PALETTE-BASED COMPOUND IMAGE COMPRESSION IN HEVC BY EXPLOITING NON-LOCAL SPATIAL CORRELATION

Wenjing Zhu[#], Oscar C. Au[#], Wei Dai[#], Haitao Yang[$], Rui Ma[#], Luheng Jia[#], Jin Zeng[#], Pengfei Wan[#]

[#] Hong Kong University of Science and Technology, [$] Huawei Technologies Co., Ltd

ABSTRACT

Non-camera captured images (also known as compound image) contain a mixture of camera-captured natural images and computer-generated graphics and texts. Nowadays, there are more and more applications calling for non-camera captured image/video compression scheme. However, current video coding standards, which are designed for natural video, treat non-camera captured video less carefully. For example, the state-of-the-art video coding standard High Efficiency Video Coding (HEVC) may blur or even remove edges in text/graphic region. A lot of schemes are proposed to preserve direction property of texts and graphics, such as palette-based intra coding. In this paper, a novel palette coding scheme is proposed for palette-based intra coding in HEVC. The palette in a block is predicted from an adaptive palette template, which records the statistical non-local spatial correlation of an image. Every block chooses its own palette using the palette template as the prediction in a rate-distortion optimized manner. Experimental results show that the proposed scheme can achieve up to 5.2% bit-rate saving compared to the state-of-the-art palette-based coding scheme in HEVC.

Index Terms— Compound image compression, palette-based coding, palette prediction, HEVC.

1. INTRODUCTION

With rapid and continuous advancements made in networking, communications, displays and devices such as tablets and smartphones, many applications call for High Efficiency Video Coding (HEVC) based coding solutions that can efficiently compress non-camera captured video content at high visual quality, which includes wireless display, tablets as second display, screen/desktop sharing and collaboration, *etc* [1]. How to efficiently compress non-camera captured images, or compound images, has become a prevalent and critical problem. However, current image coding standards (JPEG [2], JPEG 2000 [3]) and video coding standards (H.264/MPEG-4 AVC [4], HEVC [5]) aim at compressing natural camera-captured video, which will cause a lot of artifacts when used to encode and decode non-camera captured sequences [6]. Therefore, it is very important to develop efficient compound image coding tools. A lot of algorithms have been proposed to deal with this problem. In general, the algorithms can be categorized into bit-allocation approaches, layer-based approaches and block-based approaches.

The bit-allocation approaches didn't change the coding structure, they simply gave more bits or finer quantization steps to the text/graphic areas [7, 8]. But these approaches could not deal with the situation which most of the image was text/graphic region. The layer-based approaches (for example DjVu [9]) segmented an image into foreground layer and background layer using a binary mask and encoded three layers independently. These approaches could achieve significant gain over natural image coding schemes, but the performance was sensitive to segmentation method. So they could not handle the situation where the image was hard to be segmented. The block-based approaches firstly divided an image into blocks and classified the blocks into several types, each type was coded with a specific method. These approaches worked well for natural image and at the same time demonstrated significant gain for compound image. In [10], a novel method was proposed to represent a text/graphic block by several base colors (called palette) and an index map. This scheme could improve compression efficiency of compound image significantly. Algorithms in [11] and [12] followed this design and incorporated the scheme into H.264/MPEG-4 and HEVC with some modifications.

In this paper, the proposed scheme adopts the block-based approach and introduces a new palette coding scheme to fully exploit the local and non-local spatial correlation. An adaptive palette template is used to predict the palette of current block and each block chooses its own palette in a rate-distortion optimized (RDO) manner. It is introduced as a new intra coding mode into HEVC and compared with other modes in the RDO sense.

The rest of the paper is organized as follows. Section 2 gives a briefly introduction on the current palette-based coding system and lists the current drawbacks of this system. Section 3 presents the proposed palette coding scheme in detail. The experimental results are shown and analyzed in section 4. Finally, section 5 draws conclusions.

This work was done when the author Wenjing Zhu was an intern in Huawei Technologies.

