# Pattern-matching scheme for HEVC screen content coding

## Yawei Xu [1, a], Jianxin Song [1, b]

[1]Nanjing University of Posts and Telecommunications, Nanjing 210003, China.

[a]robin_xyw@163.com, [b]songjx@njupt.edu.cn

**Keywords:** Basic color, Index map, Dictionary coding, Hash-table, Fast dictionary searching.

**Abstract.** Screen Content refers to image/video generated by electronic devices, like cartoons, 3-D games or captures of computer screens or video with text overlay etc. The existing video coding techniques don't perform well on screen content. In this paper, the characteristics of screen content and coding efficiency of HEVC on screen content are analyzed. This paper proposes a pattern-matching mode aimed at improving the coding efficiency for screen content video based on HEVC. The scheme uses basic color and index map (BCIM) to convert the pixel values in a CU to index values. Then a dictionary based coding method is adopted to code the index values. In addition, a fast dictionary searching algorithm based on hash-table is proposed to reduce the string matching time. The experiment result shows the proposed mode achieves significant coding efficiency improvement on screen content compared with HEVC.

## Introduction

With the rapid development of mobile Internet and emerging of rich media applications, screen content plays an important role in many scenarios, such as remote desktop, wireless display and cloud computing. The challenging demand for screen content coding (SCC) is to achieve ultra-high visual quality and higher compression ratio. The high efficiency video coding (HEVC) standard [1] has a significant improvement on coding efficiency. However HEVC is mainly designed for nature video captured by cameras. Screen content, which is usually a mixture of computer generated content and content captured by cameras, has different characteristics compared with natural video captured by cameras. The current HEVC doesn't entirely use these characteristics and still has room for improvement in coding efficiency.

In the development of HEVC, researchers have already put forward some schemes to improve the coding efficiency of screen content. SCC using transform skip [2, 3, 4] is proposed, but it only skips the transform and the energy of screen content is still spread in the frequency domain. Furthermore, SCC using Intra Block Copy is put forward in [5, 6, 7], which is very effective to screen content, since it removes redundancy from repeating patterns which typically occur in text and graphics regions. Moreover, SCC using dictionary [8, 9, 10] is proposed by exploiting the repeated patterns in screen content. Besides, SCC using basic color and index map [11, 12] is proposed by dividing screen content into color component and structure component.

In this paper, a pattern match mode is proposed to improve the efficiency of HEVC screen content coding. The scheme uses basic color and index map (BCIM) coding method to convert CU pixels into indexes. Then indexes in CU are encoded based on dictionary compression method in the vertical scanning order. String search is conducted on the dictionary which is created by previously encoded string to find the string matching with the current index string to be encoded. If there is a matched string, it is encoded as the offset (the difference of first index's serial number between the current encoded index string and the previously encoded string) and length of continuous matching. Current encoded index string is processed with prediction coding if there's no matched string. In addition, a fast dictionary searching algorithm based on hash-table is proposed to reduce the time of string matching. Experiment result shows it has a significant coding efficiency gain compared with HEVC scheme.

**Analysis of screen content characteristics and proposed pattern-matching mode**

Screen content has different characteristics with natural video captured by cameras, as shown in Fig.1. Firstly, the color amount in local block of screen content is limited. The image block of screen content is usually occupied by several main colors. The values of several main colors may have a big difference between each other. However, the color amount of image block in a natural video is much larger and the values of pixels are very close. Secondly, screen content has many complex structures such as texts and graphics etc. The complex structure information is very important for human eyes to obtain information. And the damage of structure (such as fuzzy words) will have a huge harm to information. However the representation of information in a natural video mainly relies on color. Most non-edge area in a natural video is smooth. The energy of natural video mainly focuses on coefficients of low frequency when it is represented by frequency domain. The conventional video coding framework such as HEVC adopts the encoding method of prediction and transform and filter out small coefficients of high frequency after DCT is applied to the residuals. These coefficients of high frequency represent the details of the natural video image. These details have no significant influence to human eyes on obtaining the content of natural video image. But the energy of screen video spreads over all coefficients of frequency. Therefore the transform in HEVC doesn't have a good effect on screen content. It may damage structure information of screen video, have an influence to human eyes on obtaining text and graphic from screen content and lead to serious damage to screen video. From the above analysis, the basic color and index map method is fit for encoding screen content. It can preserve structure information of screen content and skip the transform.



