FAST MODE DECISION ALGORITHM FOR HEVC SCREEN CONTENT INTRA CODING

Wei Kuang, Sik-Ho Tsang, Yui-Lam Chan, and Wan-Chi Siu

Centre for Signal Processing, Department of Electronic and Information Engineering
The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong

ABSTRACT

Screen Content coding (SCC) is one of an extension to High Efficiency Video Coding (HEVC) developed by the Joint Collaborative Team on Video Coding (JCT-VC). It adopts two new coding tools, intra block copy (IBC) and palette (PLT) modes, to improve the compression performance for intra coding. Nevertheless, mode selection causes a substantial increase in encoding complexity. In this paper, a fast mode decision algorithm, which makes use of early mode skip decision based on the Bayesian decision rule using online learning, is proposed. The proposed algorithm is implemented in the SCC reference software SCM-7.0. Experimental results show that the proposed algorithm can achieve 23.2% complexity reduction on average with only 0.58% Bjøntegaard delta bitrate loss in All Intra (AI) configurations.

Index Terms— Bayesian decision rule, fast mode decision, High Efficiency Video Coding (HEVC), Screen Content Coding (SCC).

1. INTRODUCTION

Screen content coding (SCC) has gained more popularity nowadays and has been applied into many applications, such as remote desktop, cloud computing, and online education. Besides traditional pictorial content, screen content videos also contain computer generated content and textual content, which is noise-free with many repeated patterns, sharp edges and limited colors. Because of the different characteristics existing in screen content videos, new coding tools have been investigated to achieve high coding efficiency. SCC [1] has been included in the HEVC standard [2] as one of its requirements since January 2014. SCC adopts the same coding tree unit (CTU) partitioning structure [3, 4, 5] as HEVC, and two major coding tools: Intra block copy (IBC) [6-7] and palette (PLT) [8-10] modes have been introduced using the repeated pattern and limited color characteristics to improve the coding efficiency. A SCC encoder checks conventional intra (Cintra) mode (including 35 prediction modes), IBC mode and PLT mode when encoding a coding unit (CU). Finally, the optimal modes are selected by the rate-distortion optimization (RDO) process. Although the coding efficiency can be improved significantly by adding the new coding tools, the great computational complexity of SCC makes it impractical for real time applications.

Recently, many approaches have been designed to expedite the encoding process of SCC. The cost of Cintra mode and CU activity were used in [11] to early skip IBC search, and these algorithms have been adopted and implemented in the SCC reference software SCM. A fast algorithm speeding up the local search by checking the hash values of both current block and block candidates was investigated in [12]. Using the characteristic of noiselessness, boundary samples are used to fill all samples within the predicted CU if they have the same value in [13]. A fast CU partition algorithm was proposed based on CU entropy and coding bits in [14], which would early terminate the CU partition adaptively. In [15], CUs with zero activity or low gradient does not go through IBC mode checking to reduce the computational complexity brought by IBC mode. Temporal correlation in static regions is used to speed up the intra mode decision process in [16], where the mode information in the collocated CU is employed to predict the mode of current CU. Besides, an adaptive search step approach is applied to speed up the searching process of IBC mode.

In this paper, we propose a fast mode decision algorithm, which utilizes early mode skip decision based on online learning using the Bayesian decision rule. First, a corner point detection method is applied to roughly classify a frame into textual regions and pictorial regions. Second, to employ the Bayesian decision rule, three mode classes are defined as ω_{Cintra}, ω_{IBC} and ω_{PLT}, and the distinct color number in a CU is extracted as the feature for mode classification. To ensure the classification accuracy, all learning parameters are updated periodically for different scenes using scene change detection method.

The rest of the paper is organized as follows. Section II briefly reviews and analyzes the original mode selection process in SCC. Section III presents the proposed fast mode decision method including corner point detection, early mode skip using Bayesian decision rule and learning frame updating. Section IV gives the experimental results and discussions. Finally, Section V concludes this paper.