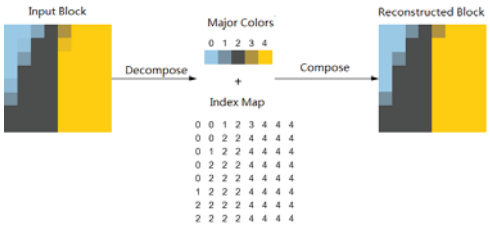

Fig. 1. BCIM representation.

2. RELATED WORK

In natural video coding, it is assumed that high frequencies are not that important to the human perception and therefore can be quantized strongly or even be removed. But the opposite is valid for non-camera captured content, assuming that an observer is more interested in text symbols than in accurate reconstruction of the background. So we have to deal with the sharp edges introduced by letters and symbols which are designed to be high contrast against the background.

Compound images have some properties that we need to consider when designing compression method. First of all, because of the strong anisotropic of compound images, transforms are not efficient. Secondly, the geometries of edges are usually complicated, the block level directional prediction in HEVC cannot adapt to the complex structure, which makes the prediction residual energy usually becomes very large and is difficult for coefficient coding algorithm in HEVC. Finally, the number of colors of a block in compound image usually is limited.

Based on the analysis, base colors and index map mode (BCIM) was proposed in [11] and incorporated into HEVC in [12] with modified multi-stage index map coding. In BCIM representation, the compound image blocks are regarded as a combination of color components and structure components. As shown in Fig. 1, the input image block is first decomposed into major colors and index map using color quantization algorithms. The index map will be entropy coded to achieve better coding performance. At the decoder side, the major colors and index map will be decoded and combined together to generate the reconstructed block.

In [13], another palette mode was proposed to HEVC as an additional intra mode. Different from BICM, the scheme in [13] also transmitted the residual together with the palette information and index map.

In general, these kinds of methods are called palette-based coding method, they always have a palette to represent the major colors of a block and an index map to specify colors of each position. Then, at decoder side, the original block is approximated using the palette and the index map with or without the help of residual. These methods can preserve the edge precisely, which make them very useful and popular in compound image compression.

In previous works, palette was generated by exploiting local spatial correlation. Each block generated the palette and index map based on its own pixel values. The palette of each block was encoded using a fix-length entropy coder. Each base color required b bits to encode where b represented the color bit-depth of the compound image. However, in compound images, different blocks may share same background colors and foreground colors, their palettes may have similar values, so there exists redundancy when coding palette directly without taking other palettes into consideration. The correlation among different palettes is called non-local spatial correlation which could be exploited to do palette prediction. In our proposed scheme, not only the local spatial correlation but also non-local spatial correlation are exploited. The palette of the current block is predicted from an adaptive palette template generated from other blocks to remove redundancy. Thus, additional coding gain can be achieved compared to the state-of-the-art non-predicted palette-based coding approaches.

3. PROPOSED PALETTE CODING SCHEME

For palette-based coding method, palette should be generated and encoded for transmission. In palette generating process, two aspects should be considered: Firstly, the palette should be chosen to well match to the pixel values of a block. Secondly, the index map corresponding to the chosen palette should be easily compressed. In this paper, we adopt the index map coding method proposed in [14], so the problem is simplified to the generation of the best palette for a particular block.

The palette of a block involves two aspects: palette length K and K base color values $p_k, k = 1, \cdots, K$. In general, the larger the palette length K is, the less the distortion of the approximation will be, but the index map will be more complex, which needs more bits to encoded. So when generating a palette, we should do a trade-off between distortion D_K and bit-rate R_K. Similar to the RDO process of HEVC, a Lagrange multiplier λ which controls the trade-off between rate R_K and distortion D_K is introduced to construct the Lagrange optimization objective function, which can be expressed as:

$$J(K, \lambda) = D_K + \lambda R_K, \tag{1}$$

where D_K and R_K are the distortion and rate given palette length K. For a fixed λ, different K will result in different J, the optimal palette length K_{opt} should satisfy

$$K_{opt} = \underset{1 \le K \le 2^b}{\arg\min}\, J(K, \lambda). \tag{2}$$

Since there are a lot of possible colors in the compound image, the number of calculating $J(K, \lambda)$ will be very large. To simplify the design, K_{max} is predefined as the upper limit of palette length to decrease complexity. The final problem can

be formulated as:

$$\begin{aligned} \min \quad & J(K,\lambda) \\ \text{s.t.} \quad & 1 \le K \le K_{max}. \end{aligned} \tag{3}$$