Fig. 1 Screen content picture (left) and natural video picture (right)

In addition, common screen content usually has many small and sharp bitmap structures, such as text, menu, icon and button. There are often many similar or same patterns in screen content picture. For example, a whole page of English only contains 26 uppercase and lowercase letters. These letters are all composed of a lower number of basic strokes and Chinese character is also composed of more than ten kinds of basic strokes. In consequence dictionary based compression coding can make good use of repeated pattern in the screen content picture.

The entire encoding flow chart of proposed pattern-matching mode in this paper is shown in Fig.2. The detailed algorithm description of pattern-matching mode is introduced in the following sections.
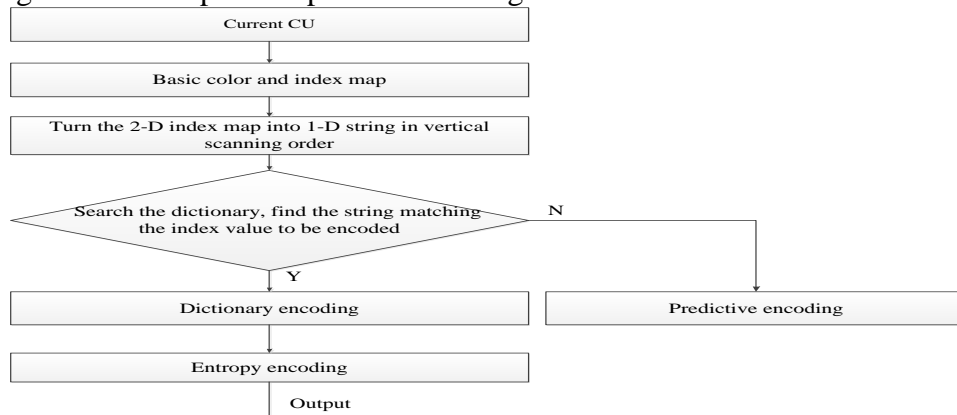


Fig. 2 Encoding flow chart of proposed pattern-matching mode

To leverage the characteristics of the screen content, three color components of input pixels are first packaged by following the format $(Y_1, U_1, V_1, Y_2, U_2, V_2 ... Y_N, U_N, V_N)$, where the $(Y_i, U_i, V_i)$ are the three color components of the i th pixel[10]. Then our scheme is performed on the packed coding units. For

each coding unit (CU), the pixel value is turned into index map according to the basic color. Then 2-D index map in CU is turned into the 1-D string adopting vertical scanning order [11].

The dictionary encoding method is adopted to encode 1-D index map value. The target of dictionary encoding is making use of the characteristic of repeated pattern in the frame. Previously encoded index map values are preserved in the reconstructed pixel buffer after reconstruction, which will be used as reference pixels in traditional predictive coding or the dictionary in pattern-matching mode. The size of dictionary equals to the size of string search window, i.e., L. For the current index map value, the dictionary is searched to find whether there is a matched string, as depicted in Fig.3. If multiple matched strings are found, the string which has the longest length should be selected. Then symbol bit 1 is sent to indicate the matched string is found, the offset (difference between the current position and reference position) and the length of successive matched indices.
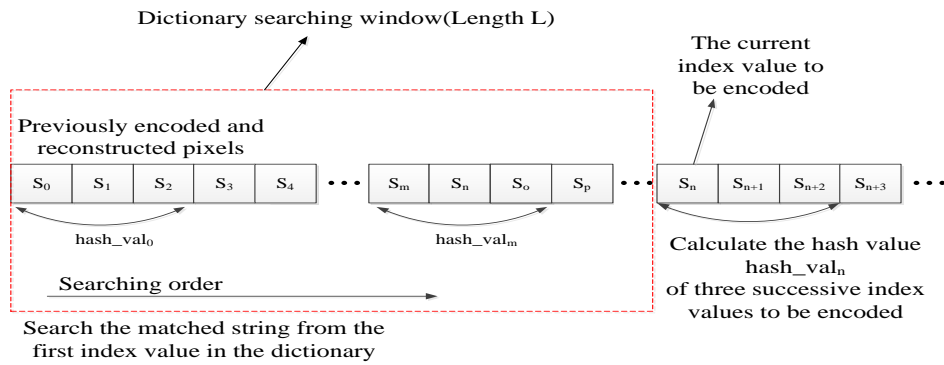


