Segmentation and Classification of Real-Time Moving Object from HEVC Video Streams

S. Sunitha
Applied Electronics
Department of Electronics and Communication Engineering
Thanthai Periyar Government Institute of Technology
Vellore-632002

Prof R. Bharathiraja, M. E
Assistant Professor
Department of Electronics and Communication Engineering
Thanthai Periyar Government Institute of Technology
Vellore-632002

Abstract— Segmentation and classification of moving object from surveillance video plays an important role for intelligent video surveillance. Compared with H.264/AVC, HEVC introduces a new coding feature for moving object segmentation and classification. In this paper, I present a real-time approach to segment and classify moving object using unique features directly extracted from the HEVC compressed domain for surveillance video streams. Firstly, motion vector interpolation for intra-coded prediction unit and MV outlier removal are employed for preprocessing. Then, blocks with non-zero motion vectors are clustered into the connected regions using the four connectivity component labelling algorithm. Object region tracking is based on temporal consistency is applied to the connected foreground regions to remove the noise regions. Moving object region boundary can be further refined by the coding unit size and prediction unit size. Finally, a person-vehicle classification model using HEVC algorithm is formed to classify the moving objects, either persons or vehicles. The simulation results demonstrate that this method provides the solid performance and can classify moving persons and vehicles accurately with increased efficiency.

Keywords— *Labelling algorithm, Removes noise, HEVC algorithm, preprocessing of moving object, Increased efficiency.)*

I. INTRODUCTION

Segmentation and classification of moving object from   surveillance video plays an important role for intelligent video surveillance. Mostly computer vision methods for moving object detection and classification assume that the original video frames are available and extract T descriptions or features from pixel domain. Most video content are received or stored in international video coding standards, such as MPEG-2, H.264/AVC and HEVC. . In video analysis at large scales, such as content analysis and search for a large surveillance network, the complexity of video decoding becomes a major bottleneck of the real-time system. The major advantage of compression-domain approaches is their low computational complexity since the full-scale decoding and reconstruction of pixels are avoided. Therefore, compressed domain methods are desired for real- time video analysis applications. In this paper, we focus on moving object detection and classification from HEVC compressed surveillance videos. Therefore, by extracting features from HEVC compressed surveillance video bit stream, the moving objects are located and classified, such as persons or vehicles.

II. SYSTEM DESIGN

In this paper, to   develop a framework for moving object segmentation and classification by using the motion vectors and associated modes directly extracted from HEVC compressed video. Specifically, I focus on surveillance videos whose cameras are stationary. Compared to existing methods in the literature, it has the following unique aspects and innovations: (1) the unique features in the HEVC compressed domain, such as coding unit and prediction unit, are employed to refine the moving object boundary; (2) the bag of words representation in the HEVC compressed domain is applied to classify the moving persons and vehicles. The overall system design consists of two stes 1. It consists of two stages: moving object segmentation and person-vehicle classification. For moving object segmentation, firstly, MV interpolation for intra-coded prediction unit (PU) and MV outlier removal are employed for preprocessing. Then, blocks with non-zero motion vectors are clustered into the connected foreground regions by using the four-connectivity component labeling algorithm . Finally, object region tracking with temporal consistency is applied to the connected foreground regions to remove the noise regions. The boundary of moving object region is further refined by using the coding unit (CU) and PU sizes of the blocks.

III. PREPROCESSING

In HEVC compressed video, one MV is associated with an inter-coded prediction unit (PU). The motion vectors are scaled appropriately to make them independent of the frame type. This is accomplished by dividing the MVs according to the difference between the corresponding frame number and the reference frame number (in display order). For example, one MV has values (4,4) for reference frame -1 while another MV in a nearby block has values (8,8) for reference frame -2, these two MV values will be corrected to both be (4,4) after the scaling process. For the PU with two motion vectors, the motion vector with larger length will be selected as the representative motion vector of the PU. In the preprocessing process, the MV interpolation for intra-coded blocks and MV outlier removal are employed before the moving object segmentation and classification.

