Compressed Domain Moving Object Segmentation and Classification from Surveillance Videos

Jini S
M.Tech Scholar
Dept. of Communication Engineering and Signal Processing Jawaharlal College of Engineering and Technology Palakkad-679301

Vysak Valsan
Assistant Professor
Dept. of Electronics and Communication Engineering
Jawaharlal College of Engineering and Technology
Palakkad-679301

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, we present a 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.*

Keywords—*Compressed domain, Moving object detection, Moving object classification*

I. 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. In this Moving object segmentation and classification from video data is one of the most important tasks for intelligent video surveillance. Moving object detection and classification in case of most computer vision methods 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[1].

For the greater part of the tracking strategy it is fundamental that object location procedure is finished all image outlines or at the time of the object first time appears in a video frame. Object detection is the system for locating an object or different objects about whether taking images from a video camera.Security and surveillance applications are based on machine vision algorithms that analyse biometric properties like faces and fingerprints or monitor sensitive areas. Gesture recognition and the tracking of facial expressions help to define new paradigms in human computer interaction. Furthermore, computer vision-based road and traffic analysis might someday lead to automatic guidance of road vehicles. Medical image analysis applied to the spatial image sequences acquired by computer tomography or magnetic resonance imaging can indicate areas of interest and provides additional information about individual patients[2]. Systems for document analysis and image/video indexing support user retrieval of relevant information from large media archives. Computer vision techniques are employed in the entertainment industry in the production process of movies and in media distribution.

Of particular interest with regard to several applications in the above-mentioned fields are capabilities for object segmentation and object recognition. The goal of this paper is to develop a classification system for video objects. By video object we mean the collection of all two-dimensional appearances (projections) of a real world object within an image sequence. A single appearance is called an object view. The classification results obtained with our system can be used to index and/or categorize videos and thus support object-based video retrieval.

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[3]. The major advantage of compression-domain approaches is their low computational complexity since the full-scale decoding and reconstruction of pixels are avoided. It can be also used for real time applications. Here focusing on moving object detection and classification from HEVC compressed surveillance videos.

II. METHODOLOGY

The block diagram of the proposed system is shown in Figure.1. The aim of this paper is to improve the performance of object detection and classification by detecting the movement of object in the images of continuous video frames. Automatic detection of objects can be the foundation for many interesting applications. An accurate and efficient detection capability at the heart of such a system is essential for building higher level vision-based intelligence.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.1. It consists of two stages: moving object segmentation and person-vehicle classification. For moving object segmentation, first the input video is preprocessed using motion vector interpolation and motion vector filtering. Motion vector filtering involves motion vector filtering, motion vector refining and isolated and small motion vector removal. After preprocessing segmentation has to be done. In segmentation, object region tracking and boundary refinement are needed to detect the moving object. 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.

In preprocessing stage, 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.

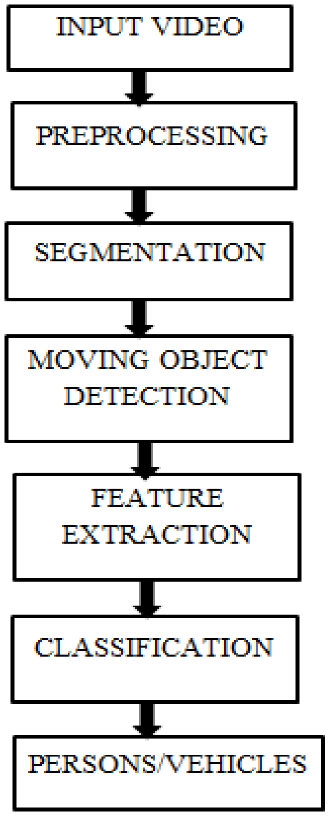

Fig. 1 Block Diagram

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

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. The flowchart of the proposed system in Fig.2.

