

# Frequency Domain Intra Pattern Copy for JPEG XS Screen Content Coding

Yao Li<sup>1</sup>, Graduate Student Member, IEEE, Zhuoyuan Li<sup>1</sup>, Graduate Student Member, IEEE, Dong Liu<sup>1</sup>, Senior Member, IEEE, and Li Li<sup>1</sup>, Senior Member, IEEE

**Abstract**—JPEG XS is a wavelet-based lightweight image coding standard that features low-complexity and low-latency. As currently there are no efficient intra-compensation prediction techniques conforming to these features, we propose a frequency domain intra-copy prediction framework named Intra Pattern Copy (IPC), to improve its coding efficiency on screen content. In IPC, prediction methods that leverage the diverse decomposed patterns of two-dimensional wavelets, including the directional and frequency characteristics, are proposed to achieve efficient predictions under low-complexity and low-latency constraints. Specifically, we perform in-band compensation predictions in a multi-band synchronized approach, with coefficients of similar pattern distributions predicted simultaneously. A coefficient grouping scheme is derived from the band characteristics to facilitate this compensation process. Based on the grouping scheme, a multi-band synchronized side information coding method is also proposed to code the pattern offset vector of coefficients. Moreover, pattern search schemes incorporating strict limitations on the search range and prediction block size are further developed. Simulation results on JPEG XS demonstrate that an average improvement of 0.75 dB and 1.99 dB in BD-PSNR can be achieved on screen content for two different wavelet decomposition configurations, respectively, with a moderate increase in complexity.

**Index Terms**—Image compression, low-latency compression, mezzanine compression, JPEG XS, screen content coding.

## I. INTRODUCTION

The growing demand for higher resolution, higher frame rate, and wider colour gamut in video applications is leading to a consistently increasing bandwidth requirement, posing challenges for video storage and near-end video transmission in video production. The vast majority of the video link used in the broadcast industry, the serial data interface (SDI), has evolved from HD-SDI (1.485 Gbps), 3G-SDI to 6G-SDI and 12G-SDI [1], to meet this growing bandwidth requirement. However, such cables are still insufficient for transmitting uncompressed videos with very high specifications (e.g., 8K60fps), and the 10 Gbps-Ethernet links used in emerging

video-over-IP protocols are similarly insufficient. Additionally, frequent access to massive data could also cause disk access time to a bottleneck in the video production workflow. Therefore, mezzanine compression has been developed to help bridge the gap between the very high data rate of uncompressed videos and the constrained bandwidth. Unlike video codecs deployed to video distribution, e.g., H.265/HEVC [2], H.266/VVC [3], AV1 [4] codecs, a mezzanine codec aims to achieve lightweight video encoding with a moderate compression rate while safeguarding the advantages of the uncompressed streams, including visually lossless quality, low-complexity, low-latency, multi-platform interoperability, and energy-efficiency.

Oriented for different applications, several mezzanine compression standards have been developed. Targeting the display interface compression, the Video Electronics Standards Association (VESA) has standardized a series of mezzanine compression standards, including DSC 1.1 [5], DSC 1.2a [6], and VDC-M [7]. These standards have been adopted across the major display interface standards such as HDMI, DisplayPort, and MIPI, backing the user experience for immersive display. A Perceptual Lossless Compression (PLL) standard [8] has also been developed by the Audio Video Coding Standard Workgroup of China (AVS) for similar application scenarios. As for Pro-AV markets, several standards and proprietary formats have also been developed, such as VC-2 [9], ProRes [10] by Apple, LLVC [11] by Sony, and TICO [12] by intoPIX. These formats favour lower logic implementation compared to display stream compression ones and are primarily designed for the film and video production industry. On the requirement of a cross-platform, low-complexity, and low-latency mezzanine compression system, the Joint Photographic Experts Group (JPEG) Committee standardized a novel compression standard called JPEG XS (where XS stands for *Extra Speed*) [13]. It aims to offer visually lossless image quality with multi-platform interoperability, allowing efficient implementation on various platforms, including CPU, GPU, FPGA and ASIC. A scalable end-to-end latency, ranging from just a few tens of lines to below one line, also makes JPEG XS well-suited for a variety of low-latency scenarios, including live video production [14]–[16], video streaming [17], [18], and real-time machine tasks [19], [20]. A third edition of JPEG XS was published by the JPEG committee recently [21]. It introduces a Temporal Differential Coding (TDC) profile, which performs wavelet domain temporal prediction to exploit the redundancy between adjacent frames.

The development of the third edition of JPEG XS primarily

Manuscript received 17 January 2025; revised 30 March 2025; accepted 15 April 2025. This work was supported in part by the National Key Research and Development Program of China under Grant 2024YFF0505702 and in part by the Natural Science Foundation of China under Grant 62021001. It was also supported by the CPU cluster built by MCC Lab of Information Science and Technology Institution, USTC. This article was recommended by Associate Editor Dr. Kai Zhang. (Corresponding author: Li Li.)

The authors are with the MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition, University of Science and Technology of China, Hefei 230027, China. Li Li is also with the IAI-ATCDI Joint R&D Center (Laboratory) on Intelligent Technologies for Comprehensive Transportation, Institute of Artificial Intelligence, Hefei Comprehensive National Science Center (e-mail: mrliyao@mail.ustc.edu.cn; zhuoyuanli@mail.ustc.edu.cn; dongliu@ustc.edu.cn; lilil@ustc.edu.cn).

Copyright ©2025 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must

be obtained from the IEEE by sending an email to [pubs\\_permissions@ieee.org](mailto:pubs_permissions@ieee.org).  
Authorized licensed use limited to: University of Science & Technology of China. Downloaded on April 22, 2025 at 03:08:35 UTC from IEEE Xplore. Restrictions apply.  
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted,

but republication/redistribution requires IEEE permission. See <https://www.ieee.org/publications/rights/index.html> for more information.

aims to improve its coding efficiency on screen content and extend its applications to keyboard-video-mouse (KVM) and remote desktop markets [22]. Considering the advantages of intra-picture coding setting for low-complexity and low-latency encoding scenarios, it is essential to develop effective intra-coding techniques for screen content. For instance, the intra-picture coding setting is capable of temporal random access, which is very crucial for latency-sensitive scenarios. Additionally, the relatively lower memory cost of intra-picture coding tools supports more cost-effective implementations of JPEG XS on memory-cost-sensitive platforms.

Given the prevalence of abundant repetitive structures in screen content images, applying pixel domain intra-compensation prediction can effectively reduce the structural redundancy and enhance compression efficiency. This principle underlies the highly efficient intra block copy (IBC) coding tool in advanced video coding standards such as HEVC [23] and VVC [24]. However, the requirement for a fine-grained prediction unit and an associated exhaustive search scheme to achieve efficient predictions limits its effectiveness in scenarios with strict complexity constraints. This limitation leads us to explore methods for achieving efficient predictions using a relatively coarse prediction unit setting. In the wavelet domain, images are decomposed into bands with different frequency components and varying texture directions. This decomposition enables targeted predictions for various patterns across different bands, facilitating effective predictions while adhering to low-complexity constraints.

Based on this insight, we propose to perform frequency domain (wavelet domain) intra-copy predictions to exploit the correlation of repeated patterns. This leads to a novel multi-band synchronized intra-compensation prediction framework. In the intra pattern copy (IPC) framework, wavelet coefficients from different bands are grouped in both frequency and spatial dimensions. Each coefficient group of a specific pattern type is predicted by reconstructed coefficients from encoded regions, with the indication of a searched pattern offset vector (PV). Additionally, the search range and prediction unit size are customized to conform to the low-complexity and low-latency requirements. In conclusion, the contributions of the paper can be summarized as follows:

- We propose a frequency domain intra-copy prediction framework to explore an effective low-complexity predictive coding framework for screen content.
- We design an effective wavelet coefficient grouping scheme that leverages correlations among coefficients at varying decomposition levels and directions to enhance prediction efficiency.
- We present efficient wavelet prediction methods that align with the low-complexity codec design and the wavelet coefficient grouping schemes. Associated encoder optimization schemes achieve a good trade-off between complexity and performance.
- An average improvement of 0.75 dB and 1.99 dB in BD-PSNR for screen content is achieved on JPEG XS under two different wavelet decomposition configurations, respectively.

The rest of the paper is organized as follows: Section II reviews the related work. Section III gives a basic idea of JPEG XS and the TDC profile. Section IV presents the motivation of our proposed IPC framework. Section V describes the IPC framework in detail. Section VI shows the experimental results and analysis. Section VII concludes this paper.

## II. RELATED WORK

In this section, we review relevant previous work from two aspects. First, we introduce the fundamentals of the pixel domain IBC technology and its development in video coding standards. Second, we introduce some methods of the in-band compensation predictions.

### A. Intra Block Copy

IBC is a prediction technology that takes advantage of the abundance of repetitive patterns in computer-generated content. In IBC, prediction is performed on a block basis with the references retrieved from the already reconstructed regions within the current frame. Block vectors (BV) are signaled to indicate the displacement between the current block and the reference block. This mechanism is similar to the block matching motion compensation (MC) techniques employed in inter-prediction, albeit with the restriction that available references are confined within the current frame.

IBC was first proposed and studied during the standardization of H.264/AVC [25], where a notable coding performance gain was reported on nature content sequences. Further, in HEVC extensions on SCC, IBC was first formally included as a coding tool for screen content coding [23]. Its implementation mirrors the HEVC inter-frame motion compensation prediction, as the same set of syntax structures for motion vector coding and residual coding are used. The available reference area of IBC in HEVC SCC is nearly the entire reconstructed portion of the current picture. This reference design may pose challenges for hardware implementation, as writing the unfiltered reconstructed picture to off-chip memory demands additional memory access bandwidth. Additionally, the potential need to refer to a just reconstructed neighbouring region may also pose tight memory writing and reading within a limited number of cycles during the decoding process [26].

Such issues are resolved in VVC, by constraining the available search range to a local area, i.e., the current coding tree unit (CTU) [27], so that the reference pixels for IBC mode can be stored in a high-speed on-chip memory with a size of  $128 \times 128$  pixels. Reference sample memory (RSM) reuse mechanism [28] was further proposed to improve the utilization of reference memory, where the available search range gradually transitions from the left adjacent CTU to the already reconstructed portion of the current CTU, aiming to provide an as relevant reference as possible. Fast algorithms have also been proposed to alleviate the computational complexity of the IBC exhaustive search, including techniques such as early CU splitting, early mode skipping [29]–[31], and hash-based motion estimation (ME) [32], [33]. However, fine-grained search blocks and a certain search range are still essential for achieving good performance in IBC for SCC.

