Special Paper
방송공학회논문지 제28권 제7호, 2023년 12월 (JBE Vol. 28, No. 7, December 2023)
https://doi.org/10.5909/JBE.2023.28.7.911
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

Performance Improvement and Modification of VVC RPR Technology for Machine Vision Systems

Ayoung Kim[a)], Eun-Vin An[a)], Soon-heung Jung[b)], Won-Sik Cheong[b)], Hyon-Gon Choo[b)], and Kwang-deok Seo[a)‡]

Abstract

Video coding technology for machines aims to enhance the compression ratio while maintaining machine inference performance. Meanwhile, Reference Picture Resampling (RPR) of Versatile Video Coding (VVC) could save the bitrate by downscaling the spatial resolution of the reference picture based on PSNR. In this paper, we propose a machine-vision-based RPR by modifying RPR from a machine-vision perspective that could reduce resulting bitrate while maintaining machine inference performance. By employing the proposed method, BD-rate reduction could be achieved by -14.48%, and BD-mAP could be improved by 3.46%.

Keywords: VCM, RPR, Spatial resampling, Machine-vision, MPEG

I. Introduction

Recently, as the proportion of media consumption by machines has increased, the interest in video coding technology for machines has grown. Accordingly, the importance of video compression technology for a machine vision systems is emerging, and research is needed to increase the bitrate compression ratio while maintaining machine inference performance. However, it is challenging to develop video compression technology considering the specific video features required for machine tasks because it is hard to explore the features. At this time, RPR, which could save bitrate by reducing the spatial resolution of reference pictures while maintaining the video features, could be a good spatial resampling technology for video coding for machines. In this paper, we propose a modified RPR to compress the video for machine inference task. The conventional RPR determines the resolution of reference pictures based on PNSR values considering a human vision

a) Division of Software, Yonsei University
b) Electronics and Telecommunications Research Institute (ETRI)
‡ Corresponding Author : Kwang-deok Seo
E-mail: kdseo@yonsei.ac.kr
Tel: +82-33-760-2788
ORCID: https://orcid.org/0000-0001-5823-2857

※ This research was supported by the Institute of Information and Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2020-0-00011, Video Coding for Machine), the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2021R1F1A1048404), and the MIST(Ministry of Science, ICT), Korea, under the National Program for Excellence in SW, supervised by the IITP in 2023 (2019-0-01219).
· Manuscript Received November 1, 2023; Revised November 16, 2023; Accepted November 16, 2023

system. On the other hand, the proposed method determines the resolution of the reference picture based on machine task performance. Furthermore, we evaluated the proposed method for compression efficiency against the performance of the machine inference task.

Ⅱ. Review of related works

1. Video Coding for Machines

ISO/IEC JTC1/SC29 Moving Picture Experts Group (MPEG) established the Video Coding for Machines(VCM) AhG at the 127th meeting in July 2019. VCM aims to compress video without hindering the performance of machine inference tasks such as the object detection and the object tracking[1,2]. There are two tracks for VCM: 1) Video coding track, 2) Feature coding track. For video coding track, a Call-for-Evidence(CfE) on VCM was released at the meeting in April 2021[3], and a Call-for-Proposal(CfP) was released in April 2022[4]. Also, the CfE for the feature coding track was published in July 2022[5]. VCM could be applied to various areas such as surveillance, intelligent transportation, smart city, intelligent industry, and intelligent content. In this scenario, a sheer of video data is not only acquired from machines but also mainly consumed by the machines. The conventional video coding technology developed in terms of the human vision system aims to compress the bitrate while providing the best quality of video. It could mean that the compression may not be optimized from a machine vision perspective. VCM compresses the video from the perspective of the machine vision system. To support machine inference tasks, VCM could receive various types of input and output such as video, descriptors, and features. Figure 1 describes the examples of possible VCM architecture[6]. Figure 1(a) shows the video coding architecture in which both machines and humans can consume the bitstreams generated by VCM, and the architecture for feature coding is illustrated in Figure 1(b).

