Emerging Technologies toward Future Video Coding

Seishi Takamura

NTT Media Intelligence Laboratories, NTT Corporation
1-1 Hikarino-oka, Yokosuka, Kanagawa, 239-0847 Japan
Keywords: Video Coding, H.264/MPEG-4 AVC, H.265/MPEG-H HEVC, Versatile ideo Coding

ABSTRACT

In this paper, we first overview the ever-advancing history of video coding technology and standardization activities as well as evolution of video communication traffic. Then we review latest standardization activity on video coding, and introduce two examples of our new approach, real-entity-oriented coding in particular, to further enhance visual quality and compression performance.

1 INTRODUCTION

1.1 Video coding technology and IP traffic trends

Video coding technology has evolved over nine decades since the emergence of AT&T Bell-Lab's video phone in 1927. First international video coding standard was established in 1984 as H.120. Since then, as shown in Fig. 1, video coding technology have produced many successful standards such as MPEG-1 [1], MPEG-2 [2], H.264/MPEG-4 AVC [3], H.265/MPEG-H HEVC [4]. Experts are currently working toward MPEG-I Versatile Video Coding (VVC) standard to publish in 2020 [5].

The amount of Internet protocol (IP) traffic is increasing rapidly worldwide. Cisco [6] reported that in 2017, the annual run rate for global IP traffic was 122 EB (exa (10^{18}) bytes) per month. They predict IP traffic to grow at a compound annual growth rate of 26% from 2017 to 2022, and to increase threefold (387 EB per month) by 2022. IP video traffic accounted for 75% of all IP traffic in 2017, which is forecast to be 82% by 2022. Such video traffic content is already compressed (mainly by MPEG standards) down to a few hundredths of the original size. Thus, it is clear that not only video-based services but also network services as a whole would collapse without video coding standards. Also, with the trend for rapid growth of video traffic, more and more powerful compression techniques are necessary.

1.2 Versatile Video Coding

In April 2018, development activity of the new standard named ISO/IEC 23090 MPEG-I Part 3 Versatile Video Coding was initiated [7]. The “I” in MPEG-I stands for Immersive Media. It is scheduled to reach the FDIS stage (technically finalized) in October 2020. Formerly its codec was called JEM [8] and is now advanced to VTM [9]. Some of the principal technologies adopted in VTM so far include chroma separate tree (CST), cross-component linear model (CCLM), adaptive loop filter (ALF), affine motion compensation (AFF), multiple transform set (MTS), and dependent quantization (DQ).