To solve the optimization problem in equation (3), D_K should be optimal given K. For an N pixels block, our purpose is to find K numbers optimally representing the N pixels in the sense of sum of square difference (SSD). Assume that $x_i, i = 1, \cdots, N$, is the pixel value of a color component or the block, x_i can be a scalar or a three dimensional vector, p_k is the k-th palette value and has as the same dimension as x_i, P_k is a set of pixels whose color are mapped to p_k, then D_K can be calculated as

$$D_K = \sum_{1 \le k \le K} \sum_{x_i \in P_k} \|x_i - p_k\|. \tag{4}$$

For R_K, it is calculated as:

$$R_K = R_p + R_i, \tag{5}$$

where R_p is the rate to encode the palette and R_i is the rate to encode the index map.

However, it is quite complicated to derive the optimal K and its corresponding palette and index map at the same time. To simplify the optimization process, the optimization problem is decomposed into two stages. At first stage, D_K is minimized respect to every K and get the palette and its index map correspondingly. Then R_p and R_i are calculated by encoding each palette and index map to get R_K. $J(K,\lambda)$ for each K is then calculated to select the K_{opt} and its palette and index map.

As discussed in the previous section, each palette is generated from its own block pixel information and is encoded independently respect to other blocks, which means the existing palette-based coding methods only consider the local spatial correlation. For a compound image, palettes of different blocks may have similar properties, such as foreground colors of different blocks may be similar or background colors of different blocks may be similar, we can call this property as non-local spatial correlation. This means that there exists redundancy among the palette of different blocks. To exploit the non-local correlation, we need to find the major values of palettes, and then construct a palette template to predict the palette of each block. To do this, assuming the size of palette template is T and its elements are $t_j, j = 1, \cdots, T$. The problem can be express as

$$\min \sum_{i=1}^{M} \sum_{k=1}^{K_i} \prod_{j=1}^{T} ||t_j - p_k^i||_0, \tag{6}$$

where M stands for the number of palette blocks in one frame, K_i represents the number of palette of block i and p_k^i represents the color of k-th palette element in i-th block. The solution can be derived by solving an equivalent problem. The equivalent problem is designed as follows. Firstly, Assemble all p_k^i into a set S. In S, the elements are categorized according to their pixel value, $S(\omega)$ means the count of pixel value ω is $S(\omega)$.

$$S(\omega) = \sum_{i=1}^{M} \sum_{k=1}^{K_i} (1 - ||\omega - p_k^i||_0). \tag{7}$$

The optimal palette template values $t_j, 1 \le j \le T$ are

$$\begin{aligned} \max \quad & \sum_{j=1}^{T} S(t_j) \\ \text{s.t.} \quad & t_m \ne t_j, \forall m, j, m \ne j. \end{aligned} \tag{8}$$

The equivalent problem is a convex problem and can be easily solved.

From the above analysis, we know that a two-pass coding is needed to construct an optimal palette template, the whole frame must be coded firstly to get all palette elements. To achieve one-pass coding, an adaptive palette template construction method is proposed. The most frequently used pixel values for all the previously coded blocks are chosen to approximate the optimal palette value. In this case, the set of palette elements just contains element of previous coded palette block, we call this set S', the constructing method of S' is same as S, so different palette block have different S'. S' can be expressed as

$$S'(\omega) = \sum_{n=1}^{i-1} \sum_{k=1}^{K_n} (1 - ||\omega - p_k^n||_0), \tag{9}$$

where i represents the index of the current palette block. Then, the palette template of block i can be derived by equation (8). But in this way, all the palette elements of previous block must be stored. To save memory, an adaptive palette template is designed to approximate the optimal palette template. For each element in the palette template, a counter is used to record the number of times it appears in the palette of previously coded blocks. The palette template and the counters are updated each time a palette coded block is processed. By this approach, only the palette template is needed to be stored. The detail algorithm is described in Algorithm 1.

