Commun.Fac.Sci.Univ.Ank.Series A2-A3
Volume 62, Number 1, Pages 1-13 (2020)
DOI: 10.33769/aupse.657684
ISSN 1303-6009 E-ISSN 2618-6462
http://communications.science.ankara.edu.tr/index.php?series=A2A3

AN ARTIFACT REDUCTION METHOD FOR BLOCK-BASED VIDEO CODING

HAKKI ALPARSLAN ILGIN and AHMET AKBULUT

ABSTRACT. Most of the image/video coding standards are based on discrete cosine transform which is a block-based coding scheme with a disadvantage of blocking effect at low bit rates. In this paper, we propose a hybrid method, which consists of downsampling/upsampling, significance map coding and local cosine transform, to reduce the disadvantage of lossy compression and is also compatible with current standards. Experimental results show that performance of the proposed method is better than the conventional approaches.

1. INTRODUCTION

Image and video coding applications require high compression ratios for low bit rate applications and poor channel conditions. However, lossless compression is not sufficient for such cases requiring lossy compression, which causes defects in the reconstructed signal at low bit rates. Image/video compression algorithms and standards mostly consist of both lossy and lossless compression techniques. Furthermore, standards based on discrete cosine transform (DCT) result in blocking artifacts when used with lossy compression schemes at low bit rates. DCT is a block-based transform that converts image into frequency domain coefficients. Lossy compression algorithms primarily neglect most of the high frequency AC coefficients of DCT blocks. Since most of the energy is compacted into DC and low frequency AC coefficients, losing some of the high frequency AC coefficients does not affect the quality of the reconstructed signal significantly. However, as more high frequency AC coefficients are flattened by quantization, reconstructed image gets blurrier. Losing more AC coefficients, especially low frequency ones with higher energy to get more compression causes severe image degradation, which appears as blocking artifact. In the literature, there are many methods to compensate

Received by the editors: December 10, 2019; Accepted: January 07, 2020.
Key word and phrases: Local cosine transform, significance map coding, image compression, video coding.

the artifacts. Some of the most frequently used methods are lapped orthogonal transform (LOT) [1], post-filtering [2], linear filtering [3], and subband/wavelet coding [4]. However, LOT is not fully compatible with the standards that include DCT. Post- and linear filtering may get blurry images at low bit rates. Wavelet or subband-based codecs are also incompatible with DCT-based standards. Deblocking filter, which is used in standards such as H.263 Annex J, H.264 [5] and HEVC [6], has its own computational complexity. Recent works with deep learning [7] have additional complexity at training process. In this paper, we use a combined method that consists of downsampling/upsampling, significance map coding (SMC) and local cosine transform (LCT) [8]. Coding is kept in the DCT domain to prevent extra operations and for compatibility with the DCT-based standards. Reducing blocking effects is discussed in the next section. Significance map coding and downsampling/upsampling scheme are applied in section 3. In the last section, conclusion and discussion are given.

2. Blocking Effect Reduction

Blocking effect is the most significant artifact of lossy DCT-based coding methods at low bit rates. Several conditions and requirements lead signals to be encoded at low bit rates. However, lossless compression techniques do not provide very high compression ratios. Therefore, lossy compression is used together with lossless techniques. The most common lossy compression method is quantization, which is used in almost all standards. When used with DCT, it causes losses in DCT coefficients. Increasing compression ratio requires more AC coefficients of DCT blocks to be neglected, namely zero. Therefore, blocking effects appear at low bit rate coding and become severe with decreasing bit rate. Being one of the methods to decrease blocking effect, LCT implements folding/unfolding and DCT operations [8]. At the encoder side of LCT, DCT is applied after folding operation. Accordingly, inverse DCT and then unfolding are applied at the decoder. Because of the folding operation, discontinuities across the neighboring block boundaries are smoothed. A bell function defined on the basis of the function for folding operation affects the reconstructed image quality or the compression ratio [8]. After folding operation is performed on all blocks, DCT of each folded block is computed.