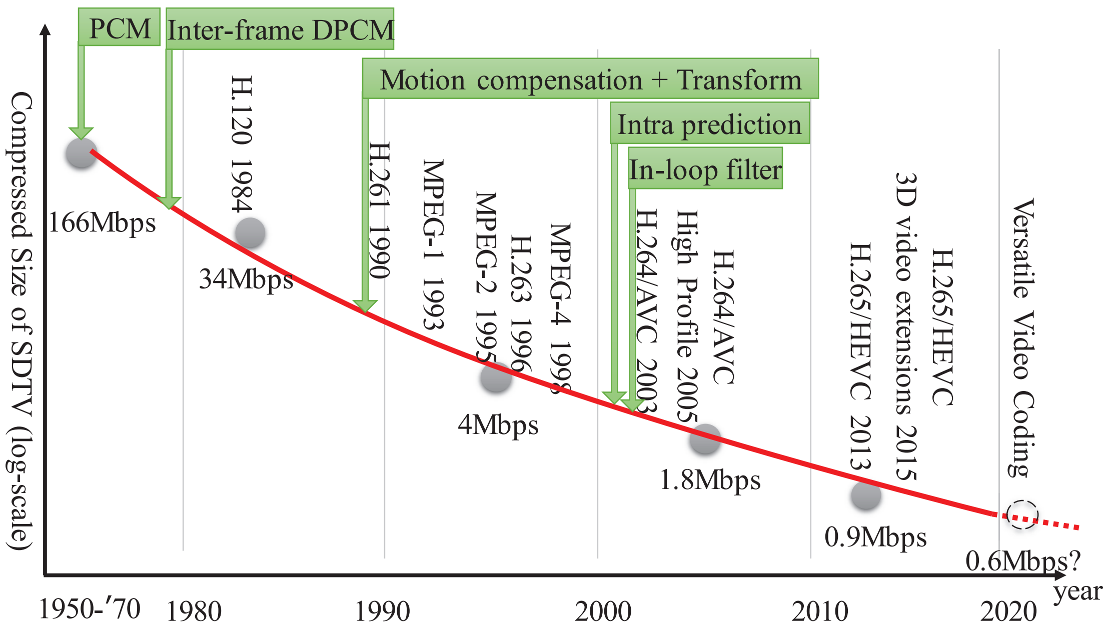

Fig. 1 Evolution of video coding technology and compression performance

Fig. 2 Comparison of decoded frames (part magnified). Left: former VVC test model (JEM7.1) [8] 559kbps, 36.4 dB; Middle: H.265/HEVC test model (HM16.19) 552kbps, 35.4dB; H.264/AVC test model (JM19.0) 1106kbps, 35.2dB

The target compression performance is a 30–50% bit-rate reduction compared to H.265/HEVC at the same subjective video quality. However, with latest VTM version 6, the compression performance is 24-35%. It is forecast that this will be almost close to final objective performance of VVC. For subjective and objective comparison, decoded frames of VVC and former standards with rates and PSNR values are shown in Fig. 2.

1.3 Real-entity-oriented coding

In conventional video coding schemes including all above standard technologies, reproducing a signal as close to the original video signal as possible has been mostly aimed at. With this approach, decoded video quality can never exceed original video signal. Generally, video capturing system inevitably offers noises, quantization, and subsampling distortions. As for fixed-camera videos, it has been reported that extraction of noise-reduced background image information by applying simple temporal filtering drastically gains coding efficiency up to 50% (measured in PSNR with original noisy signal) [10]. That work encourages us to depart from noisy pixel-oriented coding to noiseless real-entity-oriented coding (Fig. 3).

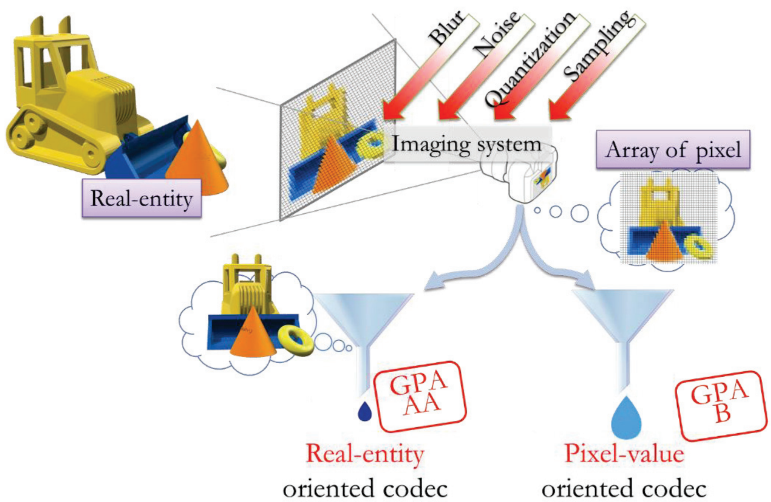

Fig. 3 Real-entity-oriented video coding vs. pixel-value-oriented (conventional) video coding

