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \left(\hat{X}_E^f(x, y, \mathrm{mv}_R^f) - \hat{X}_E^b(x, y, \mathrm{mv}_R^b)\right)^2. \tag{9}$$

Considering these features, the SI with the smaller $\mathrm{SV}_{\mathrm{mv}}$ should be selected as smaller $\mathrm{SV}_{\mathrm{mv}}$ values are typically associated with more accurate MVF and better SI quality.

2) *TC Feature:* In the $\mathrm{SI}_{\mathrm{BLC}}$ and $\mathrm{SI}_{\mathrm{ELC}}$ creation branches, the SI candidate blocks are created by using only two motion-compensated EL-decoded key frames while for $\mathrm{SI}_{\mathrm{ILC}}$ both the BL- and EL-decoded frames are exploited to create the SI candidate. Therefore, a TC metric should be a good feature to discriminate the first two SI candidates, $\mathrm{SI}_{\mathrm{BLC}}$ and $\mathrm{SI}_{\mathrm{ELC}}$, from the third SI candidate, $\mathrm{SI}_{\mathrm{ILC}}$. Thus, the TC feature, $\mathrm{SV}_{\mathrm{TC}}$, is defined as the SSD between the BL-decoded frame, $\hat{X}_B$, and its motion-compensated blocks in the BL references, $\hat{X}_B^f$, $\hat{X}_B^b$ as follows:

$$\mathrm{SV}_{\mathrm{TC}} = \left(\sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \left(\hat{X}_B(x, y) - \hat{X}_B^f(x, y, \mathrm{mv}_B^f)\right)^2 + \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \left(\hat{X}_B(x, y) - \hat{X}_B^b(x, y, \mathrm{mv}_B^b)\right)^2 \right) \Big/ 2. \tag{10}$$

Naturally, $\mathrm{SV}_{\mathrm{TC}}$ tends to be high when the TC is low. In this case, the $\mathrm{SI}_{\mathrm{BLC}}$ and $\mathrm{SI}_{\mathrm{ELC}}$ qualities, which mainly depend on the TC, will be low. Therefore, if $\mathrm{SV}_{\mathrm{TC}}$ is high, $\mathrm{SI}_{\mathrm{BLC}}$ and $\mathrm{SI}_{\mathrm{ELC}}$ should not be selected.

3) *ILC Feature:* In the proposed SI creation framework, not only the TC but also the ILC has high impact on the SI candidate quality. Therefore, proposed here is an ILC feature, $\mathrm{SV_{ILC}}$, computed as the SSD between the blocks in the motion-compensated EL-decoded key frames and the blocks in the motion-compensated BL forward and backward decoded frames in the same time instant of the EL key frames, which means

$$\begin{aligned}\mathrm{SV_{ILC}} &= \left(\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}\left(\hat{X}_E^f(x,y,\mathrm{mv}_B^f)-\hat{X}_B^f(x,y,\mathrm{mv}_B^f)\right)^2\right.\\ &\quad\left.+\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}\left(\hat{X}_E^b(x,y,\mathrm{mv}_B^b)-\hat{X}_B^b(x,y,\mathrm{mv}_B^b)\right)^2\right)\Big/2.\end{aligned} \tag{11}$$

In contrast to $\mathrm{SV_{TC}}$, if $\mathrm{SV_{ILC}}$ is small, the third SI candidate, $\mathrm{SI_{ILC}}$, created with the BL-decoded block should be selected as in this case the correlation between the BL and EL frames should be high. In this case, $\mathrm{SI_{ILC}}$ should have a better quality than the other SI candidates as the BL-decoded block is directly used in the $\mathrm{SI_{ILC}}$ definition.

4) *SI Difference Features:* Finally, the difference between the SI candidates themselves is used as a feature not only to discriminate the SIs but also to accelerate the training process as when the SI difference is large, very likely one of these SI candidates should be much better. In this case, the following SSDs between SIs are used:

$$\mathrm{SV_{SIBE}} = \sum_{x=0}^{N-1}\sum_{y=0}^{N-1}(\mathrm{SI_{BLC}}(x,y)-\mathrm{SI_{ELC}}(x,y))^2 \tag{12}$$

$$\mathrm{SV_{SIEI}} = \sum_{x=0}^{N-1}\sum_{y=0}^{N-1}(\mathrm{SI_{ELC}}(x,y)-\mathrm{SI_{ILC}}(x,y))^2 \tag{13}$$

$$\mathrm{SV_{SIBI}} = \sum_{x=0}^{N-1}\sum_{y=0}^{N-1}(\mathrm{SI_{BLC}}(x,y)-\mathrm{SI_{ILC}}(x,y))^2. \tag{14}$$

The set of the above-proposed SV features is employed in both the SVM training and classification processes.

2) SI Selection Process: The proposed SVM-based SI selection process includes three main steps: a) *SVM model training*; b) *SV feature extraction*; and c) *SVM classification*.

1) *SVM Model Training:* First, the SVM model has to be trained. Considering the training process computational complexity as well as the possibility of using original information to train the model, this paper adopts an offline instead of online training method. Although the online training method can adaptively update the SVM model, the associated computational complexity is rather high and the online solution cannot exploit the original data to create the correct decision labels as the offline method; incorrect decision labels in the training phase may lead to an unreliable SVM model. To obtain a reliable SVM model, the training samples are extracted from several training sequences (different from the test sequences), with different motion characteristics and resolutions as specified in Section VI-A. The SVM model training is performed with the following steps.
   a) *SI Candidate Creation:* First, the three SI creation branches presented previously are applied to obtain the SI candidate values for all training sequences.
   b) *SI Oracle Label Derivation:* Then, using the original data and the three SI candidates obtained from previous step, the oracle label (this means the ideal SI candidate) is identified; this would correspond to the perfect classification.
   c) *SVM Training:* Finally, a popular SVM software, $\mathrm{SVM^{light}}$ [36], is used to train the SVM model using all the data collected. The trained model will be used in the classification stage when performing real coding.
2) *Testing SV Feature Extraction:* In the testing stage, the first operation regards the extraction of the set of SV features (F) proposed above for the current test sequence block by block

$$\begin{aligned}\mathbf{F} = \{&\mathrm{SV_{mv_B}};\mathrm{SV_{mv_E}};\mathrm{SV_{mv_R}};\mathrm{SV_{TC}};\\ &\mathrm{SV_{ILC}};\mathrm{SV_{SIBE}};\mathrm{SV_{SIBI}};\mathrm{SV_{SIEI}}\}.\end{aligned}$$

3) *SVM Classification:* At this stage, the SVM classification is performed and its outcome is the selected SI candidate. SVMs were initially developed to perform binary classification, which includes only two inputs [9], [36] while the proposed SI selection includes three input SI candidates. To achieve an efficient multiclass SVM classification, this paper adopted the well-known *one-versus-one*-based multiclass SVM approach as it is able to provide the highest classification accuracy when compared with alternative approaches such as *one versus all* and *directed acyclic graph* [37]. In this approach, binary SVM classifiers for all possible SI pairs are created. Then, the SI associated with the most selected class label is adopted as the best SI. Hence, the multiclass SVM is performed in two steps.
   a) *Performing Binary SVM for Each Pair of SIs:* First, three binary SVM classification processes are performed for the three possible SI pairs $\{\mathrm{SI_{BLC}};\mathrm{SI_{ELC}}\}$, $\{\mathrm{SI_{BLC}};\mathrm{SI_{ILC}}\}$, and $\{\mathrm{SI_{ELC}};\mathrm{SI_{ILC}}\}$.
   b) *Finding the Best SI:* After the classification labels for the three classifiers are obtained, the SI candidate with the most selected class label is chosen as the final SI. If there is a tie in the selection, a tie-breaking strategy will be used. In this context, the results of the second binary SVM given previously, the SVM classification for $\mathrm{SI_{BLC}}$ and $\mathrm{SI_{ILC}}$ is used to define the selected SI as this solution typically provides the highest SI quality as determined experimentally.