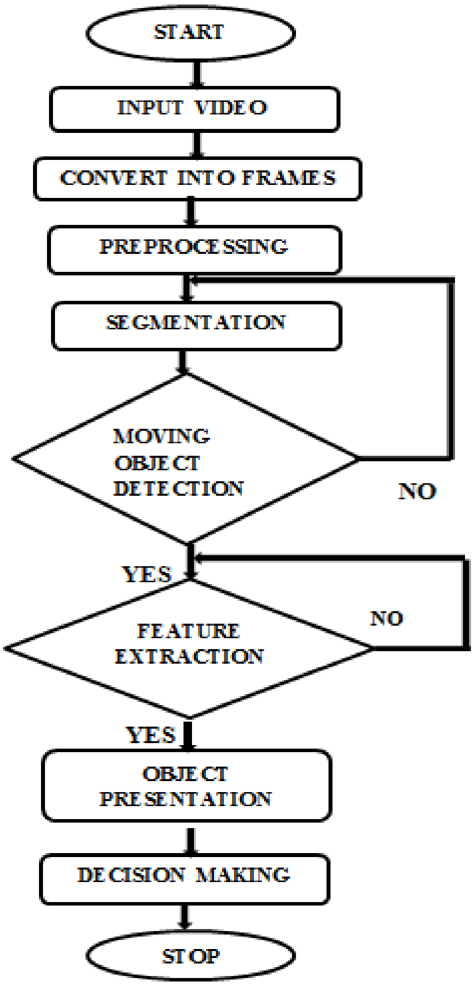

Fig.2 Flow Chart

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.

III. IMPLEMENTATION

The input video is required and video to frame conversion is employed to analyse the HEVC video and to study the characteristics of frames. Secondly the pre-processing has to be done to separate foreground and background regions from frames. In pre-processing, motion vectors are scaled down to make them independent of frame type and also motion vectors have to be assigned in order to segment foreground and background blocks. Thirdly connect foreground regions using four connectivity component labelling algorithm. It finds all connected components in an image and assign a unique label to all points in the same component. Then object region tracking algorithm used to temporally track foreground regions using motion vectors. The next step is to refine the boundary of the moving object using object boundary refinement. Moving object can be detected after segmentation process. The classification is done based on the segmented moving object and classified as either person or vehicles.

The first step in implementation is to covert the videos into frames by using a default function available in Matlab. Each frame has to be pre-processed before segmentation and classification. Each frame in the video is converted into gray level. Then divide each frame into block of 64x64 and again this block is divided into 8x8 blocks which is the sub image of 64x64 blocks. Considering one image as main image and another image as reference image we can calculate the motion vectors of both main image and reference image. The motion vectors can be assigned using the list containing neighboring motion vectors. After motion vector assigning stage then motion vector outlier removal has performed. It involves motion vector filtering, motion vector refining and isolated and small mv removal. After preprocessing, segmentation has done. At the output of preprocessing, the moving and non-moving motion vector can be separated. The foreground regions can be then connected together using connectivity labeling algorithm. The object region tracking and object boundary refinement are the next phases in segmentation. The segmented output is then given for classification purpose. The histogram of orientation gradients is the method used for extracting features like magnitude and direction from segmented moving object region and classifies them as either persons or vehicles.

IV. RESULTS AND CONCLUSION

In order to train the person-vehicle model for moving object classification, 3 training sequences are used which is showing in Fig.3. To evaluate the performance of our proposed moving object segmentation and classification scheme in HEVC compressed domain, we have collected sequences having more than one object in one frame. Example frames of the videos are shown in fig.3. 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 various bitrates, with the GOP structure IBBBB, i.e., the first frame is coded as

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

Sequence	Resolution	Number of frames
Seq_1	352x288	100
Seq_2	640x480	200

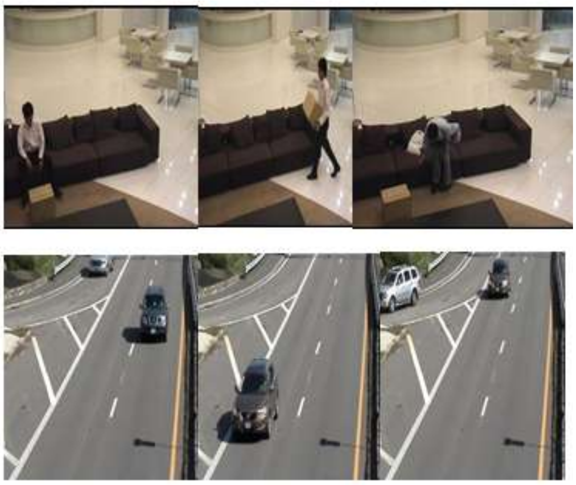

Fig.3 Example frames of testing and training videos

intra (I), and subsequent frames are coded as generalized B frames. HEVC syntax features, such as motion vectors, prediction modes, CU sizes, and PU types, are extracted from HEVC compressed bit stream.

