

Open Research Online

The Open University's repository of research publications and other research outputs

Transform domain distributed video coding using larger transform blocks

Conference or Workshop Item

How to cite:

Mahmood, Asif; Dooley, Laurence and Wong, Patrick (2018). Transform domain distributed video coding using larger transform blocks. In: 5th IEEE Global Conference on Signal and Information Processing, 14-16 Nov 2017, Montreal, Canada.

For guidance on citations see [FAQs](#).

© [not recorded]



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Accepted Manuscript

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.1109/GlobalSIP.2017.8308592>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

TRANSFORM DOMAIN DISTRIBUTED VIDEO CODING USING LARGER TRANSFORM BLOCKS

Asif Mahmood Laurence S Dooley Patrick Wong

School of Computing and Communications, The Open University, UK

ABSTRACT

Distributed Video Coding (DVC) displays promising performance at low spatial resolutions but begins to struggle as the resolution increases. One of the limiting aspects is its 4x4 block size of *Discrete Cosine Transform* (DCT) which is often impractical at higher resolutions. This paper investigates the impact of exploiting larger DCT block sizes on the performance of transform domain DVC at higher spatial resolutions. In order to utilize a larger block size in DVC, appropriate quantisers have to be selected and this has been solved by means of incorporating a content-aware quantisation mechanism to generate image specific quantisation matrix for any DCT block size. Experimental results confirm that the larger 8x8 block size consistently exhibit superior RD performance for CIF resolution sequences compared to the smaller 4x4 block sizes. Significant PSNR improvement has been observed for 16x16 block size at 4CIF resolution with up to 1.78dB average PSNR gain compared to its smaller block alternatives.

Index Terms— distributed video coding, wyner-ziv video coding, discrete cosine transform, transform block size, high resolution

1. INTRODUCTION

Distributed video coding (DVC) [1] is an alternate paradigm, which in contrast to conventional video codecs like H.264/AVC [2], shifts the computational burden of exploiting statistical redundancies from the encoder to the decoder. It is based on the work of Slepian & Wolf [3] and Wyner & Ziv [4] which established *rate-distortion* (RD) performance bounds for coding multiple correlated sources in a separate encoding-joint decoding scenario. DVC achieves competitive RD performance using simple encoders, but at the cost of more complex decoder designs, making it attractive for a wide range of emerging low-power multimedia applications including wireless surveillance networks, drones and the Internet of Things. Most effective DVC architectures feature transform domain coding using the *Discrete Cosine Transform* (DCT) [5, 6]. The RD performances of these codecs consistently surpasses that of the H.264 Intra codec, especially in low spatial resolution sequences, though a notable performance gap emerges when comparison is made with more complex inter-frame codecs [7].

Transform domain Wyner-Ziv video coding (TDWZ) solutions employ a 4x4 block based DCT followed by quantisation. The 4x4 transform block size was originally introduced in the H.264/AVC standard, with the transform applied to the prediction residuals leading to better compression performance than the more common 8x8 block size. This can however, lead to a RD performance deficit in homogeneous regions, particularly in higher resolution scenarios where the 4x4 block size is too small to be of practical use. This observation is upheld in both the transform coding designs of the H.264 Fidelity Range extension, which allows 8x8 block sizes [8] and in

the state-of-the-art H.265/HEVC standard which allows up to 32x32 block sizes [9].

Although larger DCT block sizes exhibit better compression performance, their implementation is non-trivial in the TDWZ context due to various design choices. The DCT is usually followed by quantisation which enables the flexibility of saving bandwidth at the cost of reduced quality by quantising DCT coefficients more coarsely. In conventional codecs, quantised coefficients are compressed using lossless entropy coders to further reduce the bit-rate. In contrast, quantised coefficients in TDWZ are never transmitted to the decoder, but instead are fed into a channel encoder to produce parity bits. The decoder generates an estimate of the frame and then applies the DCT and quantisation. The quantised coefficients produced act as the noisy version of original transform coefficients, which are then corrected by the channel decoder using the parity bits. Since channel coding performance is better for longer input sequences, the corresponding DCT coefficients are grouped together to form coefficient bands. Adopting a larger block size in TDWZ produces more coefficient bands, each being shorter compared to the smaller block sizes, which respectively affects the performance of the channel coder.

