MOVING OBJECT SEGMENTATION AND CLASSIFICATION USING HOG DESCRIPTOR

Jini .S[1], Vysak Valsan[2]

[1]*M.Tech Scholar, Dept. of ECE, Jawaharlal College of Engineering and Technology, Kerala, India*
[2]*Assistant Professor, Dept. of ECE, Jawaharlal College of Engineering and Technology, Kerala, India*

---***---

Abstract - *Moving object detection and their classification plays an important role in intelligent video surveillance. Compared to existing methods HEVC introduces a host of new coding features which can be further exploited for moving object segmentation and classification. Some unique features like motion vectors and associated modes are extracted from videos. In this paper, this paper presents an approach to segment and classify moving object using unique features directly extracted from the HEVC for video surveillance. In the proposed method, firstly, motion vector interpolation and MV outlier removal are employed for preprocessing. Secondly, blocks with non-zero motion vectors are clustered into the connected foreground regions. Thirdly, object region tracking based on temporal consistency is applied to the connected foreground regions to remove the noise regions. After object region tracking then object boundary refinement should have to be done in order to refine the boundaries of the moving object. Finally, a person-vehicle classification model using Histogram of Oriented Gradients is trained to classify the moving objects, either persons or vehicles.*

***Key Words*: Compression domain, object segmentation, object classification, HEVC, video surveillance.**

1. INTRODUCTION

Video is used in several different applications like Video Conferences, Medical diagnostic, Security devices etc. Video compression is needed to facilitate both storage and transmission in real time. Digital video shares all the features of other digital formats, including lossless transmission, lossless storage, and ease of editing. A video sequence will be seen as a composition of video objects instead of a set of consecutive frames. Video object segmentation, which extracts the shape information of moving object form the video sequence. The method of extracting moving objects from frame to frame is moving object segmentation. For object-based representations, it is required to compress each video object as it moves in time. One favorable aspect of compressed videos is that it already contains extra additional information such as partition modes and motion vectors can be viewed as coarse analysis of object. Moving object detection means to detect moving objects from the background image to the continuous video image. Moving object detection is required to detect moving objects and it is detected after segmentation. The frame subtraction method, the background subtraction method and the optical flow method are the currently used methods.

In this Moving object segmentation and classification from video data is one of the most important tasks for intelligent video surveillance. Most computer vision methods for moving object detection and classification assume that the original video frames are available and extract descriptions of features from pixel domain. Most video content are received or stored in compressed formats encoded with international video coding standards. To obtain the original video frame, we have to perform video decoding. 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. To address this issue, compression-domain approaches have been explored for video content analysis which extracts features directly from the bit stream syntax, such as motion vectors and block coding modes.

HEVC is the most recent video coding standard, which encodes at similar perceptual quality but with 50% bit rate reduction when compared to the previous state-of-the art coding standard H.264/AVC. Similar to the H.264/AVC standard, HEVC relies on INTRA and INTER coding modes for each pixel block in the frame. Specifically, INTRA mode codes the pixel block without referring to any other frames, while INTER mode relies on pixel block in other coded frames to perform motion estimation. Result of the estimation is coded as motion vector (MV). However, unlike H.264/AVC that divides each frame into macro blocks of fixed size (i.e., 16 × 16 pixels), HEVC considers the Coding Unit (CU), which can assume any of the four sizes ranging from 8×8 to 64×64 pixels for handling video of different spatial activities. Intuitively, a CU of smaller size (e.g., 8×8) is utilized to encode block of high spatial activity and a CU of larger size (e.g., 64×64) is utilized to encode block with low spatial activity. The major advantage of compression-domain approaches is their low computational complexity since the full-scale decoding and reconstruction of pixels are avoided and it can be also used for real time applications. Here focusing on moving object detection and classification from HEVC compressed surveillance videos. Specifically, by extracting features from HEVC compressed surveillance video bits stream, the moving objects are located and classified, such as persons or vehicles.

2. METHODOLOGY

The main steps involve pre-processing, moving object detection, feature extraction and classification. The proposed moving object segmentation and classification method is shown in Fig.4.1. 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 pre-processing. 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. For person-vehicle classification, it involves a training phase to learn the person-vehicle model using Histogram of Oriented Gradients and a testing phase to apply the learned model to test videos. For the testing phase, we first extract the magnitude and orientation information of each 4×4 block to obtain the feature descriptor. Then, the descriptors of all blocks are clustered. The foreground object is represented by a histogram. Finally, for the segmented moving object, need to apply the learned person-vehicle model to determine which category to assign.

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. 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 pre-processing process, the MV interpolation for intra-coded blocks and MV outlier removal are employed before the moving object segmentation and classification.

