Real-Time Moving Object Segmentation and Classification from HEVC Compressed Surveillance Video

A PROJECT REPORT

submitted by

SUMIT PARWAL (14BCE0338)

in partial fulfillment for the award of the degree of

B. Tech

in

Computer Science and Engineering

Vellore-632014, Tamil Nadu, India

School of Computer Science and Engineering

APRIL, 2018

School of Computer Science and Engineering

DECLARATION

I hereby declare that the project entitled **"Real-Time moving object segmentation and classification from HEVC compressed Surveillance video"** submitted by me to the School of Computer Science and Engineering, Vellore Institute of Technology, Vellore-14 towards the partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology in Computer Science and Engineering** is a record of bonafide work carried out by me under the supervision of **Shalini L, Assistant Professor Sr.** I further declare that the work reported in this project has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma of this institute or of any other institute or university.

Signature

Name : Sumit Parwal

Reg.No: 14BCE0338

School of Computer Science and Engineering

CERTIFICATE

The project report entitled "**Real-Time moving object segmentation and classification from HEVC compressed Surveillance video**" is prepared and submitted by **Candidate Sumit Parwal (Register No: 14BCE0338),** has been found satisfactory in terms of quality, scope, and presentation for partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology in Computer Science and Engineering** in Vellore Institute of Technology, Vellore-632014, India.

Guide
(Name & Signature)

Internal Examiner **External Examiner (Name &**
Signature) **(Name & Signature)**

ACKNOWLEDGEMENT

It gives me immense pleasure to convey sincere gratitude to my Project guide **Prof Shalini L**, for her helpful nature, efficient inputs, thorough guidance, and complete cooperation during my project. I would also like to thank my Head of Department, **Prof. Santhi V** for encouraging us throughout and for giving us an opportunity to present our projects at our university. Without the whole hearted cooperation of our Dean of SCOPE, this project would not have been possible. I thank our Dean, **R. Saravanan** for his constant support. Nothing is ever possible without our parents' support. Their blessings have made this project possible and I cannot thank them enough. Last but not the least I would like to thank VIT University for making me capable of being accepted in the industry and working to the best of my potential.

TABLE OF CONTENTS

`

LIST OF TABLES

`

LIST OF FIGURES

`

LIST OF ABBREVIATIONS

HEVC-High efficiency video coding

AVC-Advanced Video Communication

MV-Motion Vector

PU-Prediction Unit

CU-Coding Unit

PC-Personal Computer

CCTV-Closed Circuit Television

MRF- Markov Random Field

MPEG-Moving Pictures experts group

MVD -Motion vector difference

MVF-Motion Vector Field

GOP-Group of Pictures

`

Abstract

In the age of technological advancements, where high end cars are creating lot of ruckus on the roads and number of thefts have exponentially increased video surveillance is the need of the hour to classify moving objects. In intelligent surveillance of compressed HEVC video the segmentation and classification of moving objects plays a crucial and highly important role. Highly efficient video coding (HEVC) incorporates a number of new coding functions that can be used to segment and classify moving objects in a more efficient way as compared to H.264 / AVC . This document proposes a unique, and real-time perspective as the model first segments and then classifies the moving object from the frames extracted from surveillance video using features that are present in compressed videos.

In the approach we followed, first, the interpolation of motion vectors (MV) take place for the intra-coded prediction unit (PU) and then the elimination of outliers of MV are used for preprocessing. Second, all blocks having motion vectors greater than zero are merged and grouped in foreground areas using the four-connectivity component labeling algorithm. Third, we use temporal consistency tracking areas included in the connected foreground areas to eliminate the noise areas. Size of both coding and prediction unit are used to further define the boundary of the moving object.

Lastly, for training, we use a model named person-vehicle classification that uses a bag of HEVC syntax words that help categorize the moving object as people or vehicles. The final results show that the system can efficiently classify between person and vehicle.

`

1. Introduction

1.1 Theoretical Background

This Project consists of a MATLAB based project that segments and classify moving objects from compressed surveillance videos. The requirements to run this software is simple PC(Personal Computer) which is connected to the CCTV camera. The data is stored in the PC itself and all the process is run smoothly. Video consists of many frames depending on frame rate and size of video, moving objects are segmented from the various frames and then after refining they are classified based on person-vehicle classification model.

HEVC, a video compression standard which is also called as H.265 or MPEG-H part 2 supports resolutions up to 8192×4320. It is also one of the many successors of commonly used AVC (H.264 or MPEG-4 Part 10). H.265 provides nearly twice the data compression ratio at the identical level of video quality, it amplifies video quality at the same bit rate compared to AVC.

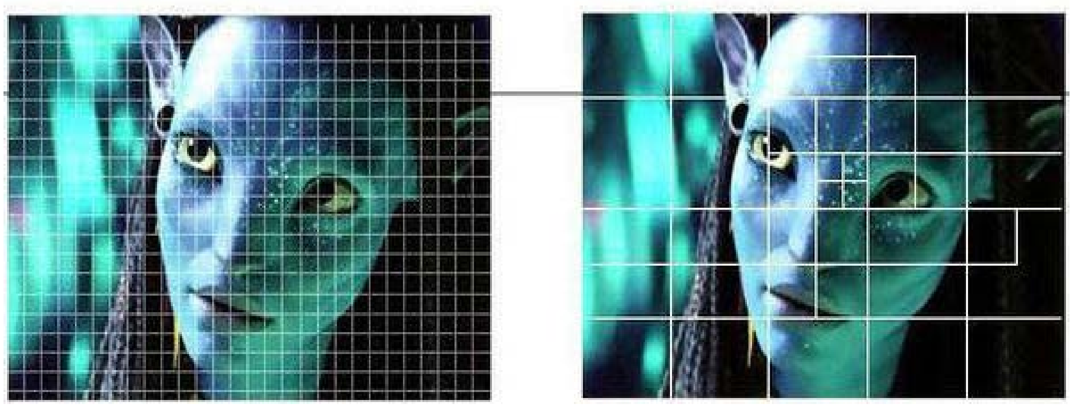

Fig 1-Differnce between H.264 and H.265(HEVC)

H.264 which is based on advanced video encoding (MPEG-4 AVC) is a video compression standard which is based on block-oriented motion compensation that supports resolutions up to 4096 × 2304. Presently, it is one of the most popular and efficient format for recording, compressing and distributing video content. But if we compare that to HEVC, the latter is much better for video surveillance because the data compression ratio is much better than the first one.

`

HEVC compression standard have enhanced video and image processing as there is a greater resolution and there is a classification model that is used to classify the moving object into a people or a vehicle and we can use "Aadhar" database to identify the person or we can also use the database to find out the owner of the car.

1.2.Motivation

With the rise of population and Income levels in India, more and more people are buying superfast cars which is the major cause of road rage accidents. We can segment and classify moving object from video data which will help save lot of lives and properties and we could also make our roads a lot more safer. Now when we are taking about safety we need to also talk about safety from thefts, terrorists and crimes that are increasing with the increase of technology, so we have to have a method which could simultaneously process the video and signal the concerned authorities to take the corrective action. Most computer vision methods that are already there for Segmentation and classification of moving objects presume that the video frames are available and that descriptions or features of the pixel domain are extracted which means they cannot process the live video. The safety and security of its citizen is utmost priority of any government of any Country and in the country like India where there are lots of terrorist attacks and is suffering from this problem surveillance is very important so that these attacks can be stopped.