Fig. 3 Hash based dictionary search

If no matched string is found, symbol bit 0 is sent to indicate there is no matched string. Meanwhile the unmatched index needs to be encoded in another way named predictive encoding. The aim of predictive encoding is to decrease the bit number of encoding unmatched index. Assuming there are $N$ basic colors, $\log_2 N$ is needed if we directly encode the unmatched index. The principle of predictive encoding is shown is Fig.4. Firstly the left and top index values $I_1, I_2$ of current index are needed to be candidates for prediction. Secondly current index value $I$ is compared with $I_1, I_2$. If $I = I_1$, symbol 10 is sent to encode current index where the first symbol bit 1 indicates the prediction is successful and the second symbol bit 0 indicates $I = I_1$. If $I = I_2$, symbol 11 is sent where the first symbol bit 1 indicates the prediction is successful and the second symbol bit 1 indicates $I = I_2$. Therefore, in the case of successful prediction, only 2 bits are needed to encode the current unmatched index. If $I \neq I_1$ and $I \neq I_2$, it means the prediction is unsuccessful. Symbol 0 which indicates the prediction is failing and additional $\log_2(N-2)$ bits are sent to encode the unmatched index.
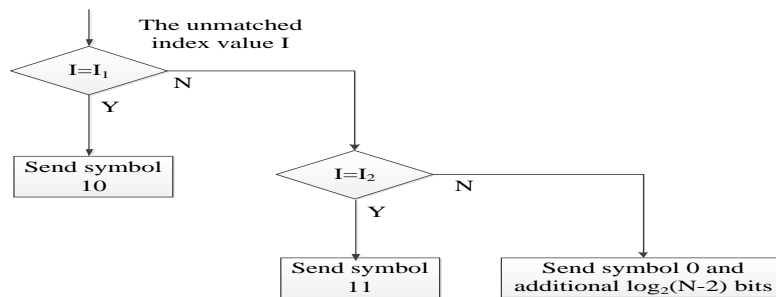


Fig. 4 Flow chart of predictive encoding

Not only the predictive encoding algorithm shows low complexity, but it can significantly decrease the encoding bit numbers. In addition, it's suitable for parallel encoding because the current index value doesn't depend on its right and top index values.

The proposed pattern-matching mode is incorporated into HEVC as an additional mode, as depicted in Fig.5. In HEVC, the input image is organized using CUs, which can vary in size, ranging from

64*64 to 8*8. The proposed mode is only implemented on 8*8 CU. The rate distortion optimization progress is adopted to determine whether current CU will use the proposed scheme. In addition 1bit is needed to indicate whether current CU uses the proposed scheme on 8*8 CU levels.
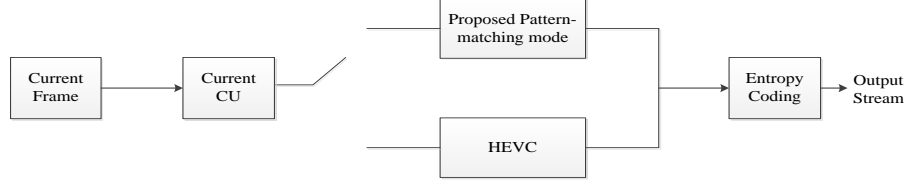


Fig. 5 Overview of the encoding framework

**Fast dictionary searching algorithm based on the hash-table**

In order to improve the compression ratio of dictionary coding, it is needed to search the matched string which has the longest length in the dictionary searching. In addition, it's required to increase the length of searching window that represents numbers of previously encoded and reconstructed pixels. For example, if we treat the whole frame as the searching window, it means all the repeated patterns in the picture will be utilized. It's no doubt that increasing the length will increase the searching time and complexity. Therefore the paper proposed a fast dictionary searching algorithm based on the hash-table to decrease searching time.

When pattern-matching mode is performed to encode the CU, vertical scanning order is used, i.e., from top to bottom and from left to right. Therefore the reference points in the searching window are numbered in the order of encoding them when we perform pattern-matching mode. Each index in the searching window corresponds to a serial number.

In the condition of the matched string length equals to 1 or 2, dictionary encoding is not only time consuming but also not significant in the effect of bit rate compression. Therefore, to improve the efficiency of dictionary encoding, dictionary search is conducted only when the length of matched string is at least 3. For the current index value to be encoded, firstly the hash value $hash\_val_n$ of three successive index values to be encoded from current position n is calculated, as shown in Fig.3.

