IMPROVING SCREEN CONTENT CODING IN HEVC BY TRANSFORM SKIPPING

Marta Mrak[*] *and Ji-Zheng Xu*[**]

[*]British Broadcasting Corporation, Research and Development, London, UK
[**] Microsoft Research Asia, Beijing, China

ABSTRACT

Screen content is nowadays a part of numerous applications - from desktop sharing to broadcasting. It consists of both camera captured content and computer generated content such as text and graphics. These two types of content have different properties requiring different processing and compression techniques. However, it is often required to compress such content with standard video coding solutions. In this paper a low-cost solution for improved screen content coding based on the upcoming video coding standard HEVC is presented. It includes specific intra and inter coding solutions that skip some of the common video coding methods, in this case transforms, enhancing the compression of screen content. Additional signalling and signal-level adjustment methods are introduced. Savings of up to 30% of the bit-rate are observed for intra coding and up to 25% for inter coding of screen content. Modest gains of up to 3% are observed for content that consists of both camera captured content and graphics..

Index Terms— screen content, high efficiency video coding, spatial transforms, quantisation

1. INTRODUCTION

With increased consumption of multimedia on different devices, the need for efficient solutions for screen content has increased. The effects of existing video coding technology, e.g. H.264 / AVC and High Efficiency Video Coding (HEVC) standard under development, may negatively influence visual quality of screen content. For example, compression of frames that include detailed structure of characters and a sharp program window boundary may not benefit from the two-dimensional transforms that are common in video coding solutions.

Because of the mixed nature of screen content it is rarely possible to efficiently use specialised compression techniques that are suitable for graphical content only. Especially in broadcasting, a single coding solution must be used. Since camera captured content often requires higher bit-rate and greater compression than graphical content, the choice for a single codec is one of the codecs suitable for such material (e.g. MPEG-2 and H.264 / AVC).

The next generation of video coders will be based on HEVC that is currently under development by the Joint Collaborative Team on Video Coding (JCT-VC). While HEVC shares the same principles of H.264 / AVC, such as block-based coding with motion compensated predictions and context-adaptive binary arithmetic coding (CABAC), it also introduces numerous new tools including new transforms, advanced intra and inter predictions, larger block structures, and improved coefficient coding. In terms of advances in transform coding, HEVC uses Residual QuadTree (RQT, [1]) for splitting a coding unit or a prediction block into transform blocks. Typically, 2 splitting levels are used, with the smallest block size of 4 × 4 pixels. DCT and DST (discrete cosine and discrete sine transforms) are applied to the transform blocks, depending on the type of prediction used for given block. The choice is based on intra prediction direction on which the properties of residues depend [2].

Recently it has been shown that although video compression may benefit from advanced transforms, it may also benefit from skipping the transform altogether. Advanced intra prediction of non-square blocks, complemented with non-square transforms, i.e. Short Distance Intra Prediction (SDIP, [3]), is beneficial for a wide spectrum of content. For intra coding of graphical content, so called Residual Scalar Quantisation (RSQ, [4]) significantly improves coding by skipping transforms in standard video coding frameworks. It has also been shown that coding of motion compensated residuals can benefit from some form of transform skipping. The motion compensated residue often consists of sharp edges or uncorrelated pixels, both in camera captured and graphical content. Such properties can be exploited in compression by skipping the transform on rows and/or columns of a given block, as defined by the inter Transform Skip Mode (TSM, [5]).

In order to enhance a standard solution for video coding with tools for suitable for screen content coding, in this paper we investigate the performance of transform skipping tools - TSM and RSQ, in the context of HEVC.

These methods are introducing minimal changes to the HEVC encoder, i.e. they do not impose burden on common video coding structure and its encoding or decoding requirements.

In Section 2 properties of typical examples of screen content are summarised. Section 3 includes descriptions of intra and inter coding tools that are based on transform skipping. The experimental setup and results are reported in Section 4, while Section 5 concludes this paper.