A. Motion Vector Interolation

In order to segment the foreground and background region, it is useful to assign a MV to an intra-coded PU. To be specific, the MVs of first-order neighboring PUs (top-left, top, top right, left, right, bottom-left, bottom, bottom-right) are employed. Fig. 2 shows an example of an intra-coded PU together with its first-order neighboring PUs. In Fig. 2, MVs of all neighboring PUs are collected and stored in *MVList*. Since one of the neighboring PUs is intracoded, *MVList* contains seven vectors

$$MVList= (MV1,MV2,MV3,MV4,MV5,MV6,MV7) \qquad (1)$$

After *MVList* is constructed, the next step is to assign a representative MV for the intra-coded PU from *MVList*. Intracoded PUs usually occur when there is a large motion in the scene. Therefore to choose the maximum MV from *MVList* as the MV of the current intra-coded PU. Specifically, when all the first-order neighboring PUs are encoded with intra mode, in order to obtain the non-zero MV within the nearby region, we extend the range of neighborhood to 16x16 (blocks), which is set empirically as being optimal. Fig.2 MV assignment for an intra-coded PU. One of the

first-order neighboring PUs is also intra-coded, and the remaining variably sized neighboring PUs have MVs.

B. MV Outlier Removal

The MVs from the compressed bitstream are determined in terms of rate-distortion, they may not represent the true object motion. In this section, to reduce the motion noise by referring to motion continuity over time and motion coherence within the spatial neighborhood. Three steps are included in the MV outlier removal, which are MV filtering, MV refining, and isolated and small MV removal.
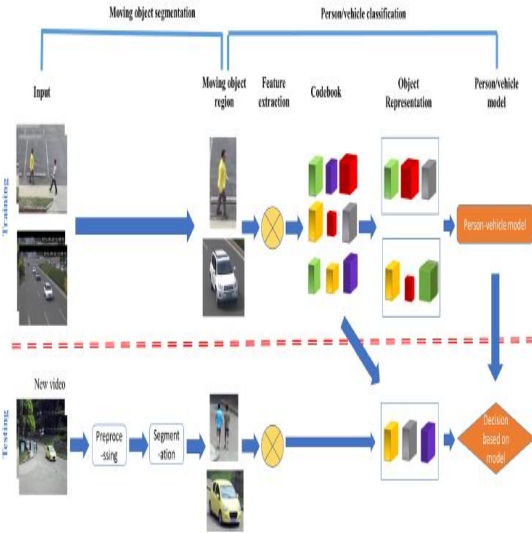
Fig.3.1 Framework for HEVC compressed domain.

1) MV Filtering

The original MVs are filtered along the temporal direction .To be specific, the original MVs at the co-located position in the m previous frames and m following frames are employed to filter the original MVs at current frame t. Since the CU and PU sizes at the same position may be different among different frames, MV filtering is operated at 4x4 block level, which is the minimum size of PU. Let and represent the original MVs along the horizontal direction and vertical direction for a 4x4 block at position at frame t.
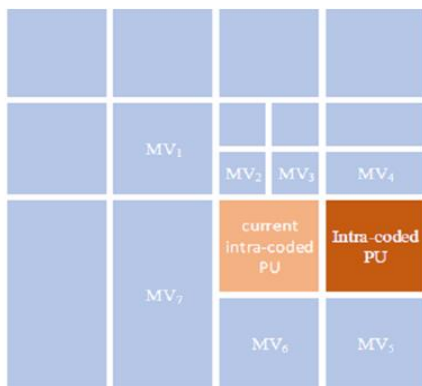

Fig.2 MV assignment for an intra-coded PU

2) MV Refining

Since the moving objects in the previous frames and following frames have a displacement relative to the moving object in current frame, the filtering process using co-located blocks in the neighboring frames may cause a few non-zero MV noise adjacent to the moving object boundary in current frame. Although the original MVs may be noisy, but if the original MV of current block and most of its

neighboring blocks are both zero, current block has a high probability to belong to background.

3) Isolated and Small MV Removal

For a foreground moving object, it usually has a connected non-zero MV region and a relatively larger filtered MV, so the PUs with isolated non-zero MVs or smaller MVs have a high probability to be the background PUs. Therefore, I propose to label the PUs with isolated non-zero MVs or small MVs as the background PUs. To be specific,I define one MV as an isolated MV when all the MVs of its spatial neighborhood are zero MV. In addition, we define one MV as a small MV when the MVs of current PU and more than half of its spatial neighboring PU are less than or equal to 1. If one PU is identified as the PU with isolated or small MV, its associated MV will be modified to zero.

