A person detection in HEVC bitstream domain based on bits density feature and YOLOv3 framework

Wahyu Wiratama, Donggyu Sim

Kwangwoon University

wiratama@kw.ac.kr, dgsim@kw.ac.kr

Abstract

This paper proposes an algorithm to detect persons in bitstream domain by skipping a reconstruction picture process in HEVC decoding. A new 3-channel feature extraction map is introduced in this paper by modelling the relationship between bits per CU density, average PU shape in CU, and total transform coefficients in CU from syntax elements. A state-of-the-art of YOLOv3 detection algorithm is used to detect and localize person on extracted feature maps. Based on the experimental results, the proposed person detection framework can achieve mAP of 0.68 and be able to find persons on feature maps. In addition, the proposed person detection can save decoding time about 60% by removing reconstruction picture process.

1. Introduction

Person detection in images is a hot topic in computer vision research filed, with application in surveillance system. Recently, various deep learning-based techniques have been proposed for person detection by achieving high detection accuracy [1][2]. However, high computational resources still become a problem, especially on real-time application. In addition, when a person in image is detected, it may be further processed for face recognition which make many people get less privacy protection. Therefore, a person detection framework with less computation complexity is still researched while keeps preserving the privacy.

In the real-practical application, an image/video is reconstructed from a bitstream through a decoding process. High Efficiency Video Coding (HEVC) is the latest standard of video codec and widely used in many applications. In the decoder side, a bitstream is decompressed using an HEVC entropy decoding to result syntax elements information. Based on those syntax elements, a picture reconstruction process is performed to generate a full reconstructed image. If a person can be found in bitstream domain by only extracting features from syntax elements information, it should be a benefit by avoiding full process, removing the reconstruction picture process. It is expected that can save about 60% [3] of full HEVC decoding process without disturbing privacy issue.

In these days, a study about face detection in bitstream domain has been proposed [4]. They proposed a 3-channel feature extraction based on intra prediction mode (IPM), prediction unit size (PU), and bin number (BN). Then, the extracted feature maps are fed into a deep learning algorithm You Only Look Once (YOLOv1) to find a face on FDDB dataset. However, this dataset contains the cropped images which cover only large persons. This is contradicting with real application of surveillance system which has multiscale of objects such as small persons as well. Therefore, this paper investigates a person detection on INRIA dataset [5] which has multiscale objects. In addition, another feature extraction approach is also introduced in this study by modelling the relationship between bits per coding unit (CU) density, prediction unit size (PUS), and transform coefficients. For detection algorithm, the successor of YOLOv1 is used, namely YOLOv3 [6].

This paper is organized as follows. In Section 2, the proposed algorithm will be presented. In Section 3, we evaluate the proposed algorithm. Finally, we summarize and conclude the papers in Section 4.

2. The proposed method

As aforementioned, HEVC decodes a bitstream through a decompressed process to produce syntax elements. Then, reconstruction picture processes such as prediction, dequantization, invers transforms, and in-loop filter are employed based on those syntax elements to generate a reconstructed image. Full process is performed in quad-tree block coding structure with ability to recursively partition a block into various coding units such as coding unit (CU), prediction unit (PU), and transform unit (TU). In the conventional person detection, the reconstructed images are fed into a detection algorithm to find the persons. Unlike conventional person detection, this study will skip the reconstruction picture process and extract feature maps based on syntax elements information to accelerate computation time.

In feature extraction process, bits density, average prediction unit size in CU (PUS), and total transform coefficients (TC) are captured during HEVC entropy decoding. 3-channel feature map (F) is constructed by the following equation

$$F_{1,i} = \frac{Totalbits_i}{CU\ size_i} \tag{1}$$

$$F_{2,i} = (1 + PUS) * \left(\frac{Totalbits_i}{CU\ size_i}\right) \tag{2}$$

$$F_{3,i} = \frac{Totalbits_i}{(TC_i + CU\ size_i)} \tag{3}$$

where each feature value is reported once per i-th CU. The visualization of feature extraction can be seen in Figure 1.