2. Reference Picture Resampling in VVC

In High-Efficiency Video Coding (HEVC), the spatial resolution in a sequence is only able to be changed at an Instantaneous Decoder Refresh (IDR) or equivalent. Meanwhile, VVC introduces RPR which makes it possible to reference the different spatial resolution frames in the decoding process[7]. Accordingly, the change of spatial resolution could occur in inter-picture prediction without

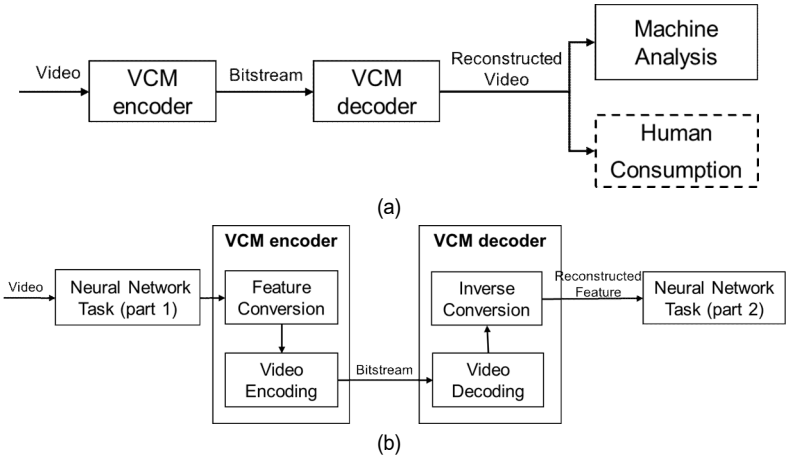

Fig. 1. Examples of possible VCM architectures

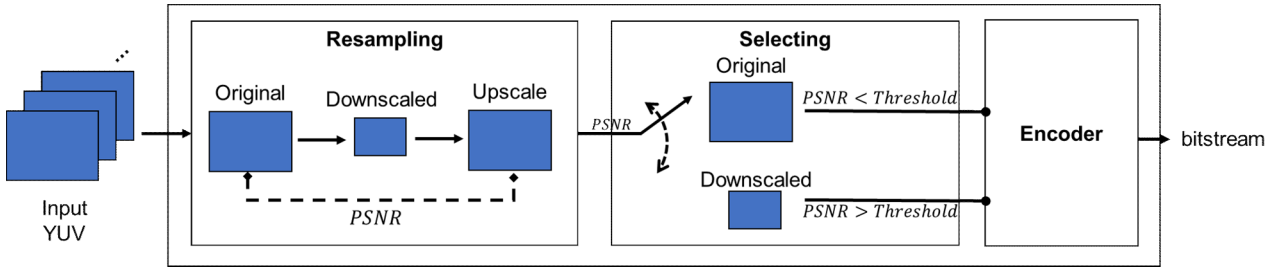

Fig. 2. Flowchart of the conventional RPR

inserting the additional IDR pictures. Figure 2 depicts the flowchart of the RPR introduced in VVC[8]. If the PSNR calculated between the original frame and the upscaled frame after downscaling exceeds a threshold, the downscaled frame is applied instead of the original frame. As a result, it could be guaranteed that there will be no significant differences in picture quality in human vision even if the reconstructed video comprises frames of different resolutions. A detailed description of RPR could be referred in[7].

Ⅲ. Machine-vision-based reference picture resampling (MV-RPR)

As mentioned above, RPR makes it possible to construct the video sequence with various resolution frames without compromising the quality of the video. However, it is needed to develop an improved RPR for machine inference tasks because the conventional RPR has been developed based on the human vision. In this paper, we have modified RPR to determine that the reference picture is downscaled based on machine inference tasks. Unlike the conventional RPR determining whether to apply it or not based on PSNR, machine-vision-based RPR (MV-RPR) is performed using the scale list generated by a threshold to keep the machine inference task performance.