Most of the content which is available and encoded with international video coding standards are received and stored in compressed formats, like HEVC, H.264/AVC and MPEG-2, for this further decoding has to be carried out to obtain the original video frame. The proposed system would help the Police and Intelligence forces and will help in saving lot of life and property.

`

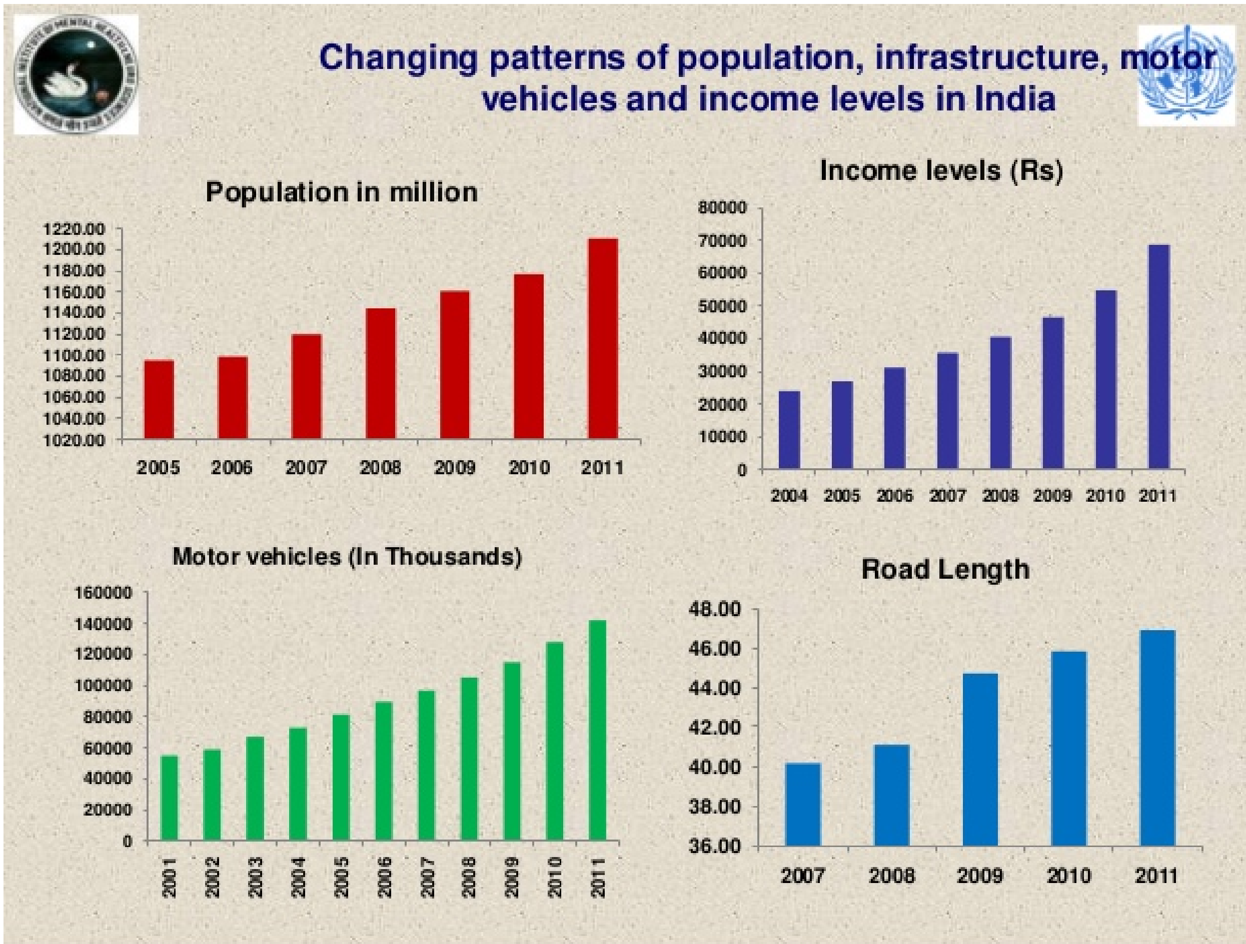

Fig 2-Changing patterns in Infrastructure, population, income levels and motor vehicles.

1.3 Aim of the proposed Work

To perform Segmentation and Classification on compressed HEVC videos in order to extract real-time moving objects.

1.4 Objective(s) of the proposed work

- To perform segmentation on each frame of the compressed HEVC video.
- To find out moving objects from various frames.

`

- To us person-vehicle classification model to classify moving objects
- To use Machine learning concepts that will make implementation of the project easier and with high success rate .

1.5 Report Organization

This report is divided into six broad units:

Introduction

Literature Survey

Overview of the Proposed System

Proposed System Analysis and Design (As Per IEEE Standard) Results and Discussion (As Per IEEE Standard)

Conclusion Limitations and Scope for future Work.

This will be preceded by an abstract, list of tables, acknowledgement, etc. The introduction introduces the project, explains the aim, motivation and objective of the project. Literature survey explains the previous related work. Overview of the proposed system explains the entire project from a architectural perspective. Proposed System Analysis and Design performs an analysis on the system and discusses in depth regarding the design. Then the results are analyzed and discussed. Finally we have conclusion and future scope of the project is presented.

`

2. Literature Survey

2.1 Survey of the Existing Models/Work

1. "DETECTION AND TRACKING OF MULTIPLE OBJECTS IN CLUTTERED BACKGROUNDS WITH OCCLUSION HANDLING Sukanyathara J and Alphonsa Kuriakose[33]"

This tracking model can efficiently track many objects under dynamic background conditions despite of occlusions. The two vital angles for surveillance systems are segmentation and tracking. Numerous boundaries, for example occlusion, camera movements, and cluttered background make the strong tracking and detection a troublesome issue, particularly if there is rise in occurrence of numerous moving items.

Object Detection with variable or troublesome luminance conditions and within the sight of camera clamor(noise) is as yet a hot territory of research. Creators, through this approach proposed a model which can proficiently track the all the objects that are moving and track them in spite of the earlier information of items in the scene and occlusion. A very efficient decision algorithm that uses a multi-background framework is utilized for segmentation.

Various objects are tracked alongside their trajectories in view of Continuous Energy Minimization with the assistance of the video object tracking. In this work, we have joined multi-foundation enrollment with a powerful detailing of multi-target tracking which is same as minimization of a continuous energy. Other than the current methodologies, it center around making utilization of a energy which compares to a more efficient portrayal of an issue, instead one that is docile to optimization. Other than picture confirm, all energy functions works on physical imperatives, for example, mutual exclusion, target dynamics, and track tirelessness(persistence). This following model can track numerous articles regardless of impediments under unique foundation conditions.

