Low complexity in-loop prediction perceptual video coding for HEVC

Low complexity in-loop prediction perceptual video coding for HEVC

Y. G. Joshi*, J. Loo*, P. Shah*, S. Rahman*, A. Tasiran*, and J. Cosmas†
*School of Science and Technology, Middlesex University, The Burroughs, Hendon, London, NW4 4BT, UK
Email: {y.joshi, j.loo, p.shah, s.rahman, a.tasiran}@mdx.ac.uk
†Multimedia and Broadcast Networks Group, Department of Electronic & Computer Engineering,
Brunel University, Uxbridge, Middlesex. UB8 3PH, United Kingdom
Email: {john.cosmas}@brunel.ac.uk

***Abstract*—This paper applies the concept of hybrid framework for perceptual video coding (PVC) during the 'in-loop' stages by extending it to the prediction stage. As low complexity environments of mobile phones and tablets are increasingly used to capture video, PVC is not occurring here due to the high complexity of perceptual algorithms. Being able to encode using PVC will enable distortion to be merited by non-linear perceptual means than by uniform cost. While ideally, existing perceptual assessments of Structural Similarity (SSIM) is used, it is not processor friendly. The hybrid framework involves applying an additional low complexity perceptual assessment on top of existing Sum of Absolute Differences (SAD) and Sum of Absolute Transform Differences (SATD) only where distortion is perceptually significant. Consequently, the results show an increase in timing of $< +4\%$ and $< +6\%$ for video encoded with low delay P and random access profiles respectively, which is complexity competitively to other PVC solutions. This also affects bit redistribution with large reductions in bits allocated to signalling, -5 to -25%, with increases in small, medium and large block sizes. Visually, the proposed encoder encourages larger blocks on perceptually homogeneous regions and more dynamic smaller block where boundaries for textures or activity is occurring. This work can be extended to allow for perceptual quantisation to enable bandwidth reduction while maintaining perceptual quality.**

I. Introduction

Increasingly video is encoded on portable devices and shared across the Internet. Conversely, video consumption represents over 75% of all Internet and over half of mobile-data traffic [1]. Similarly, there is a decline in traditional computers and a growth in portable devices, allowing greater computing mobility [2]. While advances in video codec design on being hardware friendly have allowed video based applications on devices with limited processing resources. Therefore, video encoding on phones, tablets and cameras have become commonplace in everyday environments.

Underlying all these means of extending video into new applications is the same principle of video encoding, to represent video content for a given bandwidth. Traditionally, for a hybrid block based encoder, changes are evaluated at the block level, within 'in-loop' stages of prediction and mode decision, evaluating to find the optimal sub-block candidates or combinations of sub-blocks respectively. This evaluation involves measuring distortion (D) for the level of quantisation (λ) applied to the bits used to represent the differences (R), as shown by Equation (1) [3].

$$J_{min} = D + \lambda \cdot R \tag{1}$$

Equation (1) can be represented as a convex hull curve, known as the rate-distortion (R-D) curve, and by adjusting λ the desired bit-rate or distortion level can be met [4]. This means that for higher bandwidths greater bit information is retained, which for a hybrid block based encoder largely means less quantisation and an increase in the use of smaller sub-blocks. As video content changes, the respective R-D curve would differ, causing a different response R-D curve as illustrated in Figure 1.

Part of the R-D evaluation is distortion assessment, it can influence which sub-block candidates or combinations of sub-blocks are selected. Distortion assessment in video coding standards, differences are treated with uniform cost, irrespective of content, which does not match the non-linear response of the Human Visual System (HVS) [5], [6]. The HVS is highly complex and not fully understood, however, through experiments, theories and models have been devised which provide some insight [7], [8]. This desire to understand the HVS has been fuelled in the image coding domain, leading to the several models of Contrast Sensitivity Function (CSF), Just Noticeable Distortion (JND) and a perceptual assessment of Structural Similarity (SSIM) [9], [10], [11], [12].

II. Background

The progression of video coding into low powered devices has been possible due to video coding standards integrating process friendly techniques, as seen with H.264/Advance Video Coding, (H.264/AVC). Unfortunately, perceptual models and assessments are not designed for such environments and led to solutions which offer overall low complexity across an encoded sequence but operate outside the native 'in-loop' stages.

Existing PVC solutions tend to use SSIM as perceptual means to assess distortion as is it both perceptually effective and least complex among its peers [13]. Yet, SSIM is complex relative to existing standard traditional distortion metrics (STDM) of Sum of Square Errors (SSE), Sum of

