

Screen Content Coding in HEVC and Beyond

LIN Tao, ZHAO Liping, and ZHOU Kailun
(Tongji University, Shanghai 200092, China)

1 Introduction

The screen content coding (SCC) standard [1] for high efficiency video coding (HEVC) is an international standard specially developed for screen content. It indicates the start of a new chapter in video coding research and standardization. On one hand, SCC is required by many traditional applications and an ever-increasing number of new and emerging applications as well [2]–[4]. On the other hand, screen content is very different from traditional content, thus different coding tools are needed. Furthermore, screen content is an extremely comprehensive and diverse class of content and includes traditional photosensor (e.g. CMOS or CCD sensor) captured pictures as a small subset. As a result, SCC is becoming a very active field to attract considerable attention from both academia and industry [1]–[35], and is expect to play a major role in advancing both researches and applications of video coding technology.

The Audio Video Coding Standard (AVS) Workgroup of China is also working on SCC standard, which is expected to become a national standard in China and an IEEE standard by the second half of 2016. Since SCC has much more application areas, market sectors, and customers of different requirements to serve than traditional video coding, multiple standards are needed and benefit each other to grow the market size. There are also many SCC application areas and market sectors where proprietary solutions are also acceptable.

This paper discusses the background and current status of SCC and its standardization work in HEVC and AVS. The rest of the paper is organized as follows. In section 2, application areas and requirements of SCC is presented. Section 3 describes characteristics of screen content. Section 4 is devoted to technical description and standardization of three major dedicated SCC techniques and their relation. Section 5 reports coding performance comparisons of the three SCC techniques. Fi-

Abstract

Screen content is video or picture captured from a computer screen typically by reading frame buffers or recording digital display output signals of a computer graphics device. Screen content is an extremely comprehensive and diverse class of content and includes traditional photosensor captured pictures as a small subset. Furthermore, screen content has many unique characteristics not seen in traditional content. By exploring these unique characteristics, new coding techniques can significantly improve coding performance of screen content. Today, more than ever, screen content coding (SCC) is becoming increasingly important due to the rapid growth of a variety of networked computers, clients, and devices based applications such as cloud computing and Wi-Fi display. SCC is the ultimate and most efficient way to solve the data transferring bottleneck problem in these applications. The solution is to transfer screen pixel data between these computers, clients, and devices. This paper provides an overview of the background, application areas, requirements, technical features, performance, and standardization work of SCC.

Keywords

HEVC; AVS; Screen Content Coding; String Matching; Video Coding

nally, Section 6 concludes the paper and also presents some future work of SCC.

2 Application Areas and Requirements of SCC

Almost all applications of SCC have one thing in common: display units are connected to information processing resources, including the central processing unit (CPU), graphics processing unit (GPU), and storage space, through networks.

Application areas of traditional video coding are mostly related to TV broadcasting, video content delivery or streaming, and video surveillance. However, SCC opens a huge and new application area of video coding: cloud computing platform, where CPUs, GPUs, and main storages devices are all located in a place called cloud and shared by multiple user (client) devices that are connected to the cloud through networks. As shown in **Fig. 1**, the cloud can be as big as a datacenter with thousands or tens of thousands of servers or as small as a single computer with one multi-core-CPU/GPU combo or even a smart phone. Virtual network computing (VNC), remote desktop, virtual desktop infrastructure (VDI), PC over IP (PCoIP),

This work was supported in part by National Natural Science Foundation of China under Grant No. 61201226 and 61271096, and Specialized Research Fund for the Doctoral Program under Grant No. 20130072110054.

Screen Content Coding in HEVC and Beyond

LIN Tao, ZHAO Liping, and ZHOU Kailun



▲ Figure 1. Application scenario of SCC: Cloud/cloudlet and cloud-mobile computing platform.

ultra-thin client, and zero-client are a few examples of SCC based cloud computing platform implementation. SCC based implementation has the highest graphics performance among all implementations of the cloud computing platform [2]–[4]. SCC can reduce screen pixel data bit-rate to a level that widely deployed networks can support even for screen resolution of 2560x1600 or higher at 60 Hz screen refresh rate, thus enables cloud based computing and information processing to become a mainstream model not only used by professionals but also by average people in their daily life. The daily activities that often need to handle typical screen content include web browsing, document sharing in video conferencing, remote desktop sharing and collaboration, office document editing, engineering drawing, hardware design engineering, software programming, map navigating and address direction searching, and many more. Therefore, the market of SCC based cloud computing and its variations are expected to grow exponentially and its market is becoming much bigger than traditional video coding market.

Besides cloud computing platforms, SCC has at least the following application areas:

- Cloudlet computing, a variation of cloud computing, where the cloud is a small one (cloudlet) or is split into a few small cloudlets. A client device can become a cloudlet.
- Cloud - mobile computing, a variation of cloud computing where the client devices are mobile devices such as smart phones, tablets, or notebooks
- Cloud gaming
- Wireless display, for example, Wi-Fi display, where Wi-Fi connection is used to replace a video cable that attaches a display unit such as a monitor or a TV set to a PC, a notebook, a tablet, a smart phone, a set-top-box, and so on
- Screen or desktop sharing and collaboration, where multiple users at different locations view the same desktop screen
- Video conferencing with document sharing
- Remote teaching
- Display wall

- Multi-screen display for many viewers
- Digital operating room (DiOR) or OR-over-IP.