In summary, this section has proposed a novel SI creation solution which takes into account not only the BL- and EL-decoded frames but also the decoded BL motion to efficiently generate the SI.

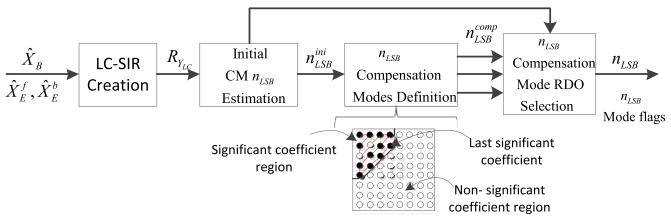

Fig. 6. Encoder ACM architecture.

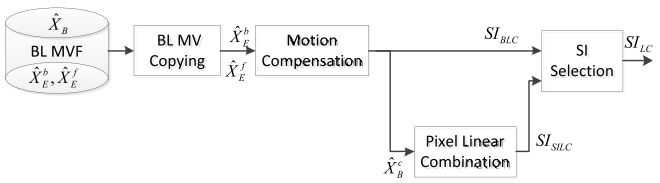

Fig. 7. LC-SI creation architecture.

V. Adaptive Correlation Modeling

The CM aims to calculate the minimum number of least significant bitplanes (n_{LSB}) of the EL residue that need to be coded and sent to the decoder to obtain an EL-reconstructed frame with a certain target quality. Therefore, a higher CM accuracy usually leads to a better DSVC RD performance. As discussed in Section II, available CM solutions, such as ECM [28], [29] and EDCM [20], still contain several limitations and drawbacks such as using a (encoder) high complexity SI and providing inaccurate correlation estimations. To overcome such problems, this paper proposes an ACM based on an initial n_{LSB} estimation mechanism [20], which is complemented with a set of n_{LSB} compensation modes, adaptively selected using an RDO approach. In the proposed ACM mechanism, the n_{LSB} compensation modes overcome the overestimations and underestimations associated with the initial CM. While the initial CM overestimations (n_{LSB} is higher than needed) imply wasting rate, underestimations (n_{LSB} is lower than needed) imply quality penalties; both effects are undesired and can be compensated.

A. Encoder Adaptive Correlation Modeling

The encoder ACM architecture and the novel n_{LSB} compensation modes are presented in the following.

1) Encoder ACM Architecture: Fig. 6 illustrates the proposed encoder ACM architecture which includes four main modules: 1) LC-SIR creation; 2) initial CM n_{LSB} estimation; 3) n_{LSB} compensation modes definition; and 4) n_{LSB} compensation mode RDO selection.

The encoder ACM proceeds as follows.

1) *LC-SIR Creation:* First, an LC-SIR, $R_{Y_{\mathrm{LC}}}$, is created. Although Section IV has proposed a powerful (decoder) SI creation solution, a lower complexity SI creation process is required, since this module also exists at the encoder. The LC-SIR creation process is described in detail in Section V-A2.
2) *Initial CM n_{LSB} Estimation:* The $R_{Y_{\mathrm{LC}}}$ created in the previous step and the decoded BL residue quantized coefficient, $\mathrm{TR}_{X_B}^{\mathrm{QP}_B}$, are used to estimate an initial n_{LSB} value, $n_{\mathrm{LSB}}^{\mathrm{ini}}$, using the CM presented in Section V-A3. This $n_{\mathrm{LSB}}^{\mathrm{ini}}$ estimation is performed in the same precise way at the encoder and decoder sides (no rate involved) using only decoded information. As such, $n_{\mathrm{LSB}}^{\mathrm{ini}}$ may involve overestimations and underestimations regarding the ideal n_{LSB} value, i.e., the n_{LSB} value that would allow us to perfectly recover the EL residue with the minimum rate.
3) *n_{LSB} Compensation Modes Definition:* After obtaining $n_{\mathrm{LSB}}^{\mathrm{ini}}$ and targeting to compensate for the overestimations and underestimations associated with $n_{\mathrm{LSB}}^{\mathrm{ini}}$, three additional block level n_{LSB} compensation modes are defined as in Section V-A4. These $n_{\mathrm{LSB}}^{\mathrm{ini}}$ compensation modes aim to mitigate the inaccurate estimations due to the use of only decoded information.
4) *n_{LSB} Compensation Mode RDO Selection:* Finally, an RDO-based selection approach is applied to the four n_{LSB} compensation modes (i.e., the initial and the three additional compensation modes) to find the best n_{LSB} compensation mode. Since the original data are required to select the best out of four n_{LSB} compensation modes for each coding block, one 2-b flag is coded and sent to the decoder.

2) LC-SIR Creation: Since the LC-SIR is simply the difference between the LC-SI frame and the corresponding BL-decoded frame, the LC-SI creation process should generate an SI frame as close as possible to the decoder SI but with lower computational complexity. Fig. 7 illustrates the LC-SI creation solution architecture.

As shown in Fig. 7, the proposed LC-SI creation solution is a simplified version of the powerful SI creation solution proposed in Section IV where one full branch is removed and another branch simplified to avoid any (encoder) ME and motion smoothing operations. In the LC-SI creation process, the first SI branch is kept where the BL MVF is used together with the two EL-decoded frames to create $\mathrm{SI}_{\mathrm{BLC}}$. Then, a pixel linear combination between the $\mathrm{SI}_{\mathrm{BLC}}$ and the BL-decoded frame, $\hat{X}_B$, is used to create a simplified IL SI, $\mathrm{SI}_{\mathrm{SILC}}$. Conceptually, the $\mathrm{SI}_{\mathrm{BLC}}$ quality mainly depends on the TC between the two EL-decoded frames while the $\mathrm{SI}_{\mathrm{ILC}}$ quality is highly dependent on the ILC between the BL and EL frames. After, a simple SI selection mechanism based on the assessment of the ILC and TC is performed to create the best LC-SI from the two computed SIs. While the TC is assessed as in (10), the ILC is assessed as in (11). A direct comparison between these two correlations will determine the final LC-SI. This means, if $\mathrm{SSD}_{\mathrm{TC}} \leq \mathrm{SSD}_{\mathrm{ILC}}$, than $\mathrm{SI}_{\mathrm{BLC}}$ is selected; otherwise, $\mathrm{SI}_{\mathrm{SILC}}$ is selected.