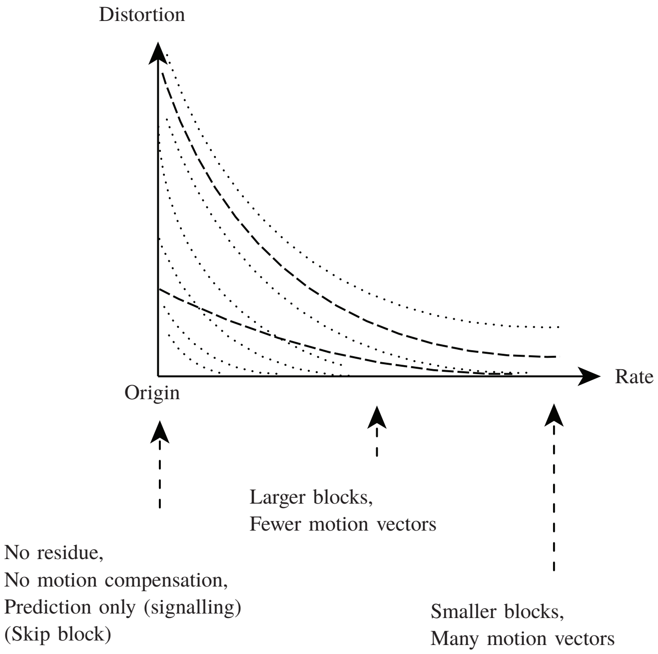


Fig. 1: Rate-Distortion curves

Absolute Difference (SAD) and Sum of Absolute Transform Difference (SATD). Consequently, existing PVC solutions are implemented out-side of the native front-end 'in-loop' stages of prediction or mode-decision as illustrated in figure 2.

Furthermore, SSIM is an index and does not provide the capabilities of a distortion metric since it does not conform to the triangle equality rule ($\triangleq$) [14]. This means that SSIM is scaled using non-linear means to be compatible with existing STDMs, which introduces further complexity, leading to PVC solutions which reside outside of the native sub-block level [15], [16]. These existing 'out-of-loop' PVC solutions operate by transforming non-perceptual scores to another value by mapping against a perceptual R-D curve as shown in figure 3.

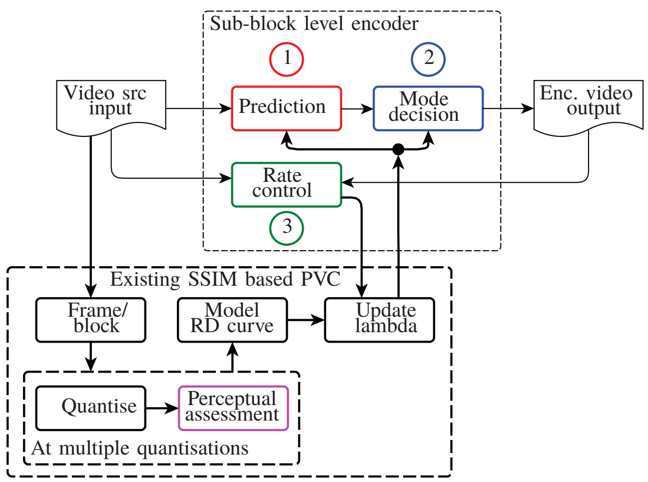


Fig. 2: Existing approach to perceptual video coding

This results in a single perceptual R-D model applicable to all mode decision choices, than the ideal of each prediction candidate or mode decision combination being evaluated on their own perceptual merits [17], [18], [19].

Ideally, distortion should be assessed perceptually, however, issues of complexity and compatibility hinder it adoption in limited processing environments. This has led investigations to understand how SSIM and STDM behave, labelling the relationship as a non-linear and geodesic-$\triangleq$ [14], [20]. When this relationship was explored at the 'in-loop' stage of prediction it revealed a shared space, revealing that perceptual assessment can vary depending upon video content [21]. Furthermore, an insight was uncovered that a component of SSIM, covariance was the key indicator to map SSIM to STDM, leading to a novel means to scale SSIM reusing existing values and low complexity techniques [22]. However, despite encouraging larger block usage, SSIM is complex relative to existing STDMs, which can justify pursuing an 'in-loop' PVC solution to influence the sub-block level choices.

To address the high complexity of perceptual assessment a framework was proposed, to apply perceptual assessment to candidates on a conditional basis, subject to whether significant perceptual distortion or activity [23]. This meant developing a new set of perceptual algorithms based upon earlier findings and existing perceptual models. Also, to keep computational demand low, pre-tests were created that sampled the sub-block candidates to determine whether perceptual assessment should occur. This meant an 'in-loop' PVC solution operated with $< +4\%$ increase in timing for medium and low activity videos, however, it lacked a solution for the most complexity sensitive stage of prediction. More recently, the complexity of previous perceptual PVC solutions is being discussed, and with JND on mode decision transform stage [24]. This is welcomed, yet it does not operate at the prediction stage which is highly complexity sensitive.