Coding with both conventional DCT and LCT are compared in terms of objective metrics. Two of the most frequently used metrics in image and video coding applications are Peak Signal-to-Noise Ratio (PSNR) and structural similarity index (SSIM) [9]. There are also specific MATLAB-based applications for image processing and image quality assessment as in [10]. Beside the objective metrics, subjective methods based on human perception are also used [11]. In this study, we use both PSNR and SSIM objective metrics to evaluate the quality of the reconstructed images when compared to the original ones. PSNR is given by

$$\mathrm{PSNR}(I,R) = 10log_2\left(\frac{\max(I)^2}{\mathrm{MSE}(I,R)}\right) \tag{1}$$

where

$$\mathrm{MSE}(I,R) = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}[I(i,j) - R(i,j)]^2 \tag{2}$$

is mean-square error between the original and reconstructed images, I and R, respectively, m and n are vertical and horizontal number of pixels and (i,j) is the coordinate of a pixel in original and reference images. The value of max(I) is 255 for 8-bit images. SSIM is correlated with the quality perception of the human visual system with the highest value of 1 and given as follows [12]:

$$SSIM(I,R) = \frac{(2\mu_I\mu_R+c_1)(2\sigma_{IR}+c_2)}{(\mu_I^2+\mu_R^2+c_1)(\sigma_I^2+\sigma_R^2+c_2)} \tag{3}$$

where μ_I and μ_R are the mean value and σ_I^2 and σ_R^2 are the variance of the original and reconstructed images, respectively. σ_{IR}^2 is the covariance of I and R. Variables c_1 and c_2 are used to stabilize the division with weak denominator and given as $c_1=(0.01\max(I))^2$ and $c_2=(0.03\max(I))^2$. If there is no correlation between the original and reconstructed images, the value of SSIM is 0. SSIM is higher for the images that are similar. SSIM is 1 when two images are the same. Comparison between codecs with DCT and LCT are given in Tab. 1.

Table 1. DCT and LCT comparison for *soccer* frame #1 and *bus* frame #1

QP	***Soccer***				***Bus***			
	PSNR (dB)		**SSIM**		**PSNR (dB)**		**SSIM**	
	DCT	**LCT**	**DCT**	**LCT**	**DCT**	**LCT**	**DCT**	**LCT**
12	30.31	31.28	0.846	0.871	29.87	31.06	0.886	0.907
15	28.93	29.71	0.808	0.834	28.16	29.31	0.853	0.879
20	27.95	28.49	0.770	0.797	26.79	27.85	0.819	0.848
25	27.02	27.37	0.733	0.759	25.50	26.45	0.784	0.815
30	26.61	26.91	0.704	0.736	24.79	25.70	0.758	0.791
35	26.08	26.33	0.678	0.711	24.00	24.78	0.730	0.766
40	25.76	26.04	0.658	0.693	23.57	24.36	0.709	0.747
45	25.41	25.74	0.641	0.676	23.02	23.76	0.686	0.726
50	25.22	25.56	0.630	0.663	22.70	23.40	0.669	0.710

FIGURE 1. Reconstructed *bus* frame #1 from coding scheme with LCT (top) and DCT (bottom) with QP=50.

In Tab. 1, first frames of 8-bit grayscale *soccer* and *bus* video sequences with 288x352 pixel resolution are encoded as intraframe with quantization parameters (QP) between 12 and 50, for both DCT and LCT coding schemes with 8x8 block size. LCT performs better than DCT for all QP values as seen in Tab. 1. As the bit rate decreases with increasing QP, artifacts become more visible. Comparison between LCT and DCT encoded *bus* frame #1 with QP=50 is given in Fig. 2. Blocking effect in the reconstructed image coded with DCT is severe as seen in Fig.

1. Also, DCT encoded reconstructed image has more blurry effects than the LCT encoded image.

3. Improving Quality By Significance Map Coding And Downsampling/Upsampling

Image quality degradation by lossy compression techniques for block-based algorithms appears block-wise at low bit rates. In the previous section, we used an effective method to decrease the blocking artifacts caused by high compression rates. In this section, we further improve image quality by applying LCT with significance map coding and downsampling in the coder and the corresponding upsampling in the decoder. Image quality is deteriorated more with higher compression rates. However, by using the hybrid approach, we aim to decrease the disadvantage of very low bit rate coding.

We first implement SMC together with LCT to improve image quality. Embedded image zerotree coding of wavelet coefficients is an efficient example of SMC [13]. This method is applied for DCT beside the wavelet transform. An improved method is set partitioning in hierarchical trees (SPIHT), which has a better compression performance [14], and still being used widespread recently [15]. SPHIT is an iterative algorithm with threshold halved at each iteration and encodes transform coefficients in decreasing order with binary output. In the first part of the algorithm, coefficients are compared with a threshold, which is the lower value of an uncertainty interval, to output bits corresponding to significance of coefficients or a hierarchical structure. In the second part, coefficients previously found to be significant are given one-bit precision. Since SPIHT can be applied for DCT, we implement the algorithm with LCT, which has DCT partly.

