FAST CODING UNIT DECISION FOR INTRA SCREEN CONTENT CODING BASED ON ENSEMBLE LEARNING

Yali Xue[1], *Xu Wang*[1*], *Linwei Zhu*[2], *Zhaoqing Pan*[3], *Sam Kwong*[4]

1 College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China
2 Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
3 School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, China
4 Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong

ABSTRACT

The Screen Content Coding (SCC) is an extension of High Efficiency Video Coding (HEVC), and it achieves significant improvement on compression ratio. However, the obtained coding efficiency is at the cost of high computational complexity. In this paper, to reduce the computation complexity, we propose to use an ensemble classifier for predicting the coding unit (CU) in intra-coding. Firstly, the L1-loss based linear support vector machine (SVM) is employed as basic classifier for its simplicity. Then, a bagging scheme is applied to train the linear classifiers and boost the prediction accuracy by ensemble learning. Compared with the reference software SCM-5.0, the proposed scheme can achieve 30% complexity reduction on average with only 1.64% bit rates increase.

Index Terms— Screen Content Coding, Linear Classification, Intra Coding, Coding Unit Decision, Ensemble Learning

1. INTRODUCTION

Screen content video (SCV) has attracted attention due to its huge potential in real-time video systems, such as wireless displays, shared screen collaboration, virtual desktop interface, online education. However, the capability of real-time encoding and transmission of SCV has been challenged by the demand of large screen resolution and frame rate. To efficiently compress SCVs, the screen content coding (SCC) extension was developed based on the High Efficiency Video Coding (HEVC) and issued in February 2016[1].

SCC inherits the quadtree based CU partition structure from HEVC. Compared to the camera captured video scenes, the SVCs contains many computer-generated contents such as text, graphic or animation in the SCVs. Since SCVs has different characteristics as it often has temporally static with sharp edges or contains few distinct colors, etc, the HEVC-SCC is developed and adopted some new coding tools, including adaptive color transform (ACT) [2], palette mode (PLT) [3], intra block copy (IBC) [4]. In HEVC-SCC test model, both traditional intra modes and SCV specific modes are encoded to find the optimal coding unit (CU) partition.

As we know, HEVC achevies 50% bit rate reduction in comparison to the H.264/AVC at the same visual quality [5], with more than 5 times complexity increase. Moreover, SCC increases the encoder complexity even more by adopting these new tools. Therefore, it is necessary to investigate the fast algorithms to speed up the SCC encoder for real-time applications.

Currently, some fast algorithms had been proposed for intra coding of SCC. For example, Kwon *et al.*[6] proposed an intra-frame motion vector search scheme based on the conventional intra mode cost and CU activity. To determine whether to split or not for the current CU, Zhang *et al.*[7] proposed a fast CU depth decision algorithm based on information entropy, which measures the relationship between the information entropy of the current block and its four sub-blocks. Tsang *et al.*[8] designed a simple intra prediction method by skipping the intra-mode or IBC mode for smooth region. Duanmu *et al.*[9] used the statistical features and sub-CU related features to train neural network based classifiers to predict the CU partition. Recently, Lei *et al.* [10] proposed a fast algorithm based on statistical observation. It classified the current CU into screen content CU or natural content CU. Then for different CU type, the encoder is modified to skip some modes checking. Up to now, there are still some room to further reduce the computation complexity.

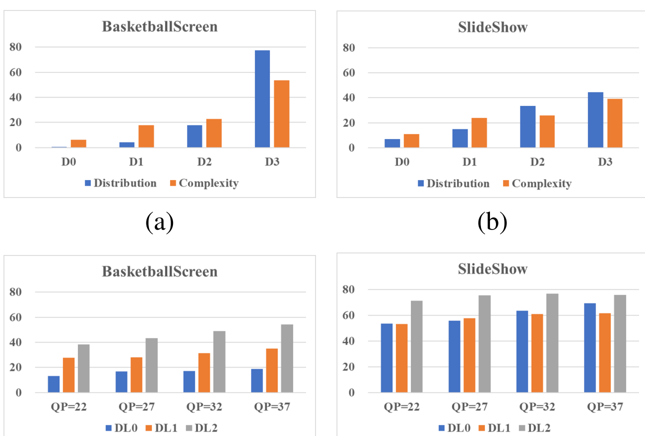
In this paper, we proposed to use an ensemble classifier for fast CU partition of SCC intra-coding, which achieves significantly performance improvement on computational complexity reduction. The main contributions are summarized as

*Corresponding author (Xu Wang, wangxu@szu.edu.cn). This work was supported in part by the National Natural Science Foundation of China under Grant 61871270, 61672443 and 61620106008, in part by the Guangdong Nature Science Foundation under Grant 2016A030310058, in part by the Shenzhen Emerging Industries of the Strategic Basic Research Project under Grants JCYJ20160226191842793, in part by the Natural Science Foundation of SZU (grant no. 827000144).

follows:

- To avoid the limitation of overhead complexity, the L1-loss linear SVM classifer is employed as a basic classifier for its efficiency on large scale data training and simplicity on inference.
- To reduce the hyper-parameter fine tuning operations of classifier and shorten the off-line training time, the bagging method is adopted to design an ensemble classifier, which boosts the prediction accuracy of classifier.
- To further improve the prediction accuracy of classifier, the feature vector correlated with mode decision were investigated. For different CU depths, the feature vectors are also different.