### B. In-Band Compensation Prediction

In-band compensation prediction schemes that perform wavelet domain motion estimation and compensation have been widely studied to reduce the temporal redundancy in wavelet-based video codecs. Initial research can be traced to the multi-resolution motion estimation (MRME) strategy in [34], which employs a coarse-to-fine ME scheme on the wavelet decomposition pyramid to obtain high-precision motion vectors and predicted residuals gradually. Subsequent work has been proposed to enhance the MRME scheme. Variable block sizes have been introduced to regulate motion vectors at different resolutions and improve the accuracy [35], [36]. All-orientation bands combined search has also been explored to reduce search complexity [37]. To ease the aperture problem and the error drifting effects in low-pass band search, several methods have also been proposed, including two-step search [38], motion vector (MV) prediction through median filtering [39], multi-candidate MV prediction [40], etc.

Another fundamental issue in wavelet-based motion compensation is the shift-variant property of discrete wavelet transforms (DWT). The decimation operation in DWT will cause frequency aliasing, resulting in coefficients shifted by odd pixels cannot be predicted. This phenomenon was discussed in [41], [42] and research has been studied to overcome this problem. [42], [43] proposed removing the decimation process in transform to prevent aliasing and facilitate motion estimation and compensation. Optimized interpolation filters have also been derived and applied to the low-pass bands of both the current frame and reference frame in [44], [45], to reduce the aliasing from a theoretical analysis. Further, in [46], a low-band-shift (LBS) method was proposed. The basic idea is to shift the low-pass band of each decomposition level along horizontal, vertical and diagonal directions to complement the all-phase bands for the reference frame. Other techniques have also been studied to address this issue, including adaptations to DWT such as complex wavelet [47], [48] and over-complete wavelet [49], [50].

### III. PRELIMINARY

JPEG XS takes a classical wavelet-based image coding paradigm. Input image lines undergo a reversible color transform (RCT), DWT, quantization, and entropy coding. JPEG XS adopts an asymmetrical transform, typically using 5 horizontal and 2 vertical decomposition levels with the LeGall 5/3 kernel in High profiles. The lower vertical decomposition level requires less pixel line buffering and could ensure the end-to-end latency of at most 32 lines [51]. Wavelet coefficients are then partitioned into precincts, which consist of coefficients from all the sub-bands that form a spatial region of the image. Precincts serve as the basic coding unit, aligning with the horizontal stripes that comprise a few lines in the image. They are further grouped into slices, which act as independent decoding units, enabling the JPEG XS error recovery, spatial random access, and coarse-grained parallel processing.

In the rate control module, truncation positions for each band in the precinct are derived from two precinct-level parameters, the base quantization parameter  $Q_p$ , and the band

refinement number  $R_p$ , along with two pre-defined arrays, the band gain table  $G$  and the band priority table  $P$ :

$$T_{p,b} = Q_p - G[b] - r_{p,b}, \quad (1)$$

where the refinement factor  $r_{p,b}$  is 1 for  $P[b] < R_p$  and otherwise 0.  $T_{p,b}$  represents the truncation position for band  $b$  in precinct  $p$ . The band gain table  $G$  reveals the importance of each band, while the priority table  $P$  indicates the priority for obtaining the available additional one-bit refinement for each band.  $Q$  and  $R$  are determined at the encoder to meet the bit budget requirement and are signaled. The entropy coder of JPEG XS takes a semi-fixed-length approach. Each group of four adjacent coefficients is combined into a coding group, and the maximum most significant bit (MSB) position is extracted to determine the coding length for all coefficients within the group.

In the third edition of JPEG XS, TDC profile is introduced to perform temporal prediction between adjacent frames. A wavelet coefficient frame buffer is incorporated and the encoder can choose to encode the temporal differential residuals. In the inter predictive coding, the encoder runs in a closed prediction loop fashion and the col-located position coefficients of the last frame are taken as the reference. Frame buffer compression is also considered for practical hardware implementation, as the frame buffer generally could not fit into the internal block RAM and the bandwidth to the external RAM is also limited. A conformance point, namely frame buffer level, is added to the third edition of JPEG XS to specify the frame buffer access bandwidth for both the encoder and the decoder.

### IV. MOTIVATION

IBC has been demonstrated as an efficient intra-coding tool for screen content coding in advanced coding standards [26], [52]. To improve the screen content coding efficiency for JPEG XS, an intuitive idea is to leverage this pixel domain IBC algorithm to eliminate the structural redundancy in the current frame. However, under the low-complexity constraints of JPEG XS, the prediction unit size can not be finely grained and the search range is also limited. These constraints significantly impact the performance of IBC. Additionally, the pixel domain predictive coding loop requires a reverse transform on the encoder side, which causes additional complexity and potentially higher latency for the JPEG XS encoder. These challenges inspire us to explore frequency domain (wavelet domain) prediction schemes, as wavelet domain analysis offers several advantageous features, especially in low-complexity and low-latency image coding scenarios. These advantageous features are also highlighted in earlier research on in-band compensation prediction [34]–[36], [53], [54].

- The time-frequency localization properties of DWT make it ideal for analyzing non-stationary image signals with varying characteristics at different regions.
- DWT provides a powerful approach for multiresolution image analysis. The low-pass band with down-sampled resolutions and blurred patterns could facilitate the matching and predictions for large blocks.

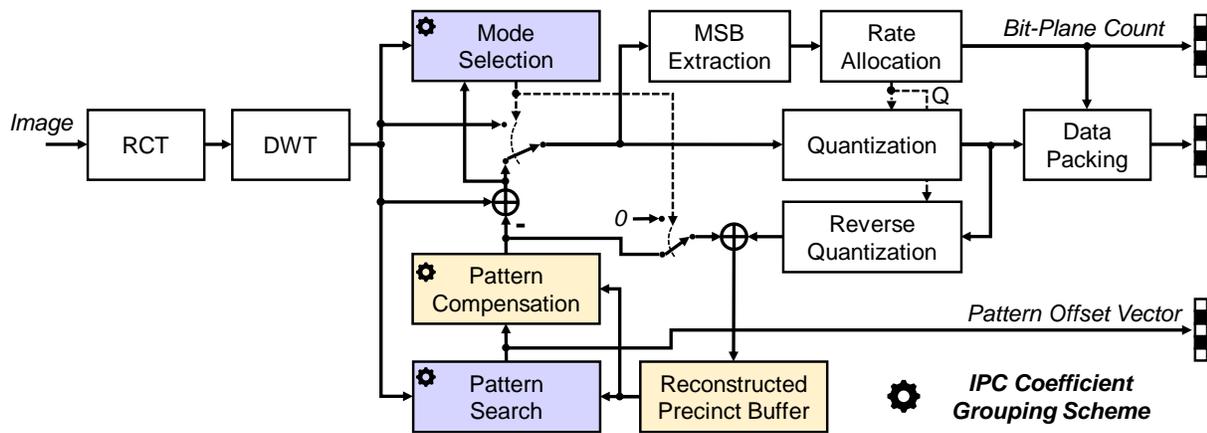


Fig. 1. Illustration of the JPEG XS encoder framework integrated with our proposed IPC framework. The coloured blocks represent newly introduced parts in IPC, with purple blocks further indicating encoder-only ones. Markers within these blocks denote that coefficient prediction and mode selection are performed under a specific IPC coefficient grouping scheme.

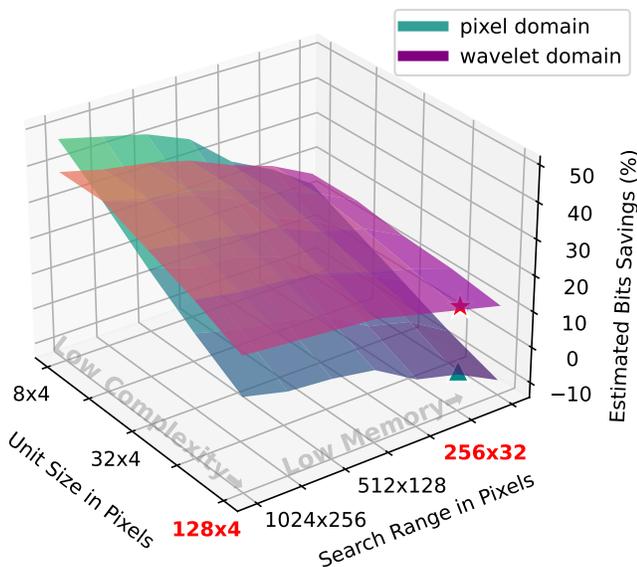


Fig. 2. Comparison of prediction efficiency between the pixel and wavelet domain prediction schemes under varying search ranges and prediction unit sizes. The wavelet domain scheme demonstrates superiority in settings with coarse unit size and limited search range settings, aligning with the low-complexity and low-memory cost requirements.

- DWT generates orthogonal patterns in different bands, which could guide the optimization of predictions for flat coding units and search ranges in low-latency scenarios.
- Wavelet domain prediction requires no encoder-side reverse transform and could save the complexity and latency.

Fig. 2 compares the prediction efficiency of pixel domain and wavelet domain prediction schemes for varying prediction unit sizes and search ranges. A pixel-domain spatial complexity feature from the video complexity analyzer (VCA) project [55] is used to estimate the coding bits for the pixel domain scheme, while the bitplane counts of half of the lower-frequency bands with higher importance on the reconstruction quality in JPEG XS are used for the wavelet domain. The savings in estimated coding bits of the prediction residuals relative to the original image are used to assess the predic-

tion efficiency. It can be observed that the wavelet domain prediction scheme outperforms the pixel domain approach for larger prediction unit sizes and limited search ranges, which aligns with low-complexity, low-memory-cost scenarios, as indicated by the markers in the figure. This insight motivated our proposed frequency domain IPC framework.