Figure 3 illustrates the simple block diagram of the proposed MV-RPR. MV-RPR receives the YUV video with the optimal scale factor list as input. The optimal scale factor list contains scale factors for each frame. Consequently, VVC with MV-RPR generates the bitstream consisting of different spatial resolutions, and the bitrate could be saved without compromising machine inference performance as the MV-RPR is used.

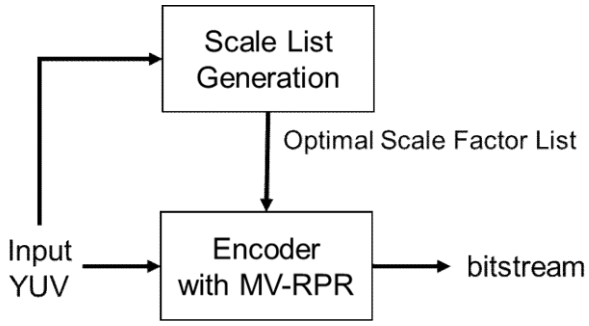

Fig. 3. Simple block diagram of the proposed MV-RPR

To confirm that performance is improved by adaptive resampling at the frame level using MV-RPR, the optimal scale factor list is generated through a simple algorithm, as shown in Figure 4. The input frame is downscaled according to each scale factor and upscaled again. Objects are detected for the original frame and upscaled frame, respectively. Then, confidence scores (hereinafter referred to as CS) for each object is utilized to determine the optimal scale factor. For each frame, the CS of the detected objects are averaged. And the difference in the average CS is calculated between the original frame and the upscaled frames for each scale factor, respectively. If the difference is less than a threshold, the scale factor is regarded as a

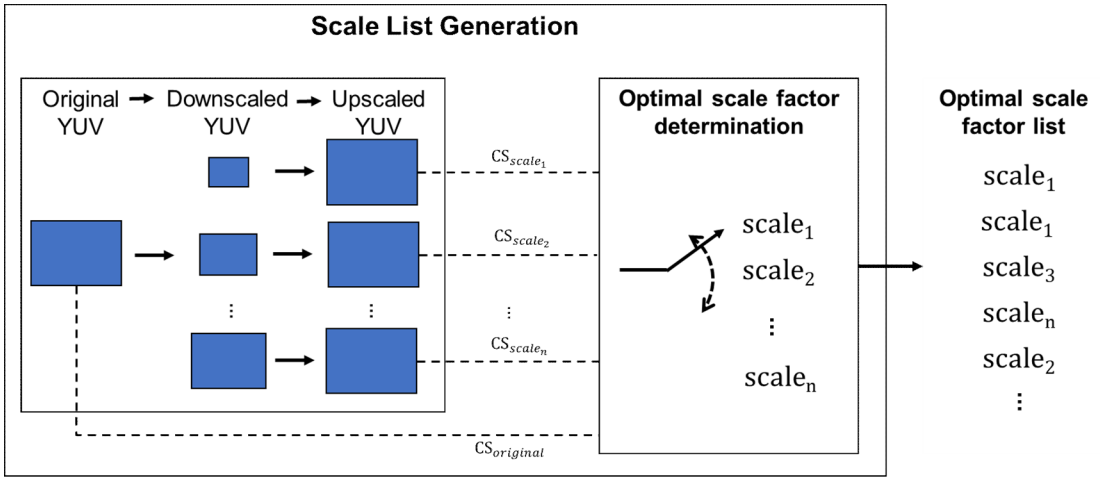

Fig. 4. Flowchart of the MV-RPR with simple scale list generation algorithm

candidate for the optimal scale factor. A minimum scale factor is determined as the optimal scale factor among the candidate scale factors.