IV. MOVING OBJECT SEGMENTATION IN HEVC

After the preprocessing of the MVs, as described in Section III, blocks with non-zero MVs are marked as foreground blocks. These foreground blocks are clustered to the connected foreground regions using the four-connectivity component labeling algorithm. For each foreground region, firstly, to examine its temporal consistency by using object region tracking. Secondly, to refine the boundary of moving object region by using CU and PU sizes of the blocks.
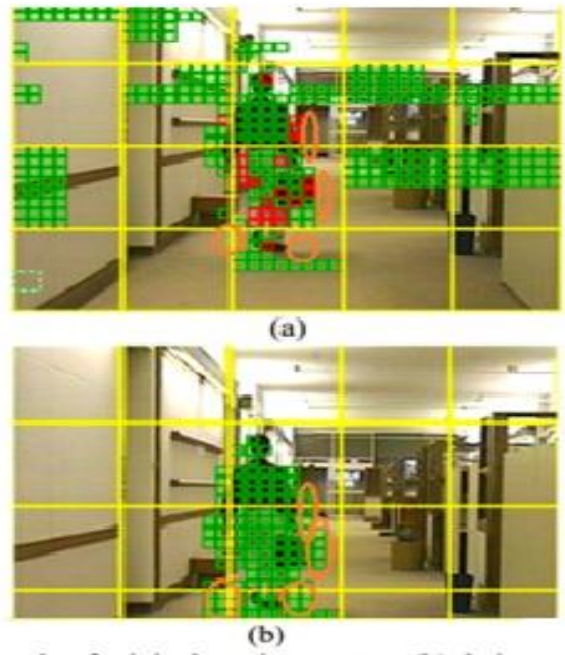

Fig.3(a) An example of original motion vector;(b) Associated filtered motion vectors.

A. Object Boundary Refinement

The below figure 3(a) and (b) show an example of block partitions for two surveillance video frames with moving persons and vehicles. Here, the largest square blocks with yellow border, smaller square blocks with green border and rectangular blocks with pink border represent coding tree unit (CTU), CU and PU respectively. It is observed that blocks within the moving person and vehicle region are encoded with smaller CU and PU sizes when compared to CU and PU sizes of the blocks within the background region.

B.Object Region Tracking

In order to examine the temporal consistency of foreground regions, these foreground regions are temporally tracked by using the MVs extracted from HEVC compressed domain. For the ith foreground region at frame t, if we find that its corresponding object region

continuously describes the same object in backward direction (from t to $t-4$) and in forward direction (from t to $t+4$), then we assume that current foreground region satisfies the condition of temporal consistency.
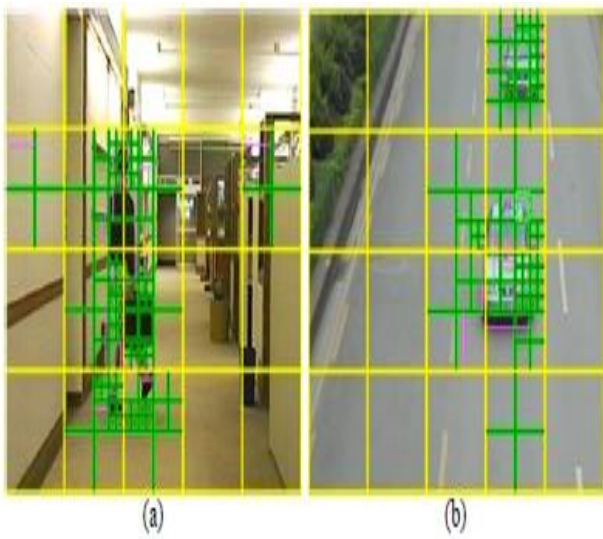