3) Initial CM n_{LSB} Estimation (Mode 1): In this step, the initial number of LSBs, $n_{\mathrm{LSB}}^{\mathrm{ini}}$, is calculated for each quantized DCT coefficient at both the encoder and decoder with the EDCM described in [20], notably by using the BL quantized residue, $\mathrm{TR}_{X_B}^{\mathrm{QP}_B}$, and the LC-SIR, $R_{\mathrm{SI}_{\mathrm{LC}}}$, obtained from the previous step. Since $\mathrm{TR}_{X_B}^{\mathrm{QP}_B}$ is the quantized DCT value of the BL residue with the BL quantization step size, QP_B, it is proposed to use the $\mathrm{TR}_{\mathrm{SI}_{\mathrm{LC}}}$ DCT value also quantized with

the BL quantization step size to compute $n_{\mathrm{LSB}}^{\mathrm{ini}}$ as

$$n_{\mathrm{LSB}}^{\mathrm{ini}} = \begin{cases} 0, & \text{if}\left(\mathrm{TR}_{X_B}^{\mathrm{QP}_B} = \mathrm{TR}_{\mathrm{SI_{LC}}}^{\mathrm{QP}_B}\right) \text{ and } \left(\mathrm{TR}_{\mathrm{SI_{LC}}}^{\mathrm{QP}_E} \neq 0\right) \\ 2 + \left\lfloor \log_2\left(\left|\mathrm{TR}_{X_B}^{\mathrm{QP}_B} - \mathrm{TR}_{\mathrm{SI_{LC}}}^{\mathrm{QP}_B}\right| + 1\right)\right\rfloor, & \text{otherwise.} \end{cases} \tag{15}$$

As $\mathrm{TR}_{\mathrm{SI_{LC}}}^{\mathrm{QP}_E} = 0$ corresponds to the case where the SI frame quality is close to the decoded BL frame, the condition $(\mathrm{TR}_{\mathrm{SI_{LC}}}^{\mathrm{QP}_E} \neq 0)$ is included to guarantee that some quality enhancement regarding the BL is always achieved.

Although the ECM could stop at this step without any encoder–decoder mismatches as $n_{\mathrm{LSB}}^{\mathrm{ini}}$ has been determined at the encoder and the decoder in the same way, for each quantized DCT coefficient, at no rate cost, a statistical analysis on the $n_{\mathrm{LSB}}^{\mathrm{ini}}$ accuracy indicates that it may be significantly improved. As the original data and the decoder SI are not simultaneously exploited, underestimations and overestimations regarding the ideal n_{LSB} value may happen and thus significant RD performance losses may occur. This accuracy analysis has been carried out by comparing the $n_{\mathrm{LSB}}^{\mathrm{ini}}$ value calculated from (15) with the oracle number of LSBs, $n_{\mathrm{LSB}}^{\mathrm{orc}}$, calculated with the encoder EL residue (quantized originals) and the decoder SIR

$$n_{\mathrm{LSB}}^{\mathrm{orc}} = \begin{cases} 0, & \text{if}\left(\mathrm{TR}_{X_E}^{\mathrm{QP}_E} = \mathrm{TR}_{\mathrm{SI_{SVM}}}^{\mathrm{QP}_E}\right) \\ 2 + \left\lfloor \log_2\left|\mathrm{TR}_{X_E}^{\mathrm{QP}_E} - \mathrm{TR}_{\mathrm{SI_{SVM}}}^{\mathrm{QP}_E}\right| \right\rfloor, & \text{otherwise.} \end{cases} \tag{16}$$

By comparing $n_{\mathrm{LSB}}^{\mathrm{ini}}$ with $n_{\mathrm{LSB}}^{\mathrm{orc}}$, three relevant situations may happen at coefficient level.

1) *Underestimation if* $n_{LSB}^{ini} - n_{LSB}^{orc} < 0$*:* In this case, the EL-reconstructed quality may decrease as the number of EL residue LSBs to be coded and sent from the encoder is less than the necessary amount of EL residue LSBs to appropriately reconstruct the EL residue.
2) *Correct Estimation if* $n_{LSB}^{ini} - n_{LSB}^{orc} = 0$*:* In this case, the number of EL residue LSBs expresses exactly the correlation between EL residue and SIR, thus allowing improvement of the DSVC RD performance without any compensation.
3) *Overestimation if* $n_{LSB}^{ini} - n_{LSB}^{orc} > 0$*:* In this case, the initial CM is overestimating n_{LSB}. While the target EL quality is fully reconstructed, there is a penalty in rate as an exaggerated number of EL residue LSBs is coded and sent to the decoder.

Fig. 8 illustrates the histogram analysis of the $n_{\mathrm{LSB}}^{\mathrm{ini}}$ and $n_{\mathrm{LSB}}^{\mathrm{orc}}$ relationship where several DCT coefficients, notably DC, AC1, AC2, AC3, AC10, and AC15 (considering a 4×4 transform), are examined for the *BlowingBubbles* sequence using the spatial and temporal resolutions of 416×240 and 30 Hz, respectively.

From the analysis of these histograms, it is possible to conclude the following.

1) The relative frequency is always much higher for the zero $n_{\mathrm{LSB}}^{\mathrm{ini}} - n_{\mathrm{LSB}}^{\mathrm{orc}}$ difference meaning that the initial CM often performs correctly.
2) The number of underestimations is typically higher than the number of overestimations for all tested DCT coefficients.

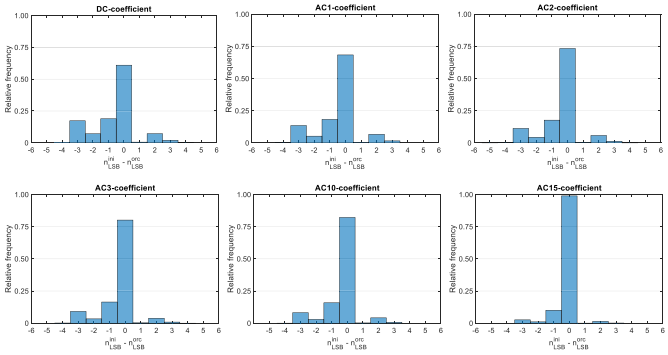

Fig. 8. Histograms of the difference between $n_{\mathrm{LSB}}^{\mathrm{ini}}$ and $n_{\mathrm{LSB}}^{\mathrm{orc}}$.

3) The number of both underestimations and overestimations is higher for the lower frequency DCT coefficients such as DC, AC1, or AC2 when compared with the higher frequency DCT coefficients such as AC10 and AC15.

4) n_{LSB} *Compensation Modes Definition:* Based on the observations, this paper proposes to define a few n_{LSB} compensation modes to mitigate the occurrence of underestimations and overestimations with the minimum rate as they directly reduce the DSVC-HEVC RD performance. To minimize the rate, the compensation will be performed at block level even if under and overestimations may coexist within the same block for different coefficients; the stronger effect will win through the RDO process. As each block typically ends (in zigzag scanning) with a number of zero coefficients, it does not make sense to compensate these *correct* zero coefficients even if it is concluded that (mostly) underestimation or overestimation is happening for the nonzero coefficients of the same block. Therefore, a segmentation process is first performed to divide each coding block into two coefficient regions, the significant and nonsignificant regions (see Fig. 6); while the significant coefficient regions will suffer n_{LSB} compensation, the nonsignificant coefficient region will still have its coefficients set to zero, independently of the compensation mode. The block segmentation procedure labels as *significant* all the coefficients up to the last nonzero coefficient in zigzag scanning while the remaining coefficients are labeled as *nonsignificant*.

Considering the statistical analysis of the $n_{\mathrm{LSB}}^{\mathrm{ini}}$ to $n_{\mathrm{LSB}}^{\mathrm{orc}}$ difference and the coding efficiency impact of the underestimation and overestimation problems, the $n_{\mathrm{LSB}}^{\mathrm{ini}}$ of the coefficients belonging to the significant group will be compensated using one of the following additional n_{LSB} compensation modes.

