

Received March 9, 2022, accepted March 21, 2022, date of publication April 18, 2022, date of current version April 26, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3168155

Efficient Partition Decision Based on Visual Perception and Machine Learning for H.266/Versatile Video Coding

MEI-JUAN CHEN¹, (Senior Member, IEEE), CHENG-AN LEE¹, YU-HSIANG TSAI¹, CHIEH-MING YANG¹, CHIA-HUNG YEH^{2,3}, (Senior Member, IEEE), LIH-JEN KAU⁴, (Senior Member, IEEE), AND CHUAN-YU CHANG⁵, (Senior Member, IEEE)

¹Department of Electrical Engineering, National Dong Hwa University, Hualien 974301, Taiwan

²Department of Electrical Engineering, National Taiwan Normal University, Taipei 106010, Taiwan

³Department of Electrical Engineering, National Sun Yat-sen University, Kaohsiung 804201, Taiwan

⁴Department of Electronic Engineering, National Taipei University of Technology, Taipei 106344, Taiwan

⁵Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Yunlin 640301, Taiwan

Corresponding author: Chia-Hung Yeh (chyeh@ntnu.edu.tw)

This work was supported by the Ministry of Science and Technology, Taiwan, under Grant MOST 109-2221-E-259-012-MY3.

ABSTRACT H.266/Versatile Video Coding (VVC) is the latest international video coding standard to encode ultra-high-definition video effectively. The quadtree with nested multi-type tree (QT-MTT) structure provides various sizes of coding tree partitioning and allows the nested binary tree (BT) split and ternary tree (TT) split at each QT level. Furthermore, numerous advanced coding tools are equipped in the H.266/VVC encoder. However, the encoding time increases tremendously. Previous researches regarding the fast coding algorithm of H.266/VVC seldom mention perceptual redundancy. This paper utilizes the human vision model of just noticeable difference to extract the visually distinguishable pixels that may affect the visual perception. We observe that the distributions acquired by the horizontal and vertical projections of visually distinguishable pixels within the coding unit are related to their corresponding MTT splitting modes. Therefore, the distributions representing the perceptual information of human vision are used to be the input features of machine learning. Fast MTT decision determined by the random forest models of machine learning is proposed to quickly select the partition for intra coding. Experimental results demonstrate that the proposed method can effectively accelerate intra coding process while maintaining good bitrate and video quality based on the properties of the visual perception. The proposed algorithm provides better performance than the previous work.

INDEX TERMS H.266, versatile video coding, intra coding, visual perception, just noticeable difference, machine learning, random forest.

I. INTRODUCTION

Video coding is one of the most important and active research topics over the past couple of decades and is closely connected to industrial applications in both hardware and software. The wide range of applications includes mobile phone, videophone, videoconferencing, video streaming, digital television, distance learning, remote medical system, etc. Videophone and videoconferencing are becoming a global trend when people need to contact each other, as well as a tool for people to provide emotional support and resources.

The associate editor coordinating the review of this manuscript and approving it for publication was Essam A. Rashed¹.

Efficient video coding technologies are quickly applied and provide many valuable services. The explosive demands for large amounts of ultra-high-definition (UHD) video bring new challenges to the capability of video compression. JVET (Joint Video Experts Team) comprised of the ITU-T VCEG (Video Coding Experts Group) and the ISO/IEC MPEG (Moving Picture Experts Group) establish the latest video coding standard H.266/versatile video coding (VVC) [1]–[3] with coding performance higher than that of the H.265/high efficiency video coding (HEVC) [4]–[10] to meet the requirements of various industrial applications including UHD television for home theater, high dynamic range video, 360-degree video, live streaming, HD video conferencing,

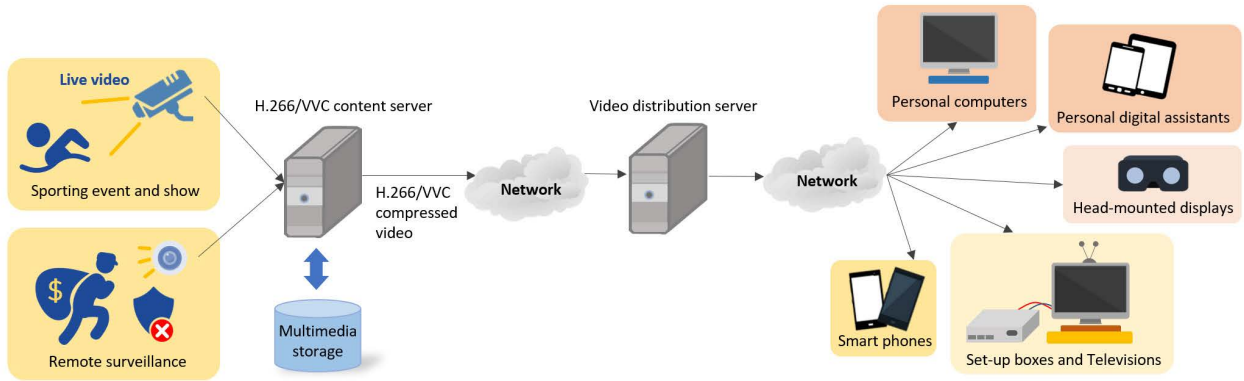


FIGURE 1. Illustration of the applications for H.266/VVC.

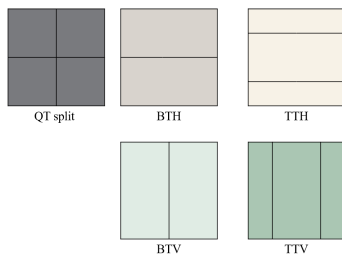


FIGURE 2. Five QT/MTT partitions in H.266/VVC.

HD video surveillance, augmented reality, virtual reality, mixed reality, etc., as shown in Fig. 1.

The intra prediction and the inter prediction are powerful solutions in video compression. Compared to the quadtree (QT) coding block structure of H.265/HEVC, more flexible coding unit (CU) partitioning can be provided by the quadtree with nested multi-type tree (QT-MTT) using binary tree (BT) and ternary tree (TT) splits in H.266/VVC. The QT-MTT is a new coding structure with BT and TT splits for the subtree partition in a QT. Each BT or TT split allows horizontal and vertical directions as indicated in Fig. 2. The input video frame is divided into coding tree units (CTUs), 128×128 blocks, under the common test conditions (CTC) [11] of H.266/VVC. In the all-intra configuration, the MTT splitting structure allows for the CUs no larger than 32×32 . The minimum size is limited to 4×4 for BT or TT split. The MTT coding structure makes the splitting of CU more diversified and increases the coding efficiency. Fig. 3 shows an example with the QT/MTT partition. However, flexible partitions have larger computational complexity, which increases encoding time. Besides, the H.266/VVC encoder equips considerable coding tools. The various CU partition and numerous coding tools improve the compression performance for different kinds of video content, also tremendously increase the encoding complexity. Undoubtedly, it is very important for the acceleration of encoding to reach real-time processing.