`

2. "Frame Sub-sampling for moving object detection in HEVC videos":

Video compression expects to evacuate spatial-temporal redundancies where the encoded bitstream, especially the motion vectors, may not speak to the genuine movements in the video. Subsequently, moving item discovery in the compressed video stream is an in fact testing assignment.

In this work, they proposed a moving object detection algorithm which uses frame sub-sampling method in the HEVC video coding standard. The number of frames is reduced with the help of (temporal) sub-sampling. The frames are re-encoded utilizing HEVC with the same settings to increase the movement of the moving object. Sub-sampling successfully builds the movement power of the objects, which can be the noteworthy sign for detecting moving objects while movements in the background still stay less.

Movement vectors and INTRA coding units of moving objects acquired by means of frame sub-sampling and re-encoding are specifically used to isolate the background and moving objects in the video. The segmented outcomes are refined and contrasted and compared with the result without performing frame sub-sampling. Results demonstrate that the sub-sampling method accomplishes higher precision, with a change more prominent than 0.35 regarding F-measure.

3. "To detect and classify Real Time Multiple Object in HEVC Videos":

We show new strategies for initializing object tracking through the automatic detection of moving objects in terms of subtraction from background. The contemporary strategies that we have proposed earlier are incorporated in real time system for tracking of objects. With our new method for updating the proposed background framework and versatile threshold retention, a mask for foreground object is provided for initialization of object tracking . The traditional background subtraction technique distinguish between motion elements by subtracting the background framework from the current image. In contrast to other basic algorithms for detecting moving objects, the subtraction technique focuses objects more precisely in the

`

foreground and detects objects in the foreground, regardless of whether objects are moving or not.

Be that as it may, one disadvantage of conventional background model is that it cannot detect natural variations, for instance, slow or unexpected changes in light. The main reason behind this disadvantage: it accept static background, which is why a background display refresh is needed for dynamic foundations.

The main problems at that point were the manner by which to refresh the background framework, and to decide a edge for arrangement of pixels in background & foreground. In this paper they have given a strategy to dynamically and automatically determine threshold, based on the pixel intensity in the present frame and a technique for learning the background framework based on the differences in the pixels in the background framework and the previous frame. Consequently, there is ample occasions to spot opportunities by reusing the structures of coding in HEVC, which can save a large amount of computational resources.

4. "HYBRID VIDEO OBJECT TRACKING IN H.265/HEVC VIDEO STREAMS by Serhan Gul, Jan Timo Meyer, Cornelius Hellge, Thomas Schierl, and Wojciech Samek[31]":

In this, they have proposed a mixture strategy for tracking which distinguishes objects that are moving in video recordings 0HEVC standard. System in this paper generally relies on types of blocks and motion vectors (MV) acquired by interpreting a video bitstream partially and sometimes utilizes pixel data to recognize the difference b/w two articles. The compressed domain technique depends on the MRF technique that catches coherence in both spatial and temporal ways for the object and is refreshed frame by frame.

The intersection of our approaches comes from the pixelated domain code, which assumes the color of fully decoded frames and is simply updated after the completion of each GOP. MRF show. The precision of our model which uses standard sequence, Displays the accuracy of the cutting edge MRF show.

`

5. "HEVC Compressed Domain Moving Object Detection and Classfication by Liang Zhao1 , Debin Zhao1 , Xiaopeng Fan1 , Zhihai He2[29]":

The segmentation and classification of objects that are moving, the compressed domain plays an important role in some running applications, such as indexing of videos and video surveillance. Contrast and the latest universal encoding video indicators, for example H.264 / AVC, HEVC presents a large group of new outstanding coding features. Therefore, the segmentation of the moving object and the order classification of the HEVC packaged records refer to another test. The document has developed a strategy for segmenting and classifying moving objects, especially humans and vehicles, in the compressed HEVC domain. First, we'll prepare a classifier to decide if an image patch has a place in terms of foreground or background using the HEVC properties.

We at that point prepare a moment grouping framework to arrange all the objects that are moving into either vehicle or people, utilizing bag of HEVC language words. The broad exploratory outcomes show that the method gives the surprising execution and it can arrange moving vehicles and people precisely and heartily.

`

2.2 Summary/Gaps identified in the Survey

EXISTING SYSTEM
EXISTING CONCEPT Several segmentation and classification algorithms for objects that are moving were reported using information of MV in the H.264 compression domain. A method to collect all MV information over time to segment the moving object. The short lived MVs that are accumulated are spatially interpolated to obtain a thick field, and the method named expectation maximization is then applied in the thick motion domain to the final segmentation.
EXISTING TECHNIQUE : • **Motion Vector**
TECHNIQUE DEFINITION: In this existing technique the information on reliability , summed up with the magnitude MV, is used for foreground objects to be segmented from the background. There are various types of MVs, such as: noise, frame, background, and foreground. After that the MVs along with their associated information on class are used for segmentation of each block. The global motion is first removed from the vector motion field, and then the motion object segmentation process is performed in the archived motion vector.
DRAWBACKS(Gaps): • The main drawback of the estimator is that even as an iterative process it cannot appreciate large shifts due to the limited efficiency of expansion of the first-order Taylor series. This is in a sharp contrast with the BM which is restricted only by the defined search space. • This is a huge problem with motion estimators that are gradient-based. Finally, all motion estimators based on gradients completely fail when the conditioning of the matrix product is very bad. This happens if the block you are looking for has no texture or if there is only one edge in the block. We can use the eigen-values and eigen-vectors of this matrix to find when these problems occur.

`

3. Overview of the Proposed System

3.1 Introduction and Related Concepts

Intelligent video surveillance is the need of the hour because of increasing road rage accidents, internal security of the country and increasing thefts in the country, for this the most important task is to perform moving object and segmentation of video data. The majority of the Computer vision methods that are available now for detection as well as classification of objects based on are they moving or not, let's say that the actual frames extracted from video are accessible, they also obtain features/descriptions from domain of pixels .

It has to be noted that most of the video content that is stored or received are of compressed type and they are coded by the international standards used for video coding, like HEVC,MPEG, and H.264. When we have to get the actual video frame we need to decode the video on the compressed version. When we are performing large scale video analysis, like when we are searching for a very big network used for surveillance and content analysis, the complexity video decoding become a very big issue of real- time systems that are present.

For solving this issue many compressed-domain procedure have been explored for analysis of video data that takes out features from the bit stream syntax, like black coding modes and motion vectors. The considerable trump card of compressed domain procedure is that they have very less complexity computationally because the complete reconstruction and decoding of the pixels are circumvented.

For this reason, the methods used in compresses area are needed for analysis of video applications. We have prioritized the detection of objects in motion and then the classification of the compressed HEVC videos. In particular, by removing the characteristics from the HEVC compressed video bitstream, all moving objects are first located and then they are classified into either people or vehicles.

`

Several algorithms that are used to classify and segment moving object were reported using MV information in the domain of H.264 compression. The MVs accumulated for a short term are spatially interpolated to get a thick field, and after that the maximization procedure is used in the thick motion field for the last segmentation.