1) *Model-Based Underestimation Compensation Mode (Mode 2):* This compensation mode aims to overcome the underestimation problem that happens more often for the lower frequency coefficients and has high impact on the EL-reconstructed WZ frame quality. The amount of n_{LSB} compensation, which means the number of bits to add to $n_{\mathrm{LSB}}^{\mathrm{ini}}$ to compensate its underestimation, is estimated using a compensation correlation model between the $n_{\mathrm{LSB}}^{\mathrm{ini}}$ obtained from (15) and the oracle number of LSBs, $n_{\mathrm{LSB}}^{\mathrm{orc}}$, obtained from (16). From intensive fitting

experiments, it has been concluded that a simple linear model may be enough to determine the compensated number of n_{LSB} to code as

$$n_{\mathrm{LSB}}^{\mathrm{model}} = \lfloor (a \times n_{\mathrm{LSB}}^{\mathrm{ini}} + b) \rfloor. \quad (17)$$

Here, a and b are the two parameters of the linear model, which are obtained by an offline fitting model process, where $n_{\mathrm{LSB}}^{\mathrm{orc}}$ and $n_{\mathrm{LSB}}^{\mathrm{ini}}$ have been computed for each coefficient in the block. To determine the best fitting, only the coefficients with an $n_{\mathrm{LSB}}^{\mathrm{ini}}$ associated with underestimation were considered. The fitting process has been carried out using several training sequences with different characteristics such as *Coastguard*, *HallMonitor* (352×288 at 30 Hz), and *RaceHorses* (416×240 at 30 Hz) and several quantization parameters (QPs) such as $\{\mathrm{QP}_B; \mathrm{QP}_E\} = \{34; 30\}, \{34; 24\}$. The estimated parameters $(a; b)$ obtained were very similar when a model was computed individually for each DCT band, and all DCT coefficients were considered. Therefore, a unified model has been adopted for all DCT coefficients with the parameters $(a; b) = (0.4819; 2.0476)$.

2) *Extreme Underestimation Compensation Mode (Mode 3):* The previous n_{LSB} compensation mode targets the underestimation problem by modeling the correlation between $n_{\mathrm{LSB}}^{\mathrm{ini}}$ and $n_{\mathrm{LSB}}^{\mathrm{orc}}$ when just this effect occurs. However, it may happen that this compensation model is not accurate enough and underestimation still remains. To overcome this situation, it is proposed to consider an extreme case of underestimation compensation where the maximum $n_{\mathrm{LSB}}^{\mathrm{ini}}$ for the coefficients in the significant region is adopted to perform the underestimation compensation for all coefficients in the significant region. In summary, the estimated number of LSBs for this n_{LSB} compensation mode is obtained by making

$$n_{\mathrm{LSB}}^{\mathrm{comp}} = \max_{k \in \mathrm{Sig\_Reg}} \left\{ n_{\mathrm{LSB}}^{\mathrm{ini}}(k) \right\}. \quad (18)$$

3) *Extreme Overestimation Compensation Mode (Mode 4):* Although overestimation happens less often than underestimation, it is still appropriate to mitigate this problem as coding rate is wasted, thus reducing the RD performance. Thus, an extreme case of overestimation compensation mode is defined where the minimum number of LSBs from all the coefficients in the significant region is adopted

$$n_{\mathrm{LSB}}^{\mathrm{comp}} = \min_{k \in \mathrm{Sig\_Reg}} \left\{ n_{\mathrm{LSB}}^{\mathrm{ini}}(k) \right\}. \quad (19)$$

5) n_{LSB} Compensation Mode RDO Selection: To be sure that moving away from the initial no-compensation mode can only bring RD performance benefits, the best n_{LSB} compensation mode from the four defined modes is adaptively selected using an RDO approach. To indicate which n_{LSB} compensation mode is selected, the encoder has to code and transmit a 2-b flag for each coding block. To perform the n_{LSB} compensation mode RDO selection, the following applies.

1) *Reconstruction:* First, for each n_{LSB} compensation mode, an EL WZ-reconstructed frame is created using the LC-SI, the estimated n_{LSB} and the associated syndrome.
2) *Distortion Assessment:* Then, the block distortion associated with each n_{LSB} compensation mode is calculated with an SSD metric.
3) *Rate Assessment:* Next, the coding bitrate associated with each n_{LSB} compensation mode is calculated using the CABAC entropy engine.
4) *Mode Selection:* Finally, the RD cost for each n_{LSB} compensation mode is determined using a Lagrangian optimization as adopted in the HEVC standard [1]. The n_{LSB} compensation mode associated with the lowest RD cost will be selected and signaled with a 2-b flag.

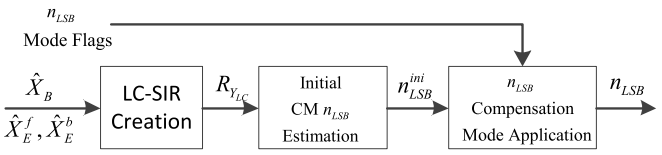

Fig. 9. Decoder CM architecture.

B. Decoder Adaptive Correlation Modeling

At the decoder, the ACM is performed to determine the number of LSBs that will be used to completely reconstruct the EL WZ frames. In this case, the n_{LSB} compensation mode flag sent from the encoder indicates which n_{LSB} compensation mode was selected at the encoder ACM and should be applied at the decoder to improve $n_{\mathrm{LSB}}^{\mathrm{ini}}$.

Fig. 9 illustrates the decoder ACM architecture which includes three main modules. While the first two modules are similar to those defined for the encoder ACM, the last module, n_{LSB} compensation mode application, is a simplified version of the encoder ACM n_{LSB} compensation mode definition module where only the signaled mode has to be considered.

The decoder ACM proceeds as follows. First, an LC-SIR, $R_{Y_{\mathrm{LC}}}$, is created as described in Section V-A2. After, $R_{Y_{\mathrm{LC}}}$ is used to create an initial n_{LSB} estimation as described in Section V-A3. Depending on which n_{LSB} compensation mode is selected at the encoder and signaled through the n_{LSB} compensation mode flag, the appropriate n_{LSB} compensation mode is applied to create the final n_{LSB} (see Section V-A).

VI. Performance Assessment

This section evaluates the novel proposed tools and the overall DSVC-HEVC codec performance under meaningful test conditions compared with the most relevant benchmarks.

A. Test Conditions

The performance assessment is carried out for four video test sequences selected from the HEVC test set [38]. These sequences were selected for their representativeness as they show a variety of motion and texture characteristics, notably *BlowingBubbles* with low local motion, *BasketballPass* with high local motion and camera movement, *Flowervase* with a zoom and brightness variations, and *BQSquare* with low local motion but lots of movements and scene changes. For SI creation, the offline SVM training method is

TABLE I
SUMMARY OF TEST CONDITIONS

Test sequences	Spatial resolution	Temporal resolution	Number of frames
BlowingBubbles (BlowB)	416×240	50 Hz	497
BasketballPass (BaskP)		50 Hz	497
Flowervase (Flow)		30 Hz	297
BQSquare (BQS)		60 Hz	599
Quantization parameters	$QP_B = \{34; 30\}$ $QP_E = QP_B - \{4; 6; 8; 10\}$		

applied to three other video sequences containing different motion characteristics and resolutions, notably *RaceHorses* (416 × 240 at 30 Hz) with high local motion, *Coastguard* (352×288 at 30 Hz) with high global motion, and *HallMonitor* (352 × 288 at 30 Hz) with low local and global motions. Table I summarizes the main sequence characteristics and the BL and EL QPs. The popular SVM^{light} [36] software has been employed for the SI candidate selection. The HEVC HM version 14.0 [39] reference software has been used to code the BL, while the SHVC SHM version 6.0 [40] reference software has been used to code the EL key frames.

In the proposed DSVC-HEVC codec, one WZ frame is created at the middle of each pair of key frames as shown in Fig. 1(a). This configuration corresponds to a group of pictures (GOP) size of 2 which can avoid a long delay when decoding the video sequence while allowing to reach a competitive DSVC-HEVC RD performance. As usual, results are presented for the luminance component and the rate includes all frames (the BL frames, the EL key frames, and the EL WZ frames).