2. SCREEN CONTENT

Being composed of both camera-captured and graphical content, screen content would benefit from a coding solution that is suitable for both types of content. Such content is increasingly used in applications such as desktop sharing, video conferencing with auxiliary material (e.g. presentation slides) and broadcasting. Sources of graphical content vary and some typical examples include a computer screen with characters, graphics, webpage scrolling, video playing, horizontal and vertical lines, as well as video game content.

Some examples of such content were studied in JCT-VC [6]. To demonstrate the variety of the content used in that study, Figure 1. includes frames and details from different sequences used for testing in this paper.

a) A frame from test sequence "BasketballDrillText"

b) A detail from "HKUST1" test sequence

c) A detail from "BJUTslide" test sequence

Figure 1: Examples of screen content

3. TRANSFORM SKIPPING

The usual implementation of a 2D transform in standard video codecs is based on integer implementation of DCT and DST - as 1D transforms applied on rows and columns of transform units. While for certain classes of signals the implementation of transforms in both directions can be an optimal choice in terms of energy compaction, for signals with different properties some other strategies may lead to better Rate-Distortion (RD) performance. It is possible, and indeed it is also common under some circumstances, for the 2D signal at the input to the transform to display different statistics along the vertical and horizontal axes. This has already been addressed in HEVC by the adoption of DST for intra residuals related to certain directional prediction modes. However, certain types of residue can benefit from completely skipping the intra transform as demonstrated by the implementation of the RSQ.

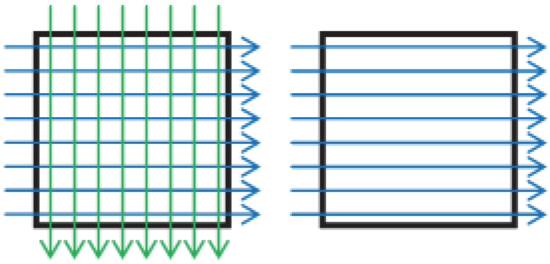
For sparser motion compensated signals that also exhibit different properties in the horizontal and vertical directions, alternative processing options can be derived. Such blocks can be processed using a method that skips transforms of blocks in the horizontal and/or vertical directions, to improve compression, as defined by TSM.

3.1. Intra Transform Skip

Intra transform skipping mode is a simple modification or extension to the current intra coding in HEVC. In HEVC, intra coding for each Transform Unit (TU) consists of several steps. First, the surrounding reconstructed pixels will be used to predict the pixels of the current TU along a certain direction. Second, a 2-D transform corresponding to the TU is applied to the predicted residue. Third, the transform coefficients are quantised and CABAC is used to code the quantised coefficients. Fourth, based on the quantised coefficients, the coded pixels are reconstructed for the current TU, which will be used to predict later TUs. Usually, since the correlation within an image is much less than that among inter images, intra prediction yields comparatively large residues, especially for those TUs with large sizes in HEVC. The intra prediction residues still preserve a certain extent of the correlation, which makes the 2-D transform reasonable and effective. However, since screen content is noiseless and usually has more regular and sharper signals, intra prediction might work well and lead to much less or even no residues. In such a case, a transform might be inefficient and even worsen the coding. For these TUs, we will skip the 2-D transform but leave all other steps unchanged to make the modification as small as possible. Modification to the entropy coding to separate the statistics between transform TUs and transform skipping TUs can further improve the coding performance.

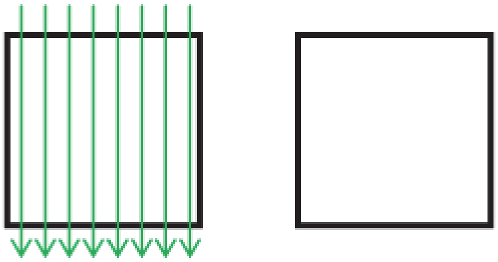
3.2. Inter Transform Skip Mode

The Transform Skip Mode (TSM) defines the transform skip in one or both directions on which a transform would be applied under normal conditions, Figure 2. As illustrated in Figure 2.a), TSM also covers the traditional approach where the transform is applied on both rows and columns of a block. A summary of TSM modes is given in Table 1.



