IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 23 February 2024, accepted 21 March 2024, date of publication 27 March 2024, date of current version 9 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3382570

RESEARCH ARTICLE

KSVM-Based Fast Intra Mode Prediction in HEVC Using Statistical Features and Sparse Autoencoder

PREETHI S. NAIR[1] **AND MADHU S. NAIR**[2], **(Senior Member, IEEE)**

[1]Department of Computer Science, University of Kerala, Kariavattom, Thiruvananthapuram, Kerala 695581, India

[2]Artificial Intelligence & Computer Vision Lab, Department of Computer Science, Cochin University of Science and Technology, Kochi, Kerala 682022, India

Corresponding author: Preethi S. Nair (preethi.snair85@gmail.com)

This work was supported by Kerala State Council for Science, Technology and Environment (KSCSTE) under Grant 079/FSHP-ENG/2013/KSCSTE, Grant 746/2014/KCSTE, Grant 960/2015/KSCSTE, Grant (M)189/2015/KSCSTE, and Grant 538/2018/KSCSTE.

ABSTRACT High Efficiency Video Coding (HEVC) is designed to deliver a video communication with better quality at reduced bit rate. For intra coding, HEVC employs an effective hierarchical quad tree partitioning and an exhaustive optimal mode search which increases the time complexity. Aiming this issue, we propose a Support Vector Machine (SVM)-based method to effectively predict the intra mode. Compared to the standard HEVC encoder HM-15.0, the proposed method could reduce 57.6% of encoding time at a bit-rate penalty of 3.3% at an average PSNR decline of only around 0.09 dB.

INDEX TERMS HEVC, intra mode prediction, RMD, SVM, Hu moments, sparse autoencoder.

I. INTRODUCTION

The explosive growth in visual communications and data transfers increases the demand for efficient video compression standards. The High Efficiency Video Coding (HEVC) [1], jointly developed by the ISO/IEC MPEG (Moving Picture Experts Group) and ITU-T VCEG (Video Coding Experts Group) achieved a world wide acceptance than its predecessors by incorporating many new techniques like quad tree partitioning, increased number of prediction modes, and exhaustive searches for optimal mode prediction in intra coding. These enabled the encoders to deliver a better output visual quality by using only one half of the bit-rate than those existed. For the past few decades, a variety of complexity reduction algorithms –(heuristic methods [2], [3], [4], [25], [26]), machine learning methods [5], [6], [7], [8]), deep learning methods [9], [10], [11] – are used to improve the quality of HEVC video coding with reduced computational complexities in terms of bit-rate, time and visual quality. This letter proposes a learned model to predict the intra mode effectively using a non-linear kernel based SVM classifier for classifying the optimal PU mode. In this work:

1) We explored statistical features to reduce the computational overhead in RMD process and to effectively represent the different sized PUs uniquely.
2) We employed a sparse auto encoder as a feature representation tool. Even though many CNN and other deep neural network based researches claim good performance in HEVC intra mode coding, sparse autoencoder based feature extraction or classification has not been used in any of the HEVC intra mode decision methods.
3) Most of the existing intra mode coding works based on machine learning classification algorithms generally classifies local as well as global feature descriptors. We used a kernel SVM classifier for intra mode prediction in RMD, which uses the features extracted from a sparse autoencoder as input.

The rest of the paper is organised as follows. Section II briefs the proposed work based on SVM. Section III details

The associate editor coordinating the review of this manuscript and approving it for publication was Zahid Akhtar.

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

the experimental analysis on the test results and Section IV concludes the paper.

II. LEARNING BASED FAST INTRA MODE PREDICTION

Feature extraction, SVM classification and Intra mode predictions form the basic foundation of our proposed model, which are explained in the succeeding sections.

A. FEATURE EXTRACTION

Statistical feature extraction method explains the local features of an image from individual pixels as well as the spatial correlations among the neighbouring pixels. Also it is computationally faster, low dimensional and good at describing simple shapes. Recently, deep learning frameworks like sparse autoencoders showed great potential to learn high-level feature representation of raw input data [12], [13], [14]. Hence, we used the statistical Hu moments features and high level latent features from sparse autoencoder in our method.