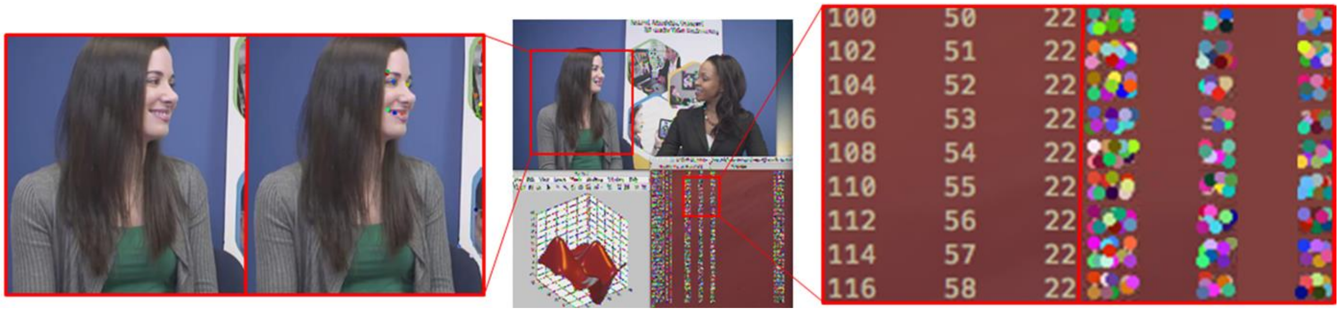

Fig.1. Detected corner points in sequence MissionControl3.

2. REVIEW ON MODE DECISION IN SCC AND ANALYSIS

Screen content videos usually contain both traditional pictorial content and computer generated content. Due to different characteristics existing in screen content videos, besides the Cintra mode in HEVC, two additional intra coding modes: IBC mode and PLT mode have been proposed for SCC.

IBC mode is similar to the motion estimation [17-19] in interframe coding and is designed based on the observation that there are many repeated patterns within one frame. IBC mode is applied for CU sizes from 64×64 down to 8×8, and it searches the reconstructed samples in the current frame to find reference blocks. IBC mode is very efficient if a good matched block can be found for the current block in the reconstructed regions.

Palette mode is designed based on the observation that textual content usually contains limited colors. PLT mode is applied for CU sizes from 32×32 down to 8×8. Several representative color values in a CU are selected to form a palette table for base colors. Those color values not in the palette table are treated as escape colors. Then, a palette index map is generated to send indices for base colors and escape colors. For base colors, only the indices in the palette are encoded. For escapes colors, the quantized color values are directly encoded. PLT mode is very effective when encoding textual blocks with few color values.

In SCC intra coding, three modes: Cintra mode, which includes 33 directional modes plus planar and DC prediction modes, IBC mode and PLT mode, will all be checked to improve the coding efficiency. However, this mode selection process leads to a dramatic increase in computational complexity of the encoder. Because IBC and PLT modes are contrived for textual content with many repeated patterns and limited colors, they are more likely to be the optimal modes in textual content, while pictorial content tends to select Cintra mode because of fewer repeated patterns and high number of base colors. Therefore, if unnecessary modes can be skipped adaptively based on the different characteristics in pictorial content and textual content, great computational complexity reduction can be achieved.

3. PROPOSED FAST MODE DECISION ALGORITHM

The mode decision process is treated as a three-class problem using the Bayesian decision rule. The proposed fast mode decision algorithm can perform early skip for the modes with low probability to be the optimal mode when encoding a CU.

3.1. Corner point detection

A corner point refers to an interest point with two dominant and different edge directions in a local neighborhood of the point. Based on the assumption that textual content contains more strong corners, while pictorial content is relatively smooth, the Shi–Tomasi corner point detection method proposed in [14] is applied to classify a frame into textual regions and pictorial regions roughly. The covariance matrix of an image patch at (u, v) and itself after shifted by (x, y) in the image I is given by

$$T(x,y) = \sum_u \sum_v w(u,v) \begin{bmatrix} {I_x}^2 & I_x I_y \\ I_x I_y & {I_y}^2 \end{bmatrix} = \begin{bmatrix} \sum_u \sum_v w(u,v)\, {I_x}^2 & \sum_u \sum_v w(u,v) I_x I_y \\ \sum_u \sum_v w(u,v)\, I_x I_y & \sum_u \sum_v w(u,v) {I_y}^2 \end{bmatrix}. \quad (1)$$