From SCC coding performance and coding quality point of view, SCC application areas and markets can be divided into the following two major segments, which have different requirements.

1) High and ultra-high video quality segment

This segment includes cloud/cloudlet/cloudlet- mobile computing platforms, enterprise IT cloud platforms, VDI, remote desktops, PCoIP, ultra-thin clients, zero-clients, cloud gaming, and more. One distinct feature of this segment is that the screen is usually viewed by one viewer at a viewing distance less than one meter. This viewing model is the same as what traditional computer users normally do. Users may include professionals, and no visual loss of screen content picture can be tolerated. Due to this feature, lossless coding or visually lossless coding with high and ultra-high picture quality is an absolute requirement in this segment. The video color format requirement for this segment is RGB or YUV 4:4:4. Another distinct feature of this segment is that human-computer interaction (HCI) is involved, and both encoding and decoding of screen content are part of the HCI process [2]. The total round-trip time from a keyboard input or a mouse click to task processing on the cloud, screen content rendering on the cloud, screen content encoding on the cloud, and finally screen content decoding on the client device should be within a limit that users can accept. Thus, the encoding and decoding time and latency are very important to get overall crisp system response time (SRT) for uncompromised excellence of HCI experiences. The total encoding and decoding latency requirement is typically 30 milliseconds or less. This is less than one frame period in 30 frames per second coding configuration. Therefore, in this segment, contrast to traditional video coding applications, peak intra-picture (all-intra) coding performance is far more critical than random-access, low-delay-P, and low-delay-B coding performance. In this segment, the highest mainstream screen resolution today and in near future is probably 2560x1600 pixels. At 60 frames per second screen refresh rate and 24 bits per pixel color precision, the raw screen pixel data bit-rate is 5626 mega bit per second (mbps). Today, advanced widely deployed networks infrastructure can probably provide sustainable bandwidth of up to 20 mbps. Therefore, the basic compression ratio requirement is close to 300:1. The compression ratio at ultra-high visually lossless picture quality is certainly very challenging, especially in all-intra coding configuration.

2) Middle and low video quality segment

This segment includes Wi-Fi display, display wall, external second display of mobile devices, multi - screen display for many viewers, video conference with document sharing, remote teaching, and more. One distinct feature of this segment is that the screen content is usually viewed by more than one viewer at a viewing distance more than one meter. This viewing model

is not much different from traditional TV viewing model. Due to this feature, lossy coding with middle (or low in some cases) picture quality is acceptable in this segment. The video color format requirement for this segment is RGB or YUV 4:4:4 or YUV 4:2:0. Another distinct feature of this segment is that human-computer interaction (HCI) is usually not involved. So, the encoding and decoding latency is not as important as in the first segment. All-intra coding performance is also not as important as in the first segment. In this segment, the highest screen resolution today and in near future is probably 4096x2160 pixels. At 60 frames per second screen refresh rate and 24 bits per pixel color precision, the raw screen pixel data bit-rate is 12,150 mbps. Today, advanced widely deployed networks infrastructure can probably provide sustainable bandwidth of up to 20 mbps. Therefore, the basic compression ratio requirement is 600:1. The compression ratio of 600:1 is certainly very challenging even at middle picture quality.

As a result, SCC requirements for compression ratio and picture quality are very challenging. Traditional coding techniques cannot meet the requirements, and new coding techniques are absolutely needed.

3 Characteristics of Screen Content

One of the most important characteristics of screen content is its diversity and comprehensiveness.

Screen content is video or picture captured from a computer screen typically by either reading frame buffers or recording digital display output signals of a computer graphics device. Computer screen content is extremely diverse and comprehensive due to the diversity and comprehensiveness of materials, data, information, and their visual bitmap representations that computers need to handle, render and display. The diversity and comprehensiveness can be seen from at least the following three aspects:

- 1) The number of distinct colors in a region (e.g. a block). The number can range from one, i.e. the entire region has only one color, to maximum, which is equal to the number of pixels in the region;
- 2) Degree of pattern matchability. A matching (either exact matching defined as having no difference or approximate matching defined as having difference within a predetermined limit) pattern set is a set of two or more patterns that have both matching shapes and matching value of pixels. Pattern matchability is the state of existence of matching pattern sets. The degree of pattern matchability can be measured by at least the following metrics:
 - The size (the number of elements) of a set of matching patterns in a predetermined range usually called searching range in an encoder or a reference range in a decoder. Big size means high degree of pattern matchability.
 - The number of matching pattern sets in a predetermined range. The number equal to 0 means the lowest degree of

pattern matchability. The degree of pattern matchability is generally related to both of the number of matching pattern sets and the average size of all matching pattern sets. In general, big number of matching pattern sets or big average size of all matching pattern sets means high degree of pattern matchability.

- The average distance between elements of a matching pattern set. Short distance usually means high degree of pattern matchability and that most elements are located closely.