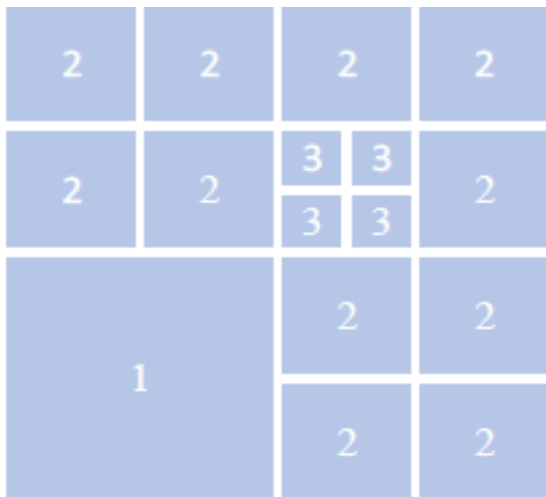Fig3.(b) Partitions of the moving person and vehicles.

Fig3.(c) Relationship between the depth level and CU size.

V. MOVING OBJECT CLASSIFICATION IN HEVC COMPRESSED DOMAIN

For object classification in surveillance videos, first classify the segmented moving objects into persons and vehicles using "bag of HEVC syntax words" in HEVC compressed domain. The "bag of words" representation has been successfully used for object classification in the pixel domain . R. V. Badu *et al.* propose to use "bag of words" representation in H.264/AVC compressed domain to classify the video content. The major contribution of this work is to establish a bag of words model in the HEVC domain for moving object classification. This proposed object classification has the following major steps: (1) describing each coding block within the moving object region using HEVC syntax features; (2) constructing a codebook using a clustering method; (3) representing each moving object using a normalized histogram of codeword from the codebook; and (4) training a binary classifier to classify the moving objects into

persons and vehicles. The main challenge is to select effective features in the HEVC compressed domain.

VI.SIMULATION AND RESULTS

In order to train the person-vehicle model for moving object classification, 4 training sequences are used, which are illustrated in Fig.4. To evaluate the performance of our proposed moving object segmentation and classification scheme in HEVC compressed domain, we have collected 2 sequences from CDNet2012 dataset (*Highway* and *Pedestrians*) HEVC syntax features, such as motion vectors, prediction modes, CU sizes, and PU types, are extracted from HEVC compressed bitstream. segmented when they are not close to each other whereas the moving person and vehicle will be segmented as one whole

Fig.4 Example frames f test videos from public dataset

object when they are close to each other. the segmentation accuracy is measured by comparing the segmented foreground and background blocks with the ground truth labels for each frame of the test sequences. Specifically, the proposed moving object segmentation algorithm is evaluated in terms of precision, recall and F-measure, which are defined. The notations *TP*, *FP* and *FN* are the total number of true positives, false positives, and false negatives respectively. Precision is defined as the number of *TP* divided by the total number of labeled 4x4 blocks. Recall is defined as the number of *TP* divided by the total number of ground truth labels. F-measure is the harmonic mean of precision and recall.

Pecision=TP/(TP+FN)

Recall=TP/(TP+FN)

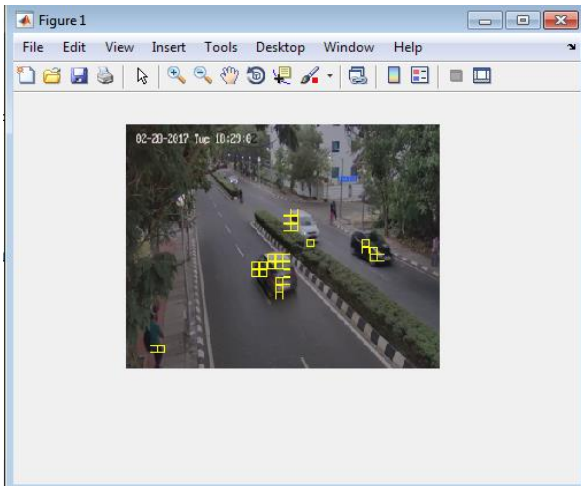F-measure=(2*Precision*Recall)/(Precision+Recall)
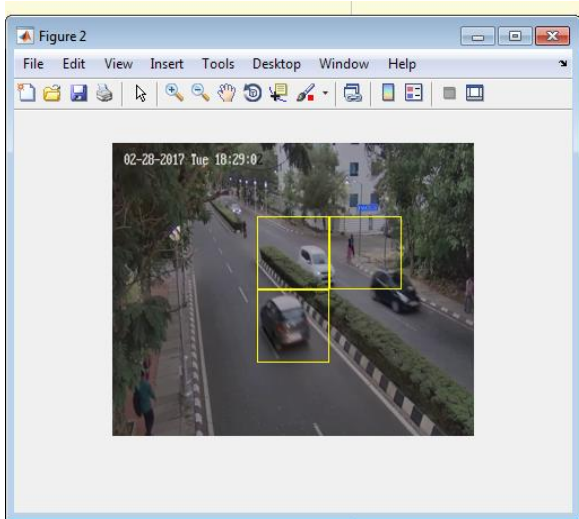

