HEVC Coding Unit Mode Based Motion Frame Analysis

Qiong Jia   Tianyu Dong   Euee S. Jang
Hanyang University

Abstract

In this paper we propose a method predict whether a video frame contains motion according to the invoking situation of the coding unit mode in HEVC. The motion prediction of video frames is conducive for use in video compression and video data extraction. In the existing technology, motion prediction is usually performed by high complexity computer vision technology. However, we proposed to analyze the motion frame based on HEVC coding unit mode which does not need to use the static background frame. And the prediction accuracy rate of motion frame analysis by our method has exceeded 80%.

1. Introduction

With the use and spread of video files, video compression and video data extraction have become a hot spot that people are more concerned about and used everywhere. However, the related research and application often only need the part containing motion. And existing motion frame analysis technologies are complex and cost a long time.

Some video data extraction technologies based on motion have been proposed and applied in recent years. Video synopsis is one of them which is used widely now. It is a technology that can realize effective video browsing and video condensation by retaining motion information and deleting unnecessary video data[1]. But in this technology, the motion frame analysis is usually by constructing a background model and comparing it with other video frames, which will generate a lot of calculations in the process. So we need an easier and quicker method to analyze the motion frame.

In HEVC encoder, a video can be treated as a sequence of pictures, where these pictures are encoded as I-(intra), P-(predicted) and B-(bi-directionally predicted) slices and the prediction type is determined according to the changes between pictures. Each slice consists of a certain number of CTU (coding tree units), while each CTU consists of some number of CU (coding units) with various sizes[2]. When decoding different pictures the CU mode called by each slice is quite different. Most of the still pictures are P-slices, and most of the CU in the P-slice are in skip mode without MV. While when moving objects appear, the I mode CU will update that motion part of the picture. The syntax of CU mode is compressed at the beginning of each CU instead of context-based entropy, so we do not need to decompress the entire CU, just look at the syntax in the head. Therefore, our method has lower computational complexity than constructing a background model and comparing it with other video frames.

In this paper, we conduct a P-test to support that there is a positive relationship between the percentage of invoking CU modes and the motion frame. Then propose a new method to predict whether a video frame contains motion based on the invoking percentage of CU modes in HEVC. Through experiments, the prediction accuracy rate of motion frame analysis by our method has exceeded 80%.

The paper is organized as follows: in Section 2 we will introduce the proposed method. Then show the results and discussions in Section 3. At last, we will conclude the paper in Section 4.

2. Proposed Method

In this process, we apply training data and test sequences to verify the method. First, we separate frames without motion from the training data and use the part of frames to conduct a P-test to support the null hypothesis: there is a positive relationship between the percentage of invoking CU mode and the motion frame.

The p-value of the P-test is the probability of obtaining test results at least as extreme as the results actually observed, under the assumption that the null hypothesis is correct[3]. Because the percentage of invoking CU mode is always greater than or equal to 0 and can only go in one direction, we use a one-tail p-value. Using the average μ and the variance σ of the unknown distribution T, we consider an observed test-statistic t from it. Then the p-value p is what the prior probability would be of observing a test-statistic value at least as "extreme" as t if null hypothesis H_0 were true. And in our test, the p was always less than 0.05 which means our null hypothesis is correct.

The p-value is defined as:

$$p = \Pr(T \geq t \mid H_0) \qquad (1)$$

In (1), t is defined as:

$$t = \mu + 2\sigma \qquad (2)$$

On this basis, we begin to determine whether an outlier exists in all training data. We use two videos as the training data then take the average of their t as the threshold(T_1).

For higher accuracy, we add the sliding window variance to analyze the motion frame together. The sliding window is on the order of 10 frames, which means we will calculate the variance of the percentage of invoking CU mode for every 10 frames. Then we get the average of all variances and set it as the threshold(T_2).

If the percentage value of invoking CU mode is more than T_1 or T_2, we consider the current frame belonging to the motion frame. On the opposite, we consider it as a frame without motion.

3. Results and Discussion

We use HM (HEVC Test Model) Version 16.20 to extract the CU mode. And the training data Road_car_1 is 1920x2160 resolution, 60 fps and about 1 minute duration and Road_car_2 is 1440x1680 resolution, 60 fps and about 1 minute duration. The results of training data are shown in Table 1 and Table 2. The prediction accuracy rate of motion frames considered as the motion frame exceeded 85.93% and 99.60%.

Table 1 Result of training Road_car_1.

	Consider as motion frame	Consider as no-motion frame
Motion frame	85.93%	14.07%
No-motion frame	15.20%	84.80%

Table 2 Result of training Road_car_2.

	Consider as motion frame	Consider as no-motion frame
Motion frame	99.60%	0.40%
No-motion frame	14.56%	85.44%

The Fig. 1 is the 1892nd frame of Road_car_1 which is considered as no-motion frame and the Fig. 2 is the 2021st frame of Road_car_1 which is considered as motion frame by our method.

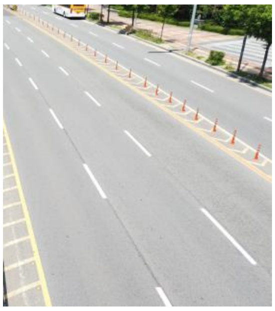

Figure 1. The 1892nd frame of Road_car_1.

Figure 2. The 2021st frame of Road_car_1.

Table 3 Result of test sequence - C0039.

	Consider as motion frame	Consider as no-motion frame
Motion frame	100.00%	0.00%
No-motion frame	38.97%	61.03%

Table 4 Result of test sequence - C0006.

	Consider as motion frame	Consider as no-motion frame
Motion frame	98.12%	1.88%
No-motion frame	26.81%	73.19%

On this basis, we conducted the noise analysis on the data of the three results and compiled the results of the noise analysis in Table 5.

Table 5 The noise analysis.

	N_c	N_m	N_f	RR	PR
Road_car_1	4279	237	484	0.948	0.898
Road_car_2	4672	13	24	0.997	0.995
C0039	906	0	32	1	0.732
C0006	1881	26	193	0.986	0.907

Note[1].

4. Conclusion

Overall, the experimental result shows that the prediction accuracy rate of motion frame analysis by our method has exceeded 80%. The result of motion frame analysis is in line with our expectations, while the analysis of non-motion frames by this method is currently not ideal. Therefore, no-motion frame analysis might prove an important area for future research.

ACKNOWLEDGMENT

This work was supported by the Technology development Program(S2877946) funded by the Ministry of SMEs and Startups(MSS, Korea)

Reference

[1]. Pritch, Yael, Alex Rav-Acha, and Shmuel Peleg. "Nonchronological video synopsis and indexing." IEEE transactions on pattern analysis and machine intelligence 30.11 (2008): 1971-1984.

[2]. Fan, Chenyou, Zehua Zhang, and David J. Crandall. "Deepdiary: Lifelogging image captioning and summarization." Journal of Visual Communication and Image Representation 55 (2018): 40-55.

[3]. Ware, James H., et al. "P values." Medical uses of statistics. CRC Press, 2019. 181-200.

[4]. Lee, Sinwook, and Euee-seon Jang. "Bitstream-Level Film Noise Cancellation for Damaged Video Playback." IEICE TRANSACTIONS on Information and Systems 97.3 (2014): 562-572.

[1] N_c: the correctly detected noise, N_m: the missed regions, N_f: the false alarmed regions, RR: recall rate, PR: Precision rate.[4]