In the next experiments, we compare the improvement by LCT with SPIHT significance map coding (LCT-SMC) to DCT-SMC for video frames encoded intraframe. Before applying significance map coding, DCT coefficients are rearranged in subband structure. We use the most common grayscale CIF videos with 288x352 pixel resolution. Improvement of LCT-SMC is clearly seen for *salesman* and *city* frame #1 in terms of PSNR and SSIM in Fig. 2. At lower bit rates, advantage of LCT-SMC over DCT-SMC is apparent as supported by PSNR and SSIM. When reconstructed image is visually almost inseparable from the original one, increase of SSIM is much slower than PSNR. LCT is very efficient for intraframe at low bit rates when used with SMC, since SMC efficiently encodes coefficients with higher energy initially, so that the degradation caused by lower energy coefficients, which are mostly related to high frequency details in the frame, does not distort image conspicuously. Besides SMC is applicable to DCT, since it is not block based algorithm, it does not have the disadvantage of blocking effect at low bit rates. Using SMC in conjunction with LCT achieves better results for reconstructed images.

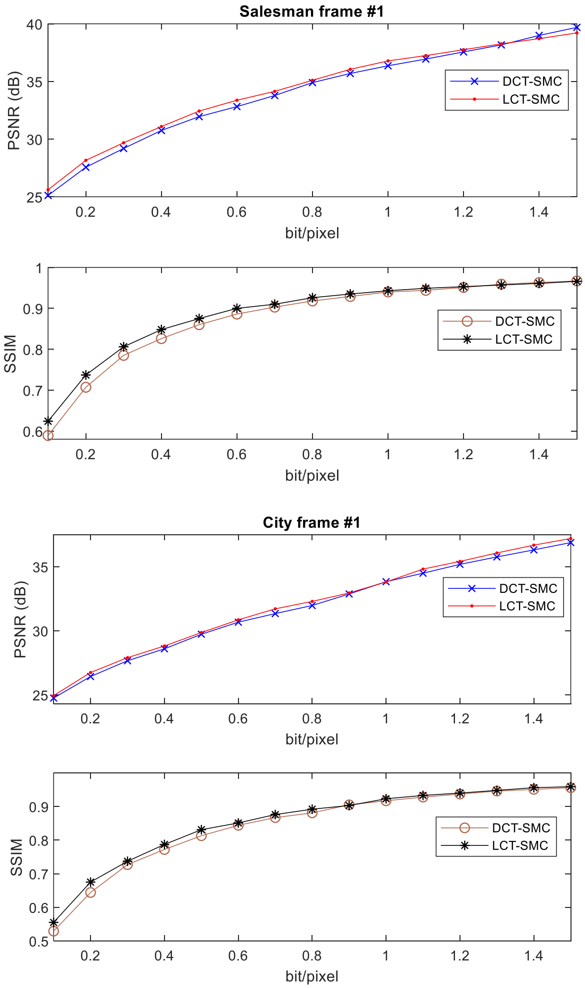


FIGURE 2. Comparison for intraframe coding at different bit rates

We also compare interframe coding performance of LCT-SMC and DCT-SMC. PSNR and SSIM per frame of both methods are given for *soccer*, *city* and *coastguard* sequences for first 30 frames in Fig. 3.