II. RELATED WORKS

To meet the high computational demand of H.266/VVC, fast computation techniques for H.266/VVC intra coding have

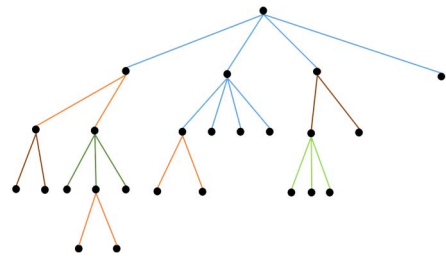
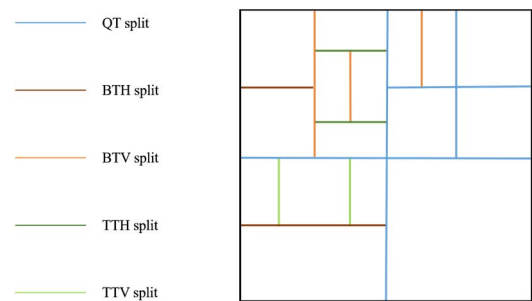


FIGURE 3. Illustration of QT/MTT partition.

been investigated previously. They can be roughly classified into probability-based [12], [13], learning-based [14]–[18], texture-based [19], gradient-based [20], [21], and texture- and gradient-based [22] methods. Fu et al. [12] jointly explore the split types and intra modes of sub-CUs to early skip vertical split (including vertical BT and TT splits) and horizontal TT split. Park and Kang [13] propose a simple early decision technique to reduce the TT encoding time of intra coding motivated by the Bayesian probability approach. Peng et al. [14] utilize support vector machine to build the horizon binary-tree decision, the vertical binary-tree decision and the quad-tree decision models for fast partition of QTBT in H.266/VVC intra coding. Yang et al. [15] select the decision tree in machine learning as the classifier for QT/MTT decision. Fast intra mode decision with gradient descent search is also introduced. Amestoy et al. [16] use the random forest classifiers in machine learning to determine for each coding block the most probable partition modes. The three binary classifiers include Split/Non-Split, QT/BT and BH/BV. Since the solution of [16] is originally designed

for the complexity reduction of QT/BT partitioning scheme, to adapt to the MTT partitioning, horizontal partition modes (BTH and TTH) and vertical partition modes (BTV and TTV) are both grouped as the output of the BH/BV classifier. Zhao *et al.* [17] design a decision tree classifier by using just noticeable difference model threshold, motion pattern and image texture features for fast CU partition decision strategy. Park and Kang [18] use a lightweight neural network model to decide whether to terminate the nested TT block structures subsequent to a quadtree based on the two kinds of features. Peng *et al.* [19] adopt texture features to classify CUs into simple, common and complex types. Early skipping partition modes for simple CUs and non-partition modes for complex CUs are designed. Unnecessary horizontal or vertical partition modes are excluded for common and complex CUs. Tang *et al.* [20] extract edge features by applying the Canny edge detector to carry out early termination and to skip vertical or horizontal partition modes. Cui *et al.* [21] calculate the gradients to skip unnecessary searches in CU size decision and pre-determine the likelihood of BT or TT partition in the horizontal or vertical direction. Fan *et al.* [22] use the variances of original pixels and the gradient features by the Sobel operator to terminate the further splitting, select the QT and choose one partition from five QT/MTT partitions. In this paper, we apply the just noticeable difference model [23] to extract the features into a machine learning algorithm for the acceleration of H.266/VVC intra coding.

III. PROPOSED METHOD

The fidelity of the decoded video is affected by not only the objective pixel distortion but also the perceptible interval of human eyes for luminance variation. Gray level variation occurs both on the edges and texture regions. Sometimes, the gradient or texture measurement may not be feasible. We need to find the regions truly sensitive to human eyes. Several literatures with perceptual video coding or processing provide promising results [24]–[30].

A. JUST NOTICEABLE DIFFERENCE (JND)

A just-noticeable distortion (JND) is employed to quantify the perceptual redundancy and the JND profile derives threshold sensitivities based on background luminance [23]. For human perception, people may just distinguish the difference once the change on luminance between the current pixel and background exceeds the perceptual threshold; below which, the differences are imperceptible. The JND visual model represents the visual distortion of luminance and the curve of visibility thresholds due to background luminance [23] is exhibited in Fig. 4. The horizontal axis indicates the background luminance pix_{avg} . Since the minimum size of a CU for BT or TT split in MTT is 4×4 , we utilize the average of 2×2 four pixels in the center of a 4×4 block, as shown in Fig. 5 and (1) to obtain pix_{avg} . Each pix_{avg} value corresponds to a JND value $JND_{pix_{avg}}$ on the vertical axis. The $JND_{pix_{avg}}$ is calculated as (2), where T_0 is 17, γ is $3/128$. For a 10-bit video sequence, the gray level of image is divided by 4 to obtain an

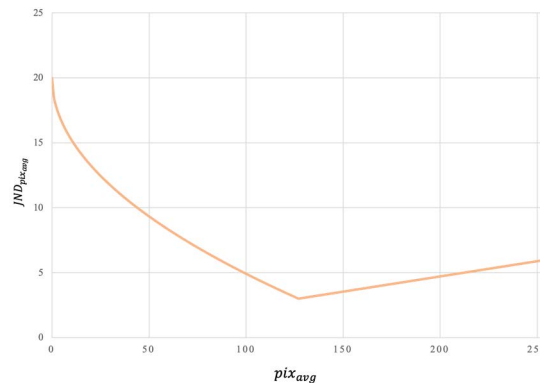


FIGURE 4. Visibility thresholds according to background luminance [23].

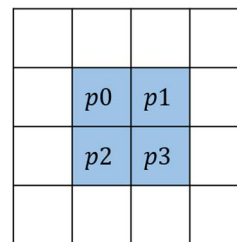


FIGURE 5. Background luminance for JND.

8-bit image and then find the JND threshold. If the absolute value of the difference between the luminance of the current pixel and the background luminance is lower than $JND_{pix_{avg}}$, it means that this luminance difference is invisible to human eyes. If the absolute value of the luminance difference is larger than or equal to $JND_{pix_{avg}}$, $VD-Pix_{i,j}$ indicates the visually distinguishable pixel for human eyes as shown in (3).

$$pix_{avg} = \frac{(p0 + p1 + p2 + p3)}{4} \tag{1}$$

$$JND_{pix_{avg}} = \begin{cases} T_0 \times \left(1 - \left(\frac{pix_{avg}}{127}\right)^{\frac{1}{2}}\right) + 3, & \text{for } pix_{avg} \leq 127 \\ \gamma \times (pix_{avg} - 127) + 3, & \text{for } pix_{avg} > 127 \end{cases} \tag{2}$$

$$VD-Pix_{i,j} = \begin{cases} 1, & \text{if } |pix_{i,j} - pix_{avg}| \geq JND_{pix_{avg}} \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

Inspired by [26], we conclude that if there is a visually distinguishable edge in the horizontal direction, there would be more possible MTT splitting modes for BTH and TTH in the process of the MTT directional split. If there is a visually distinguishable edge in the vertical direction, there would be more possible MTT splitting modes for BTV and TTV, as indicated in Fig. 6. Since the positions of visually distinguishable pixels may not be exactly on a straight line, the projection-based on MTT partition is proposed in our method.