III. Methodology

This paper uses the hybrid framework as a means to produce a perceptual 'in-loop' solution, which means that a new algorithm, a series of tests and a set of thresholds must be presented [23]. The conditional framework demonstrated that

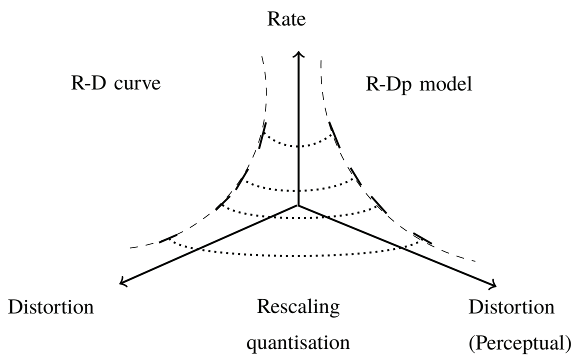


Fig. 3: Mapping R-D (STDM) curve against a R-Dp (perceptual) model to rescale quantisation

with an algorithm designed specifically for prediction the norm space is (ℓ_1) and with suitable pre-tests, both compatibility and complexity can be managed. In particular, this means that the decision on whether perceptual assessment should occur is governed by the threshold for the respective sub-block.

A. In-loop prediction algorithm design

Distortion assessment at the prediction occurs with SAD and Hadamard assessments, both of which are ℓ_1 norm space. This limits what type of perceptual choices are available for use, however, the same norm space is represented in rate-control where the hybrid framework was initially present [23]. Using the same process a new algorithm of additional pixel cost (APC) will be presented, based upon the SSIM Luma equation, as described in Equation (2),

$$SSIM_l(x,y) = \frac{2\mu_x \times \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (2)$$

where μ is mean and C_1 is a constant based upon the maximum pixel range. $C_1 = KL^2$, where K is 0.01 and L for 8 bit Luma is 255, [12]. To use SSIM Luma equation as a perceptual assessment, its scale must be extended to support the same norm-space and transformed to be tolerant of darker regions than lighter regions. This would invert the equation and match the desired norm space, however, as the rate of perceptual cost is linear, the equation should be accelerated with the perceptual model of JND using Equation (3), [11], [25].

$$JND(x,y) = \begin{cases} 17 \times (1 - (\frac{bg(x,y)}{128})^{\frac{1}{2}}) + 3 & bg(x,y) < 127 \\ \frac{3}{128} \times (bg(x,y) - 127) + 3 & bg(x,y) \geq 127 \end{cases} \quad (3)$$

where $bg(x,y)$ is the background luminance, in this case the higher of two pixel pair values. JND is a model based on the original frame only and has a non-linear response curve, having a greater tolerance for darker regions than medium to brighter regions As such, these two perceptual models of SSIM Luma and JND can be combined to form the proposed perceptual assessment. This is done by rearranging the SSIM Luma function as described above and making it in-line with common perceptual principles, labelling it as $1-SSIM_l$. Then in order to consider this as a perceptual cost, it should be scaled by the JND background luminance masking visibility threshold, producing Equation (4)

$$APC(x,y) = (2^b - 1) \times (1 - SSIM_l)^{max(JND_x, JND_y)} \quad (4)$$

where b is bit-depth. and $max(JND_x, JND_y)$ refers to using the maximum value for either the corresponding original or reconstructed pixel, which could vary depending upon the level of quantisation.

B. Proposed APC cross corner subtraction (ACCS)

From the initial paper pre-tests were used to evaluate whether perceptual assessment was required creating the perceptual pre-test technique of perceptual asymmetric side (PAS) [23]. Prediction involves evaluating different sub-block sizes, which under HEVC the largest coding unit (LCU) can be up to 64x64 and can include asymmetric sub-block sizes (width or height of 12, 24, 48). This means that perceptually assessing these varieties of sub-block sizes can represent significant proportion of processing time. Under these circumstances sub-blocks need to be evaluated faster with less perceptual accuracy. From this understanding, PAS was adapted for the sub-block corners, to form Equation (5)

$$ACCS = |(A_{TL} - B_{BR}) - (C_{TR} - D_{BL})| > ACCS_{Thresh} \quad (5)$$

where A to D denote the sub-block corners, T is top, B is bottom, R is right, L is left and $ACCS$ means APC cross calculation subtraction. ACCS, is where the respective diagonal corners are subtracted from each other based on their APC values. The equation was designed to be low complexity, with only a single absolute function being used.

The use of ACCS on the block corners is limited to four test points and was designed under an 8x8 sub-block. The choice of 8x8 is based findings that 8x8 is the optimal window value for SSIM [26]. Therefore, every whole multiple of 8x8 should undergo a further process of ACCS to ensure that perceptual cost is applied where it is suitable. This can be illustrated with Figure 4 which shows that the two stage process as outer for the sub-block corners and then as inner, for every 8x8 within.