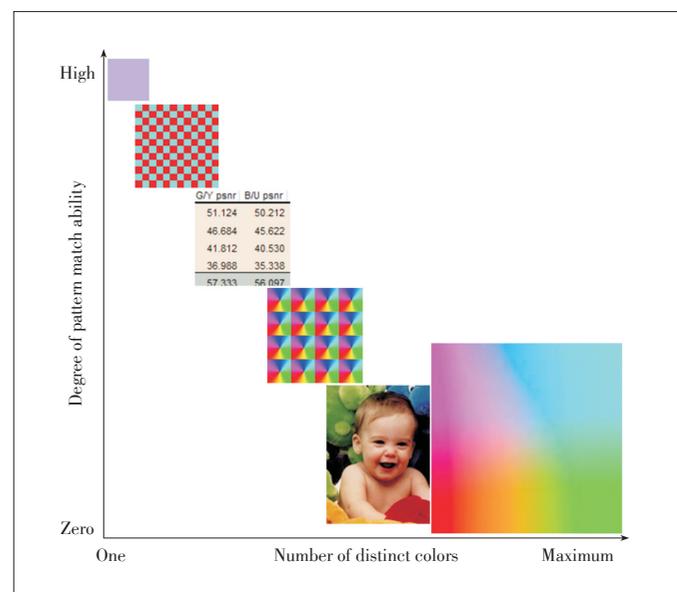
3) Shape and color of matching patterns

The shape and color of a matching pattern set can be arbitrary. Hence, the number of possible different shapes and colors is huge. In fact, matching patterns in screen content have a virtually unlimited number and variety of different shapes and colors. For example, the shapes range from simple ones such as squares, rectangles, triangles, circles, polygons, crescents, diamonds of different sizes to complex ones such as all kinds of mathematic curves, geometric shapes, and fonts of different typeface, size, weight, slope, width, and special effect.

In Fig. 2, six screen content examples illustrate the screen content diversity in terms of the number of distinct color.

As shown in Fig. 2, from top left to bottom right, the six examples are:

- 1) Single color square. The number of distinct color is one. The degree of pattern matchability is the highest.
- 2) Two color checkers. The number of distinct color is two. The degree of pattern matchability is very high.
- 3) Spreadsheet cells. The number of distinct color is about twenty. The degree of pattern matchability is high.
- 4) A color space diagram repeated 16 times. The number of distinct color is big. The degree of pattern matchability is medium.



▲ Figure 2. Diversity of screen content.

Screen Content Coding in HEVC and Beyond

LIN Tao, ZHAO Liping, and ZHOU Kailun

- 5) A photosensor (camera) captured photo, i.e. a natural picture. The number of distinct color is big. The degree of pattern matchability is very low.
- 6) A color space diagram. The number of distinct color reaches the maximum. The degree of pattern matchability is zero.

It should be noted that screen content includes the traditional photosensor captured natural content as a small subset of screen content. In fact, this special subset generally features a very large number of distinct colors and almost zero pattern matchability, as the fifth example in Fig. 2.

Screen content also includes sophisticated light-shaded and texture-mapped photorealistic scenes generated by computers. Virtual reality, 3D computer animation with lighting and shading, 3D computer graphics with lighting and shading, and computer games are examples of photorealistic screen content. From the word “photorealistic”, it can be easily seen that computer generated photorealistic screen content has almost the same properties as the traditional photosensor captured natural content and share the same features like relatively smooth edges and complicated textures. In particular, computer generated photorealistic screen content also features a very large number of distinct colors and almost zero pattern matchability.

Besides the general characteristics of diversity and comprehensiveness, typical screen content has at least the following three specific characteristics, which traditional natural content usually does not have.

- 1) Typical computer screens seen in common everyday applications are often rich in grids, window frames, window panes, table cells, slide-bars, toolbars, line charts, and so on. They feature very sharp edges, uncomplicated shapes, and thin lines with few colors, even one-pixel-wide single-color lines. Therefore, for typical screen content, the number of distinct colors is low.
- 2) Sharp and clear bitmap structures, especially small ones, such as alphanumeric characters, Asian characters, icons, buttons, graphs, charts and tables are often seen in typical computer screens. Thus, there are usually many similar or identical patterns in typical screen content. For examples, all texts are composed of a very limited number of characters, and all characters themselves are composed of a significantly further limited number of basic strokes. Therefore, typical screen content has high degree of pattern matchability.
- 3) Splitting and merging of matching patterns. A pair of matching patterns (A, B) with a pair distance $d(A, B)$ may be split into two or more pairs of small matching patterns (A_1, C_1) , (A_2, C_2) , \dots , where pattern A is split into two or more small patterns A_1, A_2, \dots , and each pair of the small matching patterns has a pair distance shorter than $d(A, B)$. On the other hand, if pattern A is split into two or more small patterns A_1, A_2, \dots , two or more pairs of matching patterns (A_1, C_1) , (A_2, C_2) , \dots may be merged into a big pair of matching pattern (A, B), whose pair distance is longer than the pair distance

of each of the small matching pattern pairs (A_1, C_1) , (A_2, C_2) , \dots . Note that in splitting and merging, it is not necessary for patterns C_1, C_2, \dots to be related to pattern B. The splitting and merging of matching patterns mean that for a piece (or block) of pixels in typical screen content, matching relation is not unique, but has multiple options available. Different options have different number of pairs and different pair distances. Pattern splitting and merging based multiple matching relation is an important characteristic which needs to be fully explored in SCC.

The first two specific characteristics are strongly related. Actually, lower number of distinct colors usually (but not always) means higher degree of pattern matchability and vice versa. On one extreme, one distinct color results in the highest degree of pattern matchability; at the other extreme, the maximum number of distinct colors results in zero degree of pattern matchability.