When the QtDepth of CU is equal to 2 or 3 and MttDepth is equal to 0, our algorithm separates the square CU into

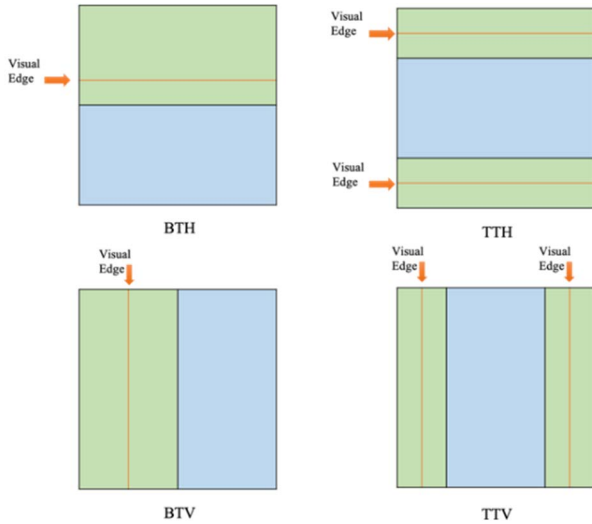


FIGURE 6. Different visual edge directions and more likely MTT splitting modes.

four vertical and horizontal sub-parts, and makes the projections to obtain the amount of visually distinguishable pixels by (4)-(5) and Fig. 7.

$$Proj_{hor,l} = \sum_{i=0}^{w-1} \sum_{j=0}^{\frac{w}{4}-1} VD-Pix_{x+i,y+\frac{w}{4} \times l+j}, \quad l = 0 \sim 3, \quad (4)$$

$$Proj_{ver,l} = \sum_{i=0}^{\frac{w}{4}-1} \sum_{j=0}^{w-1} VD-Pix_{x+\frac{w}{4} \times l+i,y+j}, \quad l = 0 \sim 3, \quad (5)$$

where $Proj_{hor,l}$ and $Proj_{ver,l}$ represent the sum of $VD-Pix_{i,j}$ in the l -th sub-part in the horizontal and the vertical directions, respectively. The (x, y) denotes the coordinate of the upper-left pixel of the CU with the size of w in MttDepth 0. The orientation of CU can be decided by collecting the distribution information of $Proj_{hor,l}$ and $Proj_{ver,l}$; then the appropriate partition direction can be determined to bypass the unnecessary CU splitting. There may be a relationship between the projection results and the partition types. Fig. 8 demonstrates the ideal cases of the relationship between the projection results and the partition types. The projection results can be calculated by (4) and (5) to obtain $Proj_{hor,l}$ and $Proj_{ver,l}$. Therefore, $Proj_{hor,l}$ and $Proj_{ver,l}$ are used for the input of random forest model of machine learning.

B. RANDOM FOREST

Random forest (RF) is an ensemble learning of the bagging method [31]. The advantage of RF is to combine multiple basic classifiers to generate better predictive ability. The basic classifier used in this paper is a decision tree (DT) [32], and the structure of DT is composed of nodes, branches and leaves, as shown in Fig. 9. The sample set in a node is

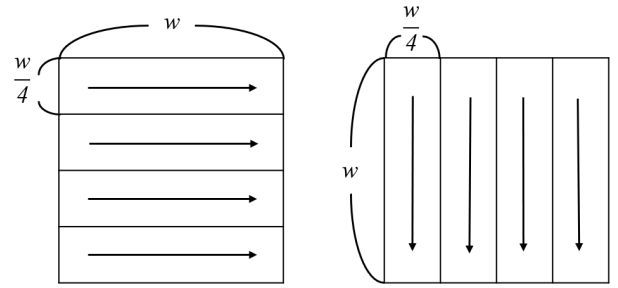


FIGURE 7. Projection of visually distinguishable pixels.

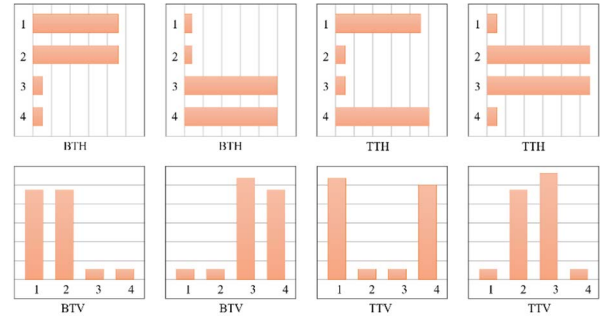


FIGURE 8. The ideal cases of the relationship of the projections and corresponding MTT splitting modes.

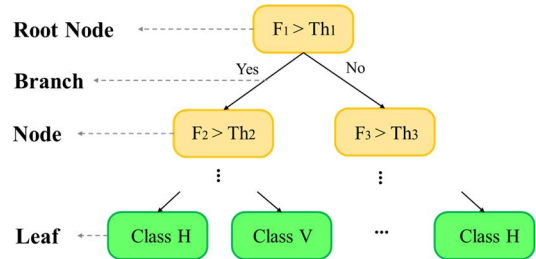


FIGURE 9. Schematic diagram of the decision tree structure.

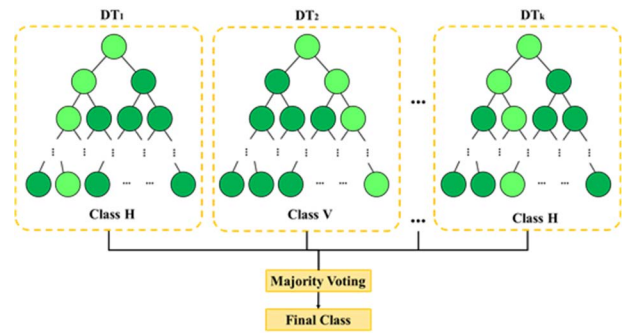


FIGURE 10. Schematic diagram of the random forest.

divided into two sub-sample sets through branching actions. Repeat the steps of setting up nodes and branches, and end the process when the samples are divided into leaves. The algorithm of the DT in this paper is CART (Classification and Regression Trees) [33]. The branch is based on the Gini impurity to select the feature and threshold. Each parent node

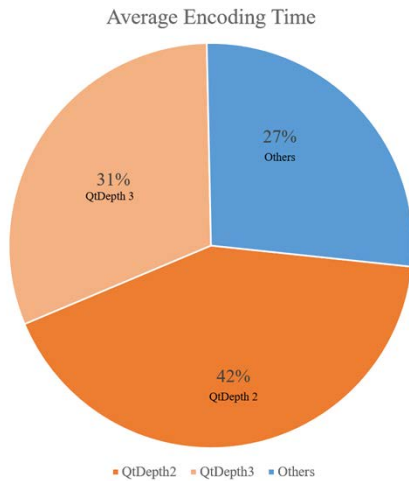


FIGURE 11. Distribution of average encoding time for various quadtree partition levels.