The rest of this paper is organized as follows. Section 2 describe the proposed algorithm in detail. Experimental results are provided in Section 3. Finally, conclusions are given in Section 4.



(a) (b)

(c) (d)

Fig. 1. Pre-encoding information for different sequences and coding parameters. (a) and (b) are distribution and computation complexity ratios of each CU depth for sequences “BasketballScreen” and “SlideShow”, respectively. (c) and (d) are the non-split ratios of each depth level for sequences “BasketballScreen” and “SlideShow”, respectively.

2. PROPOSED ALGORITHM

In this section, we will introduce the proposed algorithm in details, including the classifier designing strategy, feature extraction and selection.

2.1. Proposed Ensemble Classifier

To analyze the theoretic bound of maximum time saving for fast CU decision algorithm, two test sequences, including “BasketballScreen” and “SlideShow” are pre-encoded under different quantization parameters (QPs) for the first twenty frames. The coding information including distribution and computational complexity ratios of each CU depth are summarized in Figures 1 (a) and (b). It is observed that if the encoder can correctly determine whether to split or not split the current CU, the computational complexity will be significantly reduced without coding performance loss.

Inspired by the low complexity CU partition schemes [11] for natural scene content, the CU depth decision in SCC can be modeled as a three-layer binary classification task. For each CU depth, the classifier needs to decide the status, including split, non-split or uncertain. To extreme reduce the computational complexity under a given Rate Distortion (RD) performance loss constraint, the key problem becomes training classifiers with high prediction accuracy and small overhead on computational complexity. In this paper, instead of using radial basis function (RBF) kernel, the L1-loss linear support vector machine (L1-SVM) classifier is adopted as the basic classifier, which is efficiency on large scale data set training and simplicity for inference. Given the feature vector $x \in R^d$, where d is the feature dimension, the classifier generates a weight vector w and bias b for the model. Then the decision function y is denoted as

$$y = f(w, b, x) = \mathrm{sgn}(w^T x + b). \tag{1}$$

Considering the imbalance of sample data distribution of screen content (e.g., the ratio of the split sample is larger than non-split sample), it is hard to learn a strong model with extreme high accuracy. Thus, we design a three-output classifier based on the concept of bagging from ensemble learning. Given the large scale training set $\mathbf{A}$, we generate the sub-sets $\mathbf{S}_l$ by randomly sampling from $\mathbf{A}$. The output of weak classifier trained on the $\mathbf{S}_l$ is denoted as y_l. Suppose the number of sub-sets is n, then the voting function O is defined as

$$O = \sum_{l=1}^{n} \lambda_l y_l, \tag{2}$$

where λ_l is the weighting factor for the classifier y_l. The value of O is in the range of $[-1, 1]$.

2.2. Feature Selection

The split and non-split label of samples from intra coding is highly depended on the structural information, pre-encoding information and neighboring coding information. For example, the ratio of non-split is influenced by the scene, quantization process and CU size as shown in Figures 1 (c) and (d). To guarantee the training accuracy of classifiers, the following representative features are adopted for different CU depths, as shown in Table 1.

Structural information: To further remove the spatial redundancy by intra prediction, the encoder tends to split the current CU, which contains rich structural information such

Table 1. Feature sets of classifier for differen CU depth levels (DLs)

ID	Feature	DL0	DL1	DL2
1	X_T	✓	✓	✓
2	X_{ST}	✓	✓	✓
3	X_{no}	✓	✓	✓
4	X_{range}	✓	✓	✓
5	X_{Dis}	✓	✓	✓
6	X_{Bits}		✓	✓
7	X_{QP}	✓	✓	✓
8	X_L			✓
9	X_U			✓

as text. For CU with flat and smooth textures, the encoder may stop to avoid the bit rate overhead caused by extra header information. In the paper, the mean absolute deviation of luminance for current CU is used to measure the texture complexity X_T as

$$X_T = \frac{1}{N_B} \sum_{(i,j)\in B} |I(i,j) - \frac{1}{N_B} \sum_{(i,j)\in B} I(i,j)|, \quad (3)$$

where B represents the set of pixels in current CU, N_B represents the number of pixels in B. $I(i,j)$ is the intensity value of pixel located in (i,j). Besides, the sum of the X_T for four sub-CUs is denoted as X_{ST} and employed as a feature.