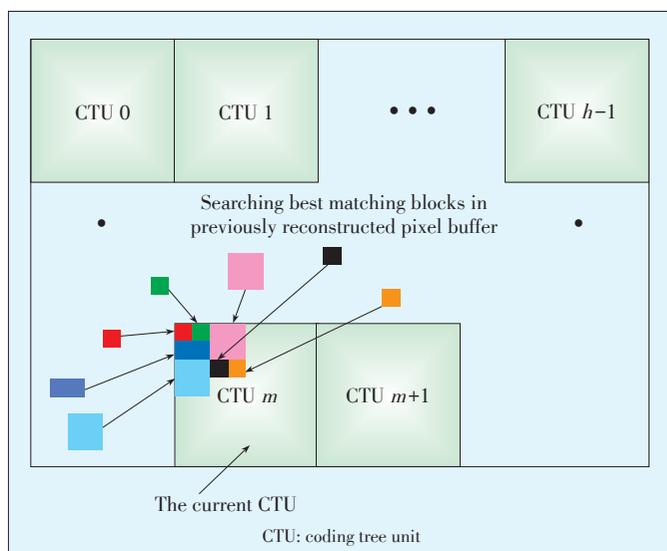
Because traditional block-matching and transform based hybrid coding technique does not take advantage of any special characteristics of screen content, dedicated SCC techniques have tremendous potential to improve coding efficiency of screen content significantly by exploring the special characteristics of screen content. Since almost all major characteristics of screen content are related to pattern matching, three major SCC techniques are all pattern matching based techniques: 1) intra picture block matching technique, also known as intra block copy (IBC) or intra motion compensation (IMC) technique; 2) intra coding unit (CU) pixel index string matching technique, also known as palette (PLT) technique; and 3) pseudo 2D string matching (P2SM or P2M) technique.

4 Technical Description and Standardization of Three SCC techniques

IBC [5]–[8] is a straightforward extension of conventional inter-prediction to intra picture coding with a few simplifications. The main simplification is to remove pixel interpolation and do only whole pixel prediction. In IBC (Fig. 3), when encoding a prediction unit (PU), the encoder searches an optimal matching block as a reference matching pattern in a pre-determined search window (reference buffer), which is usually a previously reconstructed pixel buffer. Reference matching patterns have the same shapes and sizes as PUs such as 4x8, 8x4, 8x8, 16x16, 32x32, 64x64 pixels. The search window has a pre-determined size which varies from a few CUs to the full frame. The encoding result is a motion vector (MV) and a residual block.

In IBC decoding, the decoder parses the bitstream to obtain a MV. The decoder uses the MV to locate a reference matching pattern in the previously reconstructed reference pixel buffer. The decoder then uses the values of the reference matching pattern as the predictor of the current PU being decoded.

IBC is efficient to code matching patterns of a few fixed siz-

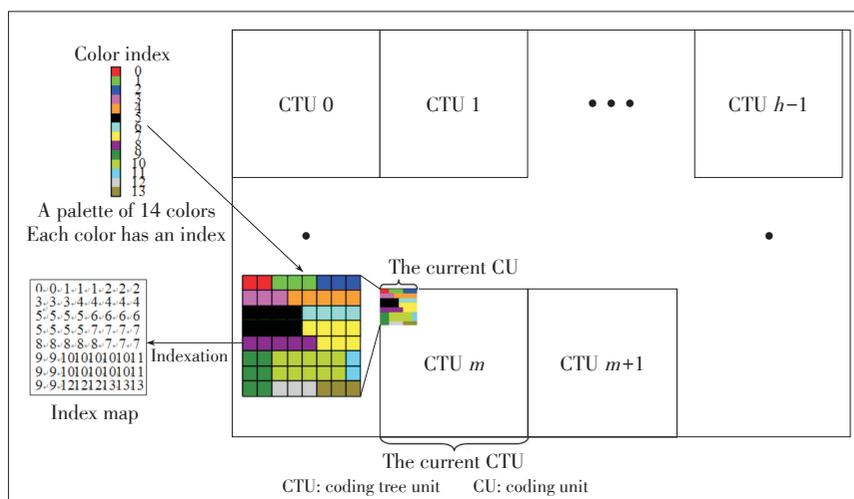


▲ Figure 3. Intra block copy coding technique.

es with rectangle or square shapes in a picture, but is not flexible enough to code matching patterns of different sizes from a few pixels to a few thousands of pixels with a variety of shapes.

IBC is adopted into the HEVC SCC draft by unification with conventional inter-prediction, i.e. specifying the current picture itself as a reference picture.

As shown in Fig. 4, when encoding a CU in PLT [9]–[11], the encoder first performs color quantization on the CU to obtain a few representative colors and puts the representative colors into a palette. Each color in the palette has an index. When the number of representative colors exceeds a limit, the last index is reserved to represent all extra colors beyond the limit. The extra colors are named escape colors. All pixels in the CU are converted into indices to build an index map. The index map is further coded by either left-string-matching or above-string-matching. The escape colors are quantized and coded into the bitstream.



▲ Figure 4. Palette coding technique.

All indices in an index map are coded string by string using two types of string matching. The first type of string matching is left-matching. The first string (0 0), second string (1 1 1), and third string (2 2 2) in the index map of Fig. 4 are examples of left-matching. In a left-matching string, all indices are identical. The second type of string matching is above-matching. In the index map of Fig. 4, string (5 5 5 5) in the 4th row, string (7 7 7) in the 5th row, string (9 9 10 10 10 10 10 11) in the 7th row, and string (9 9) in the 8th row are examples of above-matching. It is obvious that in left-matching, the reference matching string (pattern) overlaps the current string being coded, while in above-matching, the reference matching string (pattern) is the string above the current string being coded. A left-matching string has three coding parameters: string type, index, and length. An above-matching string has two coding parameters: string type and length.