## V. METHOD

Fig. 1 illustrates the encoder framework of JPEG XS integrated with our proposed intra pattern copy (IPC) framework. Wavelet coefficients generated from the revertible colour transform (RCT) and DWT first undergo spatial predictions in the IPC predictive coding loop. Prediction residual signals, rather than the original data, are then sent into the subsequent rate control, quantization, and data packing modules. In the IPC prediction loop, there are mainly four parts: pattern search, pattern compensation, mode selection, and reconstructed precinct buffer. The input wavelet coefficients are first grouped by spatial positions and wavelet decomposition band indices. For each set of coefficients, prediction signals of the same size and number of band indices are retrieved from the reconstructed precinct buffer in the pattern compensation module. This retrieval involves directly copying reconstructed signals that exhibit similar patterns to the current coding signals. Residuals are then calculated by subtracting predictions from the original values. The encoder decides whether use the original value coding mode or the IPC mode (i.e., code the IPC prediction residuals) based on the estimated coding costs. A flag and a PV are signaled to indicate the mode selection and the reference retrieval offset. The pattern search module is employed at the encoder side to find the optimal PV from all available prediction candidates. Finally, encoded wavelet coefficients are reconstructed and buffered to provide references for the subsequent coding precincts.

The entire IPC coding process, from the pattern search, pattern compensation to the mode selection, performs on a specific coefficient unit basis, as defined by the *IPC coefficient grouping scheme* in the framework. Each prediction unit consists of coefficients from different bands that correspond to a

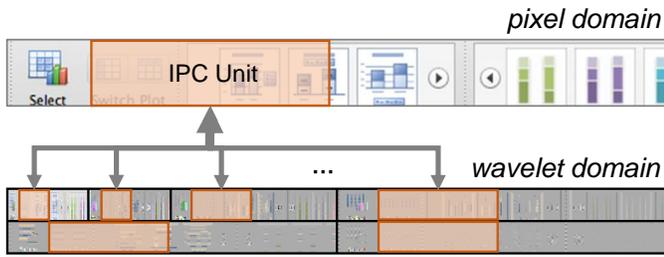


Fig. 3. Spatial mapping relationship between the pixel block and the set of wavelet coefficient blocks from various bands of an IPC Unit, illustrated with a 3 horizontal, 1 vertical decomposition.

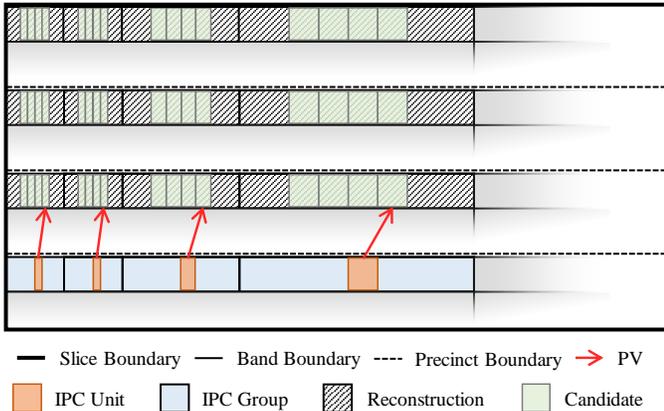


Fig. 4. Illustration of the multi-band synchronized pattern compensation process. Coefficient blocks from the same IPC Group share a PV to perform the compensation. Only several bands at higher decomposition levels are presented for conciseness and an example arrangement of reference candidates is also given.

specific spatial position. This spatial-frequency joint grouping scheme facilitates adaptive predictions for coefficients with varying characteristics at different spatial locations and decomposition bands, while ensuring the synchronization of multi-band associated predictions. We will detail this coefficient grouping scheme and the design of the IPC modules guided by this scheme in the following subsections.

### A. Pattern Compensation

In this part, we detail the cross-precinct intra pattern coefficient compensation process. This compensation process is performed in a multi-band synchronized manner according to the IPC coefficient grouping scheme. The grouping scheme divides coefficients along both spatial and frequency dimensions to facilitate the adaptive predictions. In the IPC prediction loop, two types of coefficient grouping units are introduced: the IPC Unit that performs the spatial partitioning and the IPC Group that performs the band type clustering. The intra pattern compensation process is performed on a combined basis of the IPC Unit and the IPC Group. Coefficients from different IPC Units or different IPC Groups will have their independent predictions. Moreover, due to the similar pattern characteristics of multiple bands existing in an IPC Group, coefficients from the bands within the same IPC Group and IPC Unit have shared prediction information. Fig. 3 illustrates the correspondence of a collection of coefficient blocks to be

predicted of the same IPC Unit. Different coefficient blocks within the same IPC Unit but from different bands may vary in size due to differences in decomposition levels and decimation factors. However, their relative positions within their respective bands remain consistent, as does the number of IPC Units across these bands. This alignment enables the prediction to be performed in a synchronized manner across various bands from the same IPC Group, and the reference coefficients can be retrieved as a collection of coefficient blocks from these bands.

The compensation process for the coefficient blocks of an IPC Unit is further illustrated in Fig. 4. Reference coefficients are retrieved from the co-indexed band of the reconstructed precincts in the current slice, under the indication of a PV. Following the frequency domain coefficient grouping scheme, coefficient blocks from the bands of the same IPC Group use a shared PV to conduct the compensation synchronously; specifically, their PV values are scaled according to the decomposition levels of the bands. Integer-coefficient resolution pattern compensation is performed in our proposed IPC, and we will further discuss the selection of PV resolution and its scaling relationship across different bands in the following sections.

### B. IPC Coefficient Grouping Scheme

The DWT used in JPEG XS separates the frequency components of an image into multiple bands while retaining the spatial information, enabling detailed analysis in both spatial and frequency domains. This decomposition results in distinct characteristics across various wavelet bands and different image areas, making an appropriate coefficient grouping scheme crucial for achieving efficient predictive coding. In this section, we will detail the derivation of the IPC Group, the frequency domain grouping scheme, and the configuration of the IPC Unit, which serves for the spatial coefficient grouping.

1) *IPC Group-On the Frequency Domain Pattern Variance:* Different bands of an image typically exhibit varying patterns and textures; consequently, their optimal reference patterns also differ. Fig. 5 illustrates such a variance on a JPEG XS screen content test image region. It can be observed that patterns with vertical orientation are extracted into the horizontal decomposition high-pass band, while patterns with horizontal orientation are extracted into the vertical decomposition high-pass band; these orthogonal patterns are not represented in the high-high band, as it captures the diagonal patterns. The variation in pattern orientation leads to differing levels of prediction difficulty among the various bands. Specifically, the flat shape of the JPEG XS precinct creates an arrangement of vertically adjacent reference precincts, allowing for nearby references and better prediction of vertical patterns that extend through both the current and reference precincts. In contrast, matching the horizontal patterns across different precincts is much more difficult. Additionally, such a variance is also present between the low-pass bands and the high-pass bands. The low-pass bands undergo low-pass filtering and coefficient decimation, where neighbouring pixel blending and smoothing are performed. This process results in the blurring of sharp

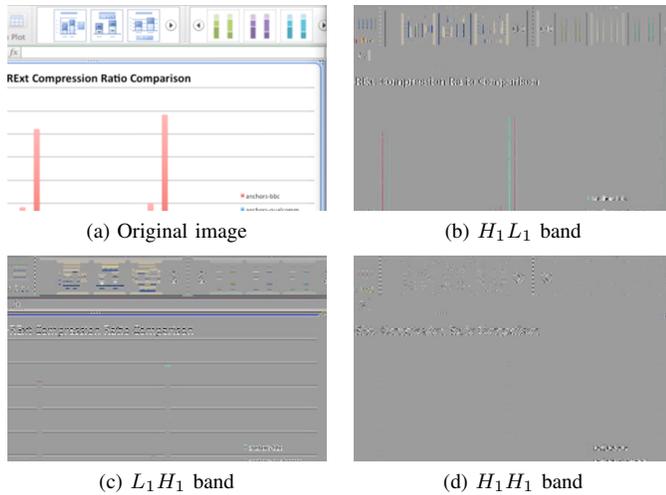


Fig. 5. Illustration of pattern variance across different bands.  $H_1L_1$  denotes the high-low band at the first decomposition level, and the rest follows. The horizontal decomposition extracts vertical patterns, and the vertical decomposition extracts horizontal patterns into their respective high-pass bands, while  $H_1H_1$  captures the diagonal details of the image.

edges in screen content images, thereby facilitating fuzzy matching and prediction of patterns and simplifying the search process. This aids in the prediction within a limited search range, where exact matching references for the patterns cannot be guaranteed.

The frequency domain coefficient grouping scheme, i.e., the band grouping strategy, should align with such a pattern distribution variance among the bands. In IPC, we employ k-means clustering to derive the IPC Group divisions. The clustering distance metric is based on the cosine similarity of the PV feature vectors from different bands, which comprise the independently searched PVs of all IPC Units within each band. This approach migrates the magnitude inconsistencies of the PVs across bands at different decomposition levels. In each iteration, the k-means clustering centroids are updated to the normalized average of the PV feature vectors of the bands in each cluster. Fig. 6 shows the clustering results of a 5 horizontal, 2 vertical decomposition case, with bands ordered from 0 to 9, traversing from the low-low frequency band at the last decomposition level to the high-high band at the first decomposition level. Independently searched PVs of all the IPC Units for each band are first flattened to vectors, and then used to calculate cosine similarity between different bands, providing the distance metric in clustering. Principal Component Analysis (PCA) is applied to extract two key features from this PV vector to visualize its distribution for different bands. Four clusters can be clearly observed from the scatter plot, as supported by the k-means clustering results. Denote  $H_1H_1$  as the high-high band at the 1st horizontal, 1st vertical decomposition level, and  $L_5L_2$  as the low-low band at the 5th horizontal, 2nd vertical decomposition level, etc. From the clustering results, we can identify four IPC Groups:  $\mathcal{G}_0 := \{L_5L_2\}$ ,  $\mathcal{G}_1 := \{H_5L_2, H_4L_2, H_3L_2, H_2L_2, H_1L_1\}$ ,  $\mathcal{G}_2 := \{L_2H_2, H_2H_2\}$ ,  $\mathcal{G}_3 := \{L_1H_1, H_1H_1\}$ , as illustrated in Fig. 7. The high-pass bands of the two vertical decomposition

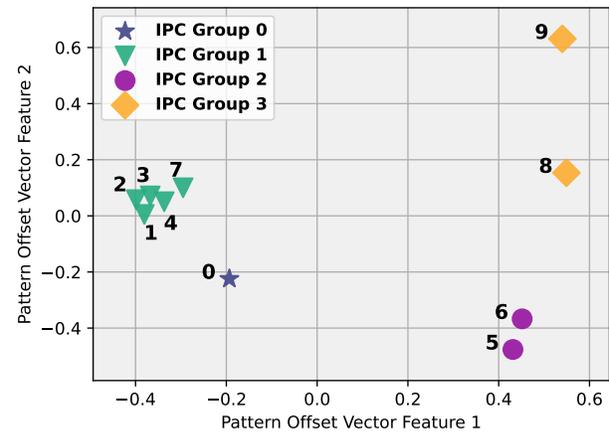


Fig. 6. Band clustering results based on their pattern offset vector (PV) distribution characteristics, under a 5 horizontal, 2 vertical decomposition. Bands are ordered from 0 to 9, traversing from  $L_5L_2$  to  $H_1H_1$  in a lower frequency band first-horizontal decomposition band first order.