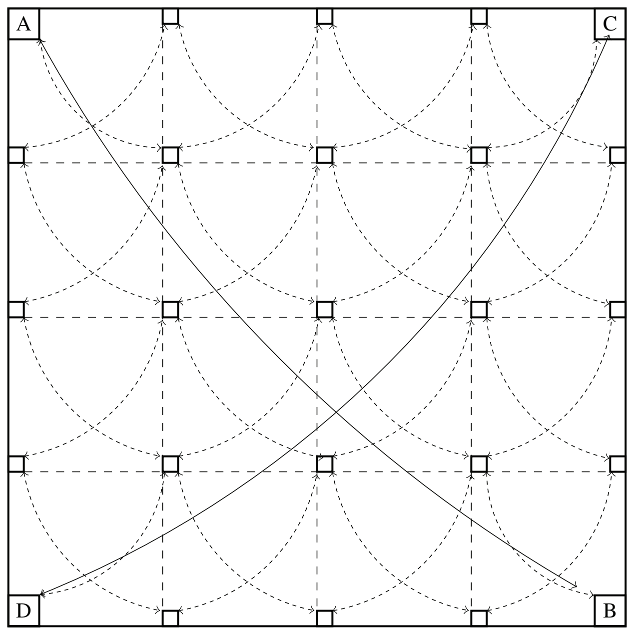


Fig. 4: APC cross calculation subtraction (ACCS) outer and inner (applicable to >8x8 sub-blocks)

C. Design for ACCS threshold in prediction

As mentioned earlier, the pre-test is threshold based, and the values must be derived to establish, however, as there are many combination of sub-blocks prediction there are multiple threshold values. Using raw Luma observations of an 8x8 sub-block the value of 128 was set, but later doubled to 256 when a high number of false triggers were observed during visual simulation. In addition, to extend this process to other sub-block sizes a non-linear scaling was formulated rounded to the nearest 8th in Equation (6).

$$ACCS_{Thresh} = (Int(log_{128}(2 \cdot blksize) \cdot 32)) \cdot 8 \quad (6)$$

where $blksize$ is the size of sub-block, width times height. Equation (6) was used to establish the thresholds for various combinations of sub-block sizes in HEVC as shown in Table I along with percentage equivalence to the initial 8x8. Equation (6) and Table I illustrate that a non-linear threshold discourages smaller sub-blocks as it has a relative lower threshold value compared to larger sub-blocks. Overall, this means the likelihood of applying additional perceptual cost is higher for smaller sub-blocks than for larger sub-blocks, which encourages the use of large sub-blocks.

IV. Results and discussion

The testing was conducted on an Intel Core i5-6600K system with 32GB of RAM using HEVC HM version 16.9. Eight video sequences were encoded, half under low delay P (LDP) and the other half under random access (RA) profile. Only the first half of each video sequence was encoded, 5 seconds worth as this has been deemed sufficient for testing [27]. Finally, each video sequence were encoded at 1, 2, 4, 8 and 16 Mbps. The encoder and decoder log files provided the Y-PSNR, timing and bit usage by block size results. While the SSIM scores were gathered using Video Quality Measurement Tool (VQMT) [28].

The results are shown in tables II and III indicating that the average PSNR losses are no more than 0.33 dB and 0.21 dB for LDP and RA respectively across all bit-rates. While the perceptual losses of 1-SSIM for both profiles are very minor < 0.003. Similarly, in terms of timing the increases for LDP and RA are $< +3\%$ and $< +6\%$ respectively including the upper standard deviation value. The changes to bit usage occurs in both LDP and RA, with significant reductions in signalling bits and increases in medium to large size blocks as shown in figure 5. The results shown are averaged across the five bit rates, however, to understand the range of differences the standard deviation (std dev) for bit changes are also shown. The std dev graphs show that the significant reduction in signalling is generally across all bit rates, with the exception of Riverbed which has the highest signalling reduction, which is probably due to the video content. In the video sequence Riverbed, the content is highly active and localised, making it very difficult to encode. Overall, the changes in bit usage is more dynamic in LDP than in RA, particularly for smaller block sizes.

Examining the decoded frames highlight which partitions are encoded with residual information. Examples comparing original and proposed for each profile are shown in Figures 6 and 7 for 1 and 16 Mbps respectively. The original frame images has partition information from the encoded bitstream is overlaid with colours of the rainbow representing each block sizes respectively, red (4x4), orange (8x8), yellow (16x16), green (32x32) and blue (64x64). In both Figures 6 and 7 white boxes have been superimposed to highlight particular features. For Figure 6 where there is low bit-rate, the proposed encoder is able allocate larger sub-blocks for homogeneous regions and smaller sub-block sizes closer to the video content. This is shown in Kimono where larger blocks are used on the background foliage, while in ParkScene smaller sub-blocks are allocated for the cyclists legs which are moving. While in Figure 7 where bandwidth is higher, the block redistribution is more dynamic. The proposed encoder identifies homogeneous regions where larger blocks or even skip blocks may be applied. This allows more smaller partitioning or sub-block sizes to be used where perceptually significant texture is present.