For each CU, the PLT encoding results are a palette, an index map coded by two types of string matching, and quantized escape colors. The encoding results are explicitly or implicitly put into the video bitstream after entropy coding.

In PLT decoding, the decoder parses the bitstream and performs other decoding steps to obtain the palette, the index map, and the escape colors, from which the decoder can complete the decoding process to reconstruct all pixels of the CU.

The palette coding technique can code matching patterns inside a CU using two types of intra-CU pixel-index string matching, but it cannot exploit non-local matching patterns outside of a CU.

PLT is adopted into the HEVC SCC draft as a CU level coding mode named palette mode.

IBC can only code matching patterns of a few fixed sizes with rectangle or square shapes efficiently. PLT can only code matching patterns completely inside a CU efficiently. However, typical screen content shows significant diversity in terms of the shape and size of matching patterns and the distance of a matching pattern pair. Therefore, IBC and PLT only partially explore the special characteristics of screen content.

P2SM has its origin in Lempel-Ziv (LZ) algorithm [12], but is more sophisticated than the original LZ algorithm. In P2SM, two reference buffers are used. One is primary reference buffer (PRB) which is typically a part of the traditional reconstructed picture buffer to provide reference string pixels for the current pixels being coded. The other is secondary reference buffer (SRB) which is a dynamically updated look-up table (LUT) storing a few of recently and frequently referenced pixels for repetitive reference by the current pixels being coded. When encoding a CU, for any starting pixel being coded, searching of optimal matching string with a variable length is performed in both PRB and

Screen Content Coding in HEVC and Beyond

LIN Tao, ZHAO Liping, and ZHOU Kailun

SRB. As a result of the searching, either a PRB string or an SRB string is selected as a reference matching pattern on a string-by-string basis. For a PRB string, an offset and a length are coded into the bitstream. For an SRB string which is really an SRB pixel color duplicated many times, an SRB address and a duplication count are coded into the bitstream. If no reference string of at least one pixel is found in PRB or SRB, the starting pixel is coded directly into the bitstream as an unmatched pixel. Thus, a CU coded by P2SM has three matching types: Match_PRB, Match_SRB, and Match_NONE. A letter S coded by P2SM is shown in Fig. 5. The size of the current CU is 8x8. The following is five examples of PRB strings or SRB strings (Fig. 5).

The 1st string marked with red “1” is a 9-pixel PRB string. The reference matching string is in PRB with offset (9, 3).

The 2nd string marked with green “2” is a 4-pixel SRB string. The reference matching string consists of the 1st SRB pixel color duplicated four times.

The 3rd string marked with red “3” is a 4-pixel PRB string. The reference matching string is in PRB with offset (0, 3).

The 4th string marked with red “4” is a 14-pixel PRB string. The reference matching string is in PRB with offset (8, -4).

The 5th string marked with green “5” is a 7-pixel SRB string. The reference matching string also consists of the 1st SRB pixel color duplicated seven times.

In P2SM decoding, the decoder parses the bitstream and performs other decoding steps to obtain the matching type, (offset, length) or (SRB address, length) or unmatched pixel, from which the decoder can complete the decoding process to reconstruct all pixels of the CU.

P2SM is adopted into the initial working draft of AVS screen mixed content coding extension as a CU level coding mode in March 2016.

It is obvious that IBC and PLT are two special cases of P2SM. In fact, IBC is a P2SM special case that restricts a PU to have only one reference matching string. PLT is also a

P2SM special case that limits all reference matching strings within the same CU being coded and allows only SRB strings (left-matching) and reference matching strings above the current strings (above-matching). The two special cases are called big string case and SRB string only case [31] in P2SM. Since P2SM is developed in a late stage of HEVC SCC project, it is not in the HEVC SCC draft. P2SM is adopted into the AVS screen mixed content coding working draft as universal string prediction (USP) tool.

5 Coding Performance Comparison of IBC, PLT, and P2SM

Coding performance comparison experiments use HM-16.6+SCM-5.2 reference software [35] and HM-16.6+P2SM software [31]. The following coding options are compared:

- 1) NoSCC implemented by disabling both IBC and PLT in HM-16.6+SCM-5.2
- 2) IBC implemented by disabling only PLT in HM-16.6+SCM-5.2
- 3) PLT implemented by disabling only IBC in HM-16.6+SCM-5.2
- 4) IBC+PLT (SCM which includes both IBC and PLT) implemented by HM-16.6+SCM-5.2
- 5) P2SM implemented in HM-16.6+P2SM.

The experimental results are generated under the common test conditions and lossy all-intra configuration defined in [34]. Fourteen test sequences are used in the experiment. The test sequences are classified into four categories: text and graphics with motion (TGM), mixed content (MC), camera captured (CC), and animation (ANI). YCbCr (YUV) color format version is used in the experiment. To evaluate the overall coding performance, the Bjøntegaard delta rate (BD-rate) metric [36], [37] is used. For each category, an average BD-rate reduction is calculated. Encoding and decoding software runtime are also compared for evaluating the complexity of the encoder and decoder.