3.2 Framework, Architecture or Module for the Proposed System(with explanation)

Overall framework of the proposed system is shown in Fig. 3. It has 2 stages, First: segmentation of moving object from the set of consecutive frames by calculating the motion vector for each frames and second: classification of moving object based on person-vehicle framework. In the first stage:

Firstly we use interpolation of MV technique for intra-coded PU, for preprocessing we employ MV outlier removal . Then, all the motion vector blocks having value greater than zero are grouped with help of a algorithm called "4 connectivity component labeling" into a connected regions of foreground objects. Finally, we apply object region tracking which has temporal consistency to the connected foreground regions to extract and remove the noise regions. The border/boundary of the moving object region is further refined by using the CU and PU sizes of the blocks.

For classification on people-vehicle framework it involves two key steps, firstly training in which the system trains itself with the help of people-vehicle framework using “bag of spatiotemporal HEVC syntax words” and secondly testing, where the trained system tests itself using videos. In testing, we get information in terms of spatial and temporal data of every block in order to extract the descriptors. After that, these extracted descriptors of each blocks are grouped into a codeword. The moving object which is in the foreground is replaced by a histogram of the codeword. Lastly, learned person-vehicle model is applied for the segmented moving object to determine that this object belongs to which category.

MV magnitude along with this reliability information is used to chunk out foreground objects from the background object. Motion vectors are of various types, such as

`

- edge
- foreground
- background
- noise

After this, all the motion vectors along with their respective class info are used to chonk out each blocks. Firstly the global MV is extracted from the moving object segmentation, and motion vector process field is implemented on the MVF.

ADVANTAGE

- Computing complexity of the proposed approach is very less compared to other approaches, many people who are working in the field of video coding have implemented this approach in order to calculate the value of MV in each pixel.

- This happens so that the effect of "moving averaging" on block manipulations is less. The proposed method is of good use, since there is a finite support region that is always used to configure the equations.

ALGORITHM

I have used three techniques :Motion Vector Interpolation, Four connectivity component labeling algorithm, Temporal consistency.

MODULES

There are basically 4 modules which are used while implementing the proposed system. These modules produce the output such that it is the input to the other module. Like we serve the video to the preprocessing module than it preprocesses the video and give it's output as input to moving object detection module, it detects object with motion vector greater than zero and extracts it and then supplies it to last module for representation and decision making which basically classifies the object as either a people or vehicle.

`

PREPROCESSING

One MV in compressed videos coded under HEVC is related with an inter-coded PU. MV's, changed acceptably so that they can be made autonomous of the type of frame. For this to happen MVs are divided on the basis of the contrast between the reference frame value(in display order) and the corresponding frame value. For Ex: let's say the value of a MV for frame- A be (2,2) while for other frame-B the value of MV be (6,6) which is adjacent block to the reference frame, both the MV values for frame-A and frame-B would be changed to both (4,4) after the procedure used to scale MVs.

The PUs which have 2 MVs, the MV with the bigger value of length would be chosen as the characteristic MV of the PU. The prior processing causes the interpolation of MV for intra-coded blocks and the elimination of outliers from MV to be used prior to the segmentation and classification of the object.

MOVING OBJECT DETECTION

After the preprocessing stage of the MVs, all the blocks having MVs with values greater than zero are called as foreground blocks. On these blocks the four-connectivity component labeling algorithm is applied so that these blocks can be grouped into the connected regions which is present in foreground. Each foreground area is first examined to determine consistency values temporally by including object space tracking. After that the boundary of the area of the moving object is refined by the CU and PU sizes of the blocks.

FEATURE EXTRACTION

For the proposed system three types of features have been identified,

- The length of prediction modes
- Motion Vector Difference (MVD)
- Motion Vectors

as an efficient features that can be used for classification of object. Firstly the value of MV length is calculated and it is related to the velocity of the object, which is a naive but a very important feature for classification of MVs either into a person or a vehicle because generally vehicles have a faster pace as compared to people. There is an observation that people generally suffer from non-rigid deformations , it is difficult to find a lookalike for every single PU in the region of people in motion.

`

The blocks which lie in and near the area of moving vehicles are coded with intra mode less as compared to the region surrounding person. That is why we can utilize the prediction unit in a efficient way.

CLASSIFICATION

For classification of objects in HEVC videos, classification of the segmented objects either into a vehicles or a persons with "bag of HEVC syntax words" in HEVC domain. The "bag of words" characterization which has been used successfully for classification of objects in the pixel domain .

This method of classification is performed by following steps:

(1) Firstly, we need to Describe each and every coding block which we have got within the region of moving object with the help of features present in HEVC syntax .

(2) Secondly, we have to use a clustering/grouping procedure to make a codebook.

(3) Thirdly, we need to represent each object which is found moving from codebook by calculating a histogram of code word which is normalized.

(4) Fourthly, we need to train a classifier using binary values so that it classifies the objects found moving either into a person or vehicle.

`

3.3 Proposed System Model(ER Diagram/UML Diagram/Mathematical Modeling)

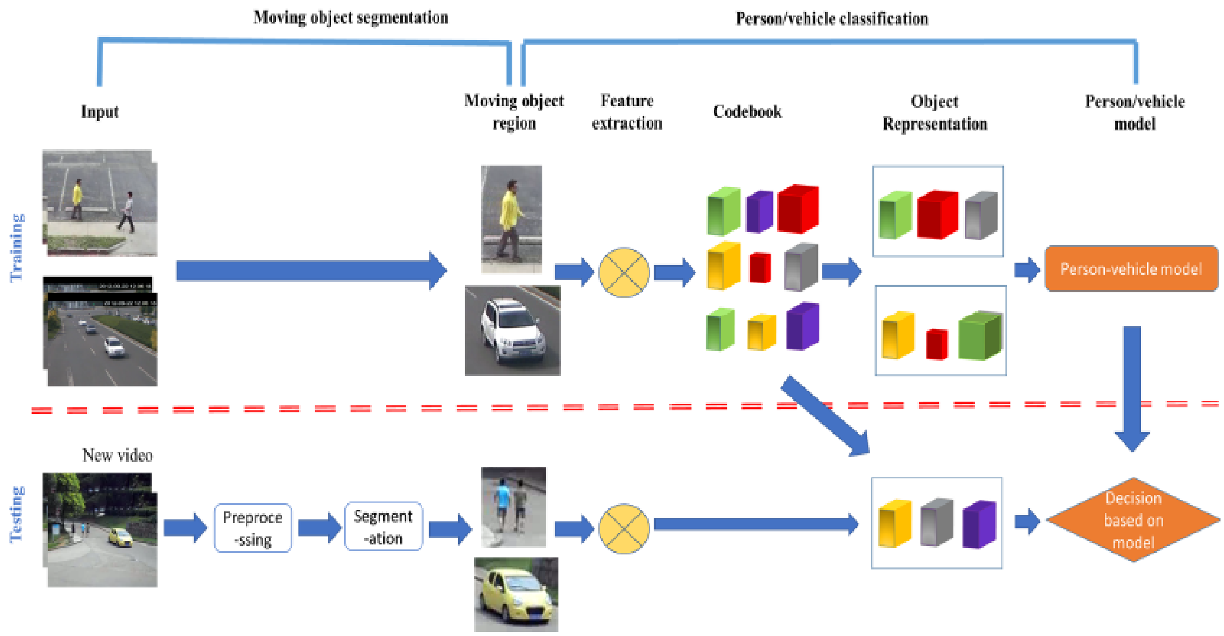

Fig 3:The proposed framework

In Fig 3, proposed framework is shown on the basis of which I am implementing the system. There are two parts training and testing .For training as soon as the video is given as a input it performs preprocessing, segmentation, feature extraction and classification at last which is seen in Fig4. It makes a codebook which is used at the time of classification in testing process.