For compound image, current block may have little relationship with the blocks far away, so time effect should be considered in the adaptive palette template constructing. As described in Algorithm 1, the counters $c[j], j = 1, \cdots, T$ are decreased by a time effect constant τ while encoding or decoding process of a palette block is finished. By this approach, the elements that haven't been used recently will have a small counter and can be easily replaced. After the old elements eliminating operation, the first step is to strengthen the elements which appear in current palette. The strengthening way is to increase the corresponding counters by a constant Δ. ϵ is a threshold to represent tolerance for assessing "same

Algorithm 1: Constructing palette template

Data: Palette of current block $P[K]$
Result: Palette template $PT[T]$
initialize palette template and counter at first block;
if *current block is palette base coded* **then**
  for $1 \leq j \leq T$ **do**
    $c[j] -= \tau$;
  for $1 \leq i \leq K$ **do**
    for $1 \leq j \leq T$ **do**
      if $\|P[i] - PT[j]\| \leq \epsilon$ **then**
        $c[j] += \Delta$.
for $1 \leq j \leq T$ **do**
  if $c[j] < \alpha$ **then**
    $PT[j]$ are replaced by new elements of $P[K]$;
    $c[j] = \Delta$;

Algorithm 2: Palette coding process

Data: Result of dynamic programming $P[K]$ and palette template $PT[T]$
Result: Array of palette prediction and prediction difference
initialization: palette length $K = 1$;
for $1 \leq K \leq K_{max}$ **do**
  for $1 \leq i \leq K$ **do**
    for $1 \leq j \leq T$ **do**
      if $\|P[i] - PT[j]\| \leq \epsilon$ **then**
        PPred[i] = 1;
        PDiff[i] = $P[i] - PT[j]$;
      else
        PPred[i] = 0;
  Calculate D_K and R_K $J_K = D_K + \lambda R_K$.
Find the optimal K which minimizes Lagrange object function J_K

element". After strengthening the frequent elements, next step is to wash out the rare elements. The elements whose counter values are less than a threshold α are recognized as rare elements and are replaced with new elements in the current palette.

As discussed before, RDO is needed when selecting the palette length K, the detailed palette length selection algorithm is given in Algorithm 2. When doing RDO process, we also take palette prediction into account. After doing palette prediction, the un-predicted palette values would be written directly into bit-stream. For the predicted palette values, because same operation of palette template constructing are performed at the encoder side and decoder side, only the prediction flag and prediction difference are written into bit-stream. By the above means, bit-saving can be achieved.

4. EXPERIMENTAL RESULT

The proposed method is integrated into the HEVC Range Extension reference software HM-10.1 + RExt-3.0 [15]. K_{max} is set to 16, λ is as the same as the Lagrange multiplier in HEVC, the size of palette template T is 8. Time elimination factor τ is set to 1, frequent element strengthening factor Δ is set to 2, Same element assessment factor ϵ is set to 3, old element assessment factor α is set to 2. In the simulation, we just choose 4 standard screen content test sequences from HEVC Range Extension common test condition [16]. BD-Rate [17] is regarded as the performance measurement of the proposed method. A negative value of the BD-Rate implied that the proposed approach achieves coding gains. The MBCIM frame work proposed in [14] and the HM-10.1 + RExt-3.0 [15] are chosen as the comparative algorithms. The simulation results are shown in Table 1. Encoding time of the three methods are the same.

Table 1. Objective comparison of comparative method and proposed method.

Test Sequence	MBCIM[14]	HM-10.1 + RExt-3.0
sc_cad_waveform	-4.9%	-56.4%
sc_pcb_layout	-5.2%	-71.3%
sc_ppt_doc_xls	-4.4%	-41.8%
sc_cg_twist_tunnel	-4.1%	-52.3%
Average	-4.7%	-55.5%

From the simulation results in Table 1, we can conclude that the proposed method outperforms current HEVC Range Extension software by 55.5% BD-Rate reduction and outperforms the state-of-the-art palette based coding method (MBCIM) by 4.7% BD-Rate reduction, while maintaining similar complexity.