Selecting the appropriate quantiser is a key design challenge in TDWZ coding. All the coefficients in a DCT coefficient band are collectively quantised before being converted into bit-planes for processing by the channel coder. As a consequence, the number of steps in the quantiser must be a power of 2 to maximise bandwidth utilisation. While some generic quantisation matrices (QM) have been proposed in [10, 11], these have been determined by trial and error for a range of test sequences and are only applicable to 4x4 block sizes. A good set of quantisers for larger block sizes ensures the best achievable output quality at any bit-rate while a bad one unnecessarily wastes bandwidth and produces low-quality output with artefacts. Thus in order to effectively support larger transform block sizes, selecting of most appropriate quantiser is a necessary design objective.

This paper critically evaluates the impact of DCT block sizes on the RD performance of the TDWZ codec. Three different DCT block sizes, 4x4, 8x8 and 16x16 are analysed and their respective bit-rate and *peak signal to noise ratio* (PSNR) performance evaluated. This TDWZ implementation, for the first time, facilitates block sizes other than of 4x4 pixels. To determine the corresponding QM for the different block sizes, a novel *content-aware quantisation* (CAQ) mechanism is introduced and integrated into the TDWZ architecture to generate a custom QM for each frame, by dynamically evaluating the DCT coefficients distribution to minimise the quantisation error. The benefit of incorporating CAQ into the TDWZ architecture is that it is able to produce a QM for any block size. Experiments have been undertaken on several popular *quarter common interchange format* (QCIF) (176x144 pixels) and *common interchange format* (CIF) (352x288 pixels) resolution test sequences, while additionally, some 4CIF (704x576 pixels) resolution sequences have

QM generated using the CAQ module were employed, with the RD performance of the former being used as the baseline comparator. The performance of 8x8 transform block size was also benchmarked for all test sequences, while the larger 16x16 transform block size was not used for QCIF resolution sequence due to unavailability of a LDPCA coder of suitable length, so this was only evaluated for the CIF and 4CIF resolution sequences. Note that the QM for both the 8x8 and 16x16 block sizes was generated by CAQ.

A set of four ABP values were input to the CAQ module to measure the TDWZ codec performance at different bit-rates for each DCT block size. The decoder selected an ABP value from this set in an analogous manner to selecting the QP in H.264/AVC in order to control the RD performance. For the 4x4 block size, the set contained 8, 16, 24, and 32, while for the 8x8 block size the constituents were 32, 64, 96, and 128. The values 128, 256, 384, and 512, were used for the 16x16 block size. These numbers indicate how many bit-planes are to be transmitted for each frame and while the ABP are comparatively much high for the larger block sizes, the resulting bit-rates are not significantly affected because the bit-plane lengths are far shorter. Bit-plane lengths for the LDPCA coder for different block sizes and spatial resolutions are summarised in Table 1.

Table 1. LDPCA coder lengths for different combinations of transform block sizes and spatial resolutions

	4x4	8x8	16x16
QCIF	1584	396	N/A
CIF	6336	1584	396
4CIF	25344	6336	1584

To equitably compare the TDWZ codec performance, some configuration parameters including block size and QM index / ABP value were preselected to dictate the transform block size and quantisation method to be used. The codec then selects the desired transform and quantisation block size according to these configuration parameters and also the LDPCA coder with the appropriate bit-plane length. All other DVC modules, including the SI generation and correlation noise model are unaffected by the configuration parameters and are fixed for all experiments thereby ensuring the only factor impacting the RD results is the transform block size and the bit-plane length.

4. RESULTS DISCUSSION

Several QCIF, CIF and 4CIF resolution test sequences sampled at 30 frames per second were applied to measure PSNR and bit-rate performance of different block sizes. The simulations were conducted with a *group of pictures* (GOP) size 2 along with varying fixed QM indices/ABP values and corresponding average bit-rate and average PSNR of WZ frames were plotted to generate RD curves in order to illustrate the performance difference between various transform block sizes.

To measure the performance of the TDWZ architecture with different block sizes at 4CIF resolution, experiments were performed on *Soccer* and *Crew* sequences whose RD performance graphs are given in Fig. 2. The nomenclature adopted in the legends, namely DCT4-F, DCT4, DCT8 and DCT16 respectively refers to 4x4 block size with fixed QM, 4x4 block size with CAQ, 8x8 block size with CAQ and 16x16 block size with CAQ. 16x16 block size was superior to 8x8 block size while both 16x16 and 8x8 block sizes performed significantly better than 4x4 block size for both sequences.