a) 2D transform

b) transform on rows only



c) transform on columns only

d) no transform

Figure 2: Transform choices enabled by TSM

Table 1: Transform skip modes

TSM mode	**Horizontal direction**	**Vertical direction**
TS0	Transformed	Transformed
TS1	Transformed	Skipped
TS2	Skipped	Transformed
TS3	Skipped	Skipped

When transform skipping is used in one direction, it is necessary to perform normalisation of the coefficients which is a generalisation of the normalisation performed with the full skip. Such a process can be embedded into the quantisation step. The quantisation (both forward and inverse) is performed using a modified quantisation factor:

$$Q' = (Q \cdot bscale) / 2^{bshift}, \qquad (1)$$

where $Q' = Q$ for the full 2D transform, Q is a quantisation factor (inverse or forward), and *bscale* and *bshift* are dependent on TSM modes and the current block dimensions. The block scale is defined as:

$$bscale = \begin{cases} 1, & \text{if TS0} \\ scale(height), & \text{if TS1} \\ scale(width), & \text{if TS2} \\ scale(height) \cdot scale(width), & \text{if TS3} \end{cases}$$

where *scale* is dependent on the implementation of the given transform. For HEVC, *scale* is defined as in Table 2, where the actual values are related to the norms of HEVC transforms.

Table 2: Scaling factors for vectors of *N* elements.

N	4	8	16	32
scale	128	181	256	362

The block shift *bshift* corresponds to the shifts of skipped transforms. Shifts are used in a practical implementation to limit the bit-depth of the coefficients after each transform step. For the inverse quantisation in HEVC, the blocks shift is defined as:

$$bshift = \begin{cases} 0, & \text{if TS0} \\ s_1, & \text{if TS1}\,. \\ s_2 - \Delta, & \text{if TS2} \\ s_1 + s_2 - \Delta, & \text{if TS3} \end{cases}$$

Δ denotes the internal bit depth increment, relative to the common 8-bit processing bit depth, and s_1 and s_2 are bit-shifts of the inverse transforms on columns and rows, respectively.

4. EXPERIMENTS AND RESULTS

The experiments are performed using 13 screen content sequences that were evaluated in JCT-VC activities [6]. The resolution and frame rate of the sequences are in the range 832 × 480 to 1920 × 1080 and 10 fps to 60 fps, see Table 3. All sequences are in 4:2:0 sampling format in order to be compatible with the current HEVC test model (HM-5.0, [7]). The bit-depth of all sequences is 8 bits. The test material covers a wide range of screen content. BJUT and HKUST sequences are based on captured desktops including website browsing, text editing, pictures, working in circuit design programs, slide show and other commonly used applications. The remaining sequences are based on camera captured sequences used in JCT-VC, with graphical overlays as in Figure 1.a.

Software used in this experiment is based on HM-5.0. The proposed intra and inter transform skipping approach with supporting tools is evaluated in comparison to the baseline HM performance. Six configurations used in JCT-VC [8] were tested: all intra (AI), random access (RA) and low-delay (LD) with high efficiency (HE) and low-complexity (LC) settings. In addition to the tools used in the LC setting, HE includes rate distortion optimised quantisation, sample adaptive offset, adaptive loop filter, chroma from luma intra prediction mode, non-square transforms and asymmetric motion partitions, [8]. Quantisation factors used in the test are 22, 27, 32 and 37.

The choice between transform skipping (1D or 2D) and performing a 2D transform is decided at the encoder side during the rate-distortion optimisation search. To keep the encoding time lower, both intra and inter RQT levels are set to 1, i.e. RQT is not used with transform skipping. However, it is used to obtain results with HM anchors. In AI coding configurations, the encoding time is about 30% higher than for HM, while for RA and LD configurations the encoding time is on average the same as the encoding time of HM. Decoding times of the tested approach and the anchor are close, with a difference that in transform skipping some expensive transform operations are not needed.