B. SI Quality Assessment

This section evaluates the performance of the proposed SI creation solution (labeled as SI_{SVM}), measured in terms of the Peak Signal-to-Noise Ratio (PSNR) metric in comparison to the SI creation solution proposed in [20] (labeled as SI as [20]) as well as to the three individual SI candidates, SI_{BLC}, SI_{ELC}, and SI_{ILC}. Table II presents the SI quality results for the proposed SI_{SVM} and the PSNR gains of the proposed SI solution regarding the relevant benchmarks.

From the results, the following conclusions may be derived.

1) *Proposed SI_{SVM} Versus Previous SI [20]:* The proposed SI creation solution, SI_{SVM}, is able to achieve a better SI quality than the recent state-of-the-art SI solution [20], notably around 1-dB PSNR improvement on average. The gains mainly result from the usage of the BL motion information and the increased accuracy of the machine learning-based SI selection solution.
2) *Proposed SI_{SVM} Versus Individual SI Candidates:* Compared with the three individual SI candidates, SI_{BLC}, SI_{ELC}, and SI_{ILC}, the proposed SI creation solution, SI_{SVM}, achieves a higher SI quality, notably with an average PSNR improvement about 2, 3.2, and 0.9 dB, respectively. These gains mainly result from the increased accuracy of the machine learning-based SI selection solution.

TABLE II
AVERAGE SIF FRAME QUALITY (PSNR) AND GAINS REGARDING SI CREATION BENCHMARK [dB]

Sequences	QP_E	SI_{SVM} [dB]	SI_{SVM} versus SI_{BLC}	SI_{SVM} versus SI_{ELC}	SI_{SVM} versus SI_{ILC}	SI_{SVM} versus SI [20]
BlowingBubbles	30	33.98	2.19	2.28	0.30	0.42
	28	35.03	2.58	2.65	0.70	0.67
	26	35.83	2.90	2.96	1.11	0.89
	24	36.47	3.19	3.24	1.53	1.10
Average SI Gain			**2.72**	**2.78**	**0.91**	**0.77**
BasketballPass	30	35.48	5.19	5.87	-0.21	0.32
	28	35.99	5.54	6.18	0.02	0.45
	26	36.39	5.82	6.42	0.27	0.56
	24	36.73	6.06	6.62	0.52	0.67
Average SI Gain			**5.65**	**6.27**	**0.15**	**0.50**
Flowervase	30	39.29	-0.24	0.58	0.53	0.46
	28	40.60	-0.33	0.77	1.06	0.76
	26	41.75	-0.42	0.95	1.68	1.09
	24	42.77	-0.51	1.16	2.38	1.45
Average SI Gain			**-0.38**	**0.87**	**1.41**	**0.94**
BQSquare	30	33.55	0.15	2.33	0.36	1.10
	28	34.40	0.14	2.67	0.78	1.68
	26	35.05	0.12	2.93	1.20	2.13
	24	35.57	0.12	3.18	1.57	2.52
Average SI Gain			**0.13**	**2.78**	**0.98**	**1.86**
Overall Average SI Gain			***2.03***	***3.17***	***0.86***	***1.02***

3) *SI Candidate Quality Comparison:* Among the individual SI candidates, SI_{ILC} outperforms SI_{BLC} and SI_{ELC} for high and middle motion sequences, such as *BasketballPass* and *BlowingBubbles*, while SI_{BLC} is more efficient for video sequences with low motion, such as *BQSquare*, and zooms, such as *Flowervase*. Although the SI_{ELC} quality is usually lower than the other SI candidates, which are designed to exploit the BL motion information, SI_{ELC} still plays an important role in the selection scheme, notably when the quality difference between BL and EL is high.
4) *SI_{SVM} Quality With Different BL and EL Quality Gaps:* The proposed SI creation solution provides a higher SI quality when the quality gap between the BL and EL increases. For example, as shown in Table II, the quality of SI_{SVM} in *BlowingBubble* always increases when the EL QP decreases (here, the BL QP is fixed). This observation is also true for other test sequences. This results from the fact that, when the EL QP decreases, the EL-decoded frames quality will increase; thus, the SI candidates quality will also increase (thus, also the final SI quality) as they are mainly created by performing MC with the EL-decoded frames.

C. Correlation Modeling Assessment

Table III presents the BD Rate gains achieved for DSVC-HEVC with the proposed ACM solution regarding the previous CMs, notably the ECM [28], [29] and the Initial CM (EDCM [20]), while Table IV shows the usage of the various ACM compensation modes in percentage [%].

The results in Table III indicate that the DSVC-HEVC RD performance with the proposed ACM outperforms both the

TABLE III
BD RATE GAINS WITH ACM REGARDING THE OTHER CMS

Sequences	Proposed ACM versus ECM [28, 29]	Proposed ACM versus Initial CM (EDCM [20])
BlowingBubbles	-5.22	-2.41
BasketballPass	-2.70	-3.46
Flowervase	-10.00	-4.68
BQSquare	-2.34	-1.55
Average	**-5.07**	**-3.03**

TABLE IV
PERCENTAGE [%] OF BLOCKS USING VARIOUS ACM MODES

Sequences	Mode1	Mode 2	Mode 3	Mode 4
BlowingBubbles	54.88	40.60	1.85	2.67
BasketballPass	60.32	28.80	1.31	9.57
Flowervase	87.81	5.83	0.29	6.07
BQSquare	62.46	32.00	1.31	4.23
Average	**66.37**	**26.81**	**1.19**	**5.64**

DSVC-HEVC with the ECM and initial CM solutions. The gains mainly come from the additional n_{LSB} compensation modes that are able to mitigate the previous CM accuracy problems. Meanwhile, the results in Table IV point out that the initial CM mode and the model-based underestimation CM compensation mode are highly selected, notably with around 66.4% and 26.8%, respectively; this shows that the two extreme compensation CM modes are really only used for extreme cases but they are still relevant. Therefore, the proposed ACM is more accurate than both the previous ECM and Initial CM solutions.

D. DSVC-HEVC RD Performance Assessment

Considering the prediction structure specified in Section III-A, notably the BL HEVC Inter-coding backward compatibility and the low encoding complexity requirements, the following benchmarks have been selected.

1) *HEVC-Simul_InterBL_IntraEL:* This non-SVC benchmark creates one independent bitstream for each required quality level. In this case, considering the EL LC requirement, the BL (lower quality) bitstream is created with the HEVC Inter-coding solution while the EL (higher quality) bitstream is created with the HEVC Intra-coding solution. The results have been obtained with the HEVC reference software HM, version 14.0 [39].
2) *HEVC-Simul_InterBL_InterEL:* This is another non-SVC solution that creates one independent bitstream for each required quality level. However, in this benchmark, both the BL and EL frames are coded with an HEVC Inter-coding solution; thus, a high RD performance can be achieved. The results have been obtained with the HEVC reference software HM, version 14.0 [39].
3) *SHVC-InterBL_IntraEL:* This is the most relevant SVC benchmark. It is backward compatible with HEVC Inter coding at the BL while providing a low encoding complexity solution at the EL with HEVC Intra coding. To obtain a similar situation to the proposed solution in terms of drift and error resilience, the EL residue is defined as the difference between the original and the EL prediction created by either the BL colocated block or the EL Intra prediction. The results for this benchmark have been obtained with the SHVC SHM reference software, version 6.0 [40].
4) *SHVC-InterBL_InterEL:* This alternative SVC benchmark corresponds to a normative SHVC solution where the Inter-coding approach is used at both BL and ELs. In this benchmark, the EL residue may be the difference between the original and an EL motion-compensated prediction. The results for this benchmark have also been obtained with the SHVC SHM reference software, version 6.0 [40].