B. HU MOMENTS FEATURES

Simply, image moments are set of statistical parameters usually used to characterize the shape of an object in an image. They are the weighted average of the pixel intensities in an image. Hu [15] proposed the 7 invariants of moment which are calculated using the normalised central moments of an image. Since HEVC followed different PU depths, each PU needs to be represented with a generic dimension. So we use Hu moments to represent each PU as a vector having seven values. Also, the encoding time of a video codec has significant impact on the real time video processing, this feature descriptor can be a better choice. Hence the statistical feature vector in our method is defined as $\{HuM_1, HuM_2, HuM_3, HuM_4, HuM_5, HuM_6, HuM_7\}$, where each HuM_i defines each of the Hu moment values.

C. SPARSE AUTOENCODER FEATURES

Autoencoders [16] are unsupervised neural networks, which can learn hidden features from the input data and regenerates it as the output. To better learn the features and their latent representation, a sparse penalty term is added to the typical autoencoder and is named as a Sparse Autoencoder (SAE). Also, under sparse constraints, the autoencoder can extract more efficient low dimensional features to better represent the data [17]. Hence we consider the SAE as a feature extraction tool in our work. Moreover, the single layer networks proved their efficiency in learning strong feature representations from unlabeled data [13], we also utilised a single layer SAE as an unsupervised method for a better feature representation. We input the Hu moments feature vector corresponding to each PU to the SAE. The encoder takes X = $\{HuM_1, HuM_2, HuM_3, HuM_4, HuM_5, HuM_6, HuM_7\}$ as the input, which is encoded by the hidden layer with seven neurons and produce high level sparse features, $\mathrm{hl}_i^{(1)}$. The high level features can be defined as $\mathrm{hl}_i^{(1)} = \sigma(W^{(1)}x_i + b^{(1)})$, is the output of the hidden layer. Here, hl_i is the hidden layer representation, σ defines the logistic sigmoid function, W and b represents the weight matrix and bias vector, respectively. Since we incorporate the sparse autoencoder for a latent feature representation rather than dimensionality reduction of features, we used seven hidden neurons to obtain the high level features having same dimension as the input vector. We used the default values for the sparsity parameter ρ, sparsity regularisation parameter β and the weight attenuation parameter λ which are 0.0001, 0.05 and 4, respectively. The latent features obtained from the SAE and the optimal mode of each PU are treated as the input for the supervised SVM classifier.

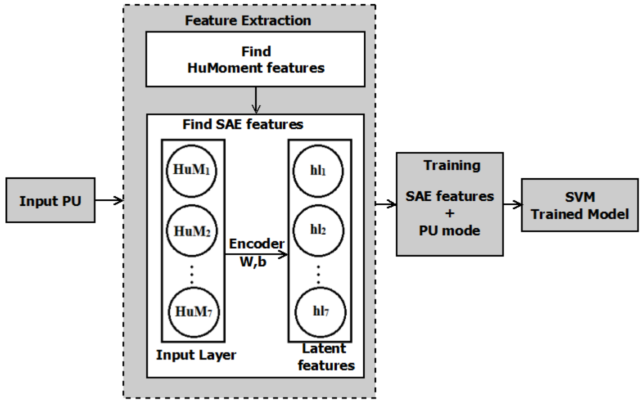

FIGURE 1. Training phase in the proposed method for a single QP value.

D. SVM CLASSIFICATION

In our work, since we have to predict an optimal intra mode from 35 mode set, SVM [18] classifier is adopted for a multi class classification. In the proposed method, each sparse autoencoder feature vector has a dimension $d = 7$ and we have a 35 class classification problem, nonlinear RBF (Radial Basis Function) kernel SVM is selected for a better classification. Also C (regularization parameter) and *gamma* (defines the curvature weight of the decision boundary) are the hyperparameters used to train the SVM model. For an optimal classification, these parameters need to be fine tuned. In our work, the training samples are separately cross validated to obtain the optimized *gamma* and C parameters for the four Quantisation Parameters(QP). For QP={22,27,32,37}, the corresponding $\{gamma, C\}$ values are {{312.5,0.50625}, {100,0.30625}, {100,0.3375}, {100,0.50625}}, respectively.