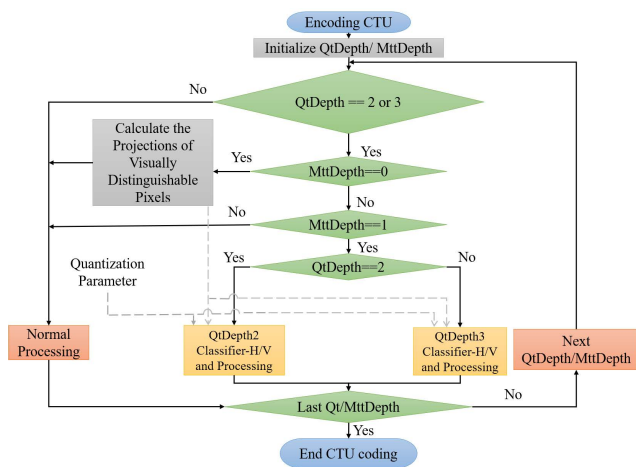


FIGURE 12. The flowchart of proposed algorithm.

is divided into two child nodes by comparing the feature value of the sample with the threshold in the parent node. When all samples in the child node belong to the same target category (horizontal or vertical), this node is a leaf.

RF generates different DTs by sampling features, and the prediction result is determined by the majority decision of all DTs, as shown in Fig. 10. In this paper, we take nine features including $Proj_{hor,l}$, $Proj_{ver,l}$ (l is 0 to 3) and the quantization parameter (QP) as the input of random forest model of machine learning to predict the suitable MTT splitting mode. The QP determines the distortion of coded video and has an important impact on the bitrate. Higher QP allows more distortion and a lower bitrate. Empirical evaluations show that the following hyperparameter setting achieves the best performance. The maximum possible depth of the decision tree is 11. The number of features to be considered when looking for the best split is 20. If the number of samples in a node is less than 3, then the node will not be split. The number of tree in the random forest is 51.

TABLE 1. Information of test sequences.

Class	Resolution	Sequence	Frame Number
Class A1	3840×2160	Tango2	194
		FoodMarket4	200
		Campfire	200
Class A2	3840×2160	CatRobot1	200
		DaylightRoad2	200
		ParkRunning3	200
		BasketballDrive	400
Class B	1920×1080	BQTerrace	500
		Cactus	400
		Kimono	140
		ParkScene	140
		BasketballDrill	400
Class C	832×480	BQMall	500
		PartyScene	400
		RaceHorses	200
		BasketballPass	400
Class D	416×240	BlowingBubbles	400
		BQSquare	500
		RaceHorses	200
		FourPeople	500
Class E	1280x720	Johnny	500
		KristenAndSara	500

C. MTT DECISION ALGORITHM

Since the MTT partition implemented in VTM 7.0 [34] is from QtDepth 2 for intra coding, we encode the first 100 frames from each of six benchmark sequences (Class A1 Campfire, Class A2 DaylightRoad2, Class B BQTerrace, Class C PartyScene, Class D BQSquare and Class E KristenAndSara) under QP 22, 27, 32 and 37 to estimate the encoding time as shown in Fig. 11. According to the distributions in Fig. 11, QtDepth 2 and QtDepth 3 occupy approximately 42% and 31% on average, respectively. In other words, the VTM consumes much coding time at QtDepth 2 and QtDepth 3. If we can determine the MTT partition in advance and terminate the following MTT encoding early, a great amount of encoding time will be saved. As a result, a fast MTT decision is desired to accelerate the partitions in QtDepth 2 and QtDepth 3 with w 32 and 16, respectively, in (4)-(5).

The training videos contain the first 100 frames for the 22 sequences of Class A1 to Class E, which are encoded under the all-intra configuration and CTC. Two random forest classifiers (Classifier-H/V for QtDepth 2 and QtDepth 3) are trained to decide horizontal or vertical split. OpenCV 3.2.0 is used to individually train the random forest models to predict the MTT partition. The training accuracies of the random forest models for QtDepth2 and QtDepth3 are 93.16% and 91.85%, respectively. If the partition for MttDepth 1 is accurately selected, unnecessary splits for MttDepth 1~3 can be omitted. Fig. 12 shows the flowchart of the proposed algorithm. During the CTU encoding, if MttDepth is 0 for QtDepth 2 or 3, the JND calculation is performed. The eight projection values of visually distinguishable pixels, $Proj_{hor,l}$, $Proj_{ver,l}$ (l is 0 to 3) and QP are input into the trained random forest model for the prediction of MttDepth 1 of QtDepth 2 and QtDepth 3. For other values of QtDepth and MttDepth, original encoding processing is performed. The

basic idea of the proposed strategy is to skip complex MTT partition structure on some unnecessary CUs by utilizing the characteristics of the human vision system and the QP as the input features of machine learning to train the random forest models.

IV. EXPERIMENTAL RESULTS

The H.266/VVC reference software version VTM 7.0 [34] is used for testing the performance under the all-intra configuration and common test conditions (CTC). The quantization parameters are set as 22, 27, 32 and 37. In our experiments, we use six classes A1, A2, B, C, D, E of the test sequences to contain natural videos with various resolutions and skip the first 100 training frames as shown in Table 1. To compare with [22], we also turn off intra sub-partitions (ISP), matrix weighted intra prediction (MIP) and low-frequency non-separable transform (LFNST) in the configuration file of VTM 7.0 as the setting in [22].

The Bjøntegaard Delta Bitrate (BDBR) [35] is measured to evaluate the coding performance. The performance of the proposed algorithm is compared with [22]. Time saving (TS) is calculated by (6) and coding gain (CG) is computed by (7).

$$\begin{aligned} \text{Time Saving (TS)} &= \frac{1}{4} \sum_{QP_i \in \{22, 27, 32, 37\}} \frac{T_{VTM}(QP_i) - T_{Proposed}(QP_i)}{T_{VTM}(QP_i)}, \end{aligned} \quad (6)$$

where T_{VTM} denotes the total encoding time of the VTM. $T_{Proposed}$ represents the total encoding time of the proposed method. QP_i indicates the QP value of the i^{th} coded video bit stream.

$$\text{Coding Gain (CG)} = \frac{\text{average TS (\%)}}{\text{average BDBR(\%)}}. \quad (7)$$

Study [22] presents good time saving performance for intra coding of H.266/VVC. We make a performance comparison of TS, BDBR and CG of the proposed method with [22] in Table 2. We additionally include UHD sequences (Class A1 and A2) with large resolutions into experiments. The horizontal and vertical projections of visually distinguishable pixels are related to the partition types. The proposed algorithm expresses good results and accurately captures the visual property in video texture. Our method has less average BDBR 1.14% than the 1.63% of [22], less BDBR variation (standard deviation: 0.39%) than the 0.99% of [22] and less TS variation (standard deviation: 1.75%) than the 7.76% of [22]. Less standard deviations for test sequences indicate that more stable and consistent performance can be obtained by our method. For Class E sequences, our approach achieves much less BDBR than that of [22]. Regarding coding gain, the 35.59 of the proposed method is superior to the 30.23 of [22]. In general, our method demonstrates better performance for test sequences with larger resolution and benefits the applications on high resolution video for H.266/VVC. Class A1 Campfire, Class A2 ParkRunning3 and Class B

TABLE 2. Performance comparison of the proposed algorithm and [22].