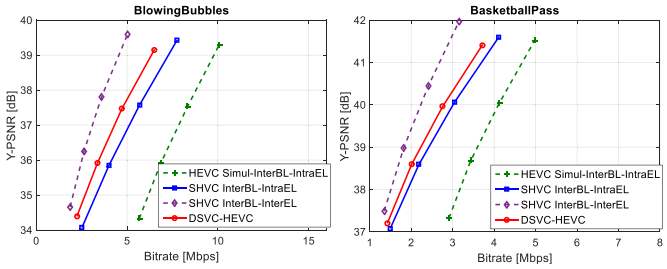

Fig. 10. DSVC-HEVC RD performance comparison ($\mathrm{QP}_B = 34$).

TABLE V
DSVC-HEVC BD-RATE REGARDING BENCHMARKS

Seq.	HEVC-Simul InterBL_IntraEL		HEVC-Simul InterBL_InterEL		SHVC-InterBL_IntraEL		SHVC-InterBL_InterEL	
	$QP_B = 34$	$QP_B = 30$	$QP_B = 34$	$QP_B = 30$	$QP_B = 34$	$QP_B = 30$	$QP_B = 34$	$QP_B = 30$
BlowB	-47.4	-46.4	-6.6	-15.5	-16.1	-8.1	38.1	29.7
BaskP	-36.7	-38.7	-7.8	-13.7	-7.7	-5.7	22.6	19.4
Flow	-52.9	-52.3	-2.9	**-7.1**	**-23.7**	**-15.3**	49.3	51.9
BQS	-50.4	-47.1	-15.5	-19.7	-10.7	-3.3	38.1	27.1
Avg.	***-46.8***	***-46.1***	***-8.2***	***-14.0***	***-14.6***	***-8.1***	***37.0***	***32.0***

To obtain a fair comparison with DSVC-HEVC, a GOP size of 2 is also adopted for the Inter configuration of the above benchmarks. To achieve the highest RD performance, the Intra refresh is set to -1 in the reference software meaning that only the first BL frame is Intra coded.

It is important to stress that the DSVC solution in [20] cannot be used as benchmark here as it addresses a completely different functional tradeoff in terms of compression efficiency, complexity, and error resilience by adopting an Intra-coded BL and a less integrated codec design.

Fig. 10 presents the RD performance for a couple of sequences while Table V shows the BD-Rate savings [41] for the proposed DSVC-HEVC solution with respect to the mentioned benchmarks for $\mathrm{QP}_B = 34$ and $\mathrm{QP}_B = 30$, respectively.

From the obtained RD performance results, the following conclusions may be derived.

1) *DSVC-HEVC Versus HEVC-Simul_InterBL_IntraEL:* The proposed DSVC-HEVC RD performance is better than HEVC-Simul_InterBL_IntraEL for all test sequences with average BD-Rate gains of about 46.8% for $\mathrm{QP}_B = 34$ and 46.1% for $\mathrm{QP}_B = 30$.

The higher BD-Rate gains are obtained for the sequences containing low motion and zooms such as *BQSquare* and *Flowervase*. Regarding HEVC-Simul_InterBL_IntraEL, the proposed DSVC-HEVC solution exploits the ILC to enhance the RD performance while HEVC-Simul_InterBL_IntraEL codes the BL and EL without any IL prediction tools.

2) *DSVC-HEVC Versus HEVC-Simul_InterBL_InterEL:* The proposed DSVC-HEVC also outperforms the HEVC-Simul_InterBL_InterEL for all test sequences with average BD-Rate gains of about 8.2% for $QP_B = 34$ and 14% for $QP_B = 30$. Although this benchmark allows exploiting the motion information at the ELs, its independent layer coding approach makes it less efficient, notably compared with other scalable coding solutions.
3) *DSVC-HEVC Versus SHVC-InterBL_IntraEL:* The DSVC-HEVC RD performance is also better than SHVC-InterBL_IntraEL (the most relevant benchmark) for all test sequences with average BD-Rate gains of 14.6% for $QP_B = 34$ and 8.1% for $QP_B = 30$. The maximum BD-Rate gain is obtained for the *Flowervase* sequence, notably about 23.7% for $QP_B = 34$ and 15.3% for $QP_B = 30$. In this sequence, higher SI quality is usually obtained since the temporal activity is low; thus, a better DSVC-HEVC compression efficiency can be achieved when compared with SHVC-InterBL_IntraEL. The coding gains mainly come from the reduction of the coded EL residue at the EL encoder and the novel coding tools such as the SI creation and ACM.
4) *DSVC-HEVC Versus SHVC-InterBL_InterEL:* As expected, the DSVC-HEVC RD performance is lower than SHVC-InterBL_InterEL where the EL encoder exploits not only the spatial correlation with the many HEVC Intra-prediction modes but also the TC with past and future EL references to create the minimum EL residue. However, this benchmark cannot be practically used for emerging applications such as video surveillance, visual sensor networks, and remote space transmission for two main reasons. First, the computational complexity associated with the ME and compensation processes in the EL Inter encoding is significantly high, making it inappropriate for applications where the power resources are limited. Second, this solution is rather sensitive to error propagation and drift, especially in error-prone environments, since the EL encoder exploits the TC between frames to minimize the EL residue.
5) *BL QP Impact:* The BL QP also impacts the DSVC-HEVC coding gains regarding relevant benchmarks. The larger improvements can be obtained for the higher BL QP as, when the BL quality decreases, the quality gap between the created SI (mainly using the EL information) and the BL-decoded frame is higher. When this quality gap is higher, the decoder SIR will have higher correlation with the encoder EL residue and a better compression efficiency is achieved.

TABLE VI
PROFILING PLATFORM FOR ENCODING COMPLEXITY ANALYSIS

Processor	Inter (R) Core (TM) i7-4790 CPU @3.6 GHz
Memory (RAM)	16.0 GB
Operating system	64-bit Window 7 Enterprise SP1
Compiler	Visual Studio C++ Compiler version 10.0

E. Encoding Complexity Analysis

As the proposed DSVC-HEVC codec claims low encoding complexity as one of its major features, this paper would not be complete without an encoding complexity assessment. Although not a perfect solution, the processing time is the most common complexity assessment approach adopted in [42]–[44]. Therefore, the encoding complexity will be assessed here by the encoding time for the full sequence, in seconds, and under controlled simulation conditions. While the encoding time is highly dependent on the used hardware and software platforms, the conclusions will be derived from their relative comparison and not from their absolute values. The overall encoding time (including the BL HEVC Inter-encoding time and the EL key and WZ frames encoding times) is the metric used since the various layers do not correspond to independent codecs and the whole codec must be assessed. The profiling platform to obtain the encoding complexity analysis in this paper is specified in Table VI. To enable accurate time measurements, nothing else was running when collecting the performance results. Under these conditions, the complexity results have a rather solid relative and comparative value, thus allowing to compare the proposed DSVC-HEVC encoding complexity with the most relevant benchmarks defined earlier.

To compare the DSVC-HEVC encoding complexity with the relevant benchmarks, the percentage of encoding time reduction (ETR) is measured as follows:

$$\mathrm{ETR}_{\mathrm{Benchmark}} = \frac{(\mathrm{EncTime}_{\mathrm{DSVC\text{-}HEVC}} - \mathrm{EncTime}_{\mathrm{Benchmark}}) \times 10\,000}{\mathrm{EncTime}_{\mathrm{Benchmark}}}. \quad (20)$$

Tables VII and VIII show the DSVC-HEVC encoding complexity reduction regarding relevant benchmarks.

From the results, the following conclusions may be derived.