In the initial hybrid framework, only two video sources were used, one for RA and another in LDP [23]. This means that comparing these results with those presented in this paper is

Blk Hght	Width (Threshold) 4	6	8	12	16	24	32	48	64
4	176	200	216	240	256	272	288	312	328
6	200	224	240	256	272	296	312	328	344
8	216	240	256	272	288	312	328	344	360
12	240	256	272	296	312	328	344	368	384
16	256	272	288	312	328	344	360	384	400
24	272	296	312	328	344	368	384	408	416
32	288	312	328	344	360	384	400	416	432
48	312	328	344	368	384	408	416	440	456
64	328	344	360	384	400	416	432	456	472
Blk Hght	**Width (as %) 4**	**6**	**8**	**12**	**16**	**24**	**32**	**48**	**64**
4	0.69	0.78	0.84	0.94	1.00	1.06	1.13	1.22	1.28
6	0.78	0.88	0.94	1.00	1.06	1.16	1.22	1.28	1.34
8	0.84	0.94	1.00	1.06	1.13	1.22	1.28	1.34	1.41
12	0.94	1.00	1.06	1.16	1.22	1.28	1.34	1.44	1.50
16	1.00	1.06	1.13	1.22	1.28	1.34	1.41	1.50	1.56
24	1.06	1.16	1.22	1.28	1.34	1.44	1.50	1.59	1.63
32	1.13	1.22	1.28	1.34	1.41	1.50	1.56	1.63	1.69
48	1.22	1.28	1.34	1.44	1.50	1.59	1.63	1.72	1.78
64	1.28	1.34	1.41	1.50	1.56	1.63	1.69	1.78	1.84

TABLE I: Non-linear threshold and percentage equivalent (with reference to 8x8) for APC cross corner subtraction (ACCS).

Mbps	ΔY-PSNR	Δ1-SSIM	ΔTime	ΔTime Std Dev
1	-0.330	-0.0021	2.55%	0.74%
2	-0.221	-0.0013	2.61%	0.75%
4	-0.123	-0.0040	2.12%	0.89%
8	-0.069	-0.0005	1.79%	0.54%
16	-0.049	-0.0004	1.80%	0.28%

TABLE II: Average changes by bit-rate for low delay p

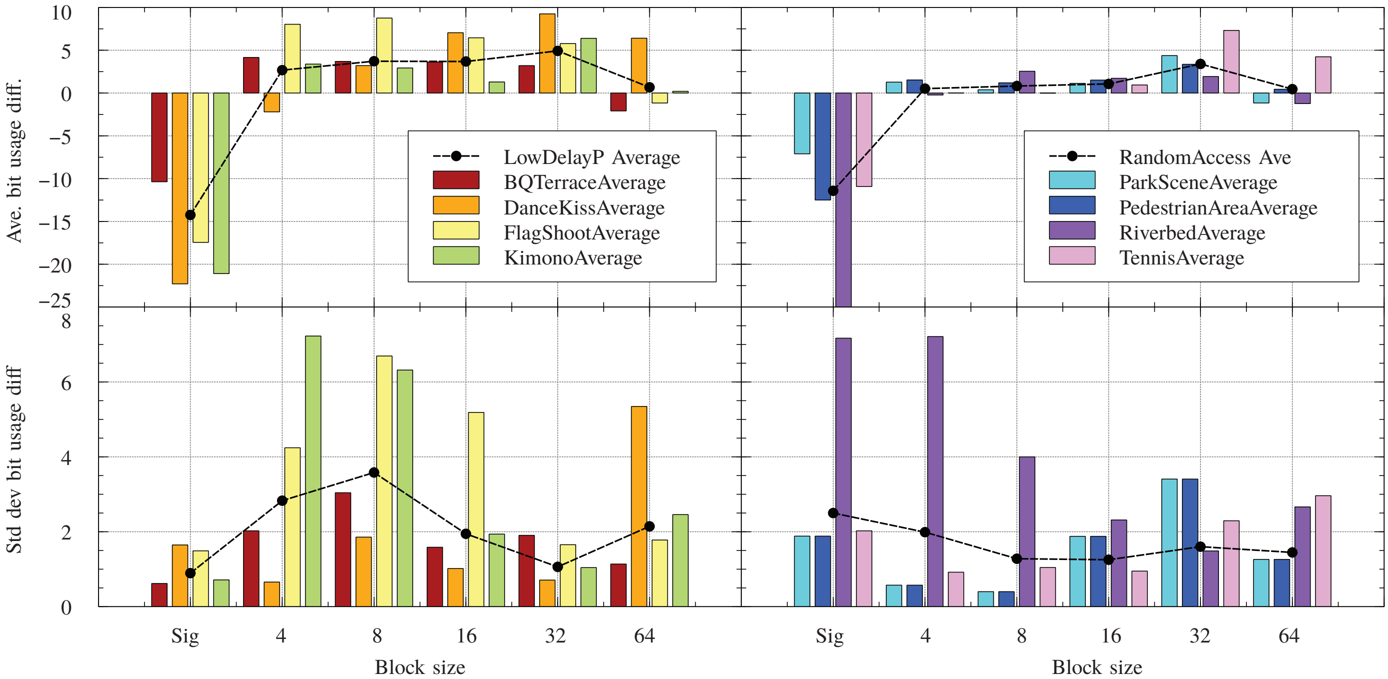
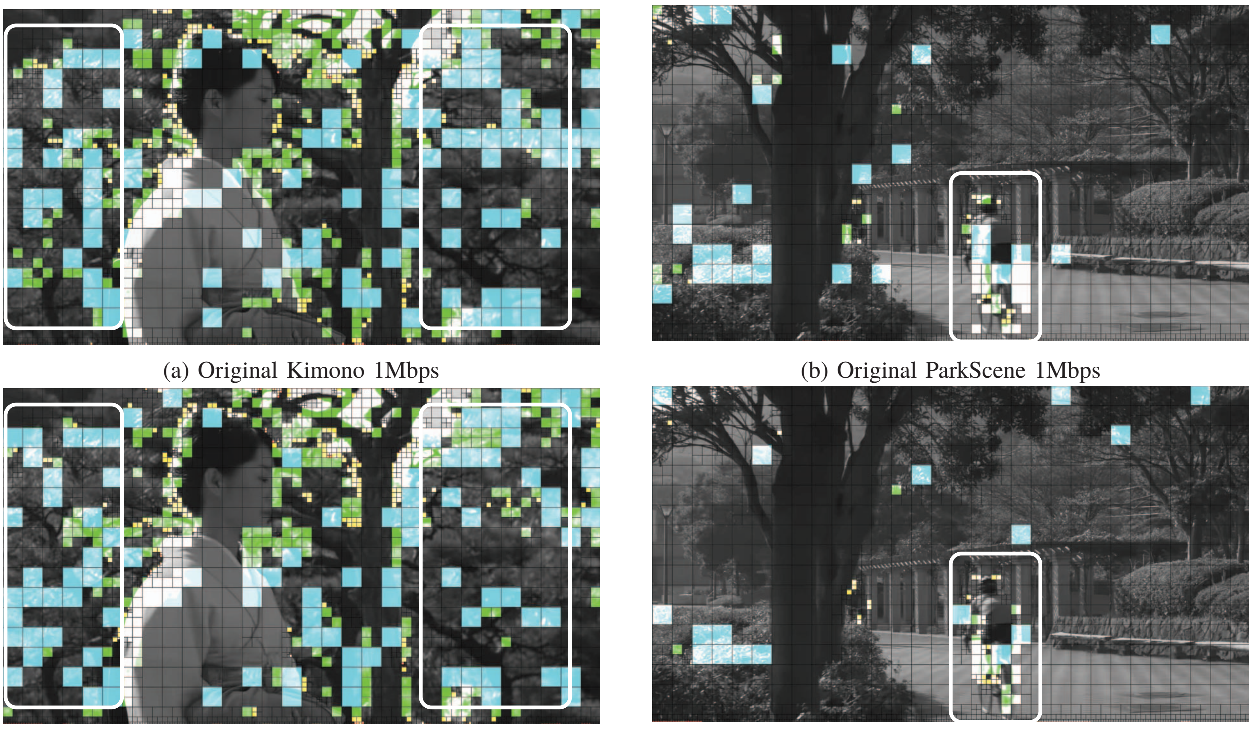