E. INTRA MODE PREDICTION

In the proposed classification based method, an off-line training is carried out separately for QP = {22,27,32,37} and four different trained models were build. These trained models are incorporated into the reference HEVC codec to further improve the coding efficiency. Video frames are randomly chosen for training, since adjacent frames hold much similar block patterns. To label each PU with

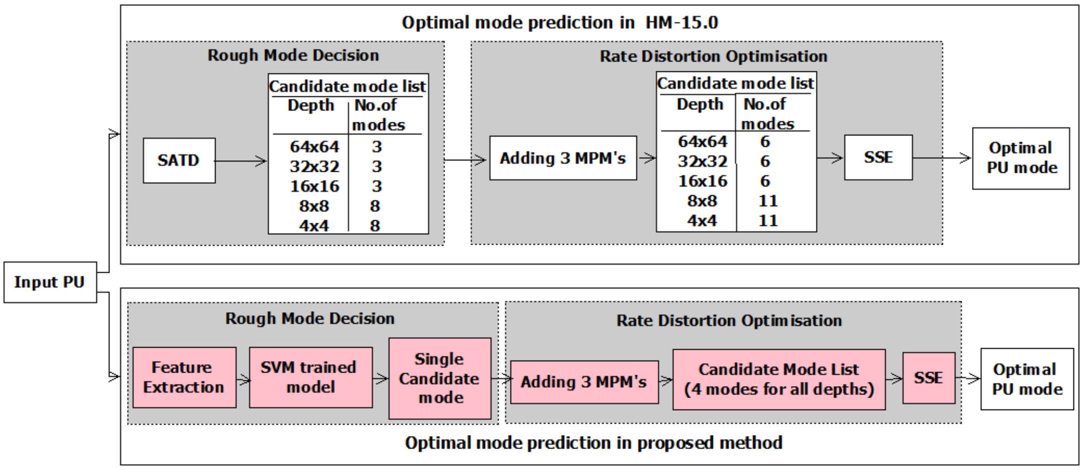

FIGURE 2. Differentiating the optimal mode prediction in HM-15.0 and the proposed SVM method.

corresponding modes, the reference codec is configured to execute with full RD optimization. The optimal intra mode for each PU chosen by the encoder is stored as actual mode of the PU.

After training, the intra mode prediction in the RMD stage is performed based on the trained SVM models. For each PU, the proposed algorithm uses a sparse auto encoder to get the feature vector, $fv = \{hl_1, hl_2, hl_3, hl_4, hl_5, hl_6, hl_7\}$ from the Hu moments features $\{HuM_1, HuM_2, HuM_3, HuM_4, HuM_5, HuM_6, HuM_7\}$ of the PU. Based on the extracted input feature vector fv, the trained SVM classifier performed an intra mode decision, which outputs a single value as the intra mode. Thus, the optimal intra mode, $L(fv)$ can be decided based on the function:

$$L(fv) = \arg\max_{m}(P(m \mid fv)), \tag{1}$$

where $L(fv) \in \{0,1,2,\ldots,34\}$. The posterior probability P_m [19] for each fv in the m^{th} class, can be described as:

$$P_m = P(m \mid fv), m = \{0, 1, 2, \ldots, 34\}. \tag{2}$$

In addition to the RMD mode, three MPMs are added to prepare the RDO candidate list. Based on the RD cost calculation for each candidate modes, an optimal intra mode coding is performed with a mode having least RD cost value. The training procedure of the proposed method is illustrated in FIGURE 1 and FIGURE 2 differentiates the optimal mode decision procedure in HM-15.0 and the proposed method. From FIGURE 2, it can be noted that, in our method, a single mode is predicted as candidate from the RMD stage instead of 8 or 3 modes in anchor codec, irrespective of the PU depth. In addition, the candidate mode list for RDO process gets filled with only 4 modes (3 MPMs plus one RMD mode), while in the anchor codec, this will be 11 or 6 with respect to PU depth. Hence the proposed model exhibits a huge reduction in the computational complexity. The pseudocode of the proposed method is outlined in Algorithm 1.

Algorithm 1 SVM-Based Intra Mode Prediction Method

1 : **Input**: High level sparse autoencoder features.
2 : **Output**: Optimal Intra mode from RMD.
3 : Start with current PU.
4 : Determine the Hu Moment features, HuM=$\{HuM_1, HuM_2, HuM_3, HuM_4, HuM_5, HuM_6, HuM_7\}$ of PU.
5 : Derive the high level sparse feature vector,fv=$\{hl_1, hl_2, hl_3, hl_4, hl_5, hl_6, hl_7\}$, from the HuM.
6 : Perform SVM based intra mode prediction by Eq. (2).
7 : Calculate RD cost for RMD mode, CF_{mode} by Eq. (1).
8 : Perform Rate Distortion Optimisation to the RDO candidate modes (3 MPM + 1 RMD mode).
9 : Select the optimal intra mode for the PU.
10 : Proceed with the next PU.