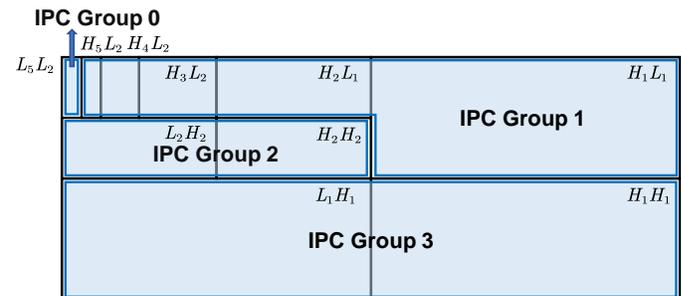


Fig. 7. Visualization of the IPC Group band division from a 5 horizontal, 2 vertical decomposition.

levels comprise two groups, respectively. Meanwhile, the remaining bands are further divided into two groups according to their horizontal decomposition frequency. Coefficients from different colour components but at the same decomposition level are also regarded as being in the same band.

2) *IPC Unit-On the Spatial Domain Pattern Variance*: The variance of the patterns and textures of different image areas is kept in the wavelet domain, and the IPC Unit is introduced to achieve the spatial content adaptive prediction. As JPEG XS is primarily tailored to low-latency compression applications, its basic coding unit, the precinct, is designed to have a flat shape, typically with a height of 4 pixels and a width equal to that of the image. The design of the IPC Unit dimensions also aligns with this insight, with a width of 128 pixels and a height of 4 pixels in the image. The coefficient blocks of the IPC Unit have sizes scaled according to the decomposition level, ranging from the size of  $64 \times 2$  at the first decomposition level to the size of  $4 \times 1$  at the final decomposition level, in a 5 horizontal and 2 vertical decomposition configuration. In this manner, the smallest coefficient block just aligns with the size of the JPEG XS entropy coding unit, and this large prediction unit design also eases the pattern search and pattern compensation complexity.

### C. Multi-Band Synchronized PV Signalling

Pattern predictions are retrieved from the reference precincts under the indication of PVs. Each IPC Group has its own PV for an IPC Unit, and the bands within the group utilize this shared PV to perform predictions in accordance with the IPC coefficient management scheme. This means that their predictions are synchronized, with the PV scaled according to the decomposition levels and band sizes. We designate an anchor band  $b_a$  from all the decomposition bands to establish this scaling relationship. The PV signalling of all IPC Groups is based on this anchor band, and the resolution of the signaled PV is also defined on this anchor band. For each PV in band  $b$ , its vertical component,  $PV_b^v$ , indicates the vertical index offset between the current and the reference precincts, and the horizontal component,  $PV_b^h$ , represents the position offset of the coefficient block within the current and reference precincts. The PV horizontal component in the anchor band has a resolution of one coefficient sample for references from the previously encoded precincts. Then, we can formulate the scaling relationship as follows:

$$\begin{cases} PV_b^v = PV_{\mathcal{G}_k}^v \\ PV_b^h = PV_{\mathcal{G}_k}^h \cdot 2^{(hlevel(\mathcal{G}_k) - hlevel(b))} \end{cases}, \quad (2)$$

where  $\mathcal{G}_k$  denotes the  $k$ -th IPC Group with  $k \in [0, 3]$ , and  $b \in \mathcal{G}_k$  represents a band type in the group;  $(PV_{\mathcal{G}_k}^h, PV_{\mathcal{G}_k}^v)$  and  $(PV_b^h, PV_b^v)$  represent the signaled PV value of the group and the actual offset value used to perform predictions for band  $b$ , respectively.  $hlevel(\cdot)$  indicates the horizontal decomposition level of the band type, with  $hlevel(\mathcal{G}_k)$  defined as that of the anchor band:

$$hlevel(\mathcal{G}_k) := hlevel(b_a). \quad (3)$$

Setting the anchor band  $b_a$  to a band from a low decomposition level will result in higher PV resolution and more accurate predictions. However, this also increases the PV signalling header bits and the complexity of pattern search. Here,  $b_a$  is set to  $L_5L_2$ , which is at the highest decomposition level. We will revisit and discuss this anchor band selection in Section VI-C. Note that predictions are performed only within the same band, ensuring that the spatial scalability feature of the DWT-based JPEG XS is not compromised. Issues arising from the shift-variant property of DWT, particularly in relation to the wavelet domain fractional-sample motion compensation [46], will also be discussed in Section V-E.

### D. In-Precinct Interweaved Pattern Compensation

The aforementioned cross-precinct pattern compensation process provides an efficient intra-predictive coding approach. However, special attention needs to be given to the prediction of horizontal patterns, as the flat-shaped precincts are less relevant in the horizontal direction. The reconstructed precincts located above may not provide sufficient relevant references. To accommodate with the issue, we propose the in-precinct pattern compensation scheme, which extracts a portion of the coefficient samples within the current precinct to serve as references for other samples in the same precinct.



Fig. 8. Illustration of the in-precinct interweaved pattern compensation. IPC Units with odd indices can reference those with even indices within the current precinct.

As shown in Fig. 8, the in-precinct pattern compensation is conducted in an interweaved manner; specifically, unquantized coefficients of the IPC Units with even indices are used to compensate those with odd indices. Long reference dependency and error accumulation can be mitigated in this interweaved approach. Consequently, a two-pass compensation process is performed, with the first pass applying regular cross-precinct compensation or direct coefficient reconstruction, supplying references for the in-precinct prediction in the second pass. Nonetheless, the overall complexity and latency are not significantly affected, as each pass processes only a subset of the IPC Units.

This in-precinct pattern compensation is applied when the parsed  $PV^v$  equals zero for those odd-indexed IPC Units. Simultaneously, the resolution of signaled  $PV^h$  is set to the coefficient block width of the IPC Unit in anchor band  $b_a$ . The encoded and reconstructed precincts are pushed into the reference buffer to provide references for the next precincts to be coded. The buffer will be reset when coding a new slice, and thus, a relatively small buffer size is required in IPC, without imposing significant implementation cost on the encoder. Algorithm 1 outlines the decoder procedure for all the precincts in an image. The cross-precinct compensation and the in-precinct interweaved compensation create two branches, which are selected for execution based on the parsed PV vertical component values.

### E. Low-Complexity Constrained Pattern Search

In this subsection, we introduce the search range setting and the PV searching metric of the IPC encoder, and discuss the impact of the shift-variant property of the DWT on the IPC.

1) *Low-Complexity Constrained Search Range*: In IPC framework, the search range has been carefully designed and limited to align with the low-complexity and low-latency features of JPEG XS, with only a few dozen candidates to be checked for each IPC Unit. In regular cross-precinct prediction, available references come from previously encoded and reconstructed precincts within the current slice. The first precinct has no available cross-precinct reference, while the last precinct has three, in a typical slice setting of four precincts. For each reference precinct, the available search range extends two IPC Units to the left and two to the right from the starting coefficient of the current IPC Unit to be predicted, constrained by the band boundaries. Meanwhile, the search step size on different bands also scales with the decomposition level, as in (2), ensuring that the number of available candidates remains consistent across all bands. The PV value resolution can be adjusted to configure a larger search range or a finer step size. Here, we set the step size on anchor band  $b_a$  (i.e.,  $L_5L_2$ ) to one coefficient sample. With

**Algorithm 1** Multi-Band Synchronized Pattern Compensation

**Input:** PVs and IPC mode flags of all the IPC Groups and IPC Units,  $\{PV_{p,u,k}\}, \{I_{p,u,k}\}$ ; Dequantized coefficients of all bands,  $\{c_{p,u,b}\}$

**Output:** Reconstructed coefficients of all bands,  $\{\hat{c}_{p,u,b}\}$

```

1: Initialize the Reconstructed Precinct Buffer,  $\{\mathcal{B}_0, \mathcal{B}_1\}$ 
2: Initialize the number of precincts in image,  $N_p$ , number of IPC Units in precinct,  $N_u$ , and the IPC Groups,  $\{\mathcal{G}_0, \mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3\}$ 
3: Initialize the anchor band index,  $b_a$ , and the IPC Unit width in the anchor band,  $w_a$ 
4: for  $p \in [0, N_p)$  do
5:   for  $iter \in [0, 2)$  do
6:     for  $k \in [0, 4)$ ,  $b \in \mathcal{G}_k$ ,  $u \in [0, N_u)$  do
7:       if  $I_{p,u,k} = False$  and  $iter = 0$  then
8:          $\hat{c}_{p,u,b} \leftarrow c_{p,u,b}$ 
9:       else
10:         $pv_b \leftarrow \text{scale\_pv}(PV_{p,u,k}, b_a, b)$  (2)
11:        if ( $pv_b^v = 0$  and  $iter = 0$ ) or ( $pv_b^v \neq 0$  and  $iter \neq 0$ ) then
12:          continue
13:        end if
14:        if  $iter = 0$  then
15:           $p_{ref} \leftarrow \text{fetch\_ref\_prec}(\mathcal{B}_0, pv_b^v)$ 
16:        else
17:           $p_{ref} \leftarrow \text{fetch\_ref\_prec}(\mathcal{B}_1, 0)$ 
18:           $pv_b^h \leftarrow pv_b^v \times w_a$ 
19:        end if
20:         $r_{p,u,b} \leftarrow \text{fetch\_ref\_val}(p_{ref}, u, b, pv_b^h)$ 
21:         $\hat{c}_{p,u,b} \leftarrow c_{p,u,b} + r_{p,u,b}$ 
22:      end if
23:    end for
24:  end for
25: return  $\{\hat{c}_{p,u,b}\}$ 

```

an IPC Unit width of four coefficients in this band, there are 16 candidates per IPC Unit on a reference precinct.