1) *DSVC-HEVC Versus SHVC-InterBL_IntraEL:* The DSVC-HEVC encoding complexity is lower than for SHVC-InterBL_IntraEL with an average ETR of about 13.6%. Important for this reduction is the DSVC-HEVC EL residue creation process, which is just the difference between the original and the BL-decoded data. For SHVC-InterBL_IntraEL, the EL residue may be the difference between the original and the EL predicted data where a highly complex RDO process needs to be performed to find the optimal prediction among the 35 directional Intra-prediction modes [1], [5].
2) *DSVC-HEVC Versus SHVC-InterBL_InterEL:* The DSVC-HEVC encoding complexity is much lower than

TABLE VII
ETR [%] REGARDING SHVC-INTERBL_INTRAEL BENCHMARK

$\{QP_B; QP_E\}$	Test sequences				Avg.
	BlowB	BaskP	Flow	BQS	
{34; 30}	-11.95	-10.51	-3.57	-12.34	-9.59
{34; 28}	-11.33	-9.85	-1.47	-11.59	-8.56
{34; 26}	-10.35	3.34	-2.48	-11.17	-5.17
{34; 24}	-10.20	3.69	-2.44	-10.28	-4.81
{30; 26}	-23.09	-7.00	-25.98	-27.77	-20.96
{30; 24}	-22.03	-6.68	-25.31	-27.17	-20.30
{30; 22}	-21.71	-4.91	-25.55	-27.15	-19.83
{30; 20}	-21.33	-4.86	-24.91	-26.13	-19.31
Seq. Avg.	**-16.50**	**-4.60**	**-13.96**	**-19.20**	***-13.57***

TABLE VIII
ETR [%] REGARDING SHVC-INTERBL_INTEREL BENCHMARK

$\{QP_B; QP_E\}$	Test sequences				Avg.
	BlowB	BaskP	Flow	BQS	
{34; 30}	-43.01	-42.44	-33.55	-30.94	-37.49
{34; 28}	-43.86	-42.70	-34.32	-32.93	-38.45
{34; 26}	-43.30	-43.09	-35.16	-34.81	-39.09
{34; 24}	-43.99	-43.57	-35.51	-35.28	-39.59
{30; 26}	-36.22	-28.05	-50.95	-51.43	-41.66
{30; 24}	-36.50	-28.48	-51.50	-52.00	-42.12
{30; 22}	-36.77	-27.74	-52.07	-52.24	-42.21
{30; 20}	-36.93	-28.02	-52.33	-51.90	-42.30
Seq. Avg.	**-40.07**	**-35.51**	**-43.17**	**-42.69**	***-40.36***

for SHVC-InterBL_InterEL with an average encoding complexity reduction of around 40.4%. Compared with the simple EL residue creation mechanism in DSVC-HEVC, the EL residue creation in SHVC-InterBL_InterEL is much more complex as it needs to perform a highly complex RDO process to find the optimal prediction among the 35 directional Intra modes, eight Inter-partition modes with 1/4 pel ME, and five Merge mode candidates [1]. Moreover, as the SHVC-InterBL_InterEL EL residue is highly dependent on the EL reference frames, errors may propagate and drift (loss of synchronization) between the encoder and decoder may occur.

As DSVC-HEVC does not require a feedback channel, no iterative Slepian–Wolf decoding is performed at the EL decoder; thus, the EL decoding complexity is not very high as it happens for most DVC decoders using a feedback channel [7], [25].

VII. CONCLUSION

This paper has presented a novel SVC solution, DSVC-HEVC, which is backward compatible with the HEVC Inter standard in the BL while adopting a DVC approach for ELs. This coding structure guarantees that a low EL encoding complexity can be achieved while temporal error propagation and drift can be avoided. Furthermore, this paper proposed two novel EL DVC coding tools: 1) a machine learning-based SI creation mechanism and 2) an ACM solution. The proposed DSVC-HEVC solution requires a lower encoding complexity than the relevant alternative coding solutions, while providing a better compression efficiency. As future work, the proposed DSVC-HEVC framework can be extended to support spatial scalability and the SI creation mechanism can be further improved by including a more powerful machine learning technique to perform the SI selection.

REFERENCES

[1] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[2] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[3] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[4] *Joint Call for Proposals on Scalable Video Coding Extensions of High Efficiency Video Coding (HEVC)*, document N12957, ISO/IEC JTC 1/SC 29/WG 11 and ITU-T SG16 WP3, Stockholm, Sweden, Jul. 2012.

[5] J. M. Boyce, Y. Yan, J. Chen, and A. K. Ramasubramonian, "Overview of SHVC: Scalable extensions of the High Efficiency Video Coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 20–34, Jan. 2016.

[6] J. Vanne, M. Viitanen, T. D. Hamalainen, and A. Hallapuro, "Comparative rate-distortion-complexity analysis of HEVC and AVC video codecs," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1885–1898, Dec. 2012.

[7] B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. IEEE*, vol. 93, no. 1, pp. 71–83, Jan. 2005.

[8] R. Puri and K. Ramchandran, "PRISM: A new robust video coding architecture based on distributed compression principles," in *Proc. 40th Allerton Conf. Commun., Control Comput.*, Urbana–Champaign, IL, USA, Oct. 2002.

[9] S. R. Gunn, "Support vector machines for classification and regression," Dept. Electron. Comput. Sci., Univ. Southampton, Southampton, U.K., Tech. Rep., May 1998.

[10] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, Mar. 2000.

[11] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.

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

[13] M. Tagliasacchi, A. Majumdar, and K. Ramchandran, "A distributed source coding based robust spatio-temporal scalable video codec," in *Proc. Picture Coding Symp.*, San Francisco, CA, USA, Dec. 2004.

[14] A. Sehgal, A. Jagmohan, and N. Ahuja, "Scalable video coding using Wyner–Ziv codes," in *Proc. Picture Coding Symp.*, San Francisco, CA, USA, Dec. 2004.

[15] H. Wang, N.-M. Cheung, and A. Ortega, "A framework for adaptive scalable video coding using Wyner–Ziv techniques," *EURASIP J. Appl. Signal Process.*, vol. 2006, p. 267, Jan. 2006.

[16] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 301–317, Mar. 2001.

[17] Q. Xu and Z. Xiong, "Layered Wyner–Ziv video coding," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3791–3803, Dec. 2006.

[18] K. Sakomizu, T. Nishi, and T. Onoye, "A hierarchical motion smoothing for distributed scalable video coding," in *Proc. Picture Coding Symp.*, Krakow, Poland, May 2012, pp. 209–212.

[19] G. Petrazzuoli, C. Macovei, I.-E. Nicolae, M. Cagnazzo, F. Dufaux, and B. Pesquet-Popescu, "Versatile layered depth video coding based on distributed video coding," in *Proc. WIAMIS*, Paris, France, Jul. 2013.

[20] X. HoangVan, J. Ascenso, and F. Pereira, "HEVC backward compatible scalability: A low encoding complexity distributed video coding based approach," *Signal Process., Image Commun.*, vol. 33, no. 4, pp. 51–70, Apr. 2015.