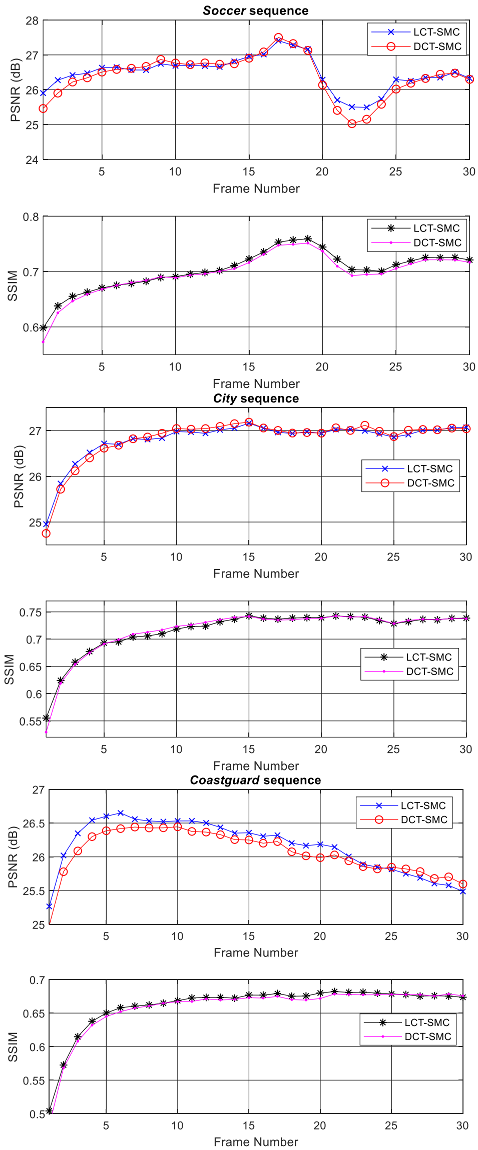


FIGURE 3. Interframe coding with LCT-SMC and DCT-SMC

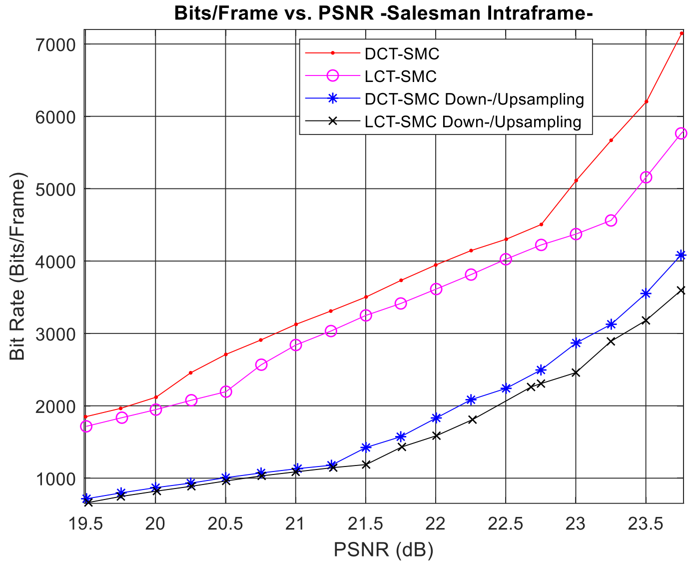
Total bits for 30 frames for each sequence is 46881. Intraframe of each sequence is encoded with 10138 bits (or 0.1 bpp) and each of the 29 interframe is encoded with just 1267 bits (or 0.0125 bpp).

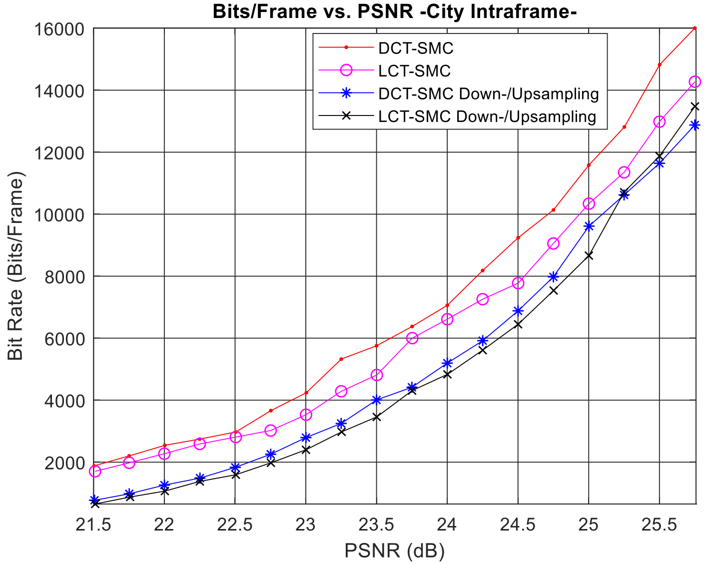
Interframe performance of LCT is not as efficient as its intraframe performance, since SMC and motion estimation/compensation help to improve the quality of the interframes [16]. However, performance of intraframe coding helps to improve coding efficiency of consecutive interframes as shown in Fig. 3, since first frame of each sequence gives higher PSNR with LCT-SMC than DCT-SMC, consequently the following interframes give high PSNR values. Since SMC is efficient for any cases, i.e. for DCT or LCT, its combination with either of the two achieves close results as seen in Fig. 3. Mean of PSNR and SSIM for 29 interframe coding in Fig. 3 are given in Tab. 2.