The major modification incorporated in this method is the usage of statistical features to get the input image details. Since statistical features are computationally efficient, we used the Hu moments to get the features of the input image. For each image, instead of considering the 35 modes one by one, our method extracted the image features using Hu moment values. This calculation reduces the no: of executions from 35 to 1. Also, we calculated the RMD cost only once for each image. But in anchor HM 15.0, it was calculated 35 times for a single image, as it considers all the 35 modes for each PU partitions.

III. EXPERIMENTAL RESULTS AND ANALYSIS

For the experimental analysis, we used the HEVC reference software, HM-15.0 and followed the common test conditions recommended by JCT-VC [20]. All-intra main configuration with QP = {22,27,32,37} are selected for the evaluation. Test sequences from Class A, B, C, D and E are used for training and testing purposes. First 100 frames from sequences SteamLocomotiveTrain, Cactus, ParkScene,

TABLE 1. Performance comparison based on BD-BR (%), BD-PSNR (dB) and ΔT (%) of our method and anchor HEVC codec HM-15.0 at QP=22,27,32,37.

Sequences	QP	HM-15.0 versus Proposed method			
		BD-BR	BD-PSNR	ΔT_{RMD}	ΔT_{RDO}
SteamLocomotive (Class A)	22	0.39	-0.0056	-58.41	-40.18
	27	0.94	-0.0025	-57.35	-37.52
	32	1.77	-0.0027	-56.19	-37.36
	37	2.20	-0.0033	-55.50	-36.40
Average		1.32	-0.0035	-56.86	-37.86
Cactus (Class B)	22	1.76	-0.1178	-59.69	-51.97
	27	2.15	-0.0439	-59.48	-49.23
	32	2.29	-0.0702	-59.60	-47.61
	37	3.94	-0.0768	-60.27	-50.83
Average		2.53	-0.0771	-59.76	-49.91
ParkScene (Class B)	22	0.78	-0.0912	-58.85	-51.10
	27	1.10	-0.0605	-58.72	-50.19
	32	1.39	-0.0735	-58.37	-34.90
	37	2.20	-0.0532	-58.94	-40.46
Average		1.34	-0.0696	-58.72	-44.16
PartyScene (Class C)	22	2.56	-0.1727	-55.31	-55.39
	27	3.09	-0.1410	-54.44	-42.70
	32	3.56	-0.1887	-58.72	-55.26
	37	4.36	-0.1514	-54.97	-48.73
Average		3.39	-0.1634	-55.86	-50.52
BlowingBubbles (Class D)	22	4.94	-0.1869	-56.74	-53.07
	27	5.69	-0.1159	-55.65	-58.75
	32	6.82	-0.1516	-56.50	-43.28
	37	6.39	-0.1399	-54.96	-50.56
Average		5.96	-0.1485	-55.96	-51.41
FourPeople (Class E)	22	3.76	-0.0655	-61.12	-51.20
	27	5.62	-0.0712	-56.49	-48.36
	32	5.74	-0.1204	-56.39	-47.41
	37	7.35	-0.1563	-59.90	-51.16
Average		5.61	-0.1036	-58.47	-49.53
Total Average		3.35	-0.09	-57.6	-47.23

PartyScene, Blowing Bubbles and FourPeople are selected for testing. The experiments are performed on Intel(R) Core(TM) i7-4510U CPU @ 2.60GHz with a memory of 8 GB, and Windows 8.1 64-bit operating system.

A. PERFORMANCE EVALUATION AND ANALYSIS

In this section, we present the coding efficiency of our method implemented in the anchor HEVC reference software HM 15.0, as compared to some of the state-of-the-art methods reported in the literature. Bjontegaard Delta rate (BD-BR), the Peak Signal to Noise Ratio (BD-PSNR) [21] were used to evaluate the performance of the proposed method. We also used the PSNR difference ($\Delta PSNR$), the percentage difference in bitrate (ΔBR) for the result comparison.

$$\Delta PSNR = PSNR_{new} - PSNR_{anchor} \tag{3}$$

$PSNR_{new}$ and $PSNR_{anchor}$ defines the PSNR values of the proposed method and HM-15.0, respectively.