While training process the supplied video is first converted into frames and moving object region is calculated with the help of motion vectors and four component connectivity labeling algorithm. After you get the moving object you extract the feature from the frame and then you refine the boundary of the moving object after that with the help of codebook you represent the object and classify it into either person or vehicle.

In the testing phase first of all you perform the same steps for segmentation, preprocessing and feature extraction . But instead of using codebook you use the object representation from the training phase and classify the object.

`

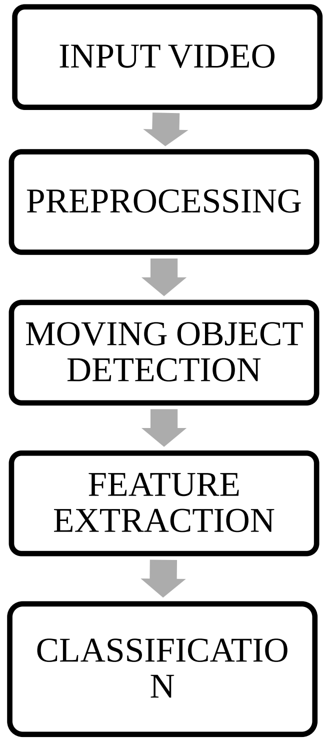

Fig 4:The module UML diagram

The module diagram contains all basic modules for the proposed model:

- Preprocessing
- Moving Object Detection
- Feature Extraction
- Classification

`

4. Proposed System Analysis and Design

4.1 Introduction

PREPROCESSING:

There are 3 steps that need to be performed for preprocessing:

A. Motion Vector Interpolation

A MV is assigned to intra-coded PUs to segment the foreground region from the background region. In this report, I suggest choosing the MV representative of the MVs of the neighboring PUs as MV. The MVs of first-order neighboring PUs are employed as:

- top-left
- top
- top-right
- left
- right
- bottom-left
- bottom
- bottom-right

MVList is created and stored for all MV neighboring PUs . *MVList* contains seven vectors, Since in this case 1 of the adjacent PUs are intra-coded, which is shown as follows:

$$MVList = (MV_1, MV_2, MV_3, MV_4, MV_5, MV_6, MV_7) .$$

Fig 5: MVList.

We need to allocate a characteristic MV for the "intra-coded PU" from list of MVs after building MVList. When there is a large amount of movement in the scene, PUs that are intra-coded usually occur. That is why we plan to select MV Lists highest MV as MV of this particular "intra-coded PU". In particular, if all adjacent first-order PUs are encoded using the intra mode to get the MV having a value greater than zero in the near region, we make the neighborhood range to 16x16 (blocks) by incrementing it, which is calculated as empirically optimal.

`

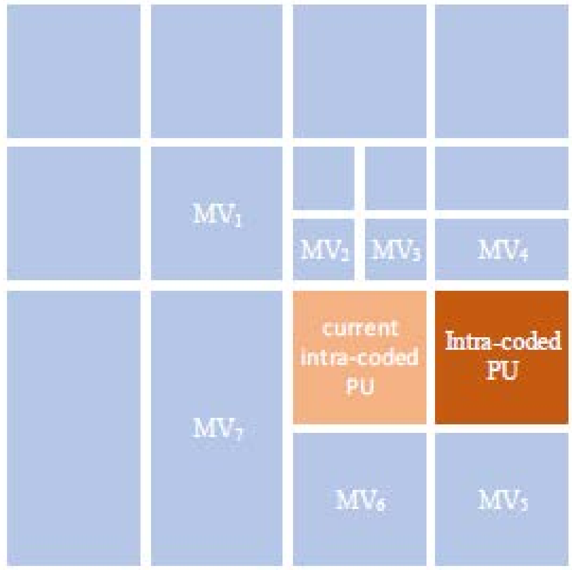

Fig 6: MV assignment for an intra-coded PU.

B. Motion Vector Outlier Removal

Compressed bit stream are calculated in terms of rate-distortion from MVs, they cannot represent the genuine motion of object that means that fields of MV maybe noisy. In this approach we have proposed to lessen the noise in MV in reference to motion coherence within the spatial neighborhood and motion continuity over time. For MV outlier removal we follow three steps that are:

- MV refining
- MV filtering
- Isolated and small MV removal.

I) Motion Vector Refining

Since we know that there is displacement relative to the moving object in current frame, between the objects that are moving in the current ad following frames, the filtering process which uses adjacent blocks in the adjacent frames can lead to value of MV which is greater than zero. Even if there is noise in original MVs but if the value is zero for original MV of the this block and most of its neighboring, then the current block has a high probability of belonging to the background.