[21] J. Ascenso, C. Brites, and F. Pereira, "Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding," in *Proc. 5th EURASIP Conf. Speech Image Process. Multimedia Commun. Services*, Smolenice, Slovakia, Jul. 2005, pp. 1–6.
[22] A. Tomé and F. Pereira, "Low delay distributed video coding with refined side information," *Signal Process., Image Commun.*, vol. 26, nos. 4–5, pp. 220–235, Apr. 2001.
[23] I. H. Tseng and A. Ortega, "Motion estimation at the decoder using maximum likelihood techniques for distributed video coding," in *Proc. Conf. Rec. 39th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Oct./Nov. 2005, pp. 756–760.
[24] A. Aaron, S. Rane, and B. Girod, "Wyner–Ziv video coding with hash-based motion compensation at the receiver," in *Proc. IEEE ICIP*, Singapore, Oct. 2004, pp. 3097–3100.
[25] R. Martins, C. Brites, J. Ascenso, and F. Pereira, "Refining side information for improved transform domain Wyner–Ziv video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 9, pp. 1327–1341, Sep. 2009.
[26] A. Abou-Elailah, F. Dufaux, J. Farah, and M. Cagnazzo, "Fusion of global and local side information using support vector machine in transform-domain DVC," in *Proc. 20th EUSIPCO*, Bucharest, Romania, Aug. 2012, pp. 574–578.
[27] C. Brites, J. Ascenso, and F. Pereira, "Side information creation for efficient Wyner–Ziv video coding: Classifying and reviewing," *Signal Process., Image Commun.*, vol. 28, no. 7, pp. 689–726, Aug. 2013.
[28] S. Milani and G. Calvagno, "A distributed video coder based on the H.264/AVC standard," in *Proc. 15th EUSIPCO*, Poznań, Poland, Sep. 2007, pp. 673–677.
[29] S. Milani, J. Wang, and K. Ramchandran, "Achieving H.264-like compression efficiency with distributed video coding," *Proc. SPIE*, vol. 6508, Jan. 2007, Art. no. 65082Z.
[30] X. HoangVan, J. Ascenso, and F. Pereira, "Correlation modeling for a distributed scalable video codec based on the HEVC standard," in *Proc. IEEE 16th Int. Workshop MMSP*, Jakarta, Indonesia, Sep. 2014, pp. 1–6.
[31] X. HoangVan, J. Ascenso, and F. Pereira, "Optimal reconstruction for a HEVC backward compatible distributed scalable video codec," in *Proc. IEEE VCIP Conf.*, Valletta, Malta, Dec. 2014, pp. 193–196.
[32] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. IEEE 6th Int. Conf. Comput. Vis.*, Mumbai, India, Jan. 1998, pp. 839–846.
[33] C. K. Chiang, W. H. Pan, C. Hwang, S. S. Zhuang, and S. H. Lai, "Fast H.264 encoding based on statistical learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 9, pp. 1304–1315, Sep. 2011.
[34] X. Shen and L. Yu, "CU splitting early termination based on weighted SVM," *EURASIP J. Image Video Process.*, vol. 2013, pp. 1–11, Jan. 2013.
[35] V. Sindhwani, P. Bhattacharya, and S. Rakshit, "Information theoretic feature crediting in multiclass support vector machines," in *Proc. SIAM Int. Conf. Data Mining*, Philadelphia, PA, USA, Dec. 2001, pp. 1–18.
[36] T. Joachims. (Aug. 14, 2008). Support vector machine: SVM-light. Cornell Univ., accessed on Jan. 31, 2015. [Online]. Available: http://svmlight.joachims.org/
[37] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.
[38] *Video Test Sequences*, accessed on Jan. 15, 2015. [Online]. Available: ftp://hevc@ftp. tnt.uni-hannover.de/testsequences/
[39] *HEVC Reference Software*, accessed on Jan. 15, 2015. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/
[40] *SHVC Reference Software*, accessed on Jan. 15, 2015. [Online]. Available: https://hevc.hhi. fraunhofer.de/svn/svn_SHVCSoftware/
[41] G. Bjøntegaard, *Calculation of Average PSNR Differences Between RD-Curves*, document VCEG-M33, 13th ITU-T VCEG Meeting, Austin, TX, USA, Apr. 2001.
[42] W. Kim, J. You, and J. Jeong, "Complexity control strategy for real-time H.264/AVC encoder," *IEEE Trans. Consum. Electron.*, vol. 56, no. 2, pp. 1137–1143, May 2010.
[43] F. Bossen, B. Bross, K. Sühring, and D. Flynn, "HEVC complexity and implementation analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1685–1696, Dec. 2012.
[44] G. Corrêa, P. Assunção, L. Agostini, and L. A. da Silva Cruz, "Performance and computational complexity assessment of high-efficiency video encoders," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1899–1909, Dec. 2012.

Xiem HoangVan received the B.Sc. degree from the Hanoi University of Science and Technology, Hanoi, Vietnam, in 2009, the M.Sc. degree from Sungkyunkwan University, Seoul, South Korea, in 2011, and the Ph.D. degree from the Instituto Superior Técnico–Universidade de Lisboa, Lisbon, Portugal, in 2015, all in electrical and computer engineering.

He is currently an Assistant Professor with the Faculty of Electronics and Telecommunication, Vietnam National University–University of Engineering and Technology, Hanoi. His current research interests include image, video processing, and coding.

Dr. HoangVan received several awards for his work on video coding, notably the Picture Coding Symposium 2015 Best Paper Award and the Fraunhofer Portugal Challenge 2015 Ph.D. Award. He has contributed more than 20 papers on video coding and has been an active reviewer for many reputed journals and conferences.

João Ascenso received the E.E., M.Sc., and Ph.D. degrees from the Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisbon, Portugal, in 1999, 2003, and 2010, respectively, all in electrical and computer engineering.

He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, Instituto Superior Técnico, and the Multimedia Signal Processing Group, Instituto de Telecomunicações, Lisbon. He coordinates the IT participation in several national and international research projects in the field of multimedia signal processing and communications. He has authored over 70 papers in international conference and journals and has more than 2000 citations over 35 papers. His current research interests include video coding, 3D imaging, indexing and searching of audio–visual content, multimedia communication systems, visual sensor networks, and social media processing.

Dr. Ascenso is an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS and acts as a member of the organizing committees of known international conferences, such as European Signal Processing Conference 2014, the IEEE International Conference on Multimedia and Expo (ICME) 2015, and International Conference on Quality of Multimedia Experience 2016. He served as a Technical Program Committee Member and Reviewer for several widely know conferences in the multimedia signal processing field, such as International Conference on Image Processing, International Conference on Acoustics, Speech, and Signal Processing, International Workshop on Multimedia Signal Processing, and ICME.

Fernando Pereira (F'08) received the B.S., M.Sc., and Ph.D. degrees from the Instituto Superior Técnico (IST), Universidade Técnica de Lisboa, Lisbon, Portugal, in 1985, 1988, and 1991, respectively, all in electrical and computer engineering.

He is currently a Professor with the Electrical and Computer Engineering Department, IST, Instituto de Telecomunicações, Lisbon, where he is responsible for the participation of IST in many national and international research projects. He acts often as a Project Evaluator and an Auditor for various organizations. He has authored over 200 papers. His current research interests include video analysis, processing, coding and description, and interactive multimedia services.

Dr. Pereira is or has been a member of the IEEE Signal Processing Society Image Technical Committees on Video and Multidimensional Signal Processing, Multimedia Signal Processing Technical Committees, the IEEE Circuits and Systems Society Technical Committees on Visual Signal Processing and Communications, and Multimedia Systems and Applications Technical Committees. He has been a member of the scientific and program committees of many international conferences. He is or has been an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON MULTIMEDIA, and the *IEEE Signal Processing Magazine*. He was an IEEE Distinguished Lecturer in 2005. He is an Area Editor of *Signal Processing: Image Communication*. He has been participating in the work of ISO/MPEG for many years, notably as the Head of the Portuguese Delegation and the Chairman of the MPEG Requirements Group, and chairing many ad hoc groups related to the MPEG-4 and MPEG-7 standards.