IMPROVED ADVANCED MOTION VECTOR PREDICTION SCHEME FOR SURVEILLANCE VIDEO CODING

Yumei Zhang[⋆] *Siwei Dong*[†] *Yaowei Wang*[⋆] *Yonghong Tian*[†] *Peiyin Xing*[†]

[⋆] School of Information and Electronics
Beijing Institute of Technology, Beijing, China, 100081
Email: yumeizhang329@163.com, yaoweiwang@bit.edu.cn
[†] National Engineering Laboratory for Video Technology
School of EE & CS
Peking University, Beijing, China, 100871
Email: {swdong, yhtian, pyxing}@pku.edu.cn

ABSTRACT

Background frame based video coding achieves a remarkable compress performance for surveillance video coding. In this paper, an improved Advanced Motion Vector Prediction (AMVP) scheme for surveillance video coding is proposed to make Motion Vector Prediction (MVP) more accurate when prediction unit (PU) references background (BG)-frame. Considering the correlations among the spatial PUs, the proposed method utilizes the Motion Vectors (MV) of spatial PUs which reference BG-frame to recalculate MVP, and reconstruct the MVP candidate list when current PU references BG-frame. Results show that our scheme averagely achieves 0.17% bit-rate saved compared to HM12.0-S.

***Index Terms*—** Surveillance video coding; HEVC; rate control; background reference

1. INTRODUCTION

Nowadays, BG-frame based video coding method achieves remarkable compress performance on surveillance video coding. Zhang et al. in [1][2] proposed to generate the more accurate BG-frame by original input frames, which achieves a higher performance compared with other methods.

MVP mainly uses the high correlation among the adjacent PUs to reduce coded bits. In H.264/AVC, this is done by doing a component wise median of three spatially neighboring motion vectors, and temporal MVPs from a co-located picture are only considered in temporal direct mode of H.264/AVC. Then, the former approach is replaced by a technique known as motion vector competition in HEVC. MVP is well predicted and stantially improves the coding performance.

In this paper, utilizing the high correlation among the adjacent PUs, we propose an improved AMVP scheme for surveillance video coding when PUs reference BG-frame, as shown in Figure 1. Firstly, a background frame is modeled from original frames and then encoded as long-term reference. Secondly, record blocks which reference BG-frame (PU$_{bgi}$), especially its MVs and positions. Then, if current block references BG-frame, set a search range for it, and find the best texture matching blocks (i.e., PU$_{bg3}$) in recorded blocks to recalculate MV. Finally, adopt zero motion vector and the new calculated MV as final MVP candidates. Experiment results manifest that our method averagely achieves 0.17% bit-rate saving.

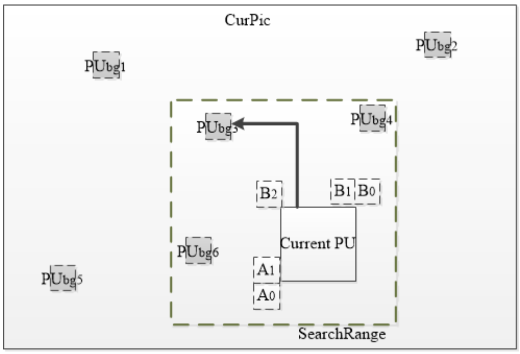

Fig. 1. The proposed improved AMVP scheme.

2. THE PROPOSED METHOD

To make more accurate MVP for blocks referencing BG-frame, the proposed method is described as follows: 1) A background frame is modeled from original frames and then encoded as long-term reference. 2) Record blocks which reference BG-frame while coding, especially its MVs, positions and pixel values. 3) If current block references BG-frame, set a search range for it, and find the best texture matching blocks in recorded blocks to recalculate MV. 4) Adopt zero motion vector and the new calculated MV as final MVP candidates. Details about the method will be referred in the following.

2.1. Background Modeling and Updating

The proposed method utilizes the low complexity Segment-and-Weight based Running Average algorithm[1] and S-GOP[2] as the background modeling and updating method.

2.2. Reconstruct MVP Candidate List

Algorithm 1 shows how to calculate the MVP though our method. After BG-frame is generated, if PU reference BG-frame and are unequal to zero motion vector, PU is divided into blocks size of 4×4. All blocks share the same MV of PU, but each block should be recorded once. Let us donate $Ax(p, q)$ and $Ay(p, q)$ as the abscissa and ordinate of the recorded motion vector, respectively.