Different from camera captured content, screen content may contain many text regions, the range of luminance in each CU is limited. Thus, the number of luminance components X_N and the range of luminance X_R in the current block are extracted as features.

Pre-encoding information: To accurately predict the split probability of current CU, an effective way to pre-encode the current CU with limited prediction mode. Since the planar mode is simple but representative, the bit rate X_R and distortion X_D information of planar mode are collected as features. Since the decision may influence by quantization process, the quantization parameter X_Q is also employed as one of features.

Neighboring coding information: Due to the spatial correlation, the coding mode of current CU is correlated to its neighbor CUs. Thus, the CU partition results X_L and X_U of the left and upper neighboring CUs are employed for CU depth prediction, respectively.

2.3. Proposed Fast CU Depth Decision Algorithm

The proposed fast CU depth partition algorithm is illustrated in Fig. 2. For the current CU (non-boundary CU), if the depth index is smaller than 3, then the encoder will check the planar mode first. After obtaining the feature vector and calling

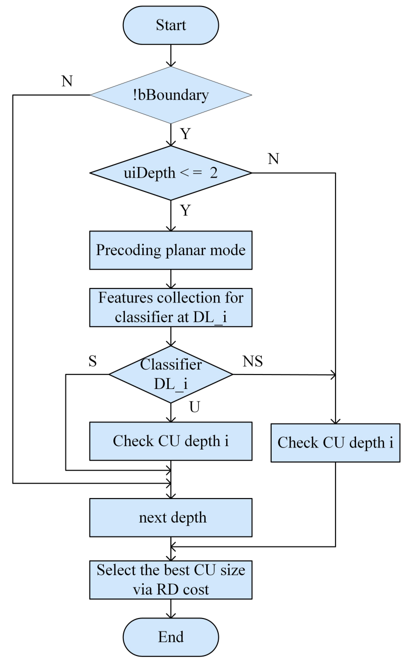


Fig. 2. Flowchart of proposed algorithm

the classifier, the encoder will make decision based on the output value of voting function. As shown in Fig. 2, if $O = -1$ (denoted as NS), then the encoder only checks the current CU depth i. Else if $O = 1$ (denoted as S), the encoder will skip the current CU and go to check the four sub-CUs at depth $i+1$. Otherwise, the encoder will check the current CU depth i, then goes to the CU depth $i+1$ and checks the four sub-CUs at depth $i+1$.

3. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed fast algorithm, SCC test model with version HM-16.6+SCM-5.0 is adopted as the software platform. The hardware platform is intel core i7-7700@3.60 GHZ, 4GB, Windows 7 64-bit professional operating system. The first ten frames of five testing sequences are pre-encoded under all-intra configuration with four different QP settings, including 22, 27, 32 and 37. For CU j of depth level i, the feature vector and its label $(x_{ij}, \hat{y}_{ij})$ is collected as a sample of data set $\mathbf{D}_i$. In this paper, the number of weak classifier for each depth level is fixed as 2. To train the classifiers, the data set $\mathbf{D}_i$ is first divided into training set $\mathbf{A}_i$ (80%) and validation set $\mathbf{V}_i$ (20%). Then, the training set $\mathbf{A}_i$ is equally divided in to two sub-sets $\mathbf{S}_{i1}$ and $\mathbf{S}_{i2}$. The weighting factor λ_1 is fixed as 0.5.

For fair comparison, we compare proposed algorithm with the state-of-the-arts algorithms, including algorithm in [7] (denoted as ZMM) and algorithm in [10] (denoted as LEI). All the algorithms are evaluated on the video sequence recommended by JCTVC-X1015. The all-intra configuration

Table 2. Summarized results on performance comparison for fast CU depth algorithms