In the interleaved in-precinct prediction that is additionally available to the even-indexed IPC Units, references include eight reconstructed IPC Units with odd indices on both the left and right. Therefore, we can conclude that, on average, each IPC Unit has 32 candidates, with 24 from the in-slice prediction and 8 from the in-precinct prediction. This moderate number does not impose significant complexity on the encoder.

2) *Efficient Pattern Search and Mode Selection Metric:*

To align with the bitplane coding in JPEG XS, the bitplane count  $BPCN$  is utilized as the metric for pattern search, where candidates that yield the prediction residual with the least bitplane counts are selected. The calculation of  $BPCN$  is performed on a coding group of four adjacent coefficients, consistent with the coefficient coding module of JPEG XS:

$$BPCN_i(c) = \max_{j \in \{0,4\}} \{ \lceil \log_2 |c_{4i+j}| \rceil \}, \quad (4)$$

where  $c$  represents a set of coefficients in a band,  $i$  indicates the index of a coding group in  $c$ , and  $c_{4i+j}$  represents the  $j$ -th

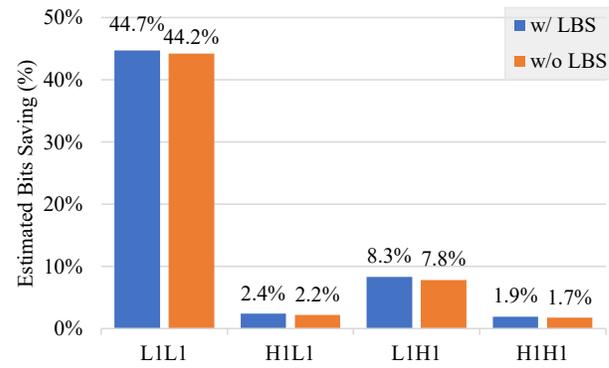


Fig. 9. Comparison of prediction efficiency between references with and without low-band-shift (LBS), illustrated with one level decomposition.

coefficient value in  $i$ -th coding group.

Similar to the pattern compensation process, the pattern search is also conducted in a multi-band synchronized approach. Best PV is selected by considering the bitplane count cost of all the bands in an IPC Group:

$$PV_{u,k} = \arg \min_{pv \in \Omega} \left\{ \sum_{b \in \mathcal{G}_k} \sum_{g \in c_{u,b}} \left( BPCN_g(c_{u,b} - \hat{c}_{u,b}^{pv}) \right) \right\}, \quad (5)$$

where  $u, k$  indicate the index of the IPC Unit and IPC Group, respectively.  $\Omega$  denotes the available search range,  $b$  represents a band in the IPC Group,  $c_{u,b}, \hat{c}_{u,b}^{pv}$  represent the original coefficients and reconstructed reference coefficients, and  $g$  stands for the coding group index. The mode selection between IPC coding and direct coding is also based on the bitplane count cost:

$$R_{IPC} = \sum_{b \in \mathcal{G}_k} \sum_{g \in c_{u,b}} \left( BPCN_g(c_{u,b} - \hat{c}_{u,b}^{PV_{u,k}}) + R_{PV_{u,k}} \right), \quad (6)$$

$$R_{Orig} = \sum_{b \in \mathcal{G}_k} \sum_{g \in c_{u,b}} (BPCN_g(c_{u,b})), \quad (7)$$

$$I_{u,k} = \begin{cases} 1 & R_{IPC} < R_{Orig} \\ 0 & \text{otherwise} \end{cases}. \quad (8)$$

The coding of PV takes a fixed-length approach, and its coding bits  $R_{PV_{u,k}}$  can be precisely calculated. Assuming the same truncation position is applied to both the prediction residual and the original coefficients, the condition in (8) can accurately reflect the bit cost relationship of the two modes. No additional calculation modules are required since the bitplane count extraction can reuse the existing modules in JPEG XS.

3) *Impact of the Shift-Variant Property of DWT:* The shift-variant property of DWT is a commonly discussed concern in the research of in-band motion estimation and motion compensation. The decimation operation in DWT leads to coefficient differences for samples with the same values but are located at misaligned positions, and the prediction efficiency can be affected. Research efforts, such as the overcomplete discrete wavelet transform (ODWT) [48] and low-band-shift (LBS) [49] have been proposed to address this issue. However, these methods either require additional memory or increased

computational complexity, potentially conflicting with the low-complexity requirement of JPEG XS. An experiment is also conducted to evaluate the performance of LBS in IPC prediction unit size and search range setting. Reconstructed reference pictures are shifted by one pixel in horizontal, vertical and both directions, to provide complementary references to overcome the shift-variant problem. As shown in Fig. 9, there is no significant degradation in prediction efficiency, allowing us to bypass this concern safely. Fractional precision prediction in the wavelet domain that relies on the ODWT is also not integrated into the IPC, considering the trade-off between performance improvement and complexity.

### F. IPC Syntax Design

An IPC subpacket syntax is added to the precinct syntax structure in JPEG XS, complementary to the existing subpackets such as significance, bitplane count, quantization level, etc. The traversal order of encoding coefficient position follows the other subpackets, proceeding first by band and then by coefficient line. IPC mode flags and PVs in an IPC Group are signaled in the subpackets corresponding to the first coefficient line in the first band of this group. Table I shows the syntax of this IPC subpacket. The IPC mode flag uses a 1-bit coding, and the PV takes a 6-bit fixed-length coding approach, with 2 bits indicating the vertical component and the remaining bits representing the horizontal offsets.

## VI. EXPERIMENTS

### A. Experimental Settings

The proposed IPC framework is integrated into the JPEG XS second edition reference software, as the updated TDC profile in the third edition is out of our intra-copy prediction research scope. Test sequences in RGB format are selected from JPEG XS AhG screen content test materials [56] and JVET SCC test sequences [57], including Class TGM (text and graphics with motion), Class M (mixed content), Class ANI (animation) and Class CC (camera-captured content). Frames with certain periods and non-repetitive scenes are extracted from these sequences\* to represent the most common screen content.

The JPEG XS encoder is configured in separate sign packing, uniform quantization, and a slice height of 16 lines. In addition to the DWT configuration of 5 horizontal and 2 vertical decomposition (5H2V), a configuration of 3 horizontal and 1 vertical decomposition (3H1V) is also tested to more comprehensively demonstrate the performance of IPC under a more lightweight encoding setting. Test target rates (bpp) are set as 0.75, 1.0, 1.5, 2.0, 3.0 and 4.0, to align with the TDC profile Core Experiment (CE) test bitrates. BD-PSNR and BD-rate [58] (RGB) compared to JPEG XS High444.12 profile are used to measure the coding efficiency. Positive values in BD-PSNR indicate the enhanced objective quality of the IPC at equivalent bitrates, while negative values in BD-rate indicate the bitrate savings relative to the JPEG XS anchor for the same reconstruction quality. To assess the complexity, the encoder

TABLE I  
IPC SYNTAX ELEMENTS

Syntax	Data
unpack_ipc( $p, s$ ) {	
for ( $b = 0; b < N_L; b = b + 1$ ) {	
for ( $\lambda = L_0[p, b]; \lambda < L_1[p, b]; \lambda = \lambda + 1$ ) {	
if ( $I[p, \lambda, b, s]$ ) {	
$k = \text{group\_idx}(b)$	
for ( $u = 0; u < N_u; u = u + 1$ ) {	
$\text{ipc\_flag}[p, k, u]$ <span style="float: right;">u(1)</span>	
if ( $\text{ipc\_flag}[p, k, u] \neq 0$ ) {	
$\text{pv}[p, k, u]$ <span style="float: right;">u(6)</span>	
}	
}	
}	
}	
}	
}	

and decoder time ratio of the IPC framework to the JPEG XS anchor, denoted as “EncT” and “DecT”, are calculated. Values greater than 100% indicate increased complexity compared to the JPEG XS anchor. Additionally, memory usage and computational complexity analysis are provided for a more comprehensive evaluation of the complexity.

### B. Overall Performance and Complexity

In this subsection, we show the performance of the IPC framework and analyze its complexity, including time complexity, memory usage, and hardware computational complexity. Several approaches for achieving more computationally efficient implementations are also discussed.

1) *Performance*: The overall performance of the proposed IPC framework is shown in Table II. For the DWT 5H2V decomposition configuration, IPC achieves an average of 0.75 dB BD-PSNR promotion and a 6.32% BD-rate reduction on Class TGM and M. For all sequences, including animation and camera-captured content, the average performance is 0.59 dB in BD-PSNR and 5.13% in BD-rate reduction. The best performance is observed in sequences such as *Desktop* and *Console*, which achieve over 1.5 dB improvement in BD-PSNR. This is mainly because of the rich explicit horizontal and vertical patterns in these sequences that can facilitate the IPC predictions. In contrast, performance on Class ANI and CC is less significant due to their non-homogeneous content.

Higher performance is observed under the DWT 3H1V decomposition configuration, where IPC achieves an average improvement of 1.99 dB in BD-PSNR and a 14.44% bitrate saving for Class TGM and M. For all sequences, the performance is 1.72 dB/-13.32%. IPC also shows effectiveness for Class ANI and CC under this configuration, with BD-PSNR promotions of 1.03 dB and 0.71 dB, respectively. The higher performance improvements of the 3H1V configuration can primarily be attributed to the larger low-pass band area at the lower decomposition level, where intra-compensation predictions are more effective due to smoother patterns. Note that despite the greater performance of IPC, the coding efficiency of the DWT 3H1V configuration remains lower than that of the DWT 5H2V configuration. That makes IPC also a favorable technology for the JPEG XS 5H2V

\*Test materials can be found in ISO/IEC JTC 1/SC 29/wg1n100906.