Here, the PSNR values of each picture is represented as the weighted sum of peak signal-to-noise ratio of the luma($PSNR_Y$) and chroma components($PSNR_U$ and $PSNR_V$) and is calculated as follows.

$$PSNR = \frac{6 \times PSNR_Y + PSNR_U + PSNR_V}{8} \tag{4}$$

Since human's eye sensitivity is higher for luma component than chroma component, usually higher weights are given to luma than chroma.

$$\Delta \text{Bit rate} = \frac{\text{Bit rate}_{new} - \text{Bit rate}_{anchor}}{\text{Bit rate}_{anchor}} \times 100 \tag{5}$$

$Bit\ rate_{new}$ and $Bit\ rate_{anchor}$ defines the bit rate of the proposed method and the reference software HM-15.0, respectively. Besides, ΔT is used to evaluate the encoding time reduction and is calculated as:

$$\Delta T = \frac{T_{new} - T_{anchor}}{Tanchor} \times 100 \tag{6}$$

Here, ΔT refers the encoding time difference between the encoded videos using the anchor HEVC software and the newly proposed one. T_{new} and T_{anchor} defines the encoding time of proposed method and anchor codec, respectively. The comparative results of our proposed method with the anchor HM 15.0 is described in TABLE 1. Here, BD-BR, BD-PSNR, ΔT_{RMD}, ΔT_{RDO} are the differences between the bit rates, PSNR values, RMD time reduction and RDO time reduction as compared with the reference HM 15.0, respectively. The results shows that our classification based method saves the encoding time by 57.6%, on an average with a bit rate and PSNR loss of about 3.35% and 0.09 dB, respectively as compared to the standard HEVC reference software. The main reason for the time savings of our method is that, we select statistical features to reduce the time in feature extraction process, as well as, the kernel SVM classifier is used to predict a single mode in the RMD process. Instead of performing the computationally complex operations like SATD, RD cost etc. on 35 iterations for each PU, the encoder only needs to perform the same for a single mode in the RMD stage and for four modes in the RDO stage. Rather than focussing on the bit rate improvement, we emphasis on reducing the computational complexity of the encoder in intra mode prediction. Our method achieved a maximum time reduction of about 59.76% and 51.41% on average in the RMD and RDO processes, respectively. The method also preserves the visual quality of the videos, since the PSNR loss is only about 0.09% on average. So, it is obvious that our method could reduce a good percentage of encoding time in both RMD and RDO stages without degrading the reconstruction quality of the output videos. The bit rate increment of our method is 3.35% on average as compared with the reference HEVC software. It is also important to note that our method presents a good performance in Class A and B, which includes high definition sequences. The average time savings for all the four QP values of SteamLocomotive, Cactus and ParkScene are 56.86%, 59.76% and 56.72%, respectively with only a negligible PSNR decline of about 0.0035dB, 0.0771dB and 0.0696dB, respectively. The aforementioned sequences

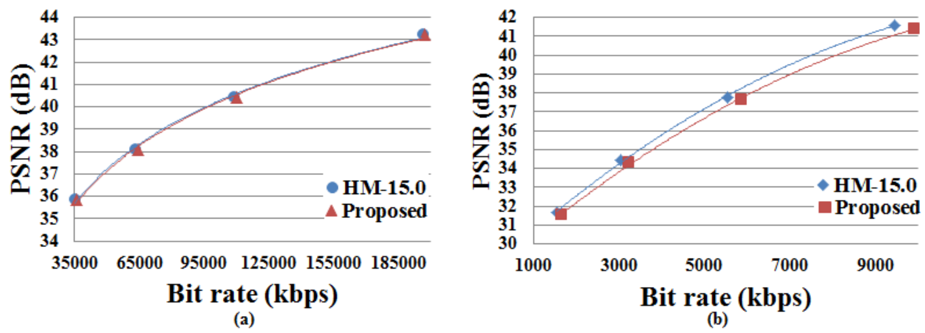

FIGURE 3. RD curve comparison of our method over HM-15.0 for QP=22,27,32,37. (a): SteamLocomotive sequence. (b): BlowingBubbles sequence.

possess only an acceptable bit rate increments of about 1.32%, 2.53% and 1.34% on average, respectively. Thus, it can be concluded that the proposed algorithm is efficient to work with high definition videos without any degradation in terms of quality. From the results, it is observed that the bit rate loss for Class D and E sequences are about 5.96% and 5.61% on average, respectively. But it is satisfactory as the sequences could achieve a considerable time savings of about 55.86% and 55.96%, respectively. While comparing the overall performance, the new method is suitable for real time videos, as it could deliver a reliable performance in terms of time savings and reconstruction quality.