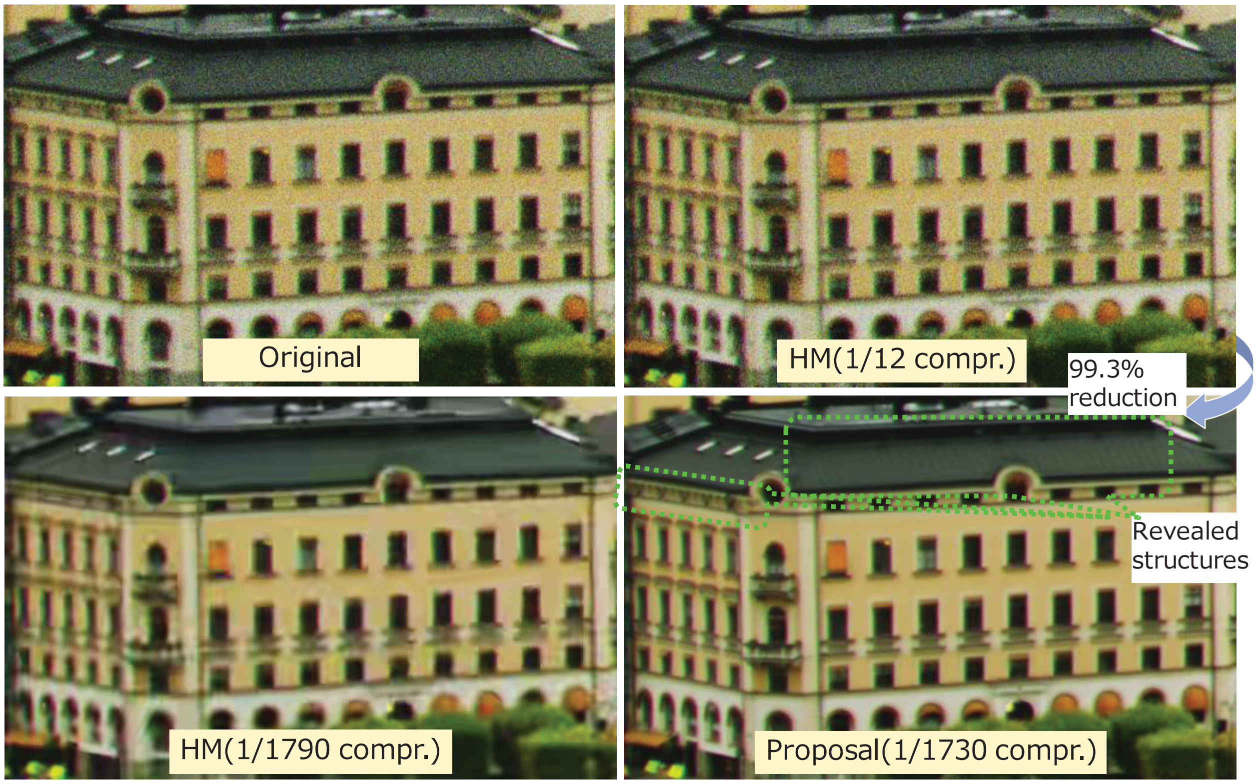

Fig. 4 Rigid object tracking-based real-entity-oriented video coding vs. conventional pixel-value-oriented video coding

2 Real-Entity-Oriented Coding Results

2.1 With rigid object tracking

With use of real-entity-oriented coding, video frames of rigid objects are registered using precise deformation and matching to successfully obtain denoised “real-entity image” and encoded and reproduced [11]. The results are shown in Fig. 4. Top-left picture is from original (uncompressed) video, bottom-left-bottom picture is from 1:1790 compressed video using HM. Bottom-right picture is from similar bit-rate, 1:1730 compressed video using proposed method. The visual quality is even better than the original, not to mention HM. And surprisingly, hidden textures (in dotted green boxes) are revealed in the proposal. Top-right picture is from high bit-rate (1:12 compression) video using HM. It is clearly observed that bits for coding noise wastes too much bits.

2.2 With water deskewing

Another target of the application of real-entity-oriented coding is shown in Fig. 5. The moving water heavily distorts the shapes of the objects under the water. By estimating the objects shape as if the water is not moving and assigning that image as additional reference frame (ARF), the coding efficiency greatly improves [12]. Fig. 6 shows the visual results of temporal median filter and proposed ARF, as well as original and still-water images. Median filter can usually eliminate unnecessary foreground objects but is quite blurred (coding efficiency is so low as expected). Our proposal looks crisp and similar to still-water image, and has undistorted shapes (cf. circular ring area and white lines in kiban1). Table 1 shows the comparison of coding efficiencies in BD-Rate [13] compared to VTM4.0 according to the ARF types, i.e., still-water, median filter, and proposal. Still-water gives the best performance, and our proposal the second with 39% bit-rate savings in average of four sequences. (Please be noted that still-water image cannot be used in real situations.)

Table 1 Water-bottom video coding gains vs. VTM4.0

ARF	Sequence	Y BD-rate
still-water (reference)	aquafish	-14.08%
	dish1	-48.19%
	kiban1	-53.36%
	pebble	-53.78%
	average	**-42.35%**
median	aquafish	5.76%
	dish1	13.77%
	kiban1	20.20%
	pebble	12.17%
	average	**12.98%**
Proposal	aquafish	-13.20%
	dish1	-45.91%
	kiban1	-47.51%
	pebble	-47.55%
	average	**-38.54%**

(a) (b) (c)

Fig. 5 Example of water-bottom video. (a) and (b): consecutive frames with moving water, (c): with still-water