TABLE II  
OVERALL RD PERFORMANCE AND RELATIVE COMPLEXITY OF IPC COMPARED TO JPEG XS HIGH444.12

Class	Sequence		DWT 5H2V				DWT 3H1V			
	Name	Resolution	BD-PSNR	BD-rate	EncT	DecT	BD-PSNR	BD-rate	EncT	DecT
TGM	<i>FlyingGraphics</i>	1920x1080	0.78 dB	-7.88%	213%	102%	1.79 dB	-15.57%	212%	104%
	<i>Desktop</i>	1920x1080	1.51 dB	-13.96%	210%	101%	2.33 dB	-19.42%	211%	102%
	<i>Console</i>	1920x1080	1.79 dB	-12.46%	210%	102%	3.18 dB	-20.49%	211%	103%
	<i>ChineseEditing</i>	1920x1080	0.32 dB	-3.67%	213%	102%	0.82 dB	-8.10%	212%	104%
	<i>IPX_XsPowerPoint3</i>	1920x1080	0.70 dB	-4.52%	210%	103%	1.91 dB	-11.66%	209%	102%
	<i>FHG_ScreenStudent</i>	1920x1080	0.77 dB	-5.93%	226%	105%	4.53 dB	-24.39%	219%	107%
	<i>WebBrowsing</i>	1280x720	0.31 dB	-3.68%	195%	105%	1.32 dB	-14.02%	192%	105%
	<i>Map</i>	1280x720	0.32 dB	-2.96%	204%	104%	1.46 dB	-11.49%	200%	103%
	<i>Programming</i>	1280x720	0.90 dB	-7.60%	201%	103%	1.97 dB	-14.28%	195%	103%
	<i>SlideShow</i>	1280x720	0.98 dB	-6.03%	204%	104%	2.53 dB	-14.02%	203%	104%
	<b>Average</b>		<b>0.84 dB</b>	<b>-6.87%</b>	<b>208%</b>	<b>103%</b>	<b>2.18 dB</b>	<b>-15.34%</b>	<b>206%</b>	<b>104%</b>
M	<i>RICHTER_ScreenContent</i>	4096x2160	1.12 dB	-9.59%	230%	103%	2.08 dB	-15.23%	231%	105%
	<i>BLENDER_TearsOfSteel</i>	4096x1714	0.20 dB	-2.58%	248%	102%	1.50 dB	-13.60%	244%	105%
	<i>BasketballScreen</i>	2560x1440	0.72 dB	-6.30%	223%	102%	1.67 dB	-12.56%	222%	104%
	<i>MissionControlClip2</i>	2560x1440	0.39 dB	-3.74%	229%	107%	1.46 dB	-11.61%	225%	104%
	<i>MissionControlClip3</i>	1920x1080	0.43 dB	-3.93%	221%	103%	1.29 dB	-10.16%	215%	105%
	<b>Average</b>		<b>0.57 dB</b>	<b>-5.23%</b>	<b>230%</b>	<b>103%</b>	<b>1.60 dB</b>	<b>-12.63%</b>	<b>227%</b>	<b>105%</b>
ANI	<i>BLENDER_Sintel2</i>	4096x1744	0.13 dB	-1.45%	233%	102%	1.31 dB	-11.16%	233%	103%
	<i>ArenaOfValor</i>	1920x1080	0.14 dB	-1.73%	218%	103%	1.11 dB	-10.70%	216%	105%
	<i>Robot</i>	1280x720	0.09 dB	-1.58%	200%	102%	0.68 dB	-9.28%	197%	105%
	<b>Average</b>		<b>0.12 dB</b>	<b>-1.59%</b>	<b>216%</b>	<b>102%</b>	<b>1.03 dB</b>	<b>-10.38%</b>	<b>215%</b>	<b>104%</b>
CC	<i>ARRI_AlexaHelicopterView</i>	3840x2160	0.09 dB	-1.72%	237%	102%	0.54 dB	-8.05%	236%	102%
	<i>EBU_PendulusWide</i>	3840x2160	0.08 dB	-1.28%	240%	103%	0.87 dB	-10.55%	237%	101%
	<b>Average</b>		<b>0.09 dB</b>	<b>-1.50%</b>	<b>238%</b>	<b>102%</b>	<b>0.71 dB</b>	<b>-9.30%</b>	<b>236%</b>	<b>102%</b>
	<b>Overall</b>		<b>0.59 dB</b>	<b>-5.13%</b>	<b>218%</b>	<b>103%</b>	<b>1.72 dB</b>	<b>-13.32%</b>	<b>215%</b>	<b>104%</b>
	<b>Overall (TGM and M)</b>		<b>0.75 dB</b>	<b>-6.32%</b>	<b>215%</b>	<b>103%</b>	<b>1.99 dB</b>	<b>-14.44%</b>	<b>213%</b>	<b>104%</b>

configuration which aims at a higher compression ratio. On the other hand, the performance of IPC under the DWT 3H1V configuration demonstrates its effectiveness under a more lightweight encoder setting. Experiment results under a further lower decomposition level are not valuable, however, as the reconstruction quality may fall below the JPEG XS working points. The coding performance improvement of IPC with high-quality reconstruction in the 5H2V configuration already demonstrates its effectiveness.

Visual quality comparison is shown in Fig. 10. In the anchor reconstruction of *FlyingGraphics*, noticeable ringing artifacts appear along the edges of the grid, and there is significant blurring in the text area. These issues are alleviated in IPC reconstruction thanks to the reduction in coefficient energy and the bit savings achieved through prediction. Similar observations can be made for *BasketballScreen*.

2) *Complexity*: Regarding the time complexity, with the strict constraint on the search range, PV resolution, and IPC Unit size, the encoder exhibits a little over double the complexity. On the decoding side, the complexity increase is less significant, with an average increase of less than five percent. Note that these figures are reported without applying any implementation optimization techniques. In practice, lower complexity can be achieved without performance degradation by utilizing more efficient computing architectures. This includes employing coarse-grained parallel computation for cost calculation and comparison among candidates, as well as instruction-level parallel computation for coefficient residual calculation within a block. These optimizations are well-suited for the block-level subtraction and accumulation operations in

the IPC framework, which can be efficiently implemented on platforms like FPGA and GPU. Additionally, further acceleration can be achieved with minor performance trade-offs by utilizing hierarchical search strategies.

For hardware complexity, we analyze the memory usage and computational complexity of the JPEG XS High444.12 profile and IPC framework, with results summarized in Table III<sup>†</sup>. The memory usage of the JPEG XS encoder mainly lies in the precinct buffer for DWT and rate allocation process, which is four lines for the 5H2V decomposition configuration. The IPC requires an extra reference precinct buffer of 12 lines to support compensation predictions. For computational complexity, we measure the number of elementary operations per coefficient, including additions and shifts across all encoder modules. The increased number of addition operations in IPC is primarily due to the enumerated candidate cost checking process. However, this complexity can be alleviated through practical and efficient encoder optimization techniques. As an example, we also include results for a simplified IPC encoder implementation, namely IPC-S, with skipped chroma component cost calculation and partial band searching within an IPC Group, in the table.

### C. Ablation Study and Performance Analysis

1) *Search Range and PV Resolution Setting*: Several search ranges and PV resolutions have been tested to find the optimal setting, as shown in Table IV. The values on the left side of

<sup>†</sup>The detailed derivation process is provided in ISO/IEC JTC 1/SC 29/wg1m105026.

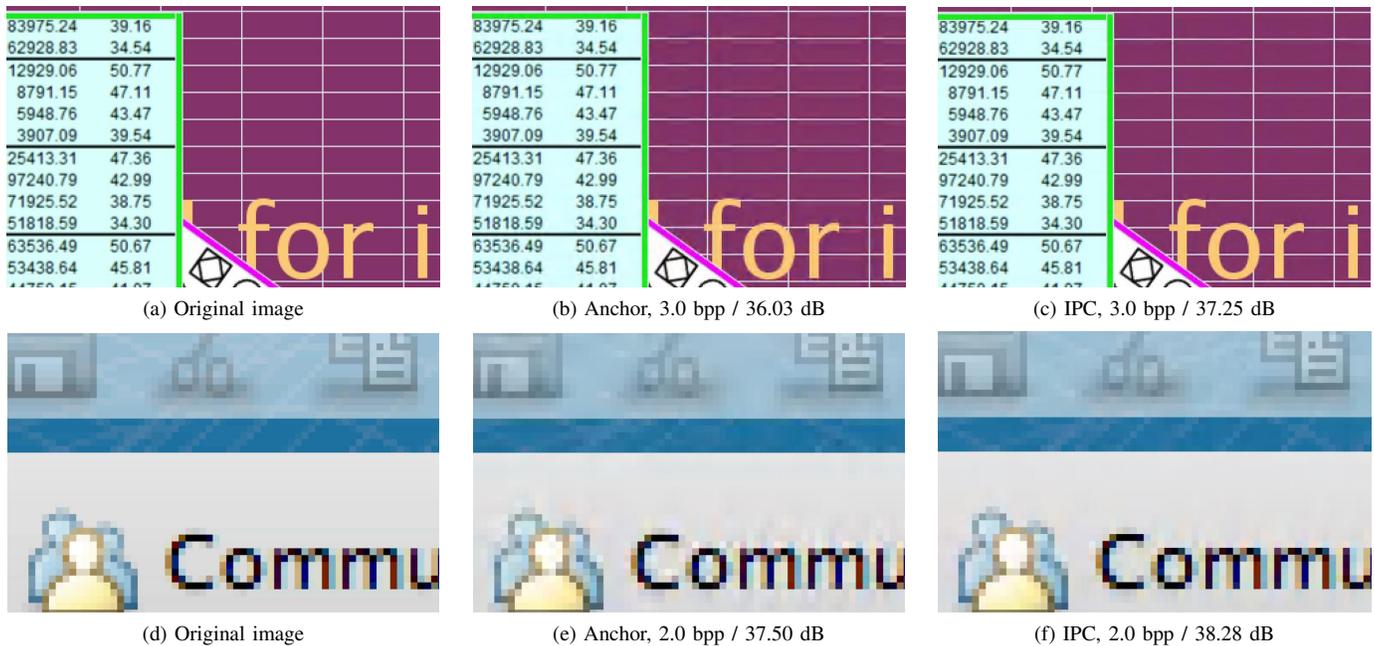


Fig. 10. Comparison of visual quality between JPEG XS anchor and with IPC. (a), (b), (c): *FlyingGraphics*; (d), (e), (f): *BasketballScreen*.