If current PU reference BG-frame, divide it into blocks size of 4×4, too. (x, y) denotes the start point for current PU and (sx, sy)

Input : The current PU is in the size of $2N \times 2N$ at position (x, y) of the current frame I. A(x,y) is the recorded MVs in 4×4 block level.
Output: MVP of the current PU.
for $m = (0 \; to \; N/2 - 1)$ **do**
  for $n = (0 \; to \; N/2 - 1)$ **do**
    $sx = (x + 4 \times m)$;
    $sy = (y + 4 \times n)$;
    for $Cq = ((x - R)/4 \; to \; (x + R)/4)$ **do**
      for $p = ((x - R)/4 \; to \; (x + R)/4)$ **do**
        if $Ax(q, p)! = 0 \; or \; Ay(q, p)! = 0$ **then**
          $p = \sum_{i,j=1}^{4} |I(sx + i, sy + j) - I(4q + i, 4p + j)|$;
          if $s < Th$ **then**
            $Th = s$;
            $Mx = Ax(q, p)$;
            $My = Ay(q, p)$;
          end
          else
            $M = (0, 0)$;
          end
        end
        else
          $M = (0, 0)$;
        end
      end
    end
  end
end

$$MVP_x = (\sum_{k=0}^{N^2/4} Mx_k) \times 4/N^2;$$

$$MVP_y = (\sum_{k=0}^{N^2/4} My_k) \times 4/N^2;$$

Algorithm 1: Algorithm of the MVP.

is the start position for a 4×4 block in PU. R is a search range with (x, y) as the center. Besides, the corresponding MV for block which starts at (sx, sy) is recorded in Ax and Ay at position $(sx/4, sy/4)$, and the search range should be $R/4$ in Ax or Ay. So in the search range, we can find some values in Ax or Ay are unequal to zero vector. But not all of the values are useful and it is the key to find the best texture match blocks with the current block. I represents the current frame. Assuming that (q, p) in Ax or Ay are unequal to zero, obtain the absolute difference of pixels between blocks which start at $(4q, 4p)$ and at (sx, sy). Record the value as MV for this block when the absolute difference is less than the threshold ($Th = 80$) and is the minimum value. Otherwise, if the absolute difference value is larger than Th or all values in the search range equal to zero vector, set the 4×4 block zero MV. Moreover, add MV of every block in PU and get a new MVP.

For most of the PUs which reference BG-frame, MVs equal zero vector. So, we adopt zero motion vector as the first candidate and our new MVP as the second candidate, thus obtaining a new MVP candidate list. As for temporal candidate, current PU references BG-frame but the two co-located PUs in AMVP may reference other frames. Besides, the recorded spatial MVs are enough for prediction. So we do not take the temporal candidates into consideration when reconstruct the new MVP candidate list.

3. EXPERIMENTAL RESULTS

The experimental platform is HM12.0-S which adopts the long-term reference mechanism[2] based on background modeling method of SWRA in [1]. We utilize surveillance videos from the dataset of PKU-SVD-A [4] [5], including four 720×576 (sd) and three 1920×1080 (hd) videos, as shown in Figure 2. For the real-time transmission of surveillance videos, the experiment is conducted on


bank-sd overbridge-sd office-sd crossroad-sd
campus-sd lake-hd snow-hd

Fig. 2. The surveillance sequences of PKU-SVD-A.

Table 1. The comparison of coding performance

Sequence	Proposed vs HM12.0	
	BD-Rate	BD-PSNR
bank-sd	-0.35%	-0.003
overbridge-sd	-0.21%	0.004
office-sd	0.10%	-0.005
crossroad-sd	-0.15%	0.003
campus-sd	-0.36%	0.011
lake-hd	-0.12%	0.001
snow-hd	-0.10%	0.003
average	**-0.17%**	**0.002**

the low-delay configuration with common test conditions [3].

As shown in Table 1, our proposed scheme achieves an average bit-rate saving of 0.17% compared to HM12.0-S. Coding performance is related to the resolutions and complexity of videos.

4. CONCLUSION

In this paper, an improved AMVP scheme for surveillance video coding is proposed to take advantage of correlations among the spatial PUs which reference BG-frame. The proposed method takes the MVs of spatial PUs which reference BG-frame into consideration, and reconstruct the MVP candidate list when current PU references BG-frame. Results show that our scheme averagely achieves 0.17% bit-rate saved compared to HM12.0-S.

5. ACKNOWLEDGMENTS

This work is partially supported by the National Basic Research Program of China under grant 2015CB351806, the National Natural Science Foundation of China under contract No.61471042 and No.61390515.

6. REFERENCES

[1] X. Zhang, Y. Tian, T. Huang, and W. Gao, Low-complexity and high-efficiency background modeling for surveillance video coding. in VCIP, 2012, pp. 16.

[2] X. Zhang, L. Liang, Q. Huang, Y. Liu, T. Huang, and W. Gao, An efficient coding scheme for surveillance videos captured by stationary cameras, in Visual Communications and Image Processing 2010. Inter- national Society for Optics and Photonics, 2010, pp. 77442A77442A.

[3] JCT-VC, HM common test conditions and software reference configurations, in JCTVC-L1100, Jan.2013.

[4] X. Zhang, T. Huang, Y. Tian, and W. Gao, Background-modeling based adaptive prediction for surveillance video coding, 2014.

[5] W. Gao, Y. Tian, T. Huang, S. Ma, and X. Zhang, IEEE 1857 standard empowering smart video surveillance systems, 2013.