Table 3 summarises the results of this evaluation. For desktop content (first 6 sequences) transform skipping achieves gains in the range of 2.4% to 29.7% BD-rate, [9]. The largest gains are observed for intra and low-complexity configurations. Due to typically linear motion in desktop content, frames can be very well predicted from intra coded frames, and therefore the gain is mostly related to intra coding. In low-complexity settings gains are larger since transform skipping partially compensates for gains of the excluded tools. Although the remaining sequences contain a relatively small amount of graphical content, the gains of transform skipping are up to 2.9% BD-rate. This gain is related to transform skipping in intra coding of graphical parts of related sequences, and to transform skipping in inter coding in all parts of the frames. Because of larger number of motion compensated frames in the low-delay configuration, inter transform skip is contributing more significantly to the overall gain in that specific configuration.

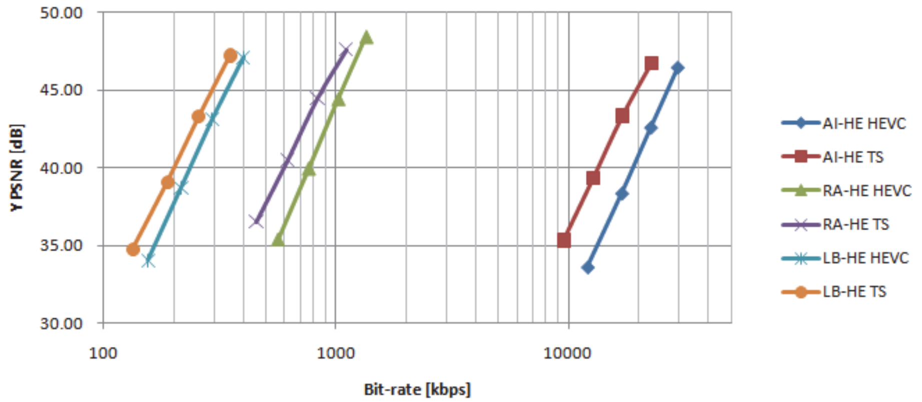
To demonstrate PSNR and bit-rate ranges used in the test, Figure 3 summarises the results for two sequences encoded in HE configurations. From the figures it can be seen that transform skipping is achieving gain both for low and high qualities of coded content.

5. CONCLUSIONS

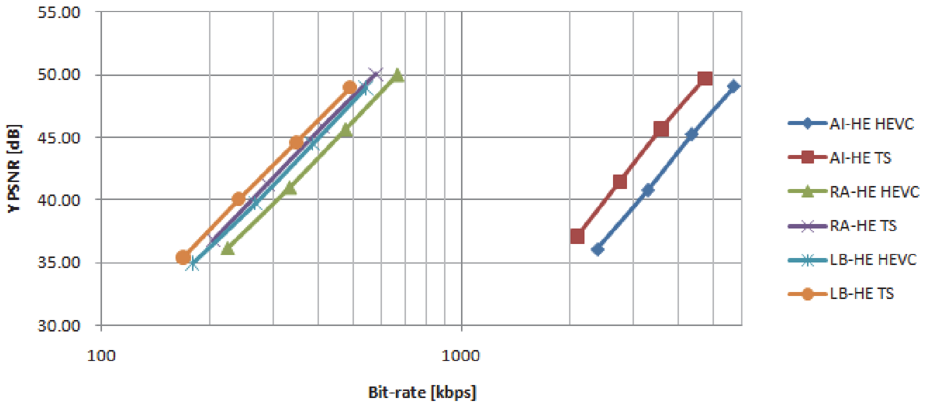
The combination of transform skipping methods for intra and inter coding is evaluated for compression of screen content. Large gains of up to 30% BD-rate have been observed in comparison to the compression with HM-5.0, showing the suitability of the tested methods for coding of screen content. Careful design of the supporting tools for transform skipping (i.e. for additional processing of the quantisation residual) is essential for achieving gain in the presence of a skipped transform. The proposed combination does not have a negative influence on the coding that requires full skipping since the skip is in both intra and inter cases made optional. In this way a codec suitable for coding mixed camera captured content and graphical content is provided.