The structure of hash table is depicted in Fig.6. In the hash table, the head node of each linked list stores the hash value $hash\_val$ corresponding to the list. Subsequent nodes store the serial number m of certain index value in the dictionary. The hash value of three successive index values from serial number m equals to the hash value stored in the head node of list, i.e., $hash\_val_m = hash\_val$.
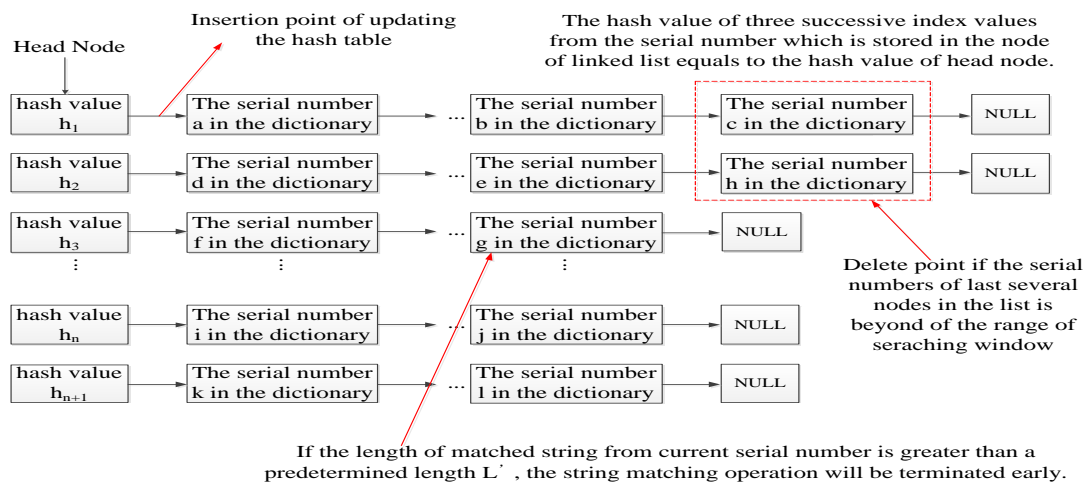


Fig. 6 Hash-based dictionary

Then the fast dictionary searching algorithm based on hash table is detailed below. First the hash value $hash\_val_n$ of three successive index values to be encoded from current position n is calculated. After that hash table is looked up to find whether the hash value exists.

If the hash value exists, it indicates there is matched string in the dictionary. The linked list whose head node equals to the hash value is traversed. For the serial number in each node of this list, the string

starting from this serial number in the dictionary is compared with the current encoded index string for string matching, which will choose the longest matched string. The offset mentioned in the dictionary searching can be easily calculated by subtracting the beginning serial number of longest matched string from the first serial number of current encoded index string. In order to reduce searching time of finding the best matched string, list traversal can be terminated in advance. If the length of longest matched string which we select currently is greater than a predetermined length $L'$, the string matching operation will be stopped early. The length of matched string selected currently is treated as the longest length. The hash table is updated simultaneously by inserting the serial number of currently encoding index value into the place after the head node, as shown in Fig.6. If the serial numbers of last several nodes in the list are beyond of the range of searching window, these nodes will be deleted from hash table at the aim of saving memory.

If the hash value doesn't exist, it indicates that there is no matched string in the dictionary, the linked list whose head node equals to this hash value is created. Then predictive encoding is conducted on the three unmatched index value to be encoded.

**Experimental results**

The proposed pattern-matching scheme is implemented into HEVC range extension reference software HM12.0RExt4.1 to evaluate its coding performance on screen content videos. All-intra high efficiency configuration is adopted in our simulations. The quantization parameters are 22, 27, 32 and 37, respectively. The screen content sequences selected by the experts in Joint Collaborative Team on Video Coding (JCTVC) group are tested. For comparison, Lin's scheme based on dictionary encoding [8] is also implemented on the same software. The proposed pattern-matching scheme and Lin are tested against each other based on the "all intra" configuration. The bit-rate saving of proposed pattern-matching scheme compared with HM and Lin's scheme is shown in table 1.