Moving Object Detection

After the pre-processing of the MVs, 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, we examine its temporal

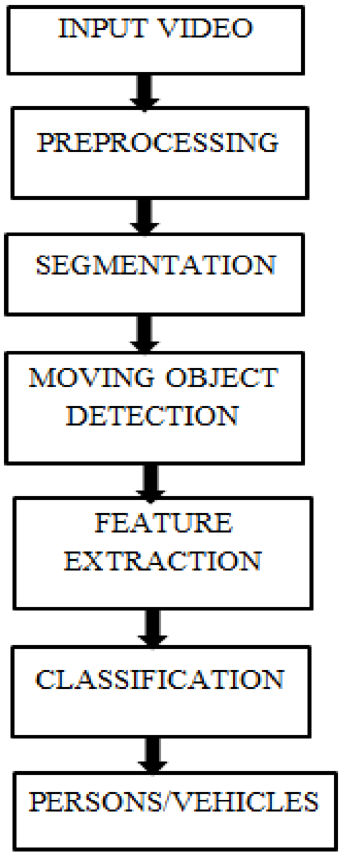

Fig -1: Block diagram

Consistency by using object region tracking. Secondly, refine the boundary of moving object region by using CU and PU sizes of the blocks.

Feature Extraction

The magnitude and orientation are effective features for object classification using Histogram of Orientation Gradient. It can be done by dividing the image into small connected regions called cells. For each cells, compiling a histogram of gradient. To improve accuracy the local histograms can be contrast normalized by calculating a measure of the intensity across a larger region of image called block and using this value to normalize all blocks within the block.

Classification

For object classification in surveillance videos, there is a need to classify the segmented moving objects into persons and vehicles using Histogram of Orientation Gradient with SVM detector. The features will be extracted from each block using HOG and represented in the form of histogram. The histograms of all blocks have to be concatenated for classification purpose. The Support Vector Machine (SVM) is used to classify the moving objects as persons or vehicles.

3. IMPLEMENTATION

The input video is acquired which is in HEVC video coding standard .Then the input video is converted into frames. Video to frame conversion is to extract valid information from video, process data efficiently, study characteristics of frames in detail and analyze properties of videos. The preprocessing has to done in order to segment foreground and background region and it reduces motion noise by referring to motion continuity over time and motion coherence within neighborhood. The preprocessing stage involves motion vector interpolation, motion vector filtering, motion vector refining and isolated and small motion vector removal. After preprocessing it is possible to separate zero and non-zero motion vectors i.e.; foreground and background. In moving object segmentation, first step is to connect all the foreground regions i.e.; regions with non-zero motion vectors using four labeling connectivity algorithm. Object region tracking and object boundary refinement should be done in order to segment foreground and background regions. The temporal consistency has to be examined using object region tracking. The object boundary refinement has to be done to refine the boundary of moving object region. The segmented moving object is shown in Fig.3 in the case of vehicles. After the moving object segmentation stage, moving object region can be detected. In next phase features has to be extracted from the moving object region and it is used for classification. Finally moving object can be classified as persons or vehicles accurately. The classified moving object is obtained as in Fig.4 in case of vehicle.

4. CONCLUSIONS

This paper presented an approach that segments and classify moving objects from videos. It is done by extracting unique features from compressed domain. The motion vectors and their associated modes are extracted from domains. In the proposed method, firstly, MV interpolation and MV outlier removal are employed for preprocessing. Secondly, blocks with non-zero motion vectors are clustered into connected foreground regions. Thirdly, object region tracking based on temporal consistency is applied to the connected foreground regions to remove the noise regions. After object region tracking then object boundary refinement should have to be done in order to refine the boundaries of the moving object. Finally, a person -vehicle classification model using Histogram of Orientation Gradients is trained to classify the moving objects, either persons or vehicles. The proposed method has a fairly low processing time, yet still provides high accuracy.

To evaluate the performance of our proposed moving object segmentation and classification scheme in HEVC compressed domain, we have collected 3 sequences. Each of the test sequences has more than one object in one frame. The resolutions and number of frames for the training and test videos are illustrated in Table I. Both the training and testing videos are encoded using the HEVC HM v10.0 encoder, at

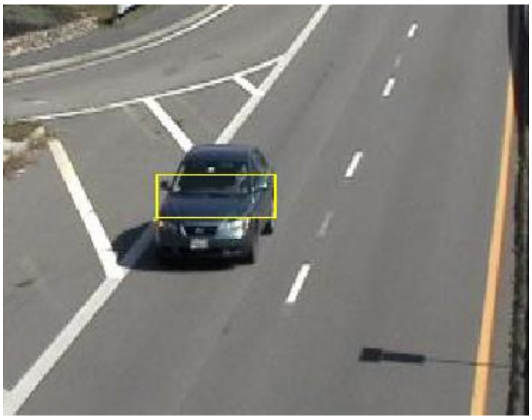

Fig -2: Moving Object Segmentation

various bitrates. HEVC syntax features, such as motion vectors, prediction modes, CU sizes, and PU types, are extracted from HEVC compressed bit stream.

Table-1: Resolution and number of frames of training and test sequences

Sequence	Resolution	Number of frames
Training_Seq_1	352x288	100
Training_Seq_2	640x480	200
Training_Seq_3	640x480	100

Moreover, 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. Table II summarize the performance of proposed segmentation algorithm in terms of Precision, Recall and F-measure.