Sequences	[22]		Proposed Method		
	BDBR (%)	TS (%)	BDBR (%)	TS (%)	
Class A1	Tango2	-	-	1.35	35.29
	FoodMarket4	-	-	1.30	41.76
	Campfire	-	-	0.85	44.33
	Average	-	-	1.17	40.46
Class A2	CatRobot1	-	-	1.38	39.38
	DaylightRoad2	-	-	0.93	40.56
	ParkRunning3	-	-	0.60	42.33
	Average	-	-	0.97	40.76
Class B	BasketballDrive	3.28	59.35	1.61	40.32
	BQTerrace	1.08	45.30	1.06	41.00
	Cactus	1.84	52.44	1.27	41.31
	Kimono	1.93	59.51	1.13	41.15
	ParkScene	1.26	51.84	0.96	42.28
	Average	1.88	53.69	1.21	41.21
Class C	BasketballDrill	1.82	48.48	1.86	40.85
	BQMall	1.87	52.47	1.27	40.01
	PartyScene	0.26	38.62	1.00	42.17
	RaceHorses	0.88	49.05	0.71	42.03
	Average	1.21	47.16	1.21	41.27
Class D	BasketballPass	1.95	47.70	1.53	40.46
	BlowingBubbles	0.47	40.35	0.95	38.34
	BQSquare	0.19	31.95	0.72	40.19
	RaceHorses	0.54	41.69	0.73	40.61
	Average	0.79	40.42	0.98	39.90
Class E	FourPeople	2.70	57.57	1.65	39.73
	Johnny	3.22	56.88	1.80	39.80
	KristenAndSara	2.78	55.11	0.42	38.66
	Average	2.90	56.52	1.29	39.40
Average	1.63	49.27	1.14	40.57	
Standard Deviation	0.99	7.76	0.39	1.75	
Coding Gain(CG)	30.23		35.59		

ParkScene sequences apparently maintain better BDBR than other sequences within the respective class.

Fig. 13 illustrates the MTT partitions based on human perception by the proposed method for the frame in Class D BQSquare sequence. The distribution of visually distinguishable pixels evidently benefits the random forest classifiers and precisely decides the splitting structure. Fig. 14 exhibits the RD curves of VTM 7.0 and the proposed scheme for Class A1 Campfire sequence. The RD curve of the proposed algorithm is very close to that of VTM 7.0, but the 44.33% of time saving is achieved by our method. Fig. 15 and Fig. 16 show the partition comparisons of VTM 7.0 and the proposed algorithm for the frame in Class B Kimono and Class C RaceHorses sequences, respectively. The CTU partition results are similar for both methods, but our proposed approach provides significant time savings. Fig. 17 demonstrates the subjective quality comparison between the VTM 7.0 and the proposed algorithm for the frame in Class A2 ParkRunning3 sequence. The video qualities of the two methods are similar. The structural similarity index measure (SSIM) [36] values are also shown in Fig. 17 to make the performance comparison as closer as possible to human perception. The similar SSIM values between VTM 7.0 and the proposed method show the good visual quality of the proposed method.

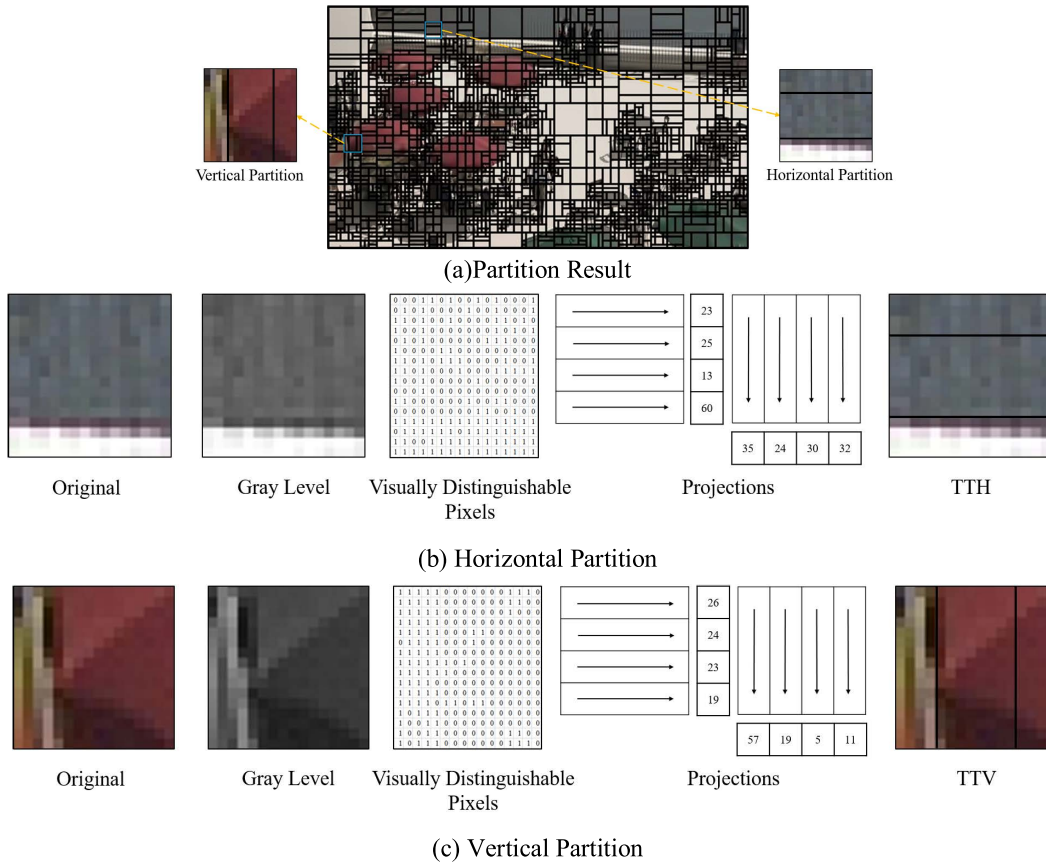


FIGURE 13. Illustration of the proposed method based on human perception for class D BQSquare sequence.