TABLE III

ENCODER MEMORY USAGE AND COMPUTATIONAL COMPLEXITY COMPARISON BETWEEN JPEG XS HIGH444.12, IPC, AND IPC-S

Profile	High444.12	IPC	IPC-S
Memory usage	4 lines	16 lines	16 lines
Operations per coefficient	12 additions 7 shifts	77 additions 8 shifts	23 additions 8 shifts
EncT	100%	215%	156%

TABLE IV

BD-PSNR (DB) PERFORMANCE AND COMPLEXITY COMPARISON FOR DIFFERENT SEARCH RANGES AND PV RESOLUTIONS ON CLASS TGM AND M

BD-PSNR/EncT	PV Resolution		
	1×	4×	16×
2× <i>IPC_Unit</i>	0.74/179%	-/-	-/-
4× <i>IPC_Unit</i>	<b>0.75/215%</b>	0.72/404%	0.70/1126%
8× <i>IPC_Unit</i>	0.73/272%	0.71/615%	-/-
16× <i>IPC_Unit</i>	0.73/350%	0.70/905%	-/-

the “/” symbol represent the BD-PSNR performance at the corresponding setting, and the values on the right side indicate the encoder time complexity measure. We vary the available search range in a reference precinct from two times IPC Unit width to 16 times IPC Unit width and the PV resolution from base resolution to 16 times base resolution by adjusting the anchor band  $b_a$  selection in (3). A larger search range and finer PV resolution lead to increased PV bit cost and higher computational complexity. The best performance is achieved at a search range of four times IPC Unit width and one base PV resolution ( $b_a$  is set to  $L_5L_2$ ), as indicated in bold in the table. Additionally, a slight performance degradation with a reduction in complexity is observed in the smaller search range, indicating potential optimization strategies for

TABLE V

ABLATION STUDY ON THE IPC GROUP SCHEME WITH BD-PSNR (DB) AND ENCODER TIME COMPLEXITY

Class	Separate	Combined	Group	Group
	Group	Group	Set 1	Set 2 (IPC)
TGM	0.73	0.64	0.83	0.84
M	0.48	0.47	0.56	0.57
ANI	0.04	0.03	0.11	0.12
CC	0.01	0.02	0.08	0.09
Avg.	0.49	0.45	0.58	0.59
Avg. TGM M	0.65	0.58	0.74	0.75
$T_{Enc}$	209%	221%	213%	215%

TABLE VI

BD-PSNR (DB) PERFORMANCE AND ENCODER TIME COMPLEXITY OF EACH IPC GROUP STEP

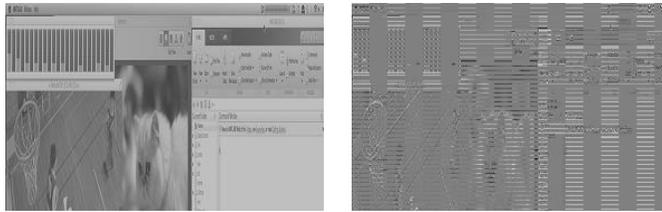
Class	$\mathcal{G}_0$	$\mathcal{G}_0, \mathcal{G}_1$	$\mathcal{G}_0, \mathcal{G}_1, \mathcal{G}_2$	$\mathcal{G}_0, \mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$
	TGM	0.04	0.67	0.75
M	0.00	0.29	0.37	0.57
ANI	0.00	0.07	0.11	0.12
CC	-0.01	0.06	0.09	0.09
Avg.	0.02	0.42	0.49	0.59
Avg. TGM M	0.03	0.54	0.62	0.75
$T_{Enc}$	125%	164%	178%	215%

IPC implementations in practice.

2) *IPC Group Design*: To verify the effectiveness of the IPC Group scheme derived by k-means clustering in Section V-B, we test several other band group schemes and compare their performance in Table V. The Separate Group scheme assigns a group to each band, while the Combined Group gathers all bands into a single group. Group Set 1 is similar to the IPC Group scheme, substituting the high-pass band divisions based on decomposition levels with those based on decomposition direction. Specifically,  $\mathcal{G}_0$  and  $\mathcal{G}_0$  remain unchanged, while

TABLE VII  
ABLATION STUDY ON THE IN-PRECINCT PREDICTION WITH BD-PSNR (DB) PERFORMANCE AND ENCODER TIME COMPLEXITY

Class	w/o In-Precinct Prediction		w/ In-Precinct Prediction	
	BD-PSNR	EncT	BD-PSNR	EncT
TGM	0.76	196%	0.84	208%
M	0.47	206%	0.57	230%
ANI	0.13	195%	0.12	216%
CC	0.10	205%	0.09	238%
Avg.	0.52	199%	0.59	218%
Avg. TGM M	0.66	199%	0.75	215%



(a) Original image

(b) Residual image

Fig. 11. Comparison of the L1 norm value distribution of the coefficient between the original image and the residual image.

$\mathcal{G}_2 := \{L_2H_2, L_1H_1\}$  and  $\mathcal{G}_3 := \{H_2H_2, H_1H_1\}$ . Experimental results in the table demonstrate the superiority of the IPC Group scheme. Differences in their memory access manner cause slight complexity fluctuations between different schemes. Table VI demonstrates the coding efficiency and complexity steps for each group in the IPC Group scheme. These four groups contribute to BD-PSNR values of 0.03 dB, 0.51 dB, 0.08 dB and 0.13 dB on Class TGM and M, respectively. These performance variances are jointly determined by the band area and band pattern characteristics in the group.

### 3) Effectiveness of the In-Precinct Pattern Prediction:

Table VII shows the ablation study results on the proposed in-precinct interweaved pattern prediction scheme. It achieves an average quality improvement of 0.09 dB on Class TGM and M at a cost of 16% encoder side complexity. The performance would be higher for sequences with prominent horizontal patterns, such as *FlyingGraphics*, which shows a 0.19 dB gain, and *BasketballScreen*, which shows a 0.16 dB gain.

4) *Visualization of the Prediction Efficiency:* To more vividly demonstrate the effectiveness of the IPC, we present a comparison of the coefficient value distribution between the original image and the predicted residual image in Fig. 11, with results shown for the  $L_3L_1$  band for conciseness. The predicted residual image exhibits lower coefficient energy than the original image in most regions. Some areas with a stripe shape still show high energy distributions, and this can be attributed to the lack of reference for those regions.

## VII. CONCLUSION

In this paper, a frequency domain intra pattern copy prediction framework is proposed for the lightweight JPEG XS codec to improve its screen content coding efficiency. Wavelet coefficients grouping schemes have been studied to facilitate pattern-adaptive predictions on different bands. The

search range, offset vector resolution, and prediction unit size have also been carefully designed to accommodate the low-complexity and low-latency requirements of JPEG XS. Simulation results show that on screen content sequences, the proposed framework achieves an average of 0.75 dB BD-PSNR improvement, 6.32% BD-rate reduction and 1.99 dB BD-PSNR improvement, 14.44% BD-rate reduction for DWT 5 horizontal-2 vertical decomposition and DWT 3 horizontal-1 vertical decomposition configurations, respectively.

In future work, we will first explore rate-distortion optimization schemes with lower complexity for pattern search. Secondly, rate allocation between coefficients direct coding and predictive coding will be considered to further enhance the coding efficiency. Thirdly, acceleration schemes will be applied to reduce the encoder-side complexity. Finally, we plan to extend the IPC prediction scheme to other wavelet-based codecs, such as the JPEG 2000, and integrate it with more advanced transforms.

## REFERENCES

- [1] SMPTE, "EG 2111-2, 3/6/12 24 Gbit/s SDI standards roadmap." [Online]. Available: <https://pub.smpte.org/doc/eg2111-2/20181203-pub/eg2111-2-2019.pdf>
- [2] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 1649–1668, 2012.
- [3] B. Bross, Y.-K. Wang, Y. Ye, S. Liu, J. Chen, G. J. Sullivan, and J.-R. Ohm, "Overview of the versatile video coding (VVC) standard and its applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, pp. 3736–3764, 2021.
- [4] Y. Chen *et al.*, "An overview of core coding tools in the AV1 video codec," in *2018 IEEE Picture Coding Symposium (PCS)*, 2018, pp. 41–45.
- [5] F. G. Walls and A. S. MacInnis, "27.4L: Late-news paper: VESA display stream compression: an overview," *SID Symposium Digest of Technical Papers*, vol. 45, pp. 360–363, 2014.
- [6] —, "VESA display stream compression for television and cinema applications," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 6, pp. 460–470, 2016.
- [7] N. Jacobson, V. Thirumalai, R. Joshi, and J. Goel, "A new display stream compression standard under development in VESA," *Applications of Digital Image Processing XL*, vol. 10396, pp. 160–171, 2017.
- [8] H. Ren *et al.*, "A novel visually-lossless compression model for low-latency interaction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, pp. 5063–5076, 2024.
- [9] SMPTE ST 2042, "VC 2 video compression," 2017.
- [10] SMPTE RDD 36, "Apple prores bitstream syntax and decoding process," 2015.
- [11] SMPTE RDD 34, "LLVC low latency video codec for network transfer," 2015.
- [12] SMPTE RDD 35, "TICO lightweight codec used in IP networked or in SDI infrastructures," 2016.
- [13] A. Descampe *et al.*, "JPEG XS—a new standard for visually lossless low-latency lightweight image coding," *Proceedings of the IEEE*, vol. 109, pp. 1559–1577, 2021.
- [14] Y. Li, X. Liao, and X. Wu, "Screen-shooting resistant watermarking with grayscale deviation simulation," *IEEE Transactions on Multimedia*, vol. 26, pp. 10908–10923, 2024.
- [15] L. Fu, X. Liao, J. Guo, L. Dong, and Z. Qin, "Waverecovery: Screen-shooting watermarking based on wavelet and recovery," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2024.
- [16] T. Kojima, J. J. Stone, J.-R. Chen, and P. N. Gardiner, "A practical approach to IP live production," *SMPTE Motion Imaging Journal*, vol. 124, no. 2, pp. 29–40, 2015.
- [17] J. Fu, Z. Chen, X. Chen, and W. Li, "Sequential reinforced 360-degree video adaptive streaming with cross-user attentive network," *IEEE Transactions on Broadcasting*, vol. 67, no. 2, pp. 383–394, 2021.