`

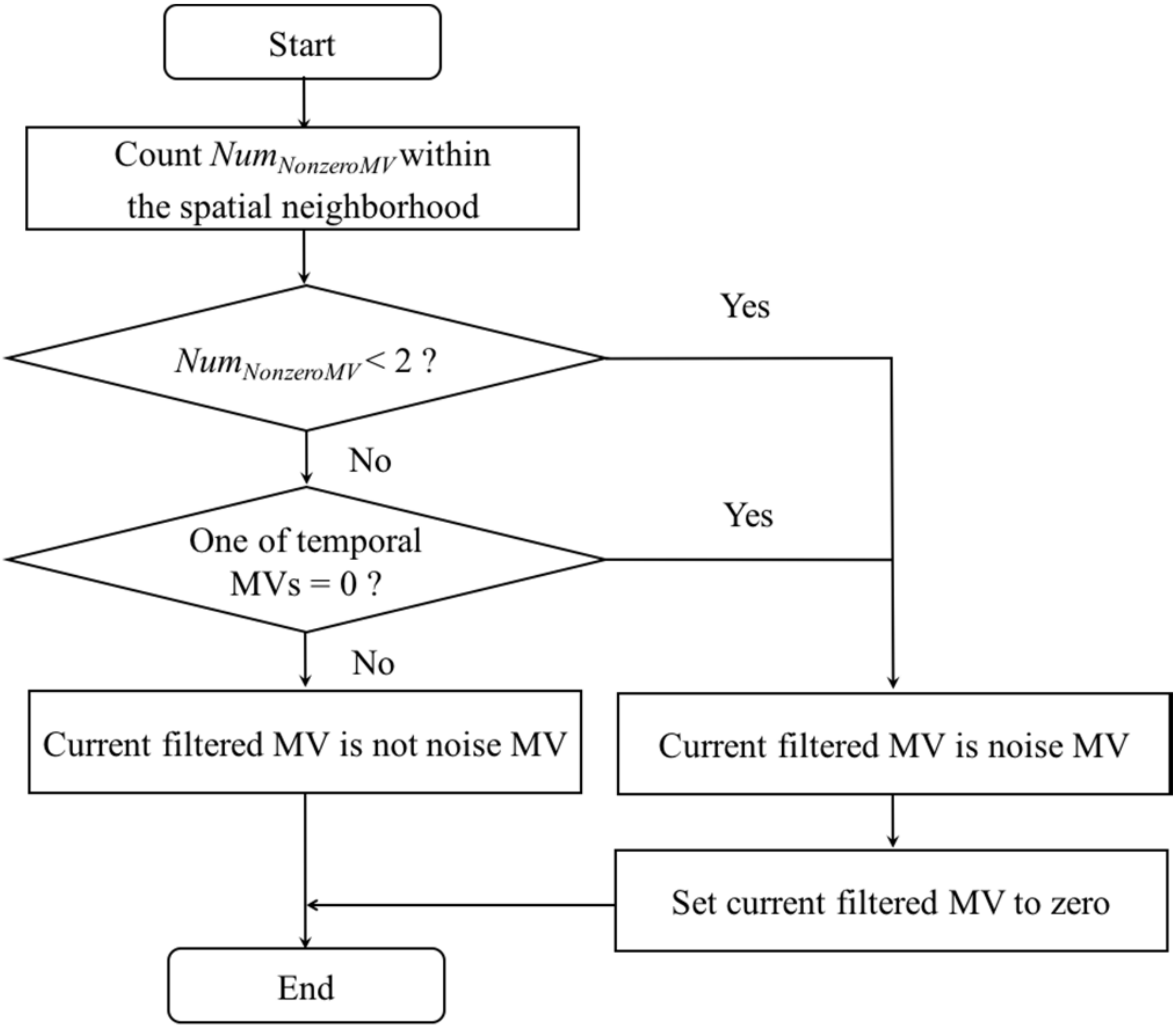

Fig 7: MV Refining

Depending on temporal continuity and spatial compactness of the MVs, temporal co-located PUs and the original MVs of spatial neighboring PUs are also included in judgment making. Original MV of that PU will be set to zero when the neighboring PU is out of the frame. In the flowchart shown in Fig. 7 of the MV refining for filtered MVs having value greater than zero. Fig.7, demonstrates that *NumNonzeroMV* denotes all the MVs having value greater than zero within the spatial neighborhood, where all the PUs are taken into account.

Firstly, if *NumNonzeroMV* has value lower than two, than we suppose that the spatial compactness condition don't satisfy.Secondly, I analyze that if the MVs in nearby temporal PUs corresponding to the two adjacent frames are greater than zero or not. If the value of even one of the MVs is found to be 0 than we suppose that the condition remains unsatisfied.

`

II) Motion Vector Filtering

Filtering of genuine MVs are done along the time-related direction. In order to filter this genuine MV at current frame *t* , MVs at the adjacent position in the *m* after frames and *m* before frames are employed. As CU and PU sizes are different among different frames at the same position, filtering of MV is used for 4x4 blocks, which is the least size of PU. MVs can be estimated by:

$$MV_t^x(k,l)' = floor\left(\frac{\sum_{i=t-m}^{t+m} MV_i^x(k,l)}{2m+1}\right),$$

$$MV_t^y(k,l)' = floor\left(\frac{\sum_{i=t-m}^{t+m} MV_i^y(k,l)}{2m+1}\right),$$

Fig 8: MV Filtering

Where;

floor(x) means that the formula will give us the integer which is highest but <= *x*.

m is set to 4.

III) Isolated and Small Motion Vector Removal

For a moving object in the foreground, it usually has a non-zero MV region and relatively larger MV filtering, so that PUs with non-zero isolated VMs or smaller MVs have a high probability of being background PUs. Therefore, we propose to label PUs with non-isolated MVs or small MVs as lower PUs. To be precise, we define an MV as an isolated MV if all MVs in their spatial neighborhood are zero MV. In addition, we define an MV as a small MV if the MV of the current PU and more than half of its adjacent spatial PU are less than or equal to one. If a PU is identified as the PU with isolated or small MV, its MV assignment is changed to zero.

"MOVING OBJECT SEGMENTATION IN HEVC COMPRESSED DOMAIN"

.A. Object Region Tracking

To examine the time-related consistency of the areas that are in foreground, these foreground areas are followed using the MVs are taken out from HEVC videos.

The flowchart of tracing the object area in the reverse direction is shown in Fig. 9, which consists of the following five main steps.

`

Step 1: The variable " *bTemporal"* is used to specify that if that foreground region and the present foreground region describes that object in reverse or not.

Step 2: Each block which is in the foreground region is predicted to the before frame *t* - 1 by using the corresponding MV. Its projected position in frame *t* – 1.

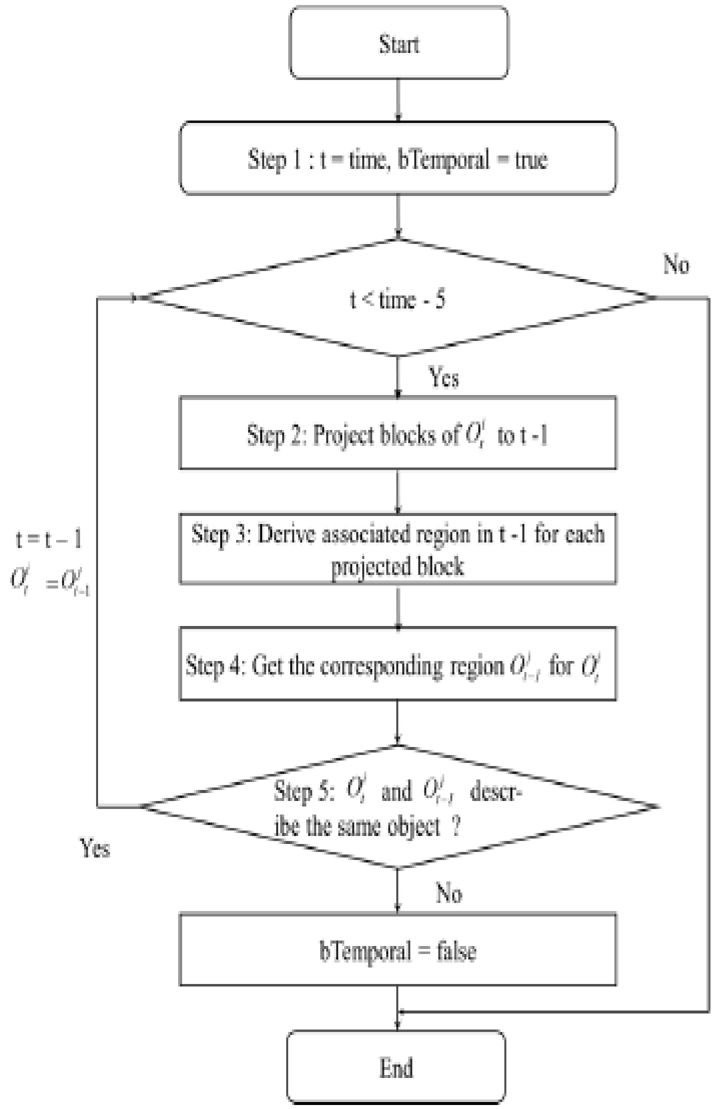

Fig 9: Object Region Tracking

Step 3: For each block that is projected at time t - 1, if it belongs to the background area, the area for is set.