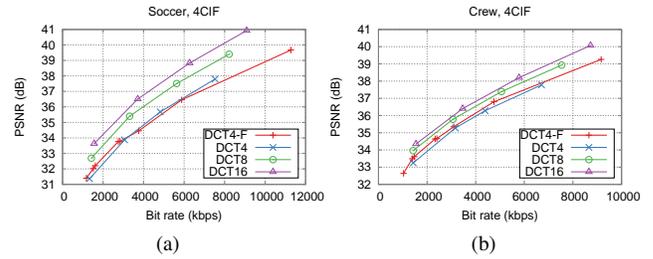


Fig. 2. RD performance for WZ frames at GOP 2 for (a) *Soccer* and (b) *Crew* at 4CIF resolution

The improved performance of larger block sizes can be explained through the compression mechanism of the codec. Compression takes place in two stages in the TDWZ architecture, during DCT and quantisation, and at the LDPCA coder. DCT accumulates the information within a block around its lower frequency components while higher frequency components can be gracefully discarded through quantisation without significantly affecting the fidelity. Therefore, DCT with larger block sizes are generally better than DCT with smaller block sizes in terms of compression performance since greater number of higher frequency components can be discarded. The LDPCA coder compresses a bit-plane by reconstructing it from corresponding SI bit-plane using parity bits, where the number of transmitted parity bits is less than the number of bits in the actual bit-plane. For instance, if the quality of the SI is good, i.e. SI is close to the original WZ frame, the correlation noise characterised by the difference between SI and WZ frame will be low. Therefore, the LDPCA coder will require very few parity bits to recover from this low noise, thus the bit-planes will be heavily compressed. Performance of the LDPCA coder also depends on the length of its input bit-plane with longer ones being superior to a shorter alternative. Respective input lengths of the LDPCA coder for different combinations of block sizes and spatial resolutions have been given in Table 1. The improved performance of larger block sizes originates from the performance advantage of larger DCT block size being greater than the performance deficit of the shorter LDPCA coder length.

The difference between RD curves measured by *Bjøntegaard Delta PSNR improvement* (BD-PSNR) and corresponding *relative bit-rate savings %* (BD-Rate) [15] for larger block sizes with respect to 4x4 block size were also computed. A positive BD-PSNR value indicates that the larger block size is able to gain higher PSNR than the smaller block size at the same bit-rate. A negative BD-Rate value refers to the bit-rate percentage of 4x4 block size saved by using a larger block size to achieve exactly the same PSNR. For *Soccer*, 16x16 block size scored 1.78dB and 40.42% for BD-PSNR and BD-Rate respectively over 4x4 block size. Corresponding performances of 8x8 block size was 1.07dB and 27.50%. For *Crew*, the BD-PSNR and BD-Rate metrics for 16x16 block size was 0.72dB and 25.53% while 8x8 block size obtained 0.46dB and 18.05% respectively.

Several CIF resolution test sequences were simulated to measure the performance of the TDWZ architecture with different block sizes at this resolution. RD performance of the codec with different block sizes and fixed QM indices/ABP values were measured, though only BD-PSNR and BD-rate metrics are presented in Table 2 due to space limitation. For all CIF sequences tested here, 8x8 block sizes provide a bit-rate saving compared to 4x4 block size to achieve the same PSNR. It also achieved better PSNR than 4x4 block size for all se-

quences except *Coastguard*, which has a slightly lower PSNR. The reason for this becomes apparent in the corresponding RD graphs for this sequence, with 4x4 block sizes found to be superior at very low bit-rates, while the 8x8 block size surpassed the 4x4 block size at both medium and high bit-rates.