- [18] M. Wien, R. Cazoulat, A. Graffunder, A. Hutter, and P. Amon, "Real-time system for adaptive video streaming based on SVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1227–1237, 2007.
- [19] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016.
- [20] J. Chen, X. Liao, W. Wang, Z. Qian, Z. Qin, and Y. Wang, "SNIS: A signal noise separation-based network for post-processed image forgery detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 2, pp. 935–951, 2023.
- [21] T. Richter and S. Foessel, "JPEG XS screen content coding extensions," in *SPIE Applications of Digital Image Processing XLV*, vol. 12226, 2022.
- [22] —, "JPEG XS profiles and levels for screen content coding," in *SPIE Applications of Digital Image Processing XLVI*, vol. 12674, 2023.
- [23] J. Xu, R. Joshi, and R. A. Cohen, "Overview of the emerging HEVC screen content coding extension," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, pp. 50–62, 2016.
- [24] T. Nguyen *et al.*, "Overview of the screen content support in VVC: Applications, coding tools, and performance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, pp. 3801–3817, 2021.
- [25] S.-L. Yu and C. Chrysafis, "New intra prediction using intra-macroblock motion compensation," in *JVT-C151, 3rd Meeting of Joint Video Team (JVT)*, 2002.
- [26] X. Xu and S. Liu, "Overview of screen content coding in recently developed video coding standards," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, pp. 839–852, 2021.
- [27] X. Xu, X. Li, and S. Liu, "CE8: Cpr mode with local search ranges (test CE8. 3.1 and CE8. 3.2)," *Joint Video Experts Team (JVET) of ITU-T SG*, 2018.
- [28] —, "Intra block copy in versatile video coding with reference sample memory reuse," in *2019 IEEE Picture Coding Symposium (PCS)*, 2019, pp. 1–5.
- [29] D.-K. Kwon and M. Budagavi, "Fast intra block copy (intra) search for HEVC screen content coding," in *2014 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2014, pp. 9–12.
- [30] H. Zhang, Q. Zhou, N. Shi, F. Yang, X. Feng, and Z. Ma, "Fast intra mode decision and block matching for HEVC screen content compression," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 1377–1381.
- [31] S.-H. Tsang, Y.-L. Chan, W. Kuang, and W.-C. Siu, "Reduced-complexity intra block copy (intra) mode with early cu splitting and pruning for HEVC screen content coding," *IEEE Transactions on Multimedia*, vol. 21, pp. 269–283, 2019.
- [32] S.-H. Tsang, Y.-L. Chan, and W.-C. Siu, "Hash based fast local search for intra block copy (intra) mode in HEVC screen content coding," in *2015 IEEE Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 2015, pp. 396–400.
- [33] W. Zhu, W. Ding, J. Xu, Y. Shi, and B. Yin, "Hash-based block matching for screen content coding," *IEEE Transactions on Multimedia*, vol. 17, pp. 935–944, 2015.
- [34] K. M. Uz, M. Vetterli, and D. LeGall, "Interpolative multiresolution coding of advanced television with compatible subchannels," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 1, pp. 86–99, 1991.
- [35] Y.-Q. Zhang and S. Zafar, "Motion-compensated wavelet transform coding for color video compression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 2, pp. 285–296, 1992.
- [36] S. Zafar, Y.-Q. Zhang, and B. Jabbari, "Multiscale video representation using multiresolution motion compensation and wavelet decomposition," *IEEE Journal on Selected Areas in Communications*, vol. 11, pp. 24–35, 1993.
- [37] S. Panchanathan, E. Chan, and X. Wang, "Fast multiresolution motion estimation scheme for a wavelet transform video coder," in *1994 SPIE Visual Communications and Image Processing (VCIP)*, vol. 2308, 1994, pp. 671–681.
- [38] J. Lee and Y. M. Ro, "Hierarchical block matching algorithm in MRME," in *2000 IEEE International Conference on Image Processing (ICIP)*, 2000, pp. 836–839.
- [39] J. Zan, M. O. Ahmad, and M. Swamy, "Wavelet-based multiresolution motion estimation through median filtering," in *2002 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 2002, pp. 3273–3276.
- [40] J. Zan, M. Ahmad, and M. Swamy, "New techniques for multi-resolution motion estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, pp. 793–802, 2002.
- [41] K. Tsunashima, J. Stampleman, and V. Bove, "A scalable motion-compensated subband image coder," *IEEE Transactions on Communications*, vol. 42, pp. 1894–1901, 1994.
- [42] A. Nosratinia and M. T. Orchard, "Multiresolution framework for backward motion compensation," *1995 SPIE Digital Video Compression: Algorithms and Technologies*, pp. 190–200, 1995.
- [43] A. Nosratinia and M. Orchard, "Multi-resolution backward video coding," in *1995 IEEE International Conference on Image Processing (ICIP)*, vol. 2, 1995, pp. 563–566.
- [44] X. Yang and K. Ramchandran, "Hierarchical backward motion compensation for wavelet video coding using optimized interpolation filters," in *1997 IEEE International Conference on Image Processing (ICIP)*, vol. 1, 1997, pp. 85–88.
- [45] —, "Scalable wavelet video coding using aliasing-reduced hierarchical motion compensation," *IEEE Transactions on Image Processing*, vol. 9, pp. 778–791, 2000.
- [46] H.-W. Park and H.-S. Kim, "Motion estimation using low-band-shift method for wavelet-based moving-picture coding," *IEEE Transactions on Image Processing*, vol. 9, pp. 577–587, 2000.
- [47] J. Magarey and N. Kingsbury, "Motion estimation using a complex-valued wavelet transform," *IEEE Transactions on Signal Processing*, vol. 46, pp. 1069–1084, 1998.
- [48] N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Applied and Computational Harmonic Analysis*, vol. 10, pp. 234–253, 2001.
- [49] Y. Andreopoulos, M. v. d. Schaar, A. Munteanu, J. Barbarien, P. Schelkens, and J. P. Cornelis, "Complete-to-overcomplete discrete wavelet transforms for scalable video coding with MCTF," in *2003 SPIE Visual Communications and Image Processing (VCIP)*, vol. 5150, 2003, pp. 719–731.
- [50] X. Li, "Scalable video compression via overcomplete motion compensated wavelet coding," *Signal Processing: Image Communication*, vol. 19, pp. 637–651, 2004.
- [51] ISO/IEC JTC 1/SC 29/WG 1, "Use cases and requirements for JPEG XS v3.2," document wg1n100633, 101st JPEG Meeting, 2023.
- [52] W. Zhu, W. Ding, J. Xu, Y. Shi, and B. Yin, "Screen content coding based on HEVC framework," *IEEE Transactions on Multimedia*, vol. 16, no. 5, pp. 1316–1326, 2014.
- [53] C. Dong, H. Ma, D. Liu, and J. W. Woods, "Wavelet-based learned scalable video coding," in *2022 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2022, pp. 3190–3194.
- [54] C. Dong, H. Ma, Z. Li, L. Li, and D. Liu, "Temporal wavelet transform-based low-complexity perceptual quality enhancement of compressed video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 5, pp. 4040–4053, 2024.
- [55] Menon, Vignesh V and Feldmann, Christian and Amirpour, Hadi and Ghanbari, Mohammad and Timmerer, Christian, "VCA: Video complexity analyzer," in *Proceedings of the 13th ACM Multimedia Systems Conference*, 2022, pp. 259–264.
- [56] ISO/IEC JTC 1/SC 29/WG 1, "Report on CE4 on JPEG XS by intoPIX," document wg1m99039, 99th JPEG Meeting, 2022.
- [57] Y.-H. Chao, Y.-C. Sun, J. Xu, and X. Xu, "JVET common test conditions and software reference configurations for non-4:2:0 colour formats," document JVET-T2013, 20th JVET Meeting, 2020.
- [58] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," document VCEG-M33, 2001.



**Yao Li** (Student Member, IEEE) received the B.S. degree in electronic engineering from Tianjin University, Tianjin, China, in 2022. He is currently pursuing the Ph.D. degree with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, China. His current research interests include image/video processing and coding.



**Zhuoyuan Li** (Student Member, IEEE) received the B.S. degree in communication engineering from Southwest Jiaotong University, Chengdu, China, in 2020. He is currently pursuing the Ph.D. degree with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, China. His research interests include video coding and processing.



**Dong Liu** (M'13–SM'19) received the B.S. and Ph.D. degrees in electrical engineering from the University of Science and Technology of China (USTC), Hefei, China, in 2004 and 2009, respectively. He was a Member of Research Staff with Nokia Research Center, Beijing, China, from 2009 to 2012. He joined USTC as a faculty member in 2012 and became a Professor in 2020.

His research interests include image and video processing, coding, analysis, and data mining. He has authored or co-authored more than 200 papers in

international journals and conferences. He has more than 30 granted patents. He has several technique proposals adopted by standardization groups. He received the 2009 IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY Best Paper Award, VCIP 2016 Best 10% Paper Award, and ISCAS 2022 Grand Challenge Top Creativity Paper Award. He and his students were winners of several technical challenges held in ICIP 2024, ISCAS 2023, ICCV 2019, ACM MM 2019, ACM MM 2018, ECCV 2018, CVPR 2018, and ICME 2016. He is a Senior Member of CCF and CSIG, an elected member of IVMSP-TC of IEEE SP Society, and an elected member of MSA-TC of IEEE CAS Society. He serves or had served as the Chair of IEEE 1857.11 Standard Working Subgroup (also known as Future Video Coding Study Group), an Associate Editor for IEEE TRANSACTIONS ON IMAGE PROCESSING, a Guest Editor for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, an Organizing Committee member for ChinaMM 2024, VCIP 2022, ICME 2021, etc.



**Li Li** (M'17–SM'25) received the B.S. and Ph.D. degrees in electronic engineering from University of Science and Technology of China (USTC), Hefei, Anhui, China, in 2011 and 2016, respectively. He was a visiting assistant professor in University of Missouri-Kansas City from 2016 to 2020. He joined the department of electronic engineering and information science of USTC as a research fellow in 2020 and became a professor in 2022.

His research interests include image/video/point cloud coding and processing. He has authored or co-

authored more than 80 papers in international journals and conferences. He has more than 20 granted patents. He has several technique proposals adopted by standardization groups. He received the Multimedia Rising Star 2023. He received the Best 10% Paper Award at the 2016 IEEE Visual Communications and Image Processing (VCIP) and the 2019 IEEE International Conference on Image Processing (ICIP). He serves as an associate editor for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and IEEE TRANSACTIONS ON MULTIMEDIA from 2024 to 2025.