The superiority of the proposed method is highlighted in FIGURE 3. It depicts the Rate-Distortion (RD) curves for Class A (high dimension) and Class D (low dimension) sequences. FIGURE 3 (a) demonstrates the Bit rate-PSNR values of the Steam Locomotive sequence, which offers the best results in our method. Also, it is clear that the bit rate increment and PSNR decline is much less of about 1.32% and 0.0035% on average, respectively for Class A sequence, where there is no notable change in both values compared to the reference HEVC encoder. FIGURE 3 (b) illustrates the Bit rate-PSNR values of the Blowing Bubbles sequence, which indicates a performance degradation in the proposed method. The bit rate loss is 5.96% and the PSNR value reduced to 0.14%, for Class D sequence and therefore the figure shows an admissible difference in the corresponding values. From the figures, it can be observed that for Steam Locomotive sequence, the difference of RD performance between the HM-15.0 and our approach is very small for all bit-rate points. It is similar for the sequences in Class B also. This shows that our approach can adapt to the different bit-rate points admirably, for the high-definition sequences. For the low definition sequence, Blowing Bubble, the RD-curve shows an admissible difference. The difference in the RD performance is getting wider with increasing QP values. This may be due to the incorrect intra mode predictions made to the PUs in the specific sequences. As it is increasingly challenging to retain a good bit rate with an increased time savings, the bit rate loss in our method is quite impressive considering an average time savings around 57%. FIGURE 4 illustrates the time savings

TABLE 2. Performance comparison of our method and other algorithms for fast intra mode decision in HEVC.

Method	BD-BR(%)	BD-PSNR(dB)	ΔT(%)
Kim et al. [22]	0.02	-0.02	-37.93
Chen et al. [23]	0.17	Not given	-21.28
Zhuang et al. [24]	1.85	-0.04	-45.04
Tariq et al. [4]	1.46	-0.08	-30.53
Tariq et al. [27]	1.52	-0.09	-32.57
Tariq et al. [28]	1.87	-0.11	-34.33
Our method	3.3	-0.09	-57.6

of our method over HM-15.0. Since we have only a single RMD mode, it will have an impact on the RDO time also. Normally, HEVC has to consider a maximum of 11 modes (based on PU depth) in RDO process. Irrespective of PU depth, this is reduced to 4 modes in our method, thereby we can achieve time savings in RDO process also. FIGURE 4(a) shows the encoding time differences in rough mode decision process of HM-15.0 and the proposed one for different QP values. FIGURE 4(b) shows the RDO time of our method and HM-15.0 at different QP values. It can be noted that, the proposed method could achieve a minimum RMD time savings of 54.96% for BlowingBubbles sequence at QP=37 and a maximum value of 60.27% for Cactus at QP=22. Also, the RDO time savings ranges between 36.4% and 58.75% for the sequences, SteamLocomotive and BlowwingBubbles at different QP values. Overall, our method achieved an average time savings of about 57.6% and 48.73% on RMD and RDO processes, respectively.

TABLE 2 compares the performance of our result with some of the HEVC fast intra mode decision algorithms. It can be observed that our method possess a bit rate drop of about 3.35%. But the loss is acceptable, since our method possess a good encoding time reduction of about 19.67%, 36.32%, 12.56%, 27.07%, 25.03% and 23.27% over the methods in [4], [22], [23], [24], [27], and [28], respectively.

We also consider the performance of the encoder by excluding the hidden feature representation by SAE from the proposed method. The method could reduce 73.53% of encoding complexities as compared to HM-15.0 and 15.93% reduction on comparing with the proposed one.

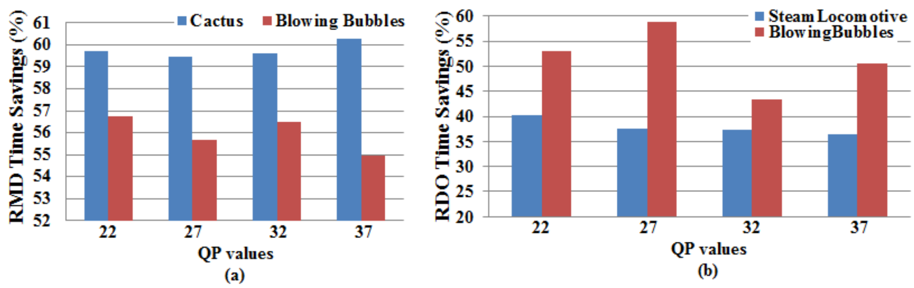

FIGURE 4. Encoding Time reduction comparison of our method over HM-15.0. (a): RMD Time Savings on Cactus and BlowingBubbles sequences. (b): RDO Time Savings on SteamLocomotive and BlowingBubbles sequences.