Fig5.1 Simulation of MV segmentation.

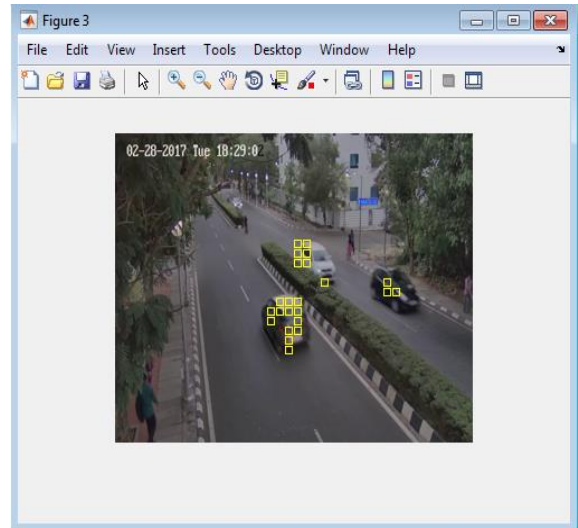Fig5.2 Moving object region after boundary refinement
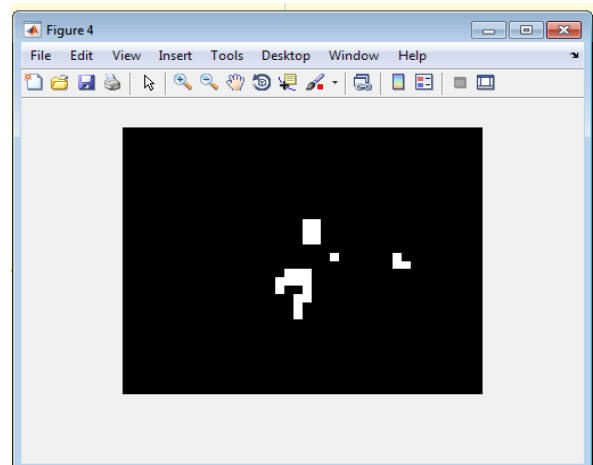

Fig5.3 Block artitions of the moving person and vehicles
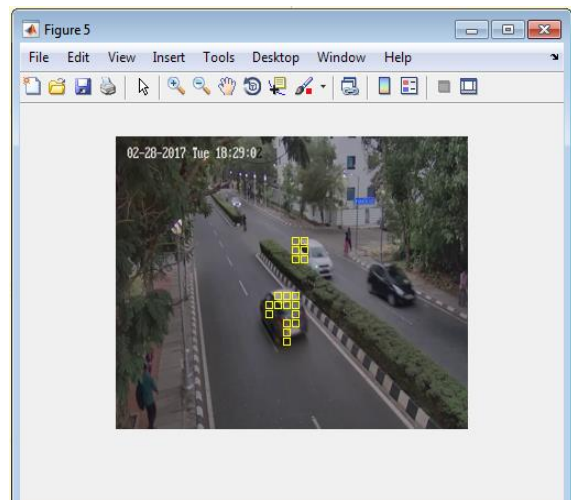