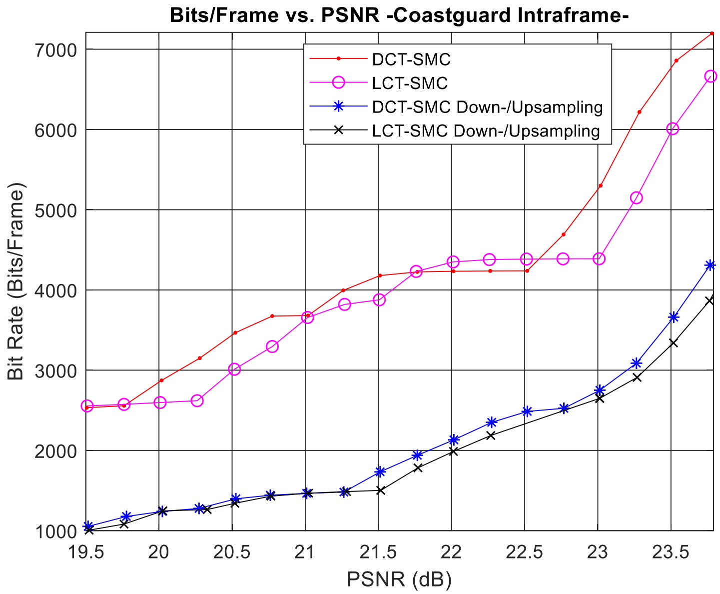
Table 2. Average PSNR and SSIM for Interframe Coding

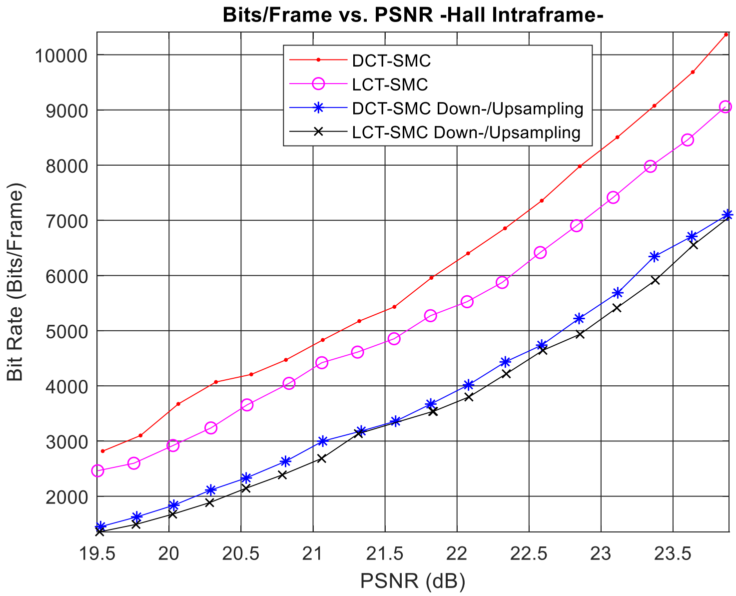
Sequence	**PSNR (dB)**		**SSIM**	
	DCT-SMC	**LCT-SMC**	**DCT-SMC**	**LCT-SMC**
Soccer	26.43	26.50	0.701	0.706
City	26.89	26.87	0.722	0.721
Coastguard	26.10	26.19	0.663	0.666

Loss in quality first appears at the details which corresponds to high frequency AC coefficients of DCT as bit rate decreases. Increasing compression ratio further makes image blurry, which is the result of degradation of more AC coefficients with lower frequency. When the compression requires very low bit rate, AC coefficients with the lowest frequencies and DC coefficient are also deteriorated. Accordingly, the resulting image will have blocking effects. Therefore, before losing information by lossy compression, we apply downsampling for each block in the frequency domain whose efficiency was shown in [17]. Another advantage of keeping downsampling in the frequency domain is that computational complexity caused by inverse DCT, decimation in the spatial domain and forward DCT is avoided. Thus, image quality is enhanced with only limited additions of computations, since proposed downsampling and upsampling are operated fully in the DCT domain. When downsampling/upsampling by 2 in the DCT domain with SMC is implemented, we obtain better results than the case of DCT-SMC. We further improve the results by using downsampling/upsampling with LCT-SMC. Results are shown in Fig. 4 for five video sequences.