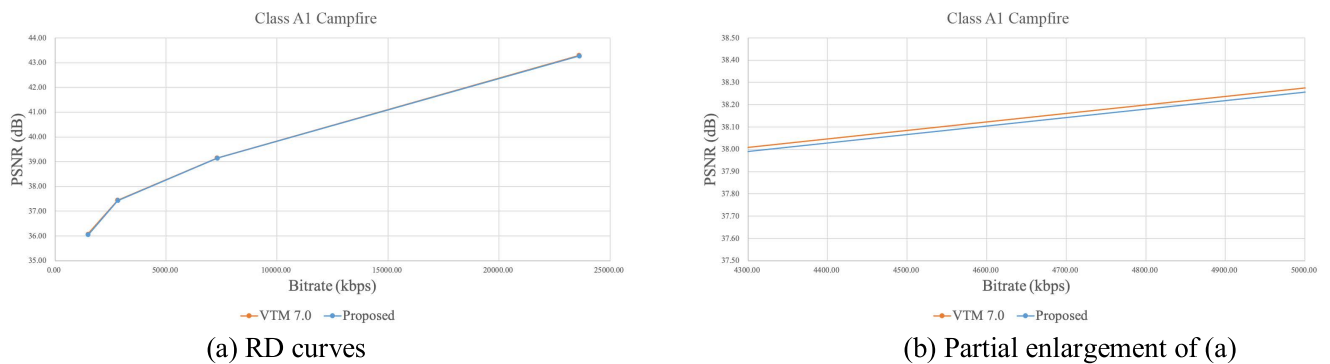


FIGURE 14. Comparison of RD curves of VTM 7.0 and the proposed method for Class A1 Campfire sequence.

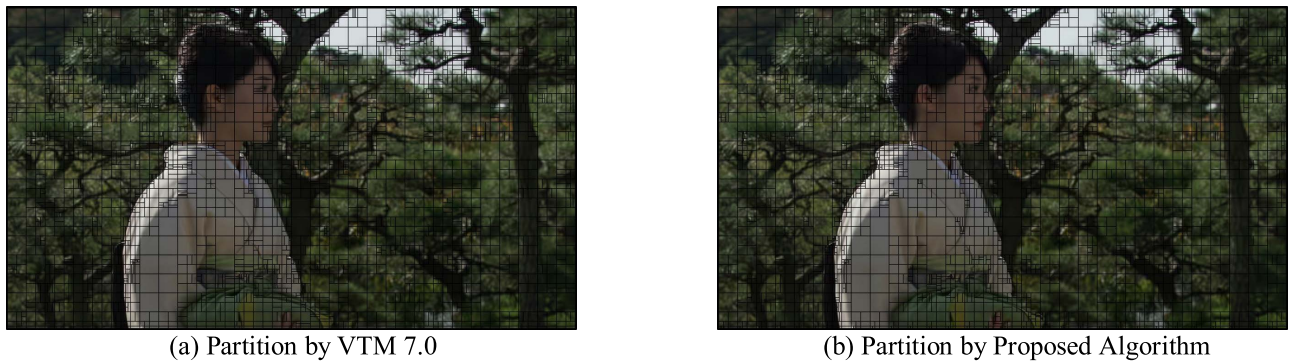
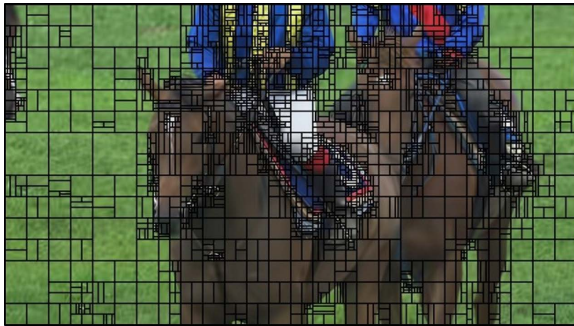


FIGURE 15. Comparison of CU partitions of VTM 7.0 and the proposed method for Class B Kimono sequence.



(a) Partition by VTM 7.0

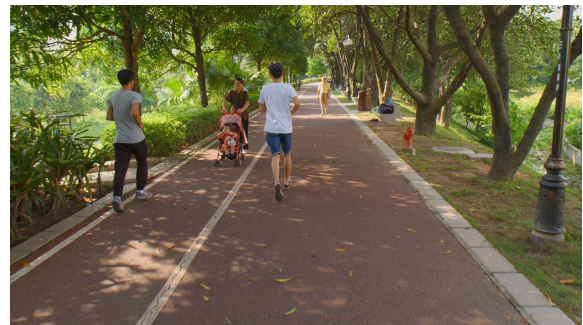


(b) Partition by Proposed Algorithm

FIGURE 16. Comparison of CU partitions of VTM 7.0 and the proposed method for Class C RaceHorses sequence.



(a) VTM 7.0, PSNR=43.5007dB, SSIM=0.97183



(b) Proposed, PSNR=43.4925dB, SSIM=0.97179, TS=42.18%

FIGURE 17. Subjective quality comparison of VTM 7.0 and the proposed method for Class A2 ParkRunning3 sequence.

These comparisons demonstrate that the proposed algorithm accelerates the coding process and maintains good video quality.

V. CONCLUSION

The new QT-MTT structure in H.266/VVC consumes a large amount of encoding time. This paper presents a fast algorithm for H.266/VVC intra coding by taking advantages of the human vision system incorporating the machine learning method. We apply the perceptual model of human eyes, just noticeable difference, to detect the visually distinguishable pixels in a CU. The horizontal and vertical projections of visually distinguishable pixels and the QP are taken as the input features of random forest machine learning models to predict the MTT partition and omit unnecessary splits. Extensive experiments show that the proposed algorithm effectively reduces the encoding time while maintaining good coding efficiency and subjective video quality.

REFERENCES

- [1] B. Bross, J. Chen, J. R. Ohm, G. J. Sullivan, and Y. K. Wang, "Developments in international video coding standardization after AVC, with an overview of versatile video coding (VVC)" *Proc. IEEE*, vol. 109, no. 9, pp. 1463–1493, Sep. 2021.
- [2] B. Bross, J. Chen, S. Liu, and Y. K. Wang, *Versatile Video Coding (Draft 10)*, document JVET-S2001, Version 17, Joint Video Experts Team (JVET), Sep. 2020.
- [3] A. Mercat, A. Mäkinen, J. Sainio, A. Lemmetti, M. Viitanen, and J. Vanne, "Comparative rate-distortion-complexity analysis of VVC and HEVC video codecs," *IEEE Access*, vol. 9, pp. 67813–67828, 2021.
- [4] 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.
- [5] K.-H. Tai, M.-J. Chen, J.-R. Lin, R.-Y. Huang, C.-H. Yeh, C.-Y. Chen, S. D. Lin, R.-M. Weng, and C.-Y. Chang, "Acceleration for HEVC encoder by bimodal segmentation of rate-distortion cost and accurate determination of early termination and early split," *IEEE Access*, vol. 7, pp. 45259–45273, 2019.
- [6] M.-J. Chen, Y.-D. Wu, C.-H. Yeh, K.-M. Lin, and S. D. Lin, "Efficient CU and PU decision based on motion information for interprediction of HEVC," *IEEE Trans. Ind. Informat.*, vol. 14, no. 11, pp. 4735–4745, Nov. 2018.
- [7] M.-J. Chen, J.-R. Lin, Y.-C. Hsu, Y.-S. Ciou, C.-H. Yeh, M.-H. Lin, L.-J. Kau, and C.-Y. Chang, "Fast 3D-HEVC depth intra coding based on boundary continuity," *IEEE Access*, vol. 9, pp. 79588–79599, 2021.
- [8] C.-H. Yeh, Z.-T. Zhang, M.-J. Chen, and C.-Y. Li, "HEVC intra frame coding based on convolutional neural network," *IEEE Access*, vol. 6, pp. 50087–50095, 2018.
- [9] C. Huang, Z. Peng, Y. Xu, F. Chen, Q. Jiang, Y. Zhang, G. Jiang, and Y.-S. Ho, "Online learning-based multi-stage complexity control for live video coding," *IEEE Trans. Image Process.*, vol. 30, pp. 641–656, 2021.
- [10] C. Huang, Z. Peng, F. Chen, Q. Jiang, G. Jiang, and Q. Hu, "Efficient CU and PU decision based on neural network and gray level co-occurrence matrix for intra prediction of screen content coding," *IEEE Access*, vol. 6, pp. 46643–46655, 2018.
- [11] F. Bossen, J. Boyce, K. Sühring, X. Li, and V. Seregin, *VTM Common Test Conditions and Software Reference Configurations for SDR Video*, document JVET-T2010, Nov. 2020.
- [12] T. Fu, H. Zhang, F. Mu, and H. Chen, "Fast CU partitioning algorithm for H.266/VVC intra-frame coding," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Shanghai, China, Jul. 2019, pp. 55–60.
- [13] S.-H. Park and J.-W. Kang, "Context-based ternary tree decision method in versatile video coding for fast intra coding," *IEEE Access*, vol. 7, pp. 172597–172605, 2019.