Tables 1–4 show the coding performance improvement (BD-rate reduction percentage in negative numbers) of IBC, PLT, IBC+PLT (SCM), and P2SM, respectively. Table 5 shows the coding performance improvement of P2SM over SCM.

Table 1. Coding performance improvement of IBC over NoSCC

	Y	U	V
TGM	-47.80%	-48.51%	-48.67%
MC	-41.23%	-42.21%	-42.34%
ANI	-1.38%	-1.72%	-1.58%
CC	-0.07%	-0.12%	-0.10%
Enc Time		158.08%	
Dec Time		78.75%	
ANI: animation	Dec: Decoding	MC: mixed content	
CC: camera captured	Enc: Encoding	TGM: text and graphics with motion	

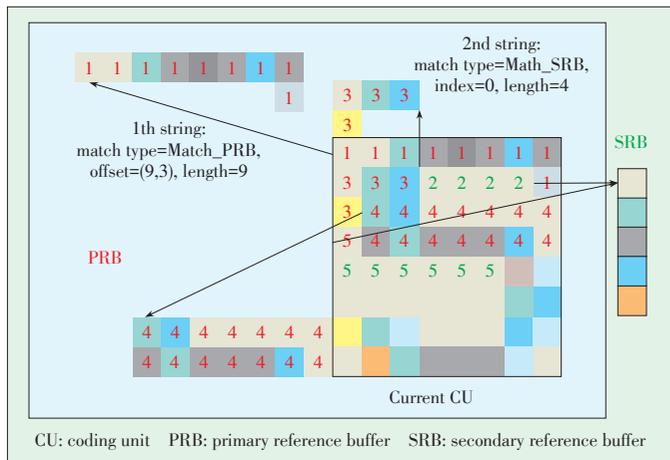


Figure 5. A letter S coded by P2SM technique.

▼ Table 2. Coding performance improvement of PLT over NoSCC

	Y	U	V
TGM	-40.83%	-46.18%	-48.15%
MC	-24.70%	-34.70%	-34.21%
ANI	0.19%	-3.04%	-2.78%
CC	0.03%	0.02%	-0.02%
Enc Time		125.48%	
Dec Time		90.89%	
ANI: animation CC: camera captured	Dec: Decoding Enc: Encoding	MC: mixed content TGM: text and graphics with motion	

▼ Table 3. Coding performance improvement of IBC+PLT (SCM) over NoSCC

	Y	U	V
TGM	-57.06%	-60.76%	-62.35%
MC	-45.57%	-50.85%	-50.90%
ANI	-1.17%	-4.40%	-4.08%
CC	-0.03%	-0.13%	-0.06%
Enc Time		172.19%	
Dec Time		80.85%	
ANI: animation CC: camera captured	Dec: Decoding Enc: Encoding	MC: mixed content TGM: text and graphics with motion	

▼ Table 4. Coding performance improvement of P2SM over NoSCC

	Y	U	V
TGM	-58.19%	-61.34%	-63.00%
MC	-46.26%	-51.24%	-51.50%
ANI	-0.01%	5.39%	2.62%
CC	0.15%	1.21%	1.60%
Enc Time		140.63%	
Dec Time		80.02%	
ANI: animation CC: camera captured	Dec: Decoding Enc: Encoding	MC: mixed content TGM: text and graphics with motion	

▼ Table 5. Coding performance improvement of P2SM over SCM

	Y	U	V
TGM	-4.74%	-4.24%	-4.20%
MC	-1.40%	-1.02%	-1.45%
ANI	1.19%	10.43%	7.00%
CC	0.17%	1.34%	1.66%
Enc Time		81.67%	
Dec Time		98.98%	
ANI: animation CC: camera captured	Dec: Decoding Enc: Encoding	MC: mixed content TGM: text and graphics with motion	

The experimental results show:

- 1) For screen content (TGM and MC categories), P2SM has higher coding performance than IBC or PLT or both combined.
- 2) IBC has higher coding performance than PLT, and both

have significant overlap.

3) For typical and common screen content (TGM), P2SM is superior to IBC and PLT combined (HM - 16.6+SCM - 5.2) by close to 5% in term of BD-rate.

Recently, it is reported [32], [33] that P2SM can achieve significant coding performance improvement for screen content rendered using sub - pixel - rendering techniques such as ClearType developed and widely applied in text rendering to achieve clear and smooth text display on an LCD panel. For a ClearType snapshot and a ClearType test sequence, P2SM can achieve 39.0% and 35.4% Y BD-rate reduction, respectively, comparing to HM-16.6+SCM-5.2.

6 Conclusions

Driven by increasing demand from both existing application areas such as Wi-Fi display and emerging application areas such as cloud computing platforms, SCC technology has made significant progress in the past three years.

Two major SCC standardization projects so far are HEVC SCC project and AVS/IEEE SCC project. Both are expected to complete by the second half of 2016. Two special cases of P2SM, i.e. IBC and PLT are adopted into the HEVC SCC draft. P2SM is adopted into the AVS screen mixed content coding working draft using the name of universal string prediction (USP).

Another technique named adaptive color transform (ACT) is also adopted into HEVC SCC. ACT is based on a prediction residual coding technique [38] and is a general technique instead of SCC dedicated. ACT is mainly effective on RGB color format sequences and has negligible effect on YUV color format sequences.