Fig. 5: Percent bit diff for low delay P and random access average (ave) and standard deviation (std dev) across bit-rates 1Mbps to 16Mbps by block size



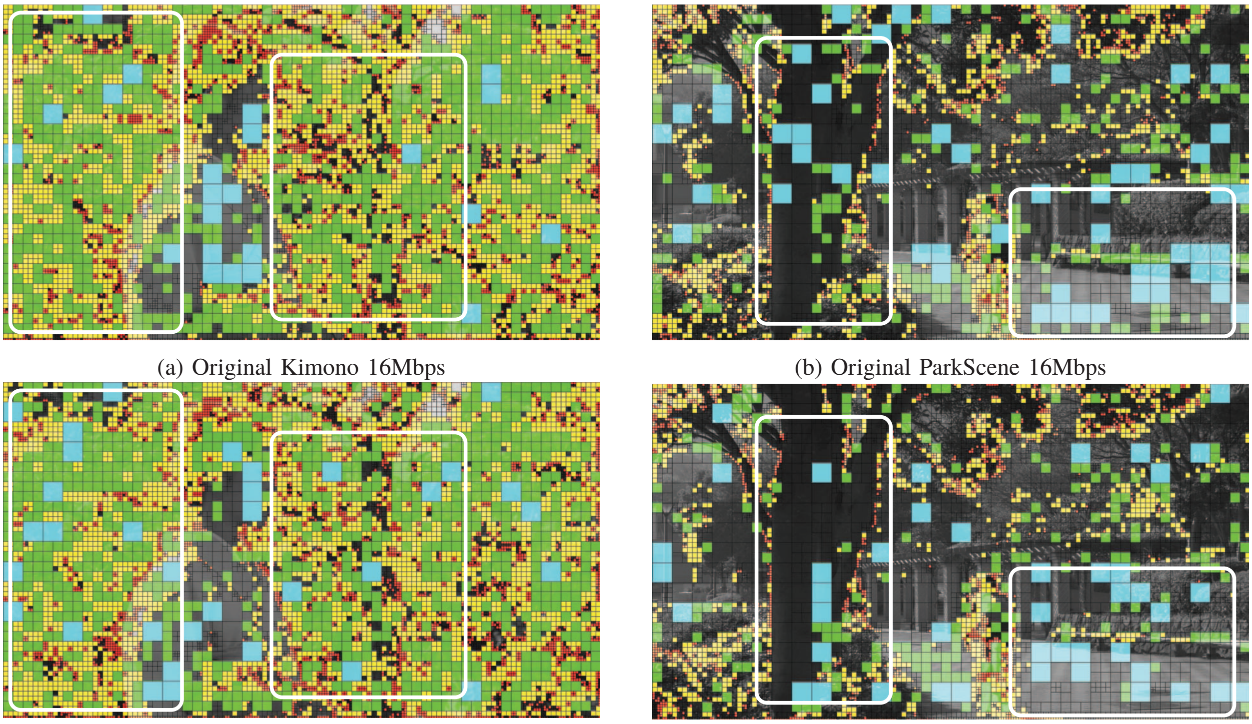
(a) Original Kimono 1Mbps

(b) Original ParkScene 1Mbps

(c) Proposed Kimono 1Mbps

(d) Proposed ParkScene 1Mbps

Fig. 6: Frame 77 for Kimono (low delay P) and ParkScene (random access) encoded sequences at 1 Mbps with highlight partition sizes



(a) Original Kimono 16Mbps

(b) Original ParkScene 16Mbps

(c) Proposed Kimono 16Mbps

(d) Proposed ParkScene 16Mbps

Fig. 7: Frame 77 for Kimono (low delay P) and ParkScene (random access) encoded sequences at 16 Mbps with highlight partition sizes

Mbps	ΔY-PSNR	Δ1-SSIM	ΔTime	ΔTime Std Dev
1	-0.205	-0.0026	4.95%	0.60%
2	-0.138	-0.0015	4.87%	0.92%
4	-0.086	-0.0008	4.64%	1.35%
8	-0.055	-0.0006	4.41%	1.17%
16	-0.035	0.0009	4.19%	1.07%

TABLE III: Average changes by bit-rate for random access

difficult. However, it is possible to say that by extending the hybrid framework to support prediction the timing increases from $< +4\%$ to $< +6\%$. In terms of PSNR and 1-SSIM losses, for LDP they are the same, while for RA they have been are reduced. This is especially true for 1-SSIM where at the lowest and medium bit-rate of 1 and 4 Mbps, the 1-SSIM losses are 1/4 of that observed previously, making the overall 1-SSIM losses virtually zero. Comparing the proposed solution with other 'out-of-loop' solutions, they offer bandwidth reductions through perceptual quantisation [15]. However, their high peak complexity makes them unsuitable for low powered applications. Where peak complexity is being addressed, the timing is still three times that shown here in this paper [24]. Therefore, the 'in-loop' solution presented here is complexity competitive to existing PVC solutions.

V. Conclusion and future work

Video encoding is increasing occurring on low powered devices, where the available processing is limited such as mobile phones and tablets. This presents a constrained complexity envelope for video coding to operate within. The proposed in-loop prediction PVC offers a low complexity means to encode video, suitable for the low powered devices. Having a in-loop PVC has shown it would allocate larger block sizes to perceptually homogeneous regions, thus allowing more dynamic partitioning on perceptually significant regions. In turn, this work can be extended from bit redistribution to bandwidth reduction with perceptual quantisation, making it more attractive solution to adopt.

References

[1] Cisco. *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2014–2019*. Cisco. Feb. 2015. URL: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html.

[2] World Bank. *World Development Report 2016: Digital Dividends*. Tech. rep. World Bank, Jan. 2016.

[3] Hugh Everett III. "Generalised Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources". In: *Operations Research* 11.3 (1963), pp. 399–417. ISSN: 0030364X.

[4] A. Ortega and K. Ramchandran. "Rate-distortion methods for image and video compression". In: *IEEE Signal Processing Magazine* 15.6 (1998), pp. 23–50. ISSN: 1053-5888.

[5] H.R Wu and K.R Rao, eds. *Digital Video Image Quality and Perceptual Coding*. CRC Press, 2005.

[6] Hong Ren Wu, Weisi Lin, and King Ngi Ngan. “Rate-perceptual-distortion optimization (RpDO) based picture coding — Issues and challenges”. In: *Digital Signal Processing (DSP), 2014 19th International Conference on*. 2014, pp. 777–782.

[7] P. Le Callet and E. Niebur. “Visual Attention and Applications in Multimedia Technologies”. In: *Proceedings of the IEEE* 101.9 (2013), pp. 2058–2067.

[8] Mathieu Carnec, Patrick Le Callet, and Dominique Barba. “Objective Quality Assessment of Color Images based on a Generic Perceptual Reduced Reference”. In: *Signal Processing: Image Communication* 23.4 (2008), pp. 239–256. ISSN: 0923-5965.

[9] J. Mannos and D.J. Sakrison. “The effects of a visual fidelity criterion of the encoding of images”. In: *Information Theory, IEEE Transactions on* 20.4 (July 1974), pp. 525–536. ISSN: 0018-9448.