5. CONCLUSION

In this paper, a palette coding method which fully exploits local and non-local spatial correlation is proposed. An adaptive palette template is constructed at the encoder and decoder side to exploit the non-local spatial correlation. Taking advantage of palette template, palette prediction can be performed, less palette elements are encoded into bitstream, which reduces the bit-rate. Simulation results show that under the same complexity, the proposed method outperforms current HEVC Range Extension software by 55.5% BD-Rate reduction and outperforms the state-of-the-art palette based coding method (MBCIM) by 4.7% BD-Rate reduction.

6. REFERENCES

[1] H. Yu, K. McCann, R. Cohen, and P. Amon, "Draft requirements for future extensions of HEVC in coding non-camera-captured content," *Document of Moving Picture Experts Group, m30477*, July 2013.

[2] G. K. Wallace, "The JPEG still picture compression standard," *Consumer Electronics, IEEE Transactions on*, vol. 38, no. 1, pp. xviii–xxxiv, 1992.

[3] M. Rabbani and R. Joshi, "An overview of the JPEG 2000 still image compression standard," *Signal processing: Image communication*, 2002.

[4] T. Wiegand, "Draft ITU-T recommendation and final draft international standard of joint video specification," *ITU-T rec. H. 264— ISO/IEC 14496-10 AVC*, 2003.

[5] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," 2012.

[6] H. Yu, X. Wang, and J. Ye, "AHG8: More investigation on screen content coding," *Document of Joint Collaborative Team on Video Coding, JCTVC-M0320*, April 2013.

[7] H. Meuel, J. Schmidt, M. Munderloh, and J. Ostermann, "Analysis of coding tools and improvement of text readability for screen content," in *Picture Coding Symposium (PCS), 2012*. IEEE, 2012, pp. 469–472.

[8] A. Zaghetto and R. L. De Queiroz, "Segmentation-driven compound document coding based on H. 264/AVC-INTRA," *Image Processing, IEEE Transactions on*, vol. 16, no. 7, pp. 1755–1760, 2007.

[9] P. Haffner, P. G. Howard, P. Simard, Y. Bengio, and Y. Lecun, "High quality document image compression with DjVu," *Journal of Electronic Imaging*, vol. 7, no. 3, pp. 410–425, 1998.

[10] W. Ding, Y. Lu, and F. Wu, "Enable efficient compound image compression in H. 264/AVC intra coding," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*. IEEE, 2007, vol. 2, pp. II–337.

[11] C. Lan, G. Shi, and F. Wu, "Compress compound images in H. 264/MPGE-4 AVC by exploiting spatial correlation," *Image Processing, IEEE Transactions on*, vol. 19, no. 4, pp. 946–957, 2010.

[12] W. Zhu, W. Ding, R. Xiong, Y. Shi, and B. Yin, "Compound image compression by multi-stage prediction," in *Visual Communications and Image Processing (VCIP), 2012 IEEE*. IEEE, 2012, pp. 1–6.

[13] L. Guo, M. Karczewicz, J. Sole, and R. Joshi, "Non-RCE3: Modified palette mode for screen content coding," *Document of Joint Collaborative Team on Video Coding, JCTVC-N0249*, July 2013.

[14] W. Zhu, J. Xu, and W. Ding, "Screen content coding with multi-stage base color and index map representation," *Document of Joint Collaborative Team on Video Coding, JCTVC-M0330*, April 2013.

[15] D. Flynn, J. Sole, and T. Suzuki, "HEVC Range Extension draft 3," *Document of Joint Collaborative Team on Video Coding, JCTVC-M1005*, April 2013.

[16] L. Guo, "HEVC Range Extension core experiment 3 (RCE3): Intra coding methods for screen content," *Document of Joint Collaborative Team on Video Coding, JCTVC-M1123*, April 2013.

[17] G. Bjontegard, "Calculation of average PSNR differences between RD-curves," *ITU-T VCEG-M33*, 2001.