Bits/Frame vs. PSNR -Salesman Intraframe-
DCT-SMC
LCT-SMC
DCT-SMC Down-/Upsampling
LCT-SMC Down-/Upsampling
Bit Rate (Bits/Frame)
PSNR (dB)
1000
2000
3000
4000
5000
6000
7000
19.5
20
20.5
21
21.5
22
22.5
23
23.5


Bits/Frame vs. PSNR -City Intraframe-
DCT-SMC
LCT-SMC
DCT-SMC Down-/Upsampling
LCT-SMC Down-/Upsampling
Bit Rate (Bits/Frame)
PSNR (dB)
2000
4000
6000
8000
10000
12000
14000
16000
21.5
22
22.5
23
23.5
24
24.5
25
25.5


Bits/Frame vs. PSNR -Coastguard Intraframe-
DCT-SMC
LCT-SMC
DCT-SMC Down-/Upsampling
LCT-SMC Down-/Upsampling
Bit Rate (Bits/Frame)
PSNR (dB)
1000
2000
3000
4000
5000
6000
7000
19.5
20
20.5
21
21.5
22
22.5
23
23.5


Bits/Frame vs. PSNR -Hall Intraframe-
DCT-SMC
LCT-SMC
DCT-SMC Down-/Upsampling
LCT-SMC Down-/Upsampling
Bit Rate (Bits/Frame)
PSNR (dB)
2000
3000
4000
5000
6000
7000
8000
9000
10000
19.5
20
20.5
21
21.5
22
22.5
23
23.5

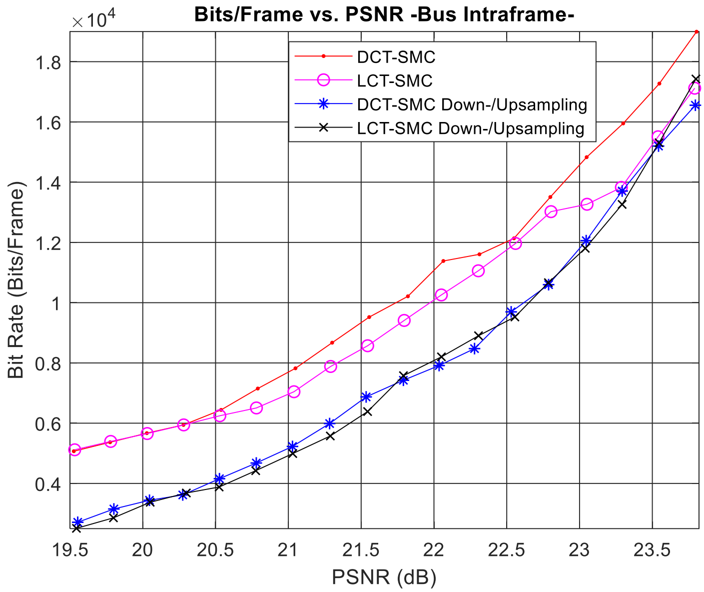


FIGURE 4. Quality improvement by downsampling/upsampling with LCT-SCM over other methods

4. Conclusion

Lossy compression of videos is required for low bandwidth and limited storage area conditions when lossless compression is not sufficient. In such cases, reconstructed video frames may have artifacts such as blocking effects and blurred scenes and objects. In order to surpass the artifacts at low bit rates, we used a hybrid method which includes LCT, SMC and downsampling/upsampling in the compressed domain. We showed that the LCT is efficient for deblocking of intraframes, while, SMC applied as SPIHT is very efficient for both intraframe and interframes. Furthermore, even though LCT has being used widely after its introduction, we utilized LCT in conjunction with SMC to decrease artifacts more efficiently. LCT-SMC is also compatible with most of the DCT-based standards still being used, since it implements DCT. Using downsampling at the coder and corresponding upsampling at the decoder increases coding performance because of energy conservation before losing DCT coefficients with high energy because of lossy compression. We implemented both downsampling and upsampling in the DCT domain with the factor of 2. It is also possible to use other factors to increase the coding performance at lower bit rates.

References

[1] Malvar, H. S., The LOT: a link between block transform coding and multirate fillter banks, *IEEE Intern. Symp. Circuits and Systems,* (1988), 835-838.