[10] J. Yogeshwar and R. J. Mammone. “A New Perceptual Model for Video Sequence Encoding”. In: *Proc. Conf. th Int Pattern Recognition*. 1990, pp. 188–193.

[11] Chun-Hsien Chou and Yun-Chin Li. “A Perceptually Tuned Subband Image Coder Based on the Measure of Just-Noticeable-Distortion Profile”. In: *IEEE Transactions on Circuits and Systems for Video Technology* 5.6 (1995), pp. 467–476.

[12] Zhou Wang, A. C. Bovik, H. R. Sheikh, et al. “Image Quality Assessment: From Error Visibility to Structural Similarity”. In: *IEEE Transactions on Image Processing* 13.4 (2004), pp. 600–612.

[13] Weisi Lin and C.-C. Jay Kuo. “Perceptual Visual Quality Metrics: A Survey”. In: *Journal of Visual Communication and Image Representation* 22.4 (2011), pp. 297–312. ISSN: 1047-3203.

[14] T. Richter. “SSIM as Global Quality Metric: A Differential Geometry View”. In: *Proc. Third Int Quality of Multimedia Experience (QoMEX) Workshop*. 2011, pp. 189–194.

[15] Yi-Hsin Huang, Tao-Sheng Ou, Po-Yen Su, et al. “Perceptual Rate-Distortion Optimization Using Structural Similarity Index as Quality Metric”. In: *IEEE Transactions on Circuits and Systems for Video Technology* 20.11 (Nov. 2010), pp. 1614–1624. ISSN: 1051-8215.

[16] Wei Dai, O.C. Au, Wenjing Zhu, et al. “SSIM-based rate-distortion optimization in H.264”. In: *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. 2014, pp. 7343–7347.

[17] Damon M Chandler. “Seven Challenges in Image Quality Assessment: Past, Present, and Future Research”. In: *ISRN Signal Processing* 2013 (Nov. 2013).

[18] Po-Yen Su, Chieh-Kai Kao, Tsung-Yau Huang, et al. “Adopting Perceptual Quality Metrics in Video Encoders: Progress and Critiques”. In: *Multimedia and Expo Workshops (ICMEW), 2012 IEEE International Conference on*. 2012, pp. 73–78.

[19] F. Zhang and D. Bull. “A Perception-based Hybrid Model for Video Quality Assessment”. In: *IEEE Transactions on Circuits and Systems for Video Technology* PP.99 (2015), p. 1. ISSN: 1051-8215.

[20] A. Horé and and D. Ziou. “Image Quality Metrics: PSNR vs. SSIM”. In: *20th International Conference on Pattern Recognition (ICPR), 2010*. Aug. 2010, pp. 2366–2369.

[21] Y.G. Joshi, P. Shah, J. Loo, et al. “Review of Standard Traditional Distortion Metrics and a need for Perceptual Distortion Metric at a (Sub) Macroblock Level”. In: *Broadband Multimedia Systems and Broadcasting (BMSB), 2013 IEEE International Symposium on*. 2013, pp. 1–6.

[22] Y.G. Joshi, J. Loo, P. Shah, et al. “A novel low complexity Local Hybrid Pseudo-SSIM-SATD distortion metric towards perceptual rate control”. In: *Broadband Multimedia Systems and Broadcasting (BMSB), 2013 IEEE International Symposium on*. 2013, pp. 1–6.

[23] Y.G. Joshi, J. Loo, P. Shah, et al. “Low complexity sub-block perceptual distortion assessment for mode decision and rate-control”. In: *Broadband Multimedia Systems and Broadcasting (BMSB), 2015 IEEE International Symposium on*. 2015, pp. 1–9.

[24] Jaeil Kim, Sung-Ho Bae, and Munchurl Kim. “An HEVC-Compliant Perceptual Video Coding Scheme Based on JND Models for Variable Block-Sized Transform Kernels”. In: *IEEE Transactions on Circuits and Systems for Video Technology* 25.11 (2015), pp. 1786–1800.

[25] Tung-Hsing Wu, Guan-Lin Wu, and Shao-Yi Chien. “Bio-inspired Perceptual Video Encoding based on H.264/AVC”. In: *Proc. IEEE Int. Symp. Circuits and Systems ISCAS 2009*. 2009, pp. 2826–2829.

[26] A.C. Brooks, XiaoNan Zhao, and T.N. Pappas. “Structural Similarity Quality Metrics in a Coding Context: Exploring the Space of Realistic Distortions”. In: *IEEE Transactions on Image Processing* 17.8 (Aug. 2008), pp. 1261–1273. ISSN: 1057-7149.

[27] F. Mercer Moss, K. Wang, F. Zhang, et al. “On the Optimal Presentation Duration for Subjective Video Quality Assessment”. In: *IEEE Transactions on Circuits and Systems for Video Technology* (to be published). Early Access.

[28] VQMT: Tool. *VQMT: Video Quality Measurement Tool | MMSPG*. 2016. URL: http://mmspg.epfl.ch/vqmt.