where $w(u, v)$ is a weight window, I_x and I_y are the part derivations of I. The eigenvectors of T are two principle directions while the eigenvalues of T represent the degrees of the change in their directions. If there is a corner, T should have two large eigenvalues θ_1 and θ_2, and the strength of a corner point can be defined as

$$R = \min(\theta_1, \theta_2). \quad (2)$$

The minimal accepted strength of corner points S_{min} in a frame is defined as

$$S_{min} = \mu \times S_{max} \quad (3)$$

where S_{max} is the largest strength of all corner points in the frame, and μ is a minimal accepted strength threshold. Only corners with strength above S_{min} can be detected. The corner point detection method is implemented based on the OpenCV2.4.9 function goodFeaturesToTrack(). Fig.1 shows the detected corner points in sequence MissionControl3.The original content and its corner points are shown in the left and right sides of the zoomed regions, respectively. It can be seen that corner points are concentrated on textual regions, while there are much less corner points in pictorial regions. In the proposed method, the threshold μ was set to 0.01 for detecting low contrast corners. Because of the different distinct color number distribution in pictorial regions and textual regions, this classification will be further refined in next sub-section. Thus, a frame is divided into two groups now: regions without corner points ($group_{NCP}$), and regions with corner points ($group_{CP}$).

3.2. Early mode skip using Bayesian decision rule

In the proposed method, three classes are defined as Cintra, IBC and PLT classes: ω_{Cintra} , ω_{IBC} and ω_{PLT} for each group. In screen content videos, pictorial content may contain sensor noise which leads to the high color number in a CU, while textual content naturally has limited color number. Therefore, distinct color number is used to preform further mode classification. To fully use the pixel value information, YUV components are combined into a 24 bits color value. After the learning phase at the beginning, the encoder would check the posteriori probability $P_{group_{j,d}}(\omega_i|n)$ for each CU during the encoding process, where n represents the distinct color number in a CU, $\epsilon\{Cintra, IBC, PLT\}$, $j\epsilon\{NCP, CP\}$, and CU depth level $d\epsilon\{0,1,2,3\}$. According to the Bayes' rule, the posteriori probability $P_{group_{j,d}}(\omega_i|n)$ for CUs in $group_j$ and depth level d is calculated as:

$$P_{group_{j,d}}(\omega_i|n) = \frac{P_{group_{j,d}}(n|\omega_i)P_{group_{j,d}}(\omega_i)}{P_{group_{j,d}}(n)} \tag{4}$$

where $P_{group_{j,d}}(n|\omega_i)$represents the conditional probability of n in class ω_i , $P_{group_{j,d}}(\omega_i)$ represents the priori probability, and $P_{group_{j,d}}(n)$ represents the total probability density of n. During the encoding process, $P_{group_{j,d}}(n|\omega_i)$ and $P_{group_{j,d}}(\omega_i)$ can be estimated by the learning frames which are encoded by the original SCC encoder, and $P_{group_{j,d}}(n)$ can be obtained by

$$P_{group_{j,d}}(n) = \sum_{\omega_i} P_{group_{j,d}}(n|\omega_i)P_{group_{j,d}}(\omega_i). \tag{5}$$

Because blocks with $n = 1$ can be encoded by all of the three modes efficiently, we exclude them for simplicity in our proposed method. The distinct color number is closely related to the CU size. Large CUs usually have higher distinct color number compared with small CUs. Therefore, the priori probability $P_{group_{j,d}}(\omega_i)$ and the conditional probability $p_{group_{j,d}}(n|\omega_i)$ are estimated and stored for CUs with different sizes from 64 × 64 down to 8 × 8 during the online learning phase.

In our proposed method, $P_{group_j}(\omega_i|n)$ estimated based on (4) and (5) will be checked for each CU of different sizes and different groups after the learning process, and mode class ω_i will be skipped if

$$P_{group_j}(\omega_i|n) < \alpha. \tag{6}$$

Thus, if the probability for a CU selecting a mode ω_i is lower than the threshold α, which is set to 0.05 in our proposed method, early mode skip for this mode will be performed.