- [14] Z. Peng, C. Huang, F. Chen, G. Jiang, X. Cui, and M. Yu, "Multiple classifier-based fast coding unit partition for intra coding in future video coding," *Signal Process., Image Commun.*, vol. 78, pp. 171–179, Oct. 2019.
- [15] H. Yang, L. Shen, X. Dong, Q. Ding, P. An, and G. Jiang, "Low-complexity CTU partition structure decision and fast intra mode decision for versatile video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 6, pp. 1668–1682, Jun. 2020.
- [16] T. Amestoy, A. Mercat, W. Hamidouche, D. Menard, and C. Bergeron, "Tunable VVC frame partitioning based on lightweight machine learning," *IEEE Trans. Image Process.*, vol. 29, pp. 1313–1328, 2019.
- [17] J. Zhao, T. Cui, and Q. Zhang, "Fast CU partition decision strategy based on human visual system perceptual quality," *IEEE Access*, vol. 9, pp. 123635–123647, 2021.
- [18] S.-H. Park and J.-W. Kang, "Fast multi-type tree partitioning for versatile video coding using a lightweight neural network," *IEEE Trans. Multimedia*, vol. 23, pp. 4388–4399, 2021.
- [19] S. Peng, Z. Peng, Y. Ren, and F. Chen, "Fast intra-frame coding algorithm for versatile video coding based on texture feature," in *Proc. IEEE Int. Conf. Real-Time Comput. Robot. (RCAR)*, Irkutsk, Russia, Aug. 2019, pp. 65–68.
- [20] N. Tang, J. Cao, F. Liang, J. Wang, H. Liu, X. Wang, and X. Du, "Fast CTU partition decision algorithm for VVC intra and inter coding," in *Proc. IEEE Asia Pacific Conf. Circuits Syst. (APCCAS)*, Bangkok, Thailand, Nov. 2019, pp. 361–364.
- [21] J. Cui, T. Zhang, C. Gu, X. Zhang, and S. Ma, "Gradient-based early termination of CU partition in VVC intra coding," in *Proc. IEEE Data Comp. Conf. (DCC)*, Snowbird, UT, USA, Mar. 2020, pp. 103–112.
- [22] Y. Fan, J. Chen, H. Sun, J. Katto, and M. Jing, "A fast QTMT partition decision strategy for VVC intra prediction," *IEEE Access*, vol. 8, pp. 107900–107911, 2020.
- [23] C.-H. Chou and Y.-C. Li, "A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 6, pp. 467–476, Dec. 1995.
- [24] H. Wang, L. Yu, J. Liang, H. Yin, T. Li, and S. Wang, "Hierarchical predictive coding-based JND estimation for image compression," *IEEE Trans. Image Process.*, vol. 30, pp. 487–500, 2020.
- [25] M.-S. Li, M.-J. Chen, C.-H. Yeh, and K.-H. Tai, "Performance improvement of multi-view video coding based on geometric prediction and human visual system," *Int. J. Imag. Syst. Technol.*, vol. 25, no. 1, pp. 41–49, Mar. 2015.
- [26] J.-R. Lin and M.-J. Chen, "Fast depth coding of 3D-HEVC based on visual perception," in *Proc. Taiwan Acad. Netw. Conf. (TANET)*, Hualien, Taiwan, Oct. 2016, pp. 368–378.
- [27] J.-R. Lin, M.-J. Chen, C.-H. Yeh, Y.-C. Chen, L.-J. Kau, C.-Y. Chang, and M.-H. Lin, "Visual perception based algorithm for fast depth intra coding of 3D-HEVC," *IEEE Trans. Multimedia*, vol. 24, pp. 1707–1720, 2022.
- [28] M.-S. Li, M.-J. Chen, K.-H. Tai, and K.-L. Sue, "Fast mode decision based on human noticeable luminance difference and rate distortion cost for H.264/AVC," *EURASIP J. Adv. Signal Process.*, vol. 2013, Mar. 2013, Art. no. 60.
- [29] M.-C. Chi, C.-H. Yeh, and M.-J. Chen, "Robust region-of-interest determination based on user attention model through visual rhythm analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 7, pp. 1025–1038, Jul. 2009.
- [30] M.-J. Chen, L.-G. Chen, and R.-M. Weng, "Error concealment of lost motion vectors with overlapped motion compensation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 3, pp. 560–563, Jun. 1997.
- [31] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
- [32] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, Mar. 1986.
- [33] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Evanston, IL, USA: Routledge, 2017.
- [34] *VVC Test Model (VTM) Software VTM-7.0*. Accessed: Nov. 13, 2019. [Online]. Available: https://vcgit.hhi.fraunhofer.de/jvet/VVCSsoftware_VTM/-tree/VTM-7.0
- [35] G. Bjontegaard, *Calculation of Average PSNR Differences Between RD-Curves*, ITU-T SG16/Q6 document VCEG-M33, Austin, TX, USA, Apr. 2001.
- [36] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.



MEI-JUAN CHEN (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from National Taiwan University, Taipei, in 1991, 1993, and 1997, respectively.