Ⅳ. Experimental results

We conducted experiments based on the VCM evaluation pipeline provided in the CTC document of VCM and VCM-RS v0.4 is used as an anchor[9]. The experiments were conducted in All-Intra configuration, and the SFU-HW-Objects dataset was used[10]. The scale factor was set to 90%, 70%, 50%, and 30%, and the threshold for the simple list generation algorithm was set to 0.05 considering that the confidence score range is between 0 and 1. BD-rate was used as an evaluation metric to check the amount of

Table 1. Experimental results of All-Intra configuration for SFU dataset

Class	Sequence	BD-rate	BD-mAP
Class A	Traffic_2560x1600_30_val	-9.06%	0.91
Class B	ParkScene_1920x1080_24_val	8.00%	-1.72
	Cactus_1920x1080_50_val	-46.95%	22.97
	BasketballDrive_1920x1080_50_val	-4.71%	0.81
	BQTerrace_1920x1080_60_val	-46.16%	8.10
Class C	BasketballDrill_832x480_50_val	-10.18%	0.94
	BQMall_832x480_60_val	-5.05%	0.57
	PartyScene_832x480_50_val	-38.06%	8.75
	RaceHorses_832x480_30_val	-6.68%	0.89
Class D	BasketballPass_416x240_50_val	-2.16%	0.12
	BQSquare_416x240_60_val	-13.94%	1.02
	BlowingBubbles_416x240_50_val	-39.98%	6.53
	RaceHorses_416x240_30_val	26.71%	-4.96
Average		-14.48%	3.46

bitrate reduction compared to the same mAP. In other words, BD-mAP to check the change in mAP compared to the same bitrate was also used as the evaluation metric.

Table 1 describes the experimental results. The average BD-rates are obtained as -9.06%, -22.35%, -14.99%, and -7.34% for classes A through D, respectively. And the BD-mAPs are improved by 0.91%, 7.54%, 2.79%, and 0.68%, respectively. The results show that the proposed method could achieve significant bitrate savings without compromising mAP performance.

V. Conclusion

In this paper, we propose a modified RPR to encode the video sequence for using machine vision systems, and evaluate its performance. The conventional RPR is designed to not interfere with the human vision system. On the other hand, the proposed method is redesigned to consider the machine inference performance, thus can be the foundation for employing RPR for the machine vision system. Throughout the experiments, it is confirmed that the performance improvement in terms of BD-rate and BD-mAP could be achieved. As a result, RPR could be employed as a competent coding technology for machine vision systems. In our further research, we plan to study on the automatic scale list generation algorithm that derives the optimal downscale factor by considering video characteristics.

References

[1] H. Kwon, S. Cheong, J. Choi, T. Lee, and J. Seo, Standardization trends in video coding for machines, Electronics and Telecommunications Trends, 35 (2020), 102-111.

[2] L. Duan, J. Liu, W. Yang, T. Huang, and W. Gao, Video coding for machines: A paradigm of collaborative compression and intelligent analytics, IEEE Trans. on Image Processing, 29, 8680-8695, 2020.

[3] Call for Evidence for Video Coding for Machines, ISO/IEC JTC1/SC29/WG2 N42, Jan. 2021.

[4] Call for Proposals for Video Coding for Machines, ISO/IEC JTC 1/SC 29/WG 2 N191, Apr. 2022.

[5] Call for Evidence on Video Coding for Machines, ISO/IEC JTC 1/SC29/WG2 N215, Jul. 2022.

[6] Use cases and Requirements for Video Coding for Machines, N00190, ISO/ IEC JTC1/SC29/WG2, Apr. 2022.

[7] B. Bross, Y. Wang, Y. Ye, S. Liu, J. Chen, G. J. Sullivan, and J. R. Ohm, Overview of the versatile video coding (VVC) standard and its applications, IEEE Trans. on Circuits and Systems for Video Technology, 31, 3736-3764, 2021.

[8] K. Andersson, J. Ström, R. Yu, P. Wennersten, and W. Ahmad, GOP-based RPR encoder control, JVET-AB0080, Joint Video Exerts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, Oct. 2022.