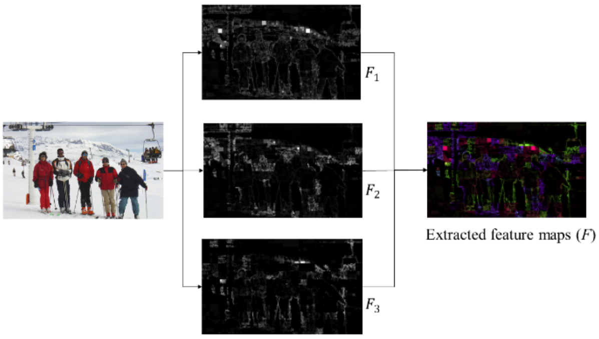

Figure 1. Extracted feature map (F) visualization.

Then, the extracted feature map is fed into a detection algorithm to find and localize the persons. This paper uses state-of-the-art of YOLOv3 as a successor of YOLOv1. It can localize the persons with bounding box detection in single pass.

3. Experimental results

In this study uses INRIA dataset [5] which contains positives and negatives images and its annotations for train and test. Firstly, all images are encoded with HEVC reference software HM 16.19 [7] under "All Intra Configuration" and QP 22. All bitstreams are then decompressed and extracted to the feature maps called "INRIA feature maps dataset". With this dataset, we use YOLOv3 pytorch implementation [8] and conduct training with configuration epochs, batch, momentum, decay, and learning rate of 272, 16, 0.9, 0.0005, and 0.001, respectively.

According to experimental result, we found that the proposed person detection can achieve the mAP of 0.68, as shown in Figure 2. The proposed can properly detect and localize persons even in larger environment, as shown in Figure 3. However, Figure 4 shows that the proposed person detection still found miss detection in several conditions and smaller objects. It is because the proposed feature extraction still cannot capture the feature clearly, especially for smaller objects.

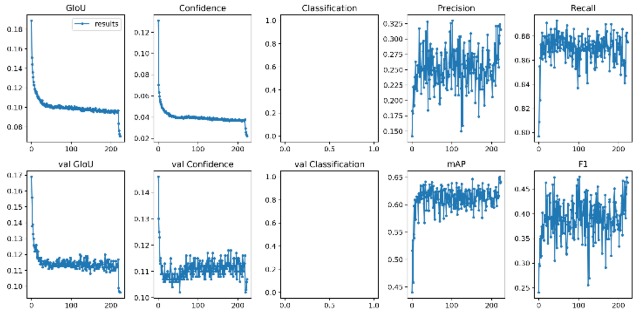

Figure 2. The performance graphs.

4. Conclusion

The proposed algorithm attempts to reduce the complexity of person detection while keeps preserving privacy issue by skipping reconstruction picture process in HEVC decoding. A new 3-channel feature maps extraction is introduced by modelling the relationship between bits density, average PU shape in Cu, and total transform coefficients in CU. A YOLOv3 is used as detection algorithm to find and localize the persons in the extracted feature maps. According to experimental results, the proposed person detection framework can yield mAP of 0.68 and be able to find the persons.

Acknowledgement

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2019-2016-0-00288) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation) and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning(NRF-2018R1A2B2008238).

References

[1] Liu, Z., Chen, Z., Li, Z. and Hu, W., "An Efficient Pedestrian Detection Method Based on YOLOv2". Mathematical Problems in Engineering, 2018.

[2] Bali, S. and Tyagi, S.S.," A Review of Vision-Based Pedestrian Detection Techniques". International Journal of Advanced Studies of Scientific Research, vol. 3, no.9, 2018.

[3] Alvar, S.R., Choi, H. and Bajic, I.V., "Can you tell a face from a HEVC bitstream?". In IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), pp. 257-261. 2018.

[4] Alvar, S.R., Choi, H. and Bajic, I.V., "April. Can you find a face in a HEVC bitstream?". In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1288-1292, 2018.

[5] Dalal, N., & Triggs, B., "Histograms of oriented gradients or human detection". In Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 886-893, 2005.

[6] Redmon, J. and Farhadi, A.," Yolov3: An incremental improvement". arXiv preprint arXiv:1804.02767, 2018.

[7] "HEVC reference software (HM 16.19)," https://hevc.hhi.fraunhofer.de/trac/hevc/browser/tags/HM-16.19, Accessed: 2019-08-27.