String matching is a superset of block matching which has been thoroughly studied for more than thirty years. P2SM provides a flexible trade-off between coding efficiency and coding complexity. String matching technology is still in its early stage of development, much like MPEG-1 stage of block matching technology, and has significant room for improvement. Therefore, future work in SCC and general video coding includes: 1) Further study of pattern matchability in screen content pictures and other types of contents, 2) improvement on string matching technology to code a variety of contents with different pattern matchability efficiently, 3) further reduction of coding complexity of string matching techniques, and 4) Optimization of string matching techniques for specific application areas with special requirement.

References

- [1] R. Joshi, S. Liu, G. Sullivan, et al., "High efficiency video coding (HEVC) screen content coding: draft 4," JCT-VC, Warsaw, Poland, JCTVC-U1005, Jun. 2015.
- [2] T. Lin, K. Zhou, and S. Wang, "Cloudlet-screen computing: a client-server architecture with top graphics performance," *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 13, no. 2, pp. 96-108, June 2013. doi: 10.1504/IJA-HUC.2013.054174.

Screen Content Coding in HEVC and Beyond

LIN Tao, ZHAO Liping, and ZHOU Kailun

- [3] Y. Lu, S. Li, and H. Shen, "Virtualized Screen: A Third Element for Cloud-Mobile Convergence," *IEEE Multimedia*, vol. 18, no. 2, pp. 4–11, Apr. 2011. doi: 10.1109/MMUL.2011.33.
- [4] T. Lin and S. Wang, "Cloudlet-screen computing: a multi-core-based, cloud-computing-oriented, traditional-computing-compatible parallel computing paradigm for the masses," in *IEEE International Conference on Multimedia and Expo*, New York, USA, Jul. 2009, pp. 1805–1808. doi: 10.1109/ICME.2009.5202873.
- [5] M. Budagavi and D. Kwon, "AHG8: video coding using Intra motion compensation," JCT-VC, Incheon, Korea, JCTVC-M0350, Apr. 2013.
- [6] D. Kwon and M. Budagavi, "Intra motion compensation with variable length intra MV coding," JCT-VC, Vienna, Austria, JCTVC- N0206, Jul. 2013.
- [7] C. Pang, J. Sole, L. Guo, M. Karczewicz, and R. Joshi, "Intra motion compensation with 2-D MVs," JCT-VC, Vienna, Austria, JCTVC- N0256, Jul. 2013.
- [8] C. Pang, J. Sole, L. Guo, R. Joshi, and M. Karczewicz, "Displacement vector signaling for intra block copying," JCT-VC, Geneva, Switzerland, JCTVC- 00154, Oct. 2013.
- [9] C. Lan, X. Peng, J. Xu, and F. Wu, "Intra and inter coding tools for screen contents," JCT-VC, Geneva, Switzerland, JCTVC-E145, Mar. 2011.
- [10] W. Zhu, W. Ding, et al., "Screen content coding based on HEVC framework," *IEEE Transaction on Multimedia*, vol.16, no.5, pp. 1316–1326, Aug. 2014.
- [11] L. Guo, W. Pu, et al., "Color palette for screen content coding," in *IEEE International Conference on Image Process*, Paris, France, Oct. 2013, pp. 5556–5560.
- [12] J. Ziv and A. Lempel, "A universal algorithm for sequential data compression," *IEEE Transactions on Information Theory*, vol. 23, no. 3, pp. 337–343, May 1977.
- [13] T. Lin, K. Zhou, X. Chen, and S. Wang, "Arbitrary shape matching for screen content coding," in *IEEE Picture Coding Symposium*, San Jose, USA, Dec. 2013, pp. 369–372. doi: 10.1109/PCS.2013.6737760.
- [14] W. Zhu, J. Xu, W. Ding, Y. Shi, and B. Yin, "Adaptive LZMA-based coding for screen content," in *IEEE Picture Coding Symposium*, San Jose, USA, Dec. 2013, pp. 373–376. doi: 10.1109/PCS.2013.6737761.
- [15] T. Lin, X. Chen, and S. Wang, "Pseudo-2-D-matching based dual-coder architecture for screen contents coding," in *IEEE International Conference on Multimedia and Expo*, San Jose, USA, Jul. 2013, pp. 1–4. doi: 10.1109/ICMEW.2013.6618315.
- [16] S. Wang and T. Lin, "Compound image compression based on unified LZ and hybrid coding," *IET Image Processing*, vol. 7, no. 5, pp. 484–499, May 2013. doi: 10.1049/iet ipr.2012.0439.
- [17] T. Lin, P. Zhang, S. Wang, K. Zhou, and X. Chen, "Mixed chroma sampling-rate high efficiency video coding for full-chroma screen content," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 1, pp. 173–185, Jan. 2013. doi: 10.1109/TCSVT.2012.2223871.
- [18] W. Zhu, W. Ding, R. Xiong, Y. Shi, and B. Yin, "Compound image compression by multi-stage prediction," in *IEEE Visual Communications and Image Processing Conference*, San Diego, USA, Nov. 2012, pp. 1–6. doi: 10.1109/VCIP.2012.6410758.
- [19] S. Wang and T. Lin, "United coding method for compound image compression," *Multimedia Tools and Applications*, vol. 71, no. 3, pp. 1263–1282, 2014. doi: 10.1007/s11042-012-1274-y.
- [20] S. Wang and T. Lin, "United coding for compound image compression," in *IEEE International Conference on Image and Signal Processing*, Yantai, China, Oct. 2010, pp. 566–570. doi: 10.1109/CISP.2010.5647270.
- [21] S. Wang and T. Lin, "A unified LZ and hybrid coding for compound image partial-lossless compression," in *IEEE International Conference on Image and Signal Processing*, Cairo, Egypt, Oct. 2009, pp. 1–5. doi: 10.1109/CISP.2009.5301019.
- [22] W. Ding, Y. Lu, and F. Wu, "Enable efficient compound image compression in h.264/AVC intra coding," in *IEEE International Conference on Image Processing*, Penang, Malaysia, vol. 2, Oct. 2007, pp. 337–340. doi: 10.1109/ICIP.2007.4379161.
- [23] L. Zhao, K. Zhou, and T. Lin, "CE3: results of test B.4.2 (minimum string length of 20 pixels) on intra string copy," JCT-VC, Geneva, Switzerland, JCTVC -T0136, Feb. 2015.
- [24] C. Hung, Y. Chang, J. Tu, C. Lin, and C. Lin, "CE3: crosscheck of CE3 test B.4.2 (JCTVC-T0136)," JCT-VC, Geneva, Switzerland, JCTVC-T0179, Feb. 2015.
- [25] L. Zhao, K. Zhou, S. Wang, and T. Lin, "Non-CE3: improvement on intra string copy," JCT-VC Doc JCTVC-T0139, Feb. 2015.
- [26] R. Liao, C. Chen, W. Peng, and H. Hang, "Crosscheck of non-CE3: improvement on intra string copy (JCTVC-T0139)," JCT-VC, Geneva, Switzerland, JCTVC-T0200, Feb. 2015.
- [27] T. Lin, K. Zhou, and L. Zhao, "Non-CE1: enhancement to palette coding by palette with pixel copy (PPC) coding," JCT-VC, Warsaw, Poland, JCTVC-U0116, Jun. 2015.
- [28] L. Zhao, W. Cai, J. Guo, and T. Lin, "Flexible coding tools to significantly improve SCC performance in cloud and mobile computing," JCT-VC, Warsaw, Poland, JCTVC-U0189, Jun. 2015.
- [29] R. Liao, C. Chen, W. Peng, et al., "Crosscheck of Non-CE1: Enhancement to palette coding by palette with pixel copy (PPC) coding," JCT-VC, Warsaw, Poland, JCTVC-U0173, Jun. 2015.
- [30] W. Wei, X. Meng, "Cross-check report of U0116," JCT-VC, Warsaw, Poland, JCTVC-U0189, June 2015.
- [31] K. Zhou, L. Zhao, and T. Lin, "Advanced SCC tool using Pseudo 2D string matching (P2SM) integrated into HM16.6," JCT-VC, Geneva, Switzerland, JCTVC-V0094, Oct. 2015.
- [32] L. Zhao, J. Guo, and T. Lin, "Significantly improving coding performance of Clear Type texts and translucently blended screen content by P2SM," JCT-VC, Geneva, Switzerland, JCTVC-V0095, Oct. 2015.
- [33] J. Guo, L. Zhao, and T. Lin, "A new SCC test sequence with ClearType text rendering for consideration," JCT-VC, Geneva, Switzerland, JCTVC-V0097, Oct. 2015.
- [34] H. Yu, R. Cohen, K. Rapaka, and J. Xu, "Common conditions for screen content coding tests," JCT-VC, Warsaw, Poland, JCTVC-U1015, Jun. 2015.
- [35] Heinrich Hertz Institute. (2015). *Rec. ITU-T H.265/ISO/IEC 23008-2 High Efficiency Video Coding* [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.6+SCM-5.2
- [36] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," ITU-T SG16 Q.6 Document, VCEG-M33, Austin, US, Apr. 2001.
- [37] G. Bjøntegaard, "Improvements of the BD-PSNR model," ITU-T SG16 Q.6 Document, VCEG-A111, Berlin, Germany, Jul. 2008.
- [38] D. Marpe, H. Kirchhoffer, V. George, P. Kauff, and T. Wiegand, "Macroblock-adaptive residual color space transforms for 4:4:4 video coding," in *IEEE International Conference on Image Processing*, Atlanta, USA, 2006, pp. 3157–3160.