[9] S. Liu, H. Zhang, and C. Rosewarne, Common test conditions for video coding for machines, N311, ISO/ IEC JTC1/SC29/WG4, Feb. 2023.

[10] H. Choi, E. Hosseini, S. Ranjbar Alvar, R. Cohen, and I. Bajić, SFU-HW-Objects-v1: Object labelled dataset on raw video sequences, 2020.
doi: https://doi.org/10.17632/hwm673bv4m

Introduction Authors

Ayoung Kim

- Feb. 2016 : B.S. degree, Division of Computer and Telecommunications, Yonsei University
- Mar. 2016 ~ currently : Ph.D. Candidate, Division of Software, Yonsei University
- ORCID : https://orcid.org/0000-0002-3793-1365
- Research interests : Video coding for machine, immersive media, multimedia communication systems

Eun-Vin An

- Aug. 2016 : B.S. degree, Division of Computer and Telecommunications, Yonsei University
- Mar. 2017 ~ currently : Ph.D. Candidate, Division of Software, Yonsei University
- ORCID : https://orcid.org/0000-0001-7681-4682
- Research interests : Video coding for machine, immersive media, multimedia communication systems

Soon-heung Jung

- Feb. 2001 : B.S., Electronics, Pusan National University
- Feb. 2003 : M.S., Electrical Engineering, KAIST
- Feb. 2016 : Ph.D., Electrical Engineering, KAIST
- Mar. 2003 ~ Mar. 2005 : Assistant Researcher, LG Electronics
- Aug. 2019 ~ Aug. 2020 : Visiting Scholar, Indiana University Bloomington
- Apr. 2005 ~ currently : Principal Researcher, ETRI
- ORCID : https://orcid.org/0000-0003-2041-5222
- Research interests : Immersive media, computer vision, video coding, realistic broadcasting system

Won-Sik Cheong

- Feb. 1992 : B.S., Department of Electronic and Electrical Engineering, Kyungpook National University
- Feb. 1994 : M.S., Department of Electronic and Electrical Engineering, Kyungpook National University
- Feb. 2000 : Ph.D., Department of Electronic and Electrical Engineering, Kyungpook National University
- May 2000 ~ currently : Principal member of research staff, ETRI
- ORCID : http://orcid.org/0000-0001-5430-29697
- Research interests : 3DTV broadcasting system, light field imaging, video and image coding, video coding for machines and deep learning based signal processing.

Hyon-Gon Choo

- Feb. 1998 : B.S., Department of Electronic engineering, Hanyang University
- Feb. 2000 : M.S., Department of Electronic engineering, Hanyang University
- Feb. 2005 : Ph.D., Department of Electronic communication engineering, Hanyang University
- Feb. 2005 ~ currently : Principal Researcher, Electronics and Telecommunications Research Institute (ETRI)
- Jan. 2015 ~ Feb. 2017 : Director of the Digital Holography Section, ETRI
- Sep. 2017 ~ Aug. 2018 : Visiting Researcher, Warsaw University of Technology, Poland
- Feb. 2023 ~ currently :Director of the Digital Holography Section, ETRI
- ORCID : https://orcid.org/0000-0002-0742-5429
- Research interests : video coding for machines, holography, multimedia protection and 3D broadcasting technologies.

Introduction Authors

Kwang-deok Seo

- Feb. 1996 : B.S., Department of Electrical Engineering, KAIST
- Feb. 1998 : M.S., Department of Electrical Engineering, KAIST
- Aug. 2002 : Ph.D., Department of Electrical Engineering, KAIST
- Aug. 2002 ~ Feb. 2005 : Senior research engineer, LG Electronics
- Sep. 2012 ~ Aug. 2013 : Courtesy Professor, Univ. of Florida, USA
- Mar. 2005 ~ currently : Professor, Yonsei University
- ORCID : http://orcid.org/0000-0001-5823-2857
- Research interests : Video coding, Visual communication, digital broadcasting, multimedia communication system