[8] YOLOv3 Pytorch implementation, "https://github.com/ultralytics/yolov3", Accessed: 2019-08-27.

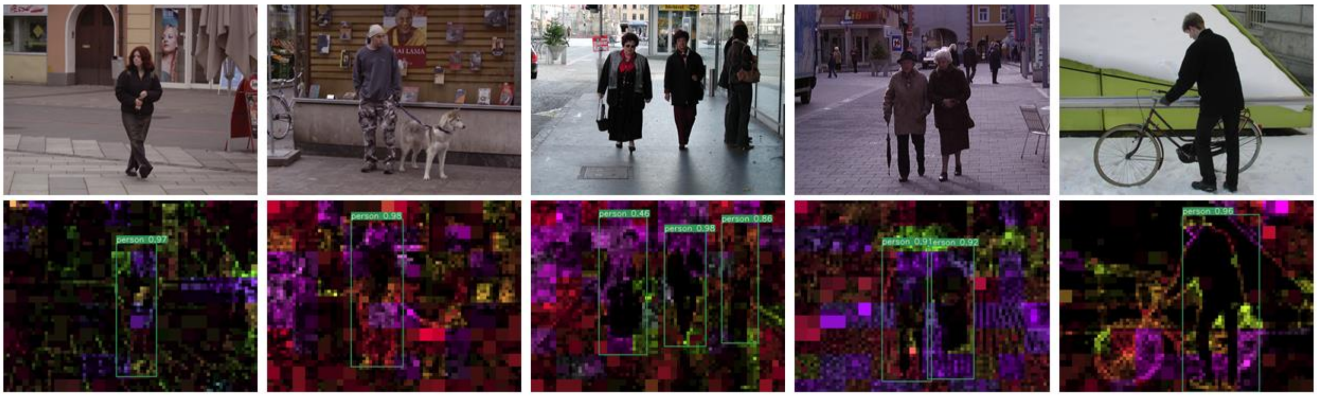

Figure 3. Detection and localization result on the proposed feature maps (top row is original image and bottom row is detection result on extracted feature map).

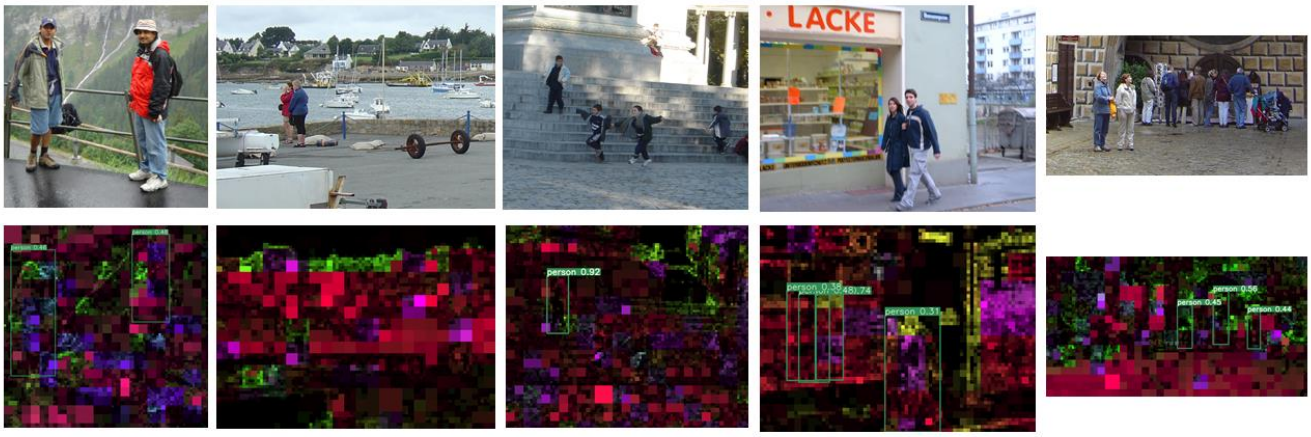

Figure 4. False alarm detection and localization result on the proposed feature maps (top row is original image and bottom row is detection result on extracted feature map).