Table 2. BD-PSNR (dB) and BD-Rate (%) metrics of 8x8 and 16x16 block sizes over 4x4 block size at CIF resolution

Sequence	8x8		16x16	
	BD-PSNR	BD-Rate	BD-PSNR	BD-Rate
<i>Bus</i>	0.16	-4.57	-0.48	16.75
<i>Coastguard</i>	0.10	-2.27	-1.02	66.42
<i>Crew</i>	0.48	-18.91	0.16	-8.81
<i>Football</i>	0.35	-7.26	0.38	-7.72
<i>Foreman</i>	0.33	-23.97	-0.83	62.38
<i>Hall</i>	0.05	-1.16	-2.23	637.07
<i>Mother</i>	0.21	-28.58	-1.96	867.25
<i>Soccer</i>	0.33	-9.19	0.39	-10.58
<i>Stefan</i>	0.67	-16.75	0.45	-10.95

Performance gap between length-1584 LDPCA coder used for 8x8 block size and length-6336 LDPCA coder used for 4x4 block size is not very significant and was compensated by the superior performance of the 8x8 block based DCT over its 4x4 block counterpart. Even for *Hall*, where the SI is generally very close to the original WZ frame due to low object motion activity and consequently, the compression performance of the LDPCA coder is nearly maximal, 8x8 block based DCT was able to recover from the performance deficit of its shorter-length LDPCA coder. Conversely, for *Stefan* which contains high object motion, the compression performance of the LDPCA coder is very poor due to a low-quality SI, so there is little advantage remaining of using a longer LDPCA coder length leading to a much higher PSNR gain.

BD-PSNR and BD-Rate metrics for 16x16 block sizes are also presented in Table 2 for CIF resolution sequences. Similar to the 8x8 block size, superior RD performance compared to the 4x4 block size was attained for sequences characterised by high and complex motion activity. It surpassed the performance of the 8x8 block size for *Football* and *Soccer* though its comparative performance dropped noticeably in low-motion sequences due to the significant performance deficit of length-396 LDPCA coder from length-6336 LDPCA coder paired with 16x16 and 4x4 block sizes respectively.

For completeness, Some QCIF resolution test sequences were also simulated though only 8x8 block size was analysed whose BD-PSNR and BD-rate metrics are given in Table 3. N/A values in the BD-Rate column means that there were no common PSNR range covered by both the respective RD curves. All entries in Table 3 shows lesser performance of 8x8 block size with respect to 4x4 block size for QCIF resolution sequences. The performance gap between length-396 LDPCA coder used for 8x8 block size and length-1584 LDPCA coder used for 4x4 block size is much greater than the performance gap between longer LDPCA coders used in the experiments of higher resolution sequences. The performance advantage of larger DCT block size was not able to compensate for this significant deficit leading to the lesser performance of 8x8 block size than 4x4 block size.

Overall, the TDWZ codec performance results for various DCT block sizes reveal that while the 4x4 block size is the best choice for QCIF sequences, for higher resolution CIF, 8x8 block sizes perform better especially in sequences exhibiting more complex object motion and occlusion characteristics. The 16x16 block size is generally

Table 3. BD-PSNR (dB) and BD-Rate (%) metrics of 8x8 block size over 4x4 block size at QCIF resolution

Sequence	BD-PSNR	BD-Rate
<i>Akiyo</i>	-2.58	N/A
<i>Coastguard</i>	-0.82	252.57
<i>Crew</i>	-0.04	1.13
<i>Foreman</i>	-0.88	92.38
<i>Hall</i>	-2.12	N/A
<i>Mother</i>	-1.59	N/A
<i>Soccer</i>	-0.31	14.067
<i>Mobile</i>	-1.21	529.12

not suitable for either QCIF or CIF sequences except for cases containing high motion activities. This originates from the decoder being unable to generate a sufficiently high-quality SI thereby leading to poor LDPCA coder performance which is overcome by superior performance of larger DCT block size. Thus, larger block sizes are better for such sequences as well as at higher GOP scenarios, where generating high-quality SI is more challenging. It is also advantageous in latency constrained scenarios where SI is generated from only temporally preceding frames using frame extrapolation techniques [16]. In contrast, with more advanced SI generation techniques [5, 6] that produce high quality SI, smaller block sizes are a better alternative. In the experiments conducted on higher resolution 4CIF sequences, 16x16 block size consistently outperformed both 8x8 and 4x4 block sizes, though these initial findings needs more rigorous investigation to validate the TDWZ codec performance at high spatial resolutions.

In addition to the consistent RD performance improvement, using larger DCT block sizes has another implicit advantage in parallel processing scenario [7], where multiple bit-planes can be simultaneously decoded. Concurrently decoding two bit-planes from the same DCT band is not pragmatic because it limits the intermediate refinement options, however since DCT coefficients are statistically independent, decoding bit-planes from different bands does not impose such a constraint so these architectures can capitalise on the larger DCT block sizes because they produce more coefficient bands.