She was an Assistant Professor, from 1997 to 2000, and an Associate Professor, from 2000 to 2005, at the Department of Electrical Engineering, National Dong Hwa University, Hualien, Taiwan, where she has been a Professor, since August 2005. She also served as the Chair for the Department, from 2005 to 2006. Her research interests include image/video processing, video compression, motion estimation, error concealment, video transcoding, and artificial intelligence. In 1993 and 1997, she was the recipient of the Dragon Paper Awards. She was also the recipient of the 2005 K.T. Li Young Researcher Award from ACM Taipei/Taiwan Chapter for her contribution to video signal codec technique. This award is given annually to only one person under the age of 36, conducting research in Taiwan. In 2006, she received the Outstanding Young Electrical Engineer Award from the Chinese Institute of Electrical Engineering, Taiwan. In 2006, 2016, and 2017, she also received the Excellent Ph.D./Master Thesis Supervision Awards from the Institute of Information and Computer Machinery. In 2005 and 2012, she received the Jun S. Huang Memorial Foundation Best Paper Awards. She received IPPR Society Best Paper Awards in 2013, 2015, and 2016. She received the Best and Excellent Master Thesis Supervision Awards from Taiwan Institute of Electrical and Electronic Engineering in 2011, 2014, 2015, and 2016. In 2014, she received the Best Paper Award from National Symposium on Telecommunication. In 2016 and 2017, she received the Best Paper Awards from Taiwan Academic Network Conference. In 2017, she won the First Place of Best Paper Award in 2017 IEEE International Conference on Consumer Electronics-Taiwan. She became a Fellow of IET in 2019. She received the Outstanding Electrical Engineering Professor Award from the Chinese Institute of Electrical Engineering, Taiwan, in 2020. She won the Second Best Paper Award from 2020 International Computer Symposium. She serves as an Associate Editor for *IEEE ACCESS*, *EURASIP Journal on Advances in Signal Processing*, and *International Journal of Electrical Engineering*.



CHENG-AN LEE received the M.S. degree in electrical engineering from National Dong Hwa University, Hualien, Taiwan, in 2020. His research interests include video coding and machine learning.



YU-HSIANG TSAI received the B.S. degree in electrical engineering from National Dong Hwa University, Hualien, Taiwan, in 2020, where he is currently pursuing the M.S. degree in electrical engineering. His research interests include image/video coding and processing, machine learning, and medical image signal processing.



CHIEH-MING YANG received the B.S. degree in electronic engineering from Ming Chuan University, Taoyuan, Taiwan, in 2015, and the M.S. degree in electrical engineering from National Dong Hwa University, Hualien, Taiwan, in 2018. He is a full-time Research Assistant with the Department of Electrical Engineering, National Dong Hwa University. He is a Technical Manager with the Eastern Taiwan AI and AIOT Academia Industry Alliance. He has coauthored conference and journal papers related to signal processing and artificial intelligence. His research interests include image/video coding and processing, medical signal processing, and big data analysis in artificial intelligence.



LIH-JEN KAU (Senior Member, IEEE) received the B.S. degree in control engineering, and the M.S. and Ph.D. degrees in electrical and control engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1991, 1997, and 2008, respectively. From 1996 to 1998, he was a Senior Technician at Chunghwa Telecom, Taipei, Taiwan. He was an Instructor at the Department of Computer and Communication Engineering, Dahan Institute of Technology, Hualien, Taiwan, in August 1998. Since February 2006, he has been an Assistant Professor with the Dahan Institute of Technology. Since August 2009, he has been with the Department of Electronic Engineering (EE), National Taipei University of Technology (NTUT), Taipei, where he is currently an Associate Professor. His research interests include image or video coding and processing, healthcare and biomedical information technology, and autonomous vehicle system integration. He was a recipient of the 2003 ZyXEL Scholarship Award from ZyXEL Communications Corporation, Hsinchu, the Best Live Demonstration Award from the IEEE Biomedical Circuits and Systems Conference, in 2016, and the Best Paper Award from the IEEE International Conference of Intelligent Applied Systems on Engineering, in 2019.



CHIA-HUNG YEH (Senior Member, IEEE) received the B.S. and Ph.D. degrees from the Department of Electrical Engineering, National Chung Cheng University, Chiayi, Taiwan, in 1997 and 2002, respectively. From August 2002 to December 2004, he was a Postdoctoral Fellow at the Department of Electrical Engineering-Systems, University of Southern California, Los Angeles, CA, USA. He was an Assistant Professor, from 2007 to 2010, an Associate Professor, from 2010 to 2013, and a Professor, from 2013 to 2017, at the Department of Electrical Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan. He is currently a Distinguished Professor with National Taiwan Normal University, Taipei, Taiwan, and the Vice Dean of the College of Technology and Engineering. He has coauthored more than 250 technical international conferences and journal papers and held 47 patents in the USA, Taiwan, and China. His research interests include multimedia, video communication, three-dimensional reconstruction, video coding, image or video processing, and big data. He became a fellow of IET in 2017. He is an Associate Editor of the *Journal of Visual Communication and Image Representation*, *EURASIP Journal on Advances in Signal Processing*, and *APSIPA Transactions on Signal and Information Processing*. He is an active TC Member of the IEEE Communication Society on Multimedia Communication, APSIPA, and IWAIT. He is also one of Founding Member of ACM SIGMM Taiwan Chapter. He is on the Best Paper Award Committee of JVCi and APSIPA. He was the recipient of the 2007 Young Researcher Award of NSYSU, the 2011 Outstanding Young Electrical Engineer Award from the Chinese Institute of Electrical Engineering, the 2013 Distinguished Young Researcher Award of NSYSU, the 2013 IEEE MMSP Top 10% Paper Award, the 2014 IEEE GCCE Outstanding Poster Award, the 2015 APSIPA Distinguished Lecturer, the 2016 NAR Labs Technical Achievement Award, the Superior Achievement Award, the 2017 IEEE SPS Tainan Section Chair, the 2017 Distinguished Professor Award of NTNU, the IEEE Outstanding Technical Achievement Award (IEEE Tainan Section), and the Second Best Paper Award from 2020 International Computer Symposium.



CHUAN-YU CHANG (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from National Cheng Kung University, Tainan, Taiwan, in 2000. From 2009 to 2011, he was the Chair of the Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology (YunTech), Taiwan. From 2011 to 2019, he was the Dean of the Research and Development, the Director of the Incubation Center for Academia-Industry Collaboration and Intellectual Property, YunTech. He is currently the Deputy General Director of the Service Systems Technology Center, Industrial Technology Research Institute, Taiwan. He is also a Distinguished Professor with the Department of Computer Science and Information Engineering, YunTech. He has authored or coauthored more than 200 publications in journals and conference proceedings in his research fields. His research interests include computational intelligence and their applications to medical image processing, automated optical inspection, emotion recognition, and pattern recognition. He is an IET Fellow and a Life Member of IPPR and TAAI. He was the Program Co-Chair of TAAI 2007, CVGIP 2009, 2010–2019 International Workshop on Intelligent Sensors and Smart Environments, and the third International Conference on Robot, Vision and Signal Processing (RVSP 2015). He was the General Co-Chair of 2012 International Conference on Information Security and Intelligent Control, from 2011 to 2013 Workshop on Digital Life Technologies, CVGIP 2017, WIC 2018, ICS 2018, and WIC 2019. From 2015 to 2017, he was the Chair of the IEEE Signal Processing Society Tainan Chapter and the Representative for Region 10 of IEEE SPS Chapters Committee. He received the National Award for Distinguished Contribution to Industry-Academia Cooperation from Ministry of Education, and the Outstanding Electrical Engineering Professor Award from the Chinese Institute of Electrical Engineering, Taiwan, in 2021. He is also the President of the Chinese Image Processing and Pattern Recognition Society.

...