Fig5.4 Simulation of gray code from RGB

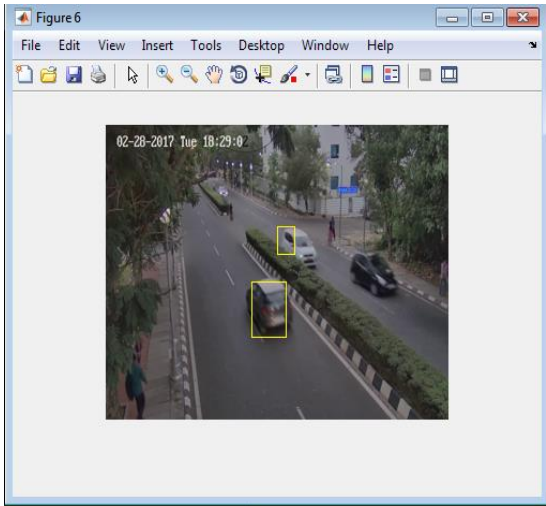Fig5.5 Simulation of block into subimage.

Fig5.6 Moving object region after boundary refinement

VII.CONCLUSION

In this project, I proposed a novel approach to segment and classify the moving objects from HEVC compressed surveillance video. Only the motion vectors and the associated coding modes from the compressed stream are used in the proposed method. Firstly, MV interpolation for intra-coded PU and MV outlier removal are employed for preprocessing. Secondly, blocks with non-zero motion vectors are clustered into connected foreground regions by the four-connectivity component labeling algorithm. Thirdly, object region tracking based on temporal consistency is applied to the connected foreground regions to remove the noise regions. The boundary of moving object region is further refined by the coding unit size and prediction unit size. Finally, a person- vehicle classification model using bags of spatial-temporal HEVC syntax words is trained to classify the moving objects, either persons or vehicles. It has a fairly low processing time and provides high accuracy. The accuracy and efficiency of the image pixel can be improved.

REFERENCES

[1] R. V. Babu, K. R. Ramakrishnan, H. S. Srinivasan, "Video object segmentation: a compression domain approach," IEEE Trans. Circuits Syst. Video Technol., vol. 14, no. 4, pp. 462–474, Apr. 2004.

[2] F. Porikli, F. Bashir, and H. Sun, "Compression domain Video Object Segmentation," IEEE Trans. Circuits Syst. Video Technol., vol. 20, no. 1, pp. 2–14, Jan. 2010.

[3] M. Grundmann, V. Kwatra, M. Han, and I. Essa, "Efficient hierarchical graph-based video segmentation," in Proc. IEEE Conf. Comput. Vis. And Pattern Recognit., pp. 2141–2148, Jun. 2010.

[4] Y. Chen, I. V. Bajic, and P. Saeedi, "Moving region segmentation from compressed video using global motion estimation and Markov random fields," IEEE Trans. Multimedia, vol. 13, no. 3, pp. 421–431, Jun. 2011.

[5] W. Lin, M. Sun, H. Li, Z. Chen, W. Li, and B. Zhou, "Macroblock classification method for video applications involving motions," IEEE Trans. Broadcasting, vol. 58, no. 1, pp. 34–46, Mar. 2012.

[6] H. Sabirin and M. Kim, "Moving object detection and tracking using a spatio-temporal graph in H.264/AVC bitstreams for video surveillance," IEEE Trans. Multimedia, vol. 14, no. 3, pp. 657–668, Jun. 2012.

[7] J. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards -including high efficiency video coding (HEVC)", IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 12, pp. 1669–1684, Dec. 2012.

[8] H. Sakaino, "Video-based tracking, learning, and recognition method for multiple moving objects", IEEE Trans. Circuits Syst. Video Technol., vol. 14, no. 5, pp. 1661–1674, Oct. 2013.s

[9] F. Wang, Z. Sun, Y. Jiang, and C. Ngo, "Video event detection using motion relativity and feature selection," IEEE Trans. Multimedia, vol. 16, no. 5, pp. 1303–1315, Aug. 2014.

[10] B. Dey, and M. K. Kundu, "Efficient foreground extraction from HEVC compressed video for application to real-time analysis of surveillance 'big' data," IEEE Trans. Image Process., vol. 24, no. 11, pp. 3574–3585, Nov. 2015.

[11] S. Giil, J. T. Meyer, T. Schierl, C. Hellge, and W. Samek, "Hybrid video object tracking in H.265/HEVC video streams," IEEE Int. conf on MMSP, 2016.