Table-2 : Comparison of Several Methods In Terms Of Precision, Recall And F-Measure For Segmentation

Method	Precision	Recall	F-measure
Proposed	63.9	88.1	73.9
[20]	72.8	82.4	78.1
[14]	27.9	91.9	37.3
[10]	15.6	90.1	22.9

At the meantime, the running speed of our segmentation method is much faster. For the test video with resolution 352x288, the processing speed of the existing method is about 100 frames per second (fps), whereas the processing speed of our proposed method is over 410 fps.

Table-3: Processing Speed of Proposed Segmentation and Classification Method

Sequence	Resolution	Running Speed (fps)
Hall Monitor	352x288	390
Highway	320x240	535
Pedestrians	360x240	459

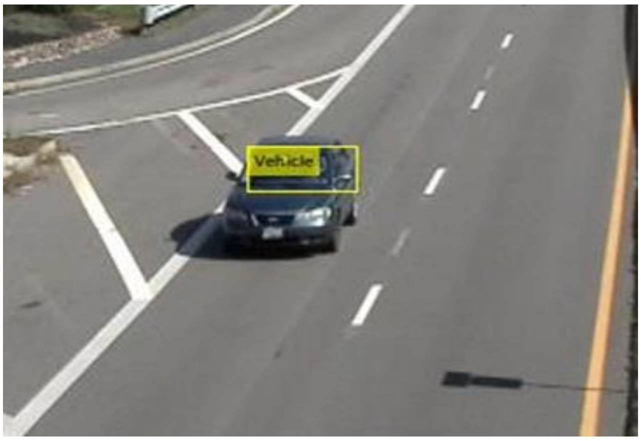

Fig -3: Classified Output

The moving object region containing just one kind of object is used for classification whereas the moving object region containing both the person and vehicle is not used for classification. Since the input of our classification algorithm is the moving object region extracted from the video frame by our segmentation algorithm, the classification accuracy of our classification algorithm depends on the segmentation accuracy of our segmentation algorithm. In sequences the objects are not close to each other, our classification algorithm can classify them with about 95% accuracy. In Highway sequence, the vehicles are near to each other, our segmentation algorithm fails to segment them correctly, even though, our algorithm can also classify them with about 85% accuracy.

Table-4: Accuracy of the Proposed Classification Algorithm

Hall Monitor	Highway	Pedestrians
94	71	88
96	76	94
94	80	83
91	81	87

ACKNOWLEDGEMENT

I would like to thank Prof. C Venugopal, Professor and Head of the Department and my project guide Mr.Vysak Valsan, Assistant Professor, Dept. of Electronics and Communication Engineering for his technical support and guidance and steering me to successful completion of this work.

REFERENCES

[1] 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.

[2] S. Chien, W. Chan, Y. Tseng, and H. Chen, “Video object segmentation and tracking framework with improved threshold decision and diffusion distance,” IEEE Trans. Circuits Syst. Video Technol., vol. 23, no. 6, pp. 921–634, Jun. 2013.

[3] 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。

[4] “Generic Coding of Moving Pictures and Associated Audio Information - Part 2: Video,” ITU-T and ISO/IEC JTC 1, ITU-T Recommendation H.262 and ISO/IEC 13 818-2 (MPEG-2), 1994.

[5] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 7, pp. 560–576, Jul. 2003.

[6] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, “Overview of the High Efficiency Video Coding (HEVC) standard”, IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 12, pp. 1649-1668, Dec. 2012.

[7] R. V. Babu, M. Tom, and P. Wadekar, “A survey on compressed domain video analysis techniques,” Multimedia Tools and Applications, vol. 75, no. 2, pp. 1043–1078, Jan. 2013.

[8] R. V. Babu, K. R. Ramakrishna, 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.

[9] S. D. Bruyne, C. Poppe, S. Verstockt, P. Lambert, and R. V. D. Walle, “Estimating motion reliability to improve moving object detection in the H.264/AVC domain,” In Proc. IEEE Int. Conf. Multimedia Expo., pp. 330–333, Jun. 2009.

[10] W. Zeng, J. Du, W. Gao, and Q. Huang, “Robust moving object segmentation on H.264/AVC compressed

video using the block-based MRF model," Real-Time Imaging, vol. 11, no. 4, pp. 290–299, Aug. 2005.

[11] 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.

[12] Z. Liu, Y. Lu, and Z. Zhang, "Real-time spatiotemporal segmentation of video objects in the H.264 compression domain," J. Visual Commun. Image Represent. vol. 18, no. 3, pp. 275–290, Jun. 2005.

[13] Y. Chen and I. V. Bajic, "Motion segmentation from a coarsely sampled motion vector field,"IEEE Trans. Circuits Syst. Video Technol., vol. 21, no. 9, pp. 1316–1328, Sep. 2011.

[14] C. Poppe, S. D. Bruyne, T. Paridaens, P. Lambert, and R. V. D. Walle, "Moving object detection in the H.264/AVC compression domain for video surveillance applications," J. Visual Commun. Image Represent. vol. 20, no. 6, pp. 428–437, Aug. 2009.