3.3. Learning frame updating

There can be many scene changes in screen content videos, such as documents opening or closing, slideshow playing, etc., and it would lead to wrong classification due to the inaccurate estimation of learning parameters. Thus, the learning parameters should be updated by selecting new learning frames if a new scene is appeared. A typical correlation measurement method, histogram of difference (HOD) introduced in [20] is used to perform scene detection in our proposed approach. The histogram of difference between two adjacent frames, $F^a - F^b$, is defined by $hod(l)$, where $l \in [-q+1, q-1]$. The further the histogram of difference is distributed from the origin of $hod(l)$, the more different the frames are. Therefore, the HOD is defined as

$$HOD(F^a, F^b) = \frac{\sum_{l \notin [-\tau,\tau]} hod(l)}{\sum_{l=-q+1}^{q-1} hod(l)} \tag{7}$$

where τ is a threshold to determine the closeness to zero. If the value of HOD is larger than a threshold φ, a different scene is regarded to be appeared, and another group of m frames are selected as learning frames to update the learning parameters. Considering the proposed fast mode and CU size decision method cannot be applied to the learning frames, increasing the number of learning frames would decrease the computational complexity reduction. As suggested by [21], τ, φ and m is set to 32, 0.2 and 2, respectively.

As a summary of the proposed algorithm, the flowchart of our proposed method is shown in Fig. 2.

4. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, the performance of the proposed fast mode decision method is evaluated in terms of the encoding time reduction and Bjøntegaard delta bitrate (BDBR) [22]. The proposed method is implemented in the HEVC reference software SCM-7.0 [23]. The experimental conditions are based on All Intra (AI) configurations and strictly follow the Common Test Conditions (CTC). The test sequences are the typical SCC test sequences including both screen content (SC) videos and camera-capture content (CC) videos, which are selected by the experts in the JCT-VC group [24]. Those

Table 1. Performance comparison with [14].

Category	Sequence	Proposed		Zhang *et al*. [14]	
		BDBR (%)	$\Delta Time$ (%)	BDBR (%)	$\Delta Time$ (%)
SC-TGM	FlyingGraphics, 1920×1080	0.21	-13.7	0.54	-4.0
	Desktop, 1920×1080	0.72	-27.3	0.67	-5.9
	Console, 1920×1080	0.23	-29.1	2.64	-8.0
	ChineseEditing, 1920×1080	0.25	-25.4	0.14	-5.0
	WebBrowsing, 1280×720	-0.12	-29.6	0.27	-7.5
	Map, 1280×720,	0.78	-18.9	0.96	-11.5
	Programming, 1280×720,	0.99	-14.3	0.44	-19.2
	SlideShow, 1280×720	2.32	-12.9	0.36	-47.2
SC-M	BasketballScreen, 2560×1440	0.43	-17.8	0.45	-13.2
	MissionControlClip2, 2560×1440	0.77	-13.5	0.40	-21.4
	MissionControlClip3, 1920×1080	0.30	-17.7	0.37	-12.1
SC-A	Robot, 1280×720	0.90	-29.4	0.43	-18.8
CC	EBURainFruits, 1920×1080	0.25	-36.2	0.21	-18.9
	Kimono1, 1920×1080	0.09	-38.8	0.14	-26.4
Average (SC)		0.65	-20.8	0.64	-14.5
Average (CC)		0.17	-37.5	0.18	-22.7
Average (ALL)		0.58	-23.2	0.57	-15.7

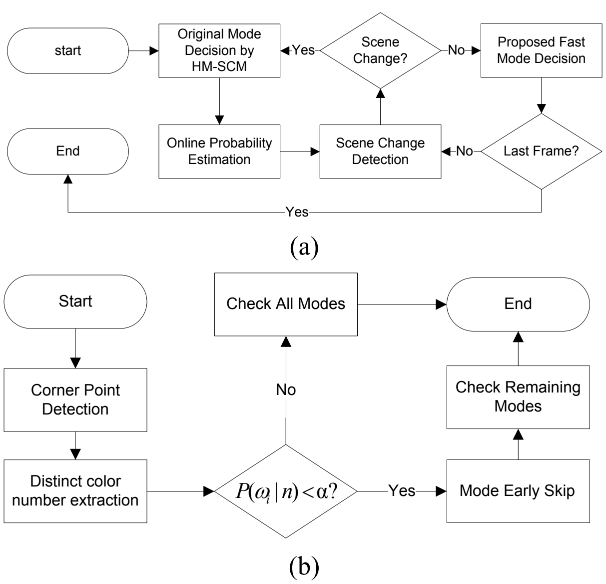

Fig. 2. Flowchart of the proposed overall algorithm, (a) learning frames updating, (b) fast mode decision after learning phase.