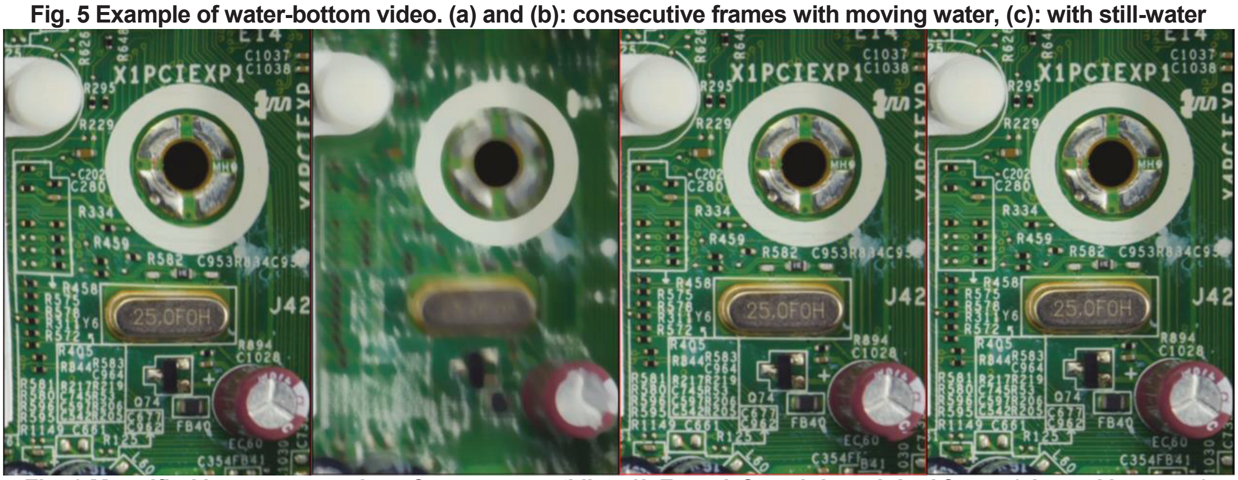

Fig. 6 Magnified image comparison for sequence 'kiban1'. From left to right, original frame (skewed by water), median filter (non-skewed but blurry), proposed Additional Reference Frame (ARF), and still-water (for reference)

3 CONCLUSIONS

Real-entity-oriented approach presented in this paper is still in early-stage but a steady departure from conventional pixel-oriented technology, and will change the video quality metric because the reproduced video has higher quality than the original. The impact of the new approach is promising. It will help reduce the amount of huge video traffic with several digits of order, which supports the network services as a whole and also help incubating totally new video services over the network.

REFERENCES

[1] ISO/IEC 11172-2:1993, "Information technology -- Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s -- Part 2: Video", (last reviewed and confirmed in 2015)

[2] ISO/IEC 13818-2:2013, "Information technology -- Generic coding of moving pictures and associated audio information: Video" (2013)

[3] ISO/IEC 14496-10:2014 "Information technology -- Coding of audio-visual objects -- Part 10: Advanced Video Coding" (2014)

[4] ISO/IEC 23008-2:2015 "Information technology -- High efficiency coding and media delivery in heterogeneous environments -- Part 2: High efficiency video coding" (2015)

[5] ISO/IEC 23090-3 "Information technology -- Coded representation of immersive media -- Part 3: Versatile Video Coding," to be published in 2020

[6] Cisco Visual Networking Index: Forecast and Trends, 2017–2022 White Paper, https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html

[7] S. Takamura, "Versatile Video Coding: a Next-generation Video Coding Standard", Global Standardization Activities, NTT Technical Review (2019)

[8] Joint Video Exploration Team, JVET JEM Reference Software, https://hevc.hhi.fraunhofer.de/trac/jem [online]

[9] Joint Video Experts Team, JVET VTM Software, https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/ [online]

[10] S. Takamura, A. Shimizu: "Simple and efficient H.265/HEVC coding of fixed camera videos," Proc. ICIP 2016, pp. 804–808 (2016)

[11] S. Takamura, A. Shimizu, "Efficient Video Coding Using Rigid Object Tracking", APSIPA ASC 2017, FP-04.5 (2017)

[12] S. Takamura, A. Shimizu, "Water-Bottom Video Coding Based on Coding-Oriented Reference Frame Generation", VCIP 2019 (2019, to appear)

[13] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," ITU-T VCEG-M33 (2001)