Table 3: BD rate results for luminance component for transform skipping compared to HEVC without transform skipping; negative values indicate bit-rate savings

			Coding configuration					
Sequence	Resolution	Frame rate	AI-HE	AI-LC	RA-HE	RA-LC	LD-HE	LD-LC
BJUTdoc	1280x720	10	-19.9%	-22.7%	-14.6%	-20.2%	-11.3%	-17.9%
BJUTweb	1280x720	10	-19.9%	-23.7%	-13.0%	-20.5%	-11.7%	-17.1%
BJUTslide	1280x720	20	-5.3%	-7.1%	-5.0%	-8.5%	-2.4%	-4.4%
HKUST1	1280x720	30	-28.4%	-29.7%	-20.5%	-25.5%	-14.6%	-22.0%
HKUST2	1280x720	30	-26.0%	-27.7%	-17.8%	-22.6%	-6.0%	-13.2%
HKUST3	1280x720	30	-20.6%	-20.2%	-19.7%	-23.6%	-14.9%	-18.1%
RaceHorsesText	832x480	30	-1.5%	-2.1%	-0.4%	-1.0%	-0.1%	-0.1%
BasketballDrillText	832x480	50	-1.9%	-2.9%	-0.8%	-1.7%	-0.2%	-0.4%
BasketballDriveText	1920x1080	50	-2.0%	-2.9%	-0.5%	-1.3%	0.0%	-0.2%
PartySceneText	832x480	50	-1.3%	-1.9%	-0.5%	-1.3%	-0.1%	-0.1%
CactusText	1920x1080	50	-1.5%	-2.2%	-0.6%	-1.3%	-0.1%	-0.3%
BQTerraceText	1920x1080	60	-1.1%	-1.8%	-0.4%	-1.4%	-0.1%	-0.3%
BQMallText	832x480	60	-1.6%	-2.3%	-0.5%	-1.2%	0.0%	-0.3%



a) HKUST1



b) BJUTdoc

Figure 3: Example of the comparison of transform skipping (TS) and HM-5.0 (HEVC) performance for HKUST1 and BJUTdoc sequences

6. ACKNOWLEDGMENTS

The authors would like to thank Beijing University of Technology (BJUT), Hong Kong University of Science and Technology (HKUST), and Huawei for providing screen content material used for experiments presented in this paper.

7. REFERENCES

[1] Martin Winken, Philipp Helle, Detlev Marpe, Heiko Schwarz, Thomas Wiegand, "Transform Coding in the HEVC Test Model," in Proc. IEEE Int'l Conf. on Image Processing, September 2011.

[2] Chuohao Yeo, Yih Han Tan and Zhengguo Li, "Low-complexity mode-dependent KLT for block-based intra coding," in Proc. IEEE Int'l Conf. on Image Processing, September 2011.

[3] Xiaoran Cao et al. "CE6.c Report on Simplification of Short Distance Intra Prediction (SDIP) Method," document JCTVC-G556 of Joint Collaborative Team on Video Coding, November 2011

[4] Cuiling Lan, Guangming Shi, Feng Wu, "Compress Compound Images in H.264/MPGE-4 AVC by Exploiting Spatial Correlation," IEEE Transactions on Image Processing, Vol. 19, Iss. 4, pp. 946 - 957, April 2010.

[5] Andrea Gabriellini, Matteo Naccari, Marta Mrak, David Flynn, "Spatial Transform Skip in the Emerging High Efficiency Video Coding Standard," submitted to IEEE Int'l Conf. on Image Processing 2012.

[6] Oscar C. Au, Jizheng Xu, Haoping Yu, "JCT-VC AHG report: Screen Content Coding," document JCTVC-F013 of Joint Collaborative Team on Video Coding, July 2011.

[7] HEVC Test Model version 5.0 (HM-5.0), available from http://hevc.kw.bbc.co.uk/trac/browser/tags/HM-5.0

[8] Frank Bossen, "Common test conditions and software reference configurations," document JCTVC-G1200 of Joint Collaborative Team on Video Coding, November 2011.

[9] Gisle Bjontegaard, "Calculation of Average PSNR Differences between RD curves," document VCEG- M33 of ITU-T SC16/Q6, Video Coding Experts Group (VCEG), April 2001.