Manuscript received: 2015-11-12

Biographies

LIN Tao (lintao@tongji.edu.cn) received his MS and PhD degrees from Tohoku University, Japan in 1985 and 1989. He has been with VLSI Lab, Tongji University, China since 2003. In 2005, he was awarded "Chang Jiang Scholars", the highest honor given by China Ministry of Education. From 1988 to 2002, he was a postdoctoral researcher with University of California, Berkeley, and developed multimedia ICs and products at several companies in Silicon Valley, including Integrated Device Technology, Inc., PMC-Sierra Inc., Cypress Semiconductor Corp., and NeoMagic Corp. He has been granted 24 US patents and 14 China patents. His current research interests include cloud-mobile computing, digital signal processing, audiovisual coding, and multimedia SoC design.

ZHAO Liping (vlsi@tongji.edu.cn) received her MS degree in computer science and technology from Hunan University, China in 2009. She is currently a PhD candidate in control science and engineering at VLSI lab of Tongji University, China. Her current research interests include screen content coding and video coding.

ZHOU Kailun (vlsi@tongji.edu.cn) received his MS degree from Shanghai Jiaotong University, China in 2003. He is currently pursuing the PhD degree with Tongji University, China. His current research interests include embedded system design, video coding, and ASIC architecture, design and verification.