`

Step 4: Count the number of all the predicted blocks in foreground region in the t-1. Then, the foreground region which is having the most predicted blocks is recognized as the corresponding region for the t box.

Step 5: The junction of and its respective foreground region is calculated and contrasted to an threshold. The size of object determines the numeric value of foreground blocks in that particular object of the region and the value is set t0 50%. If condition 5 is not met, than the process sets the value to false for bTemporal and the current process ends.

If not, we assume that the corresponding foreground region should describe the same object area. Then go back to step 1 and do perform every step again.

$$\frac{size(O_t^j \cap O_{t-1}^j)}{size(O_t^j)} > \partial \,.$$

Fig 10: Formula Object Region Tracking

B. Object Boundary Refinement

The main steps are described as follows:

Step I: Firstly we need to put the values and calculate using formula in fig 11 the average depth of the foreground region. The depth of CU in the position (k,l) is denoted by "CUDepth (k, l)", depth of PU in the position (k,l) is denoted by "PUDepth (k, l)" and depth of the block at position (k, l) is denoted by depth (k, l).

foregBlockNum- number of blocks that are there within the foreground region

floor (x) implies that the formula will give the result where the answer would be smallest integer >= x. Therefore, the avg depth formula is:

$$depth(k,l) = CUdepth(k,l) + PUdepth(k,l)\,,$$

$$AvgDepth = floor\Big(\sum_{\substack{V0 \le k \le V1 \\ H0 \le l \le H1}} \frac{depth(k,l)}{foregBlockNum} + 0.5\Big) - 1\,.$$

Fig 11: Object Boundary Refinement

`

Step II: The maximum depth is comparable to AvgDepth for each row of blocks within the region of the moving object and is given by:

$$MaxDepth(k) = \max_{H0 \leq l \leq H1}(depth(k,l)) .$$

Fig 12: Object Boundary Refinement(max depth)

If (12) is satisfied, then that particular row would be selected so that it has to be modified to background blocks.

$$MaxDepth(k) \leq AvgDepth .$$

Fig 13: Object Boundary Refinement(relation b/w max and avg depth)

Step III: Every row of candidate blocks, we also need to check that whether they are real background blocks or not. IndexMin (k) is calculated by:

$$IndexMin\,(k) = \operatorname*{argmin}_{l}(depth(k,\,l)) .$$

Fig 14: Object Boundary Refinement(Index Min)

Relation between Maxdepth and Avg depth is that max depth should be less than or equal to Avgdepth.

$$MaxDepth(MaxRowTop) \leq AvgDepth ,$$

$$MaxDepth(MinRowBottom) \leq AvgDepth .$$

Fig 15: Object Boundary Refinement(min)

`

MOVING OBJECT CLASSIFICATION IN HEVC COMPRESSED DOMAIN

For Classification we will use motion vector difference which is equal to zero if neighboring motion vector or the current motion vector comes out to be zero. If not then the value of motion vector difference of the current block would be the absolute value that comes after subtracting the motion vector of current block from the motion vector of neighboring block.

$$CurrMVD_t^i = \begin{cases} 0, & \text{if } NeighMV_t^i \text{ or } CurrMV = 0 \\ abs(NeighMV_t^i - CurrMV_t), & else \end{cases}$$

$$MaxCurrMVD_t = max(CurrMVD_t^i),$$

Fig 16: Moving Object Classification

The classification which is based on person vehicle framework is found out based on the above formula and if the value is zero than it is a person otherwise a vehicle so it is then labeled as either person or vehicle in the frames .The frames after processing are stitched to together in a video with the help of the MATLAB code .

`

4.2 Requirement Analysis

Requirement analysis or requirements engineering, is the process of determining what user wants in the product. The requirements must be relevant, detailed and measurable. There requirements are basically divided into four broad types:

4.2.1 Functional Requirements

Functional requirements basically give a detailed description about the functions that should be present in the system then only the system could be called as fully functional. Function is basically set of inputs, behavior and outputs. Functional requirements can be based :

4.2.1.1 Product Perspective

The product is made as a Capstone Final year Project . It is a Digital Image Processing Project Implemented in MATLAB for Segmentation and classification of Objects from a Surveillance video.

The system is run on MATLAB and have a user friendly GUI. The system consist of many options in which the moving object is detected. The detected moving object is person or vehicle depends on the classification through "bag of words" in HEVC compressed format. Images can sometimes doesn’t show correct results due to pixel size and other criteria therefore enhancement is used for this purpose. An initial training is required for running this software and make it friendlier.

The output we get from the system is in terms of frames, which is converted into video by MATLAB so the person who uses this software have some knowledge regarding MATLAB.

4.2.1.2 Product features

The features are:

- There all lot of interesting features in the HEVC domain, like CU and PU which are used to refine the boundary of the object moving.
- The bag of words which is available in the compressed domain of HEVC videos are used to classify the persons and vehicles which are moving.
- Computing complexity of the proposed approach is very less compared to other approaches, many people who are working in the field of video coding have implemented this approach in order to calculate the value of MV in each pixel.

`

- This happens so that the effect of "moving averaging" on block manipulations is less. The proposed method is of good use, since there is a finite support region that is always used to configure the equations.

4.2.1.3 User characteristics

It is assumed that the user of the system whether a administrator or a normal user should know the basic approach to use MATLAB and to have access to the surveillance videos. The user is anticipated to be familiar with the GUI and other basic functions of the MATLAB. There are basically 2 types of user :

Administrator: one who knows about the algorithm of the system, how the MATLAB is working and how everything is happening behind the scenes. Which changes will lead to what results is also known by administrator.

Normal user: they have the basic knowledge of how to use the software ,they are basically security officers who are continuously tracking everything to keep the security of their jurisdiction up to date.

4.2.1.4 Assumption & Dependencies

This performance of the product is completely dependent on the type and version of MATLAB which is installed on the system where the product is installed . It should be greater than or equal to R2015 and we have assumed that the system has a RAM of at least 4 GB and the videos are HEVC. User has to know about basic features of MATLAB as it will help the user to work efficiently. The system should be able to run 24*7.

4.2.1.5 Domain Requirements

These requirements basically shows the environment in which our product operates ,here we have a MATLAB software and a VLC media Player. the system is coded in MATLAB and after the processing and classification of the input video takes place then VLC media player is used to play the video.

`

4.2.1.6 User Requirements

It is assumed that the user of the system whether a administrator or a normal user should know the basic approach to use MATLAB and to have access to the surveillance videos. The user is anticipated to be familiar with the GUI and other basic functions of the MATLAB. There are basically 2 types of user :

Administrator: one who knows about the algorithm of the system, how the MATLAB is working and how everything is happening behind the scenes. Which changes will lead to what results is also known by administrator.