sequences are classified into four categories: text and graphics with motion (SC-TGM), mixed content (SC-M), animation (SC-A), and CC. Each sequence is encoded with 4 quantization parameters (QPs) at 22, 27, 32 and 37. $\Delta Time$ is used to measure the time encoding time reduction, which is defined as

$$\Delta Time = \frac{Time_{proposed} - Time_{reference}}{Time_{reference}} \tag{8}$$

where $Time_{proposed}$ represents the encoding time of the proposed fast mode decision method, and $Time_{reference}$ represents that of HEVC reference software SCM-7.0. It should be noted that negative value of $\Delta Time$ denotes encoding time decreasing. The experimental results for our proposed method compared with SCM-7.0 in terms of BDBR and $\Delta Time$ are given in Table 1. Besides, the performance of the fast CU partition method in [14] is also provided for comparison. As shown in Table 1, the proposed fast mode decision method can achieve up to 29.6% and 38.8% encoding time reduction for SC and CC videos, respectively. On average, 20.8%, 37.5% and 23.2% encoding time reduction can be achieved with negligible loss of video quality for SC, CC and all videos, respectively. Compared with Zhang *et al*.'s method [14], our proposed method can achieve 7.5% more encoding time reduction with only 0.01% BDBR increment.

5. CONCLUSIONS

In this paper, a fast mode decision algorithm based on the Bayesian decision rule using online learning is proposed to reduce the computational complexity of SCC. A corner point detection method is first applied to classify a frame into pictorial and textual regions roughly. Then, distinct color number in a CU is extracted as the feature for mode classification. Besides, scene detection is applied to update the learning parameters for different scene. Compared with the reference software SCM-7.0, the proposed method can achieve 23.2% encoding time reduction with only 0.58% BDBR increase for AI configurations.

6. ACKNOWLEDGEMENTS

This work was supported by the Centre for Signal Processing, Department of Electronic and Information Engineering, The Hong Kong Polytechnic University (PolyU), and a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Grant No. PolyU 152052/15E).

7. REFERENCES

[1] J. Xu, R. Joshi, and R. A. Cohen, "Overview of the emerging HEVC screen content coding extension," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 50-62, January 2016.

[2] G. J. Sullivan, J. 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.

[3] J. Ohm, G. J. Sullivan, H. Schwarz, T. T. Keng, and T. Wiegand, "Comparison of the coding efficiency of video coding standards—including high efficiency video coding (HEVC)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1669–1684, Dec. 2012.

[4] T.-K. Lee, Y.-L. Chan, and W.-C. Siu, "Adaptive Search Range for HEVC Motion Estimation based on Depth Information," *IEEE Trans. Circuits Syst. Video Technol.*, pp. 1-15, Early Access Articles.

[5] T.-K. Lee, Y.-L. Chan, and W.-C. Siu, "Adaptive Search Range by Neighbouring Depth Intensity Weighted Sum for HEVC Texture Coding," *IET Electron. Lett.*, vol. 52, no. 12, pp. 1018-1020, June 2016.

[6] J. Sole, and S. Liu, "HEVC screen content coding core experiment 1 (SCCE1): intra block copying extensions," *Joint Collaborative Team on Video Coding (JCT-VC)*, JCTVC-Q1121, Valencia, Spain, Mar. 2014.

[7] C. Pang *et al*, "CE2 Test1: Intra block copy and inter signaling unification," *Joint Collaborative Team on Video Coding (JCT-VC)*, JCTVC-T0094, Geneva, Switzerland, Feb. 2015.

[8] L. Guo, J. Sole, and M. Karczewicz, "Palette mode for screen content coding," *Joint Collaborative Team on Video Coding (JCT-VC)*, JCTVC- M0323, Incheon, Korea, Apr. 2013.