Seqences	ZHANG [7]			LEI [10]			Proposed		
	$\Delta D\,(dB)$	$\Delta R\,(\%)$	ΔT	$\Delta D\,(dB)$	$\Delta R\,(\%)$	ΔT	$\Delta D\,(dB)$	$\Delta R\,(\%)$	ΔT
BasketballScreen	-0.06	0.62	0.12	-0.30	3.04	0.27	-0.14	1.38	0.29
ChinaSpeed	-0.03	0.30	0.17	-0.23	2.76	0.36	-0.08	0.90	0.28
ChineseEditing	-0.03	0.19	0.04	-0.11	0.67	0.19	-0.46	2.87	0.32
MissionControlClip2	-0.02	0.27	0.21	-0.18	2.04	0.35	-0.09	0.94	0.31
MissionControlClip3	-0.06	0.50	0.13	-0.26	2.21	0.26	-0.14	1.19	0.30
Console	-0.44	1.83	0.05	-0.43	1.83	0.18	-0.48	2.00	0.24
Desktop	-0.18	0.81	0.04	-0.40	1.86	0.22	-0.47	2.20	0.34
FlyingGraphics	-0.09	0.61	0.03	-0.14	0.91	0.18	-0.14	0.90	0.30
Map	-0.04	0.46	0.10	-0.09	1.00	0.24	-0.10	1.07	0.26
Programming	-0.09	0.70	0.12	-0.35	2.70	0.26	-0.24	1.89	0.32
Robot	-0.01	0.31	0.17	-0.28	6.54	0.56	-0.02	0.44	0.19
WebBrowsing	-0.15	1.12	0.07	-0.47	3.51	0.26	-0.52	3.98	0.30
SlideShow	-0.06	0.70	0.43	-0.34	4.07	0.53	-0.14	1.61	0.40
avg	**-0.10**	**0.65**	**0.13**	**-0.28**	**2.55**	**0.30**	**-0.23**	**1.64**	**0.30**

is enabled. The results of the HM-16.6+SCM-5.0 encoder is used as benchmark. To measure the performance of the algorithm, three metrics including BDBR (ΔR), BDPSNR (ΔD) [12] and average time saving (ΔT) are employed. The value of ΔT indicates the ratio of computation complexity reduction. The value of ΔD and ΔR indicate the quality gain and ratio of increased bit rate.

Detailed experimental results are summarized in Table 2. It is observed that our proposed algorithm can achieve the best performance among all the algorithms. For example, although both our proposed algorithm and scheme LEI can save 30% execution time on average, the ΔD value of the proposed algorithm is larger than that of scheme LEI. For sequences "BasketballScreen", "MissionControlClip3", "FlyingGraphics" and "Programming", the performance of proposed algorithm is outstanding.

4. CONCLUSION

In this paper, we propose a fast CU depth decision scheme based on the concept of ensemble learning for SCC intra coding. The encoder can make a decision based on the output of ensemble classifier for executing branch subroutine, including split, non-split or uncertain. Extensive experiments demonstrate that the proposed scheme can achieve 30% complexity reduction on average with only 1.64% bit rate increase over the test sequences under all-intra configuration.

5. REFERENCES

[1] 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, no. 1, pp. 50–62, 2016.

[2] L. Zhang, J. Chen, J. Sole, M. Karczewicz, X. Xiu, and J. Xu, "Adaptive color-space transform for HEVC screen content coding," in *Data Compression Conference*, 2015, pp. 233–242.

[3] W. Pu, M. Karczewicz, R. Joshi, V. Seregin, F. Zou, J. Sole, Y.C. Sun, T.Z. Chuang, P. Lai, and S. Liu, "Palette mode coding in HEVC screen content coding extension," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 6, pp. 1–13, 2016.

[4] X. Xu, S. Liu, T. Chuang, and S. Lei, "Block vector prediction for intra block copying in HEVC screen content coding," in *2015 Data Compression Conference*, 2015, pp. 233–242.

[5] G. J. Sullivan, J.-R. Ohm, W. 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, 2013.

[6] D. K. Kwon and M. Budagavi, "Fast intra block copy (IntraBC) search for HEVC screen content coding," in *International Symposium on Circuits and Systems*. IEEE, 2014, pp. 9–12.

[7] M. Zhang, Y. Guo, and H. Bai, "Fast intra partition algorithm for HEVC screen content coding," in *Visual Communications and Image Processing Conference*, 2014, pp. 390–393.

[8] S. Tsang, Y. Chan, and W. Siu, "Fast and efficient intra coding techniques for smooth regions in screen content

coding based on boundary prediction samples," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 1409–1413.

[9] F. Duanmu, Z. Ma, and Y. Wang, "Fast CU partition decision using machine learning for screen content compression," in *International Conference on Image Processing (ICIP)*. IEEE, 2015, pp. 4972–4976.

[10] J. Lei, Y. Li, Q. Pan, Z. Sun, S. Kwong, and C. Hou, "Fast intra prediction based on content property analysis for low complexity HEVC-based screen content coding," *IEEE Transactions on Broadcasting*, vol. 63, pp. 48–58, 2017.

[11] Y. Zhang, S. Kwong, X. Wang, H. Yuan, Z. Pan, and L. Xu, "Machine learning-based coding unit depth decisions for flexible complexity allocation in high efficiency video coding," *IEEE Transactions on Image Processing*, vol. 24, pp. 2225–2238, 2015.

[12] G. Bjontegaard, "Calculation of average psnr differences between RD curves (VCEGM33)," document ITU-T SG16 Q6, VCEG Meeting, Austin, TX, USA, April 2001.