Table 1 BITRATE SAVING OF THE PROPOSED PATTERN-MATCHING SCHEME

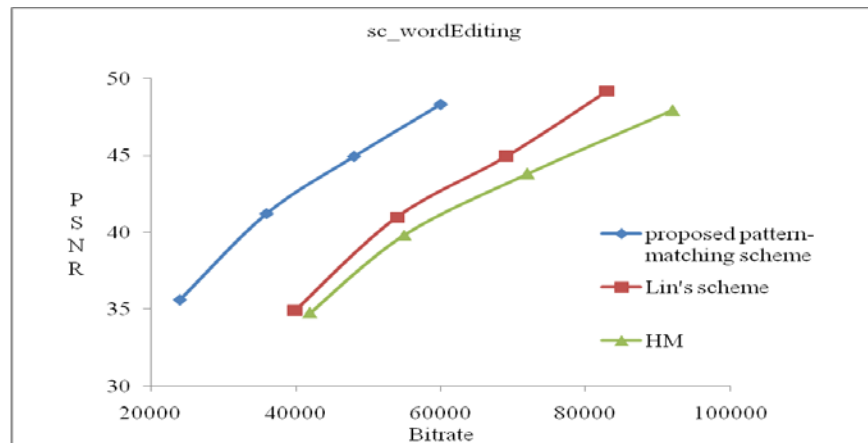| Sequences | vs.HM | | | vs..Lin | | |
|---|---|---|---|---|---|---|
| | Y | U | V | Y | U | V |
| Class F | -4.7% | -3.8% | -4.0% | -4.5% | -3.8% | -4.0% |
| SC RGB 444 | -22.9% | -22.5% | -22.4% | -3.7% | -3.2% | -2.6% |
| SC YUV 444 | -24.1% | -23.9% | -23.7% | -15.4% | -13.5% | -12.6% |



Fig. 7 R-D curve of sc_wordEditing

As we can see, the proposed scheme achieves 4.7%, 22.9% and 24.1% bit-rate saving on Class F, RGB and YCbCr classes respectively compared with HM. It also achieves 4.5%, 3.7% and 15.4% bit-rate saving respectively compared with Lin's scheme. Fig.7 depicts the R-D curve of sc_wordEditing sequence. The experimental result indicates significant improvement of encoding performance on screen videos compared with HEVC scheme.

## Conclusions

   With the fast development of mobile internet and emergency of rich video applications, screen content has become an important kind of video source types. The paper proposed a pattern-matching scheme which combines the basic color and index map encoding method with dictionary encoding, in order to utilize the characteristic that the image block of screen content is usually dominated by several major colors, contains complex structures and has repeated patterns. In addition a fast dictionary searching algorithm based on hash table is proposed to reduce the time of string matching. The experimental result shows significant improvement of encoding performance on screen videos compared with HEVC scheme.

## References

[1]  Sullivan G J, Ohm J, Han W J, et al. Overview of the high efficiency video coding (HEVC) standard[J]. Circuits and Systems for Video Technology, IEEE Transactions on, 2012, 22(12): 1649-1668.

[2]  Lan C, Shi G, Wu F. Compress compound images in H. 264/MPGE-4 AVC by exploiting spatial correlation[J]. Image Processing, IEEE Transactions on, 2010, 19(4): 946-957.

[3]  Mrak M, Xu J Z. Improving screen content coding in HEVC by transform skipping[C]//Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European. IEEE, 2012: 1209-1213.

[4]  Nam J, Sim D, Bajic I V. HEVC-based adaptive quantization for screen content videos[C]//Broadband Multimedia Systems and Broadcasting (BMSB), 2012 IEEE International Symposium on. IEEE, 2012: 1-4.

[5]  Yu S L, Chrysafis C. New intra prediction using intra-macroblock motion compensation[C]//JVT-C151, 3rd meeting of Joint Video Team (JVT). 2002, 1.

[6]  Kwon D K, Budagavi M. Fast intra block copy (IntraBC) search for HEVC screen content coding[C]//Circuits and Systems (ISCAS), 2014 IEEE International Symposium on. IEEE, 2014: 9-12.

[7]  Chen C C, Xu X, Liao R L, et al. Screen content coding using non-square intra block copy for HEVC[C]//Multimedia and Expo (ICME), 2014 IEEE International Conference on. IEEE, 2014: 1-6.

[8]  Lin T, Zhang P, Wang S, et al. Mixed chroma sampling-rate high efficiency video coding for full-chroma screen content[J]. Circuits and Systems for Video Technology, IEEE Transactions on, 2013, 23(1): 173-185.

[9]  Zhu W, Xu J, Ding W, et al. Adaptive LZMA-based coding for screen content[C]//2013 Picture Coding Symposium (PCS). 2013.

[10] Zhu W, Ding W, Xu J, et al. 2-D Dictionary Based Video Coding for Screen Contents[C]//Data Compression Conference (DCC), 2014. IEEE, 2014: 43-52.

[11] Ma Z, Feng X, Xu M. Advanced screen content coding using color table and index map[C]//Multimedia and Expo Workshops (ICMEW), 2014 IEEE International Conference on. IEEE, 2014: 1-6.

[12] Zhu W, Ding W, Xu J, et al. Screen content coding based on HEVC framework[J]. 2014.