5. CONCLUSION

This paper has investigated the impact of exploiting larger transform block sizes on the performance of the *transform domain Wyner-Ziv video coding* (TDWZ) architecture. TDWZ solutions have traditionally only used 4x4 block sizes for the *discrete cosine transform* (DCT) together with set of fixed *quantisation matrices* (QM) for quantising the DCT coefficients. The RD performance of 4x4 block based DCT is inferior to larger block sizes in homogeneous regions and is often impractical for higher resolution sequences. In order to utilize a larger block size in TDWZ, appropriate quantisers have to be selected and this has been solved by means of incorporating a content aware quantisation mechanism to generate image specific quantisation matrix for any DCT block size. Experimental results confirm that the larger 8x8 block size consistently exhibit superior RD performance for CIF resolution sequences compared to the smaller 4x4 block sizes. Significant PSNR improvement has been observed for 16x16 block size at 4CIF resolution with up to 1.78dB average PSNR gain compared to its smaller block alternatives.

6. REFERENCES

- [1] Bernd Girod, Anne Margot Aaron, Shantanu Rane, and David Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 71–83, 2005.
- [2] Thomas Wiegand, Gary J Sullivan, Gisle Bjøntegaard, and Ajay Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.
- [3] David Slepian and Jack Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, 1973.
- [4] Aaron Wyner and Jacob Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, 1976.
- [5] Huynh Van Luong, Lars Lau Raket, and Søren Forchhammer, "Re-estimation of motion and reconstruction for distributed video coding," *IEEE Transactions on Image Processing*, vol. 23, no. 7, pp. 2804–2819, 2014.
- [6] Abdalbassir Abou-Elailah, Frederic Dufaux, Joumana Farah, Marco Cagnazzo, Anuj Srivastava, and Béatrice Pesquet-Popescu, "Fusion of global and local motion estimation using foreground objects for distributed video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 6, pp. 973–987, 2015.
- [7] Yun-Chung Shen, Han-Ping Cheng, Ji-Ciao Luo, Yu-Hsun Lin, and Ja-Ling Wu, "Efficient real-time distributed video coding by parallel progressive side information regeneration," *IEEE Sensors Journal*, vol. 17, no. 6, pp. 1872–1883, 2017.
- [8] Gary J Sullivan, Pankaj N Topiwala, and Ajay Luthra, "The H.264/AVC advanced video coding standard: overview and introduction to the fidelity range extensions," in *Optical Science and Technology, the SPIE 49th Annual Meeting*. International Society for Optics and Photonics, 2004, pp. 454–474.
- [9] Gary J Sullivan, Jens Ohm, Woo-Jin Han, and Thomas Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.
- [10] Anne Aaron, Shantanu D Rane, Eric Setton, and Bernd Girod, "Transform-domain Wyner-Ziv codec for video," in *Proceedings of SPIE*, 2004, vol. 5308, pp. 520–528.
- [11] Catarina Brites, Joao Ascenso, and Fernando Pereira, "Improving transform domain Wyner-Ziv video coding performance," in *2006 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2006, vol. 2, pp. II–II.
- [12] Xavi Artigas, Joao Ascenso, Marco Dalai, Sven Klomp, Denis Kubasov, and Mourad Ouaret, "The DISCOVER codec: architecture, techniques and evaluation," in *Picture Coding Symposium*, 2007.
- [13] David Varodayan, Anne Aaron, and Bernd Girod, "Rate-adaptive codes for distributed source coding," *Signal Processing*, vol. 86, no. 11, pp. 3123–3130, 2006.
- [14] Joao Ascenso, Catarina Brites, and Fernando Pereira, "Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding," in *5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services*. Smolenice, Slovak Republic, 2005, pp. 1–6.
- [15] Gisle Bjøntegaard, "Calculation of average PSNR differences between RD-curves," *Doc. VCEG-M33 ITU-T Q6/16, Austin, TX, USA, 2-4 April 2001*, 2001.
- [16] Ran Li, Zongliang Gan, Ziguan Cui, Minghu Wu, and Xiuchang Zhu, "Side information extrapolation using motion-aligned auto regressive model for compressed sensing based Wyner-Ziv codec," *KSII Transactions on Internet & Information Systems*, vol. 7, no. 2, 2013.