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. 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. When compared to existing methods, our method can achieve similar segmentation performance.

Table II. Processing speed of proposed method

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

At the meantime, the running speed of our segmentation method is much faster. For the test video with resolution 352x288, the processing speed of our proposed method is over 410 fps. The processing speed of our proposed method for each sequence is illustrated in Table II.

Acknowledgments

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.

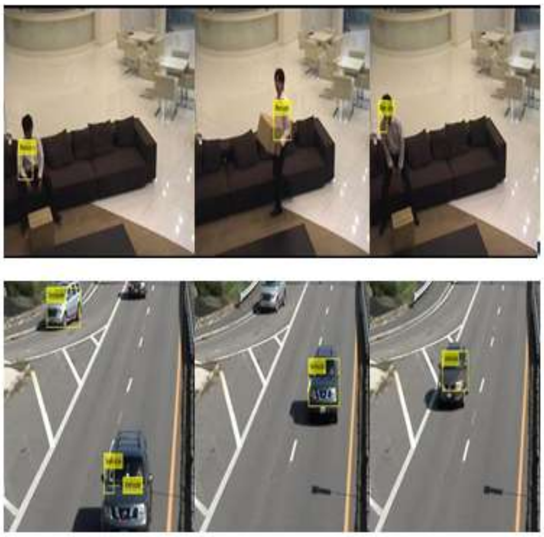

Fig.4. Segmented and classified output of proposed system

References

[1] P. Dong, Y. Xia, L. Zhuo, and D. Feng, "Real-time moving object segmentation and tracking for H.264/AVC surveillance videos," in *Proc. IEEE Int. Conf. Image Processing*, Brussels, pp. 11–14, Sep. 2011.

.

[2] S. H. Khatoonabadi and I. V. Baji´c, "Video object tracking in the compression domain using spatio-temporal Markov random fields," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 300–313, Jan. 2013.

[3] 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.

[4] H. Li, Y. Zhang, M. Yang, Y. Men, and H. Chao, "A rapid abnormal event detection method for surveillance video based on a novel feature in compressed domain of HEVC," in *Proc. IEEE Int. Conf. Multimedia and Expo*. (*ICME*), pp.1-6, Jul. 2014.

[5] 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.

[6] D. Park, D. Lee, and S. Oh, "Object tracking in HEVC bitstreams," *Journal of Broadcast Engineering*, vol. 20, no. 3, pp.449-463, May 2015.

[7] L. He, Y. Chao, and K. Suzuki, "A run-based two-scan labeling algorithm," *IEEE Trans. Image Process.*, vol. 17, no. 5, pp. 749–756, May 2008.

[8] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray. "Visual categorization with bags of keypoints," In *Workshop on statistical learning in computer vision, ECCV*, pp. 1-22, May 2004.

[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] S. Biswas, and R. V. Babu. "H. 264 compressed video classification using histogram of oriented motion vectors (HOMV) ," in *Proc. IEEE Conf. Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2040-2044, May 2013.

[11] N. Dalal, and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit.*, pp. 886-893, Jun. 2005.

[12] M. Moriyama, K. Minemura and K.S. Wong, "Moving object detection in HEVC video by frame sub-sampling," *IEEE Int. Symposium on ISPACS*,

[13] 48-53, Nov. 2015.

[14] S. Gül, 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.

[15] S. Pulare and S. S. Tale, "Implementation Implementation of Real Time Multiple Object Detection and Classification of HEVC Videos," *Int. Journal for IRST*, vol. 2, no. 11, pp. 248-254, 2016