On comparing with the anchor, this method shows a BD-BR loss of around 5.89% on average. The PSNR value is also degraded to 0.14 dB, which is 0.09 dB for the proposed method. So, it can be concluded that the proposed method outperforms the later method by reducing the average bit rate loss to 3.35% with an improved PSNR value of 0.09 dB.

From the experimental analysis and comparisons, the proposed kernel SVM-based fast intra mode prediction method proved its efficiency in terms of time savings and visual quality with acceptable bit rate drop. Also, it is important to note that our method could achieve impressive results on high dimensional input sequences. Therefore, the method is more suitable for real time applications with high dimensional quality.

IV. CONCLUSION

In summary, research on the complexity reduction in HEVC intra coding methods still needs to be more deepened. On the grounds, a novel classification method for intra mode decision is proposed in this letter. Instead of using the low level features directly, we exploited sparse autoencoder for better feature representation. Then we replaced the exhaustive search in the intra mode selection of standard HEVC with an RBF kernel SVM-based intra mode prediction. The purpose of this approach is to reduce the extensive computational time in the RMD stage by following a single mode prediction rather than predicting a candidate mode list. The proposed method exhibits a better performance in terms of computational time reduction of about 57.6% on average, without compromising any degradation in the visual quality of the output.

Even though we could achieve an increased time reduction without compromising the video quality, we could not reduce the bit rate to a greater extent for some sequences. This may be due to the incorrect mode predictions for some sequences, particularly low definition sequences. Here we only considered the reduction in number of intra modes. We can perform the same by considering the PU partitions also. If we consider an optimized CU partition, the same method can be used for the mode prediction to reduce execution time to a better extend. i.e, for each image, if we could select the single relevant PU, then we only want to perform the mode prediction for that PU, which will reduce the execution time further. Also, we plan to include deep learning methods to reduce the intra modes in the RMD as well as RDO modules in future.

REFERENCES

[1] T. Wiegand, J. R. Ohm, G. J. Sullivan, W. J. Han, R. Joshi, T. K. Tan, and K. Ugur, "Special section on the joint call for proposals on high efficiency video coding (HEVC) standardization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, pp. 1661–1666, Dec. 2010.

[2] J. Wang, F. Lin, J. Zhao, and Y. Long, "Texture grouping and statistical optimization based mode prediction decision algorithm for fast HEVC intra coding," Res. Square, Tech. Rep., 2021.

[3] J. Tariq, A. Armghan, A. Ijaz, and I. Ashraf, "Light weight model for intra mode selection in HEVC," *Multimedia Tools Appl.*, vol. 80, no. 14, pp. 21449–21464, Jun. 2021.

[4] J. Tariq, A. Alfalou, A. Ijaz, H. Ali, I. Ashraf, H. Rahman, A. Armghan, I. Mashood, and S. Rehman, "Fast intra mode selection in HEVC using statistical model," *Comput., Mater. Continua*, vol. 70, no. 2, pp. 3903–3918, Jan. 2022.

[5] P. S. Nair, K. R. Rao, and M. S. Nair, "A machine learning approach for fast mode decision in HEVC intra prediction based on statistical features," *J. Intell. Fuzzy Syst.*, vol. 36, no. 3, pp. 2095–2106, Mar. 2019.

[6] I. Werda, A. Maraoui, F. Belghith, E. F. Sayadi, and N. Masmoudi, "HEVC coding unit decision based on machine learning," *Signal, Image Video Process.*, vol. 16, no. 5, pp. 1345–1353, Jul. 2022.

[7] A. Feng, C. Gao, L. Li, D. Liu, and F. Wu, "CNN-based depth map prediction for fast block partitioning in HEVC intra coding," in *Proc. IEEE Int. Conf. Multimedia Expo. (ICME)*, Jul. 2021, pp. 1–6.

[8] J. Tariq, A. Ijaz, A. Armghan, H. Rahman, H. Ali, and F. Alenezi, "HEVC's intra mode process expedited using histogram of oriented gradients," *J. Vis. Commun. Image Represent.*, vol. 88, Oct. 2022, Art. no. 103594.

[9] F. Zaki, A. E. Mohamed, and S. G. Sayed, "CtuNet: A deep learning-based framework for fast CTU partitioning of H265/HEVC intra-coding," *Ain Shams Eng. J.*, vol. 12, no. 2, pp. 1859–1866, Jun. 2021.

[10] H. Man, X. Fan, R. Xiong, and D. Zhao, "Tree-structured data clustering-driven neural network for intra prediction in video coding," 2021, *arXiv:2106.05481*.