[9] X. Xiu *et al*, "Palette-based coding in the screen content coding extension of the HEVC standard," in *Proc. of Data Compression Conf*, Snowbird, UT, U.S.A., pp 253-262, Apr. 2015.

[10] S.-H. Tsang, Y.-L. Chan, and W.-C. Siu, "Exploiting Inter-Layer Correlations in Scalable HEVC for the Support of Screen Content Videos," in *Proc. of Int. Conf. on Digital Signal Process. (DSP)*, pp.888-892, Hong Kong, August 2014.

[11] D.-K. Kwon, and M. Budagavi, "Fast intra block copy (IntraBC) search for HEVC screen content coding," in *Proc. of IEEE Int. Symposium on Circuits and Systems (ISCAS)*, Melbourne, VIC, Australia, pp. 9-12, Jun. 2014.

[12] S.-H. Tsang, Y.-L. Chan, and W.-C. Siu, "Hash based fast local search for intra block copy (IntraBC) mode in HEVC screen content coding," in *Proc of APSIPA Annual Summit and Conference (APSIPA ASC)*, Hong Kong, China, pp. 396-400, Dec. 2015.

[13] S.-H. Tsang, Y.-L. Chan, and W.-C. Siu, "Fast and efficient intra coding techniques for smooth regions in screen content coding based on boundary prediction samples," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Brisbane, Australia, pp.1409-1413, Apr. 2015.

[14] M. Zhang, Y. Guo, and H. Bai, "Fast intra partition algorithm for HEVC screen content coding," in *Proc. of IEEE Vis. Commun. Image Process. (VCIP)*, Valletta, Malta, pp. 390-393, Dec. 2014.

[15] S.-H. Tsang, W. Kuang, Y.-L. Chan and W.-C. Siu, "Fast HEVC screen content coding by skipping unnecessary checking of intra block copy mode based on CU activity and gradient," in *Proc of APSIPA Annual Summit and Conference (APSIPA ASC)*, Jeju, Korea, pp.1-5, Dec. 2016.

[16] H. Zhang *et al*, "Fast intra mode decision and block matching for HEVC screen content compression," in *Proc. of IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Shanghai, China, pp.1377-1381, Mar. 2016.

[17] Y.-L. Chan, K.-C. Hui, and W. C. Siu, "Adaptive Partial Distortion Search for Block Motion Estimation," *Journal of Visual Communication and Image Representation*, vol. 15, issue.4, pp. 489-506, Dec. 2004.

[18] K.-C. Hui, W.-C. Siu, and Y.-L. Chan, "Fast Motion Estimation of Arbitrarily Shaped Video Objects in MPEG-4," *Signal Process.: Image Commun.*, vol. 18, pp.33-50, January 2003.

[19] Y.-L. Chan, and W.-C. Siu, "On Block Motion Estimation Using a Novel Search Strategy for an Improved Adaptive Pixel Decimation," *J. of Visual Commun. and Image Represent.*, vol. 9, no. 2, pp. 139-159, May 1998.

[20] J.-B, Shi, and C. Tomasi, "Good features to track," in *Proc. of IEEE Int. Conf. Comp. Vis. Pattern Recognit. (CVPR)*, pp. 593-600, Jun. 1994.

[21] H.-S. Kim and R.-H. Park, "Fast CU partitioning algorithm for HEVC using an online-learning-based bayesian decision rule," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 130-138, Jan. 2016.

[22] G. Bjøntegaard, "Improvements of the BD-PSNR model, document," *Video Coding Experts Group (VCEG)*, VCEG-AI11, ITU-T SG16/Q6, Jul. 2008.

[23] R. Joshi, J. Xu, R. Cohen, S. Liu, and Y. Ye, "Screen content coding test model 7 encoder description," *Joint Collaborative Team on Video Coding (JCT-VC)*, JCTVC-W1014, San Diego, U.S.A., Feb. 2016.

[24] H. -P. Yu, R. Cohen, K. Rapaka, and J. -Z Xu, "Common test conditions for screen content coding," *Joint Collaborative Team on Video Coding (JCT-VC)*, JCTVC-U1015, Warsaw, Poland, Jun. 2015.