[2] Tzou, K.-H., Post-filtering of transform-coded images, *Applications of Digital Image Proc.,* vol. 974, (1988), 121-126.

[3] Awil, C. and Nguyeri-Trsng, T., Linear filtering for reducing blocking effects in orthogonal transform image coding, Journal of Electronic Imaging 1(2) , (April 1992), 183-191.

[4] Wu, Po-Cheng, Chen, Liang-Gee and Lai, Yeong-Kang, A block shifting method for reduction of blocking effects in subband/wavelet image coding, IEEE Transactions on Consumer Electronics, vol. 44, no. 1, (Feb. 1998), 170-177. doi: 10.1109/30.663744

[5] Joint Video Team of ITU-T and ISO/IEC JTC 1, Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC), Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT-G050r1, May 2003.

[6] Sullivan, G. J., Ohm, J. R. and Han W. J., Overview of the high efficiency video coding (HEVC) standard, *IEEE Trans. Circuits Syst. Video Technol.,* vol. 22 no. 12 (Dec. 2012) pp. 1649-1668.

[7] Goodfellow, Ian, Bengio, Yoshua and Courville, Aaron, Deep Learning, MIT Press, 2016.

[8] Aharoni, G., Averbuch, A., Coifman, R. and Israeli, M., Local Cosine Transform - A Method for the Reduction of the Blocking Effect in JPEG, *Journal of Mathematical Imaging and Vision,* (1993), pp. 7-38.

[9] Horé, A. and Ziou, D., Image Quality Metrics: PSNR vs. SSIM, *2010 20th International Conference on Pattern Recognition*, Istanbul, (2010), 2366-2369. doi: 10.1109/ICPR.2010.579

[10] Krasula, L., Klima, M., Rogard, E. and Jeanblanc, E., MATLAB based Applications for Image Processing and Image Quality Assessment – Part I: Software Description, *Radioengineering,* Vol. 20, No. 1, (Dec. 2011) 1009-1015.

[11] Richter, T. and Larabi, C., Subjective and Objective Assesment of Visual Image Quality Metrics and Still Image Codecs, *Data Compression Conference (dcc 2008)*, Snowbird, UT, (2008), 541-541. doi: 10.1109/DCC.2008.44

[12] Wang, Zhou, Bovik, A. C., Sheikh, H. R. and Simoncelli, E. P., Image quality assessment: from error visibility to structural similarity, *IEEE Transactions on Image Processing*, vol. 13, no. 4, (April 2004), pp. 600-612. doi: 10.1109/TIP.2003. 819861

[13] Shapiro, J. M., Embedded image coding using zerotrees of wavelet coefficients, *IEEE Transactions on Signal Processing*, vol. 41, no. 12, (Dec. 1993), 3445-3462. doi: 10.1109/78.258085

[14] Said, A. and Pearlman, W. A., A new, fast, and efficient image codec based on set partitioning in hierarchical trees, IEEE Transactions on Circuits and Systems for Video Technology, vol. 6, no. 3, (June 1996), 243-250. doi: 10.1109/76.49

[15] Bini, S. P. and Abirami, S., Secure image deduplication using SPIHT compression, *2017 Int. Conf. on Commu. and Signal Proc.* (ICCSP), Chennai, 2017, 276-280. doi: 10.1109/ICCSP.2017.828636
[16] Ilgin, H. A., Akbulut, A. and Kalaycıoğlu, A., Image and Video Compression Artifact Reduction at Low Bit Rates Using Local Cosine Transform in Conjuction with Significance Map Coding, *11th International Conference on Electronics, Computers and Artificial Intelligence*, ECAI 2019, June 2019.
[17] Ilgin ,H. A. and Chaparro, L. F., Low Bit Rate Video Coding Using DCT-Based Fast Decimation/Interpolation and Embedded Zerotree Coding, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 7, July 2007, 833-844. doi: 10.1109/TCSVT.2007.898655

Current Address: Hakkı Alparslan ILGIN (Corresponding author): Ankara University, Electrical and Electronics Engineering Department, Ankara, Turkey
E-mail: ilgin@eng.ankara.edu.tr
Orcid ID: https://orcid.org/0000-0003-0112-4833

Current Address: Ahmet AKBULUT: Ankara University, Electrical and Electronics Engineering Department, Ankara, Turkey
E-mail: aakbulut@ankara.edu.tr
Orcid ID: https://orcid.org/0000-0001-8868-7385