[11] W. Imen, M. Amna, B. Fatma, S. F. Ezahra, and N. Masmoudi, "Fast HEVC intra-CU decision partition algorithm with modified LeNet-5 and AlexNet," *Signal, Image Video Process.*, vol. 16, no. 7, pp. 1811–1819, Oct. 2022.

[12] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[13] E. Li, P. Du, A. Samat, Y. Meng, and M. Che, "Mid-level feature representation via sparse autoencoder for remotely sensed scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 3, pp. 1068–1081, Mar. 2017.

[14] B. Abraham and M. S. Nair, "Computer-aided classification of prostate cancer grade groups from MRI images using texture features and stacked sparse autoencoder," *Computerized Med. Imag. Graph.*, vol. 69, pp. 60–68, Nov. 2018.

[15] M.-K. Hu, "Visual pattern recognition by moment invariants," *IEEE Trans. Inf. Theory*, vol. IT-8, no. 2, pp. 179–187, Feb. 1962.

[16] D. Bank, N. Koenigstein, and R. Giryes, "Autoencoders," 2020, *arXiv:2003.05991*.
[17] B. Yan and G. Han, "Effective feature extraction via stacked sparse autoencoder to improve intrusion detection system," *IEEE Access*, vol. 6, pp. 41238–41248, 2018.
[18] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 988–999, Sep. 1999.
[19] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *J. Mach. Learn. Res.*, vol. 5, pp. 975–1005, Aug. 2004.
[20] F. Bossen, "Common test conditions and software reference configurations," in *Proc. 5th Meeting Joint Collaborative Team Video Coding*, Jan. 2011, p. 1.
[21] G. Bjontegaard, *Calculation of Average PSNR Differences Between RD-Curves*, Standard VCEG-M33, 2001.
[22] T. S. Kim and M. H. Sunwoo, "A fast mode decision algorithm using hierarchical and skip methods for intra prediction in HEVC," *J. Signal Process. Syst.*, vol. 91, no. 6, pp. 663–671, Jun. 2019.
[23] Z. Chen, J. Shi, and W. Li, "Learned fast HEVC intra coding," *IEEE Trans. Image Process.*, vol. 29, pp. 5431–5446, 2020.
[24] H. Zhuang, D. Fan, L. Liu, R. Wang, and C. Sui, "A fast intra prediction decision algorithm for HEVC based on image texture complexity," in *Proc. 9th Int. Conf. Commun. Broadband Netw.*, Feb. 2021, pp. 60–65.
[25] N. Najafabadi and M. Ramezanpour, "Mass center direction-based decision method for intraprediction in HEVC standard," *J. Real-Time Image Process.*, vol. 17, no. 5, pp. 1153–1168, Oct. 2020.
[26] B. Heidari and M. Ramezanpour, "Reduction of intra-coding time for HEVC based on temporary direction map," *J. Real-Time Image Process.*, vol. 17, no. 3, pp. 567–579, Jun. 2020.
[27] J. Tariq, M. Javed, B. Ayub, A. Armghan, A. Ijaz, F. Alenezi, H. Rahman, and A. Zulfiqar, "Nature inspired algorithm based fast intra mode decision in HEVC," *Multimedia Tools Appl.*, vol. 82, no. 19, pp. 29789–29804, Aug. 2023.
[28] J. Tariq, M. Javed, H. Rahman, A. Armghan, and A. Ijaz, "AI application in video: Spiral optimizer based fast intra mode selection in HEVC," *Multimedia Tools Appl.*, pp. 1–16, Jan. 2024.

PREETHI S. NAIR received the M.Sc. degree in computer science from the University of Kerala, India. She is currently a Research Scholar with the Department of Computer Science, University of Kerala. Her research interests include image processing, machine learning, and video compression.

MADHU S. NAIR (Senior Member, IEEE) received the B.C.A. and M.C.A. degrees (Hons.) from Mahatma Gandhi University, in 2000 and 2003, respectively, the M.Tech. degree (Hons.) in computer science from the University of Kerala, in 2008, and the Ph.D. degree in computer science from Mahatma Gandhi University, in 2013. He had also qualified UGC-NET Examination for Lectureship and GATE Examination conducted by IIT. He has published 105 research papers in reputed international journals and conference proceedings published by IEEE, Springer, Elsevier, Wiley, and IOS Press. His research interests include digital image processing, pattern recognition, computer vision, data compression, soft computing, and scientometrics. He is a Senior Member of ACM, an Associate Life Member of CSI, and a member of IAENG. He has received prestigious international and national awards from different professional societies.