Normal user: they have the basic knowledge of how to use the software ,they are basically security officers who are continuously tracking everything to keep the security of their jurisdiction up to date.

4.2.2 Non Functional Requirements

There are various kinds of non-functional requirements:

4.2.2.1 Product Requirements

It is robust.

Less frequency of failures and application crashing.

The application will be available 24*7.

Reliable.

Can be maintained easily.

Additional features and customization can be adjusted in the application.

4.2.2.1.1 Efficiency (in terms of Time and Space)

Efficiency of time and space depends on the size of video given as a input to the product.

For eg a video named "vision traffic" is broken into 529 frames which approximately takes 1.30 minutes for execution .

4.2.2.1.2 Reliability

Since the system is used for intelligence video surveillance it should be reliable at all times and it should classify correctly at all times. The current system is very reliable and has a accuracy of more than 85 % in average.

`

4.2.2.1.3 Portability

Portability means the usability of the same system in different environments and since MATLAB works on all desktop operating systems which implies this product is portable to all operating systems.

4.2.2.1.4 Usability

The usability of the system now totally depends on the knowledge of the user as the system is created in such a way so that it is very user friendly and the system has no delays. the system reacts accordingly and the speed at which it travel between different module is fast as well. The proposed system is very user friendly and it is very efficient.

4.2.3 Engineering Standard Requirements

Economic

This system is built on MATLAB codes. So apart from MATLAB software it doesn’t require any other software. One system in which the MATLAB is running is enough. So economically our software has very less cost. It only requires time to code, to build the system. Apart from that it does not cost for too many softwares. According to the time it takes to build – rate can be decided which is not too expensive.

Environmental

This system does not affect environment by any means. It saves people, it gives surety to the people affected by road rage, terrorism and security issues that there is someone who is using this software and they are safe. So, basically it helps people and helps society. It doesn’t have any hardware equipment to harm environment. It just require videos from cameras to help save people. So this work is for saving human life, it can’t affect environment direct or indirectly.

Social

This system identifies moving object in the surveillance video . Social implications of this device are huge, it can save so many lives by detecting and classifying any such activity. It can save so much time and lives and can also stop some not so harmonious thing before it happens.

`

Political

This software is simply a moving object detection and classification software which helps the security personnel in order to increase the saftey and security of the general public. So there is no issue of political amendments required to build and run this software. The algorithms we used are common, no specific dark web algorithms are used. In India this type of work is praised. So no political issues involved.

Ethical

From a moral point of view, the software can't give wrong results. The end clients gets information specifically from the software by means of frames and graphs which cannot be messed around with. So the end user can get the result correctly. So if the result has issues, it is not from the system, it is from the end-user. The image might be corrupted or the the user might seen something else. But this can't affect the system, it always gives correct results.

Health and Safety

This product is extremely safe for all the users. It is simply a software installed in a computer. So it can harm only as much as the computer harms you. It will not cause any bodily harm to anyone using it. Also, it is used for the purpose of detection of moving objects. So basically it saves people not harm people. It has widespread applications in the field of internal security.

Sustainability

The field of video surveillance classification is relatively new in the world. Not many products have been built around it. The ones that have been built are expensive or do not boast of latest technology. This field is more popular in developed countries than in developing countries. The aim of this project is to make a product which is affordable and usable for common people in low income countries as well. So, the product will sustain in the market once it is launched.

Legality

This product is legal in all sense. It doesn't use anything which is illegal and prohibited for use. All the products are genuine and tested. Also, none of the features are used in the system are banned in any country so people will not have any problem in acquiring the result from surveillance videos.

`

Inspectabilty

The software shows images which are real-time video data from surveillance videos which are updated at a regular interval of time. Depends on the type of moving object one can define what is the moving object.

4.2.3 System Requirements

These are the basic system requirements that the system should have for successful implementation and working of the project.

4.2.3.1 H/W Requirements(details about Application Specific Hardware)

Processor	:	Pentium Dual Core 2.00GHZ
Hard Disk	:	500 Gigabytes
Random Access Memory	:	Four Gigabytes (minimum)
Keyboard	:	110 keys enhancement.

4.2.3.2 S/W Requirements(details about Application Specific Software)

MATLAB 8.6 Version R2015b & VLC media Player

MATLAB is a very helpful language in this particular project as it has video processing toolbox present which helps in image processing and neural networks toolbox and built in commands also help a lot in this particular project.

VLC media player is used for playing the videos.

`

5. Results and Discussion

Testing is considered to be the final step of the procedure that is software development. In software development we first get the user requirements, upon getting the requirements we make a general design, after general design we make a detailed design after the designing is over we go to coding part where we code each and every module after coding we need to do some testing.

Before we start outlining tests, it is imperative to have some reasonable comprehension of the reason of why we were testing. The understanding of which impacts which tests we stretch and how early we should start testing. When all is said and done we test for four reason:

When we need to discover bugs in our programming, the best way out is testing in order to mitigate risks for both client and organization.

To lessen development and upkeep costs so that the performance of the system can be enhanced.

Reduce costs: The costing of testing is low if the bug is found earlier. Once a product is finished and dispatched than the cost of finding and removing bug is too high.

To enhance performance: for performance of the system to be enhanced, we need to find and remove out dead code and also the code that reduce the efficiency of the system. We should guarantee that the product utilizes the maximum capacity of the equipment and accordingly stays away from the feared "hardware re-spin."

`

5.1 Sample Test Case

The test cases are designed to test the functioning of the device and ensure that the device performs as it is expected to. They help us in making qualitative judgments with respect to the degree to which the product (system) meets, surpasses or fails to meet stated goals. These test cases ensure that the device remains robust across varying conditions to which it is subjected to. Some of the test cases are specified in the following tables:

Test Case ID:	1.1	Test Case Engineer:	Sumit
Product Module:	Classification of Moving objects from HEVC Videos	Test Date:	21/03/2018
Product Version:	1	Testing Cycle:	1
Revision History:		Status:	Pass
Purpose:	To check if the all the algorithms are working fine		
Assumptions:	MATLAB is running		
Pre-conditions:	Database of surveillance video		
Steps to Reproduce:	1. Input the video. 2. Check algorithm one by one		
Expected Results:	Preprocessing		
Actual Outcome:	Preprocessing of frame takes place		

Table 1: Test Case A

`

Test Case ID	1.2	Test Case Engineer	Sumit
Product Module:	Classification of Moving objects from HEVC Videos	Test Date:	21st march 2018
Product Version:	1	Testing Cycle:	1
Revision History:		Status:	Pass
Purpose:	To check if the all the algorithms are working fine		
Assumptions:	MATLAB is running		
Pre-conditions:	Database of surveillance video		
Steps to reproduce:	1. Input the video. 2. Check algorithm one by one		
Expected Results:	Frame is segmented		
Actual Outcome:	Segmentation of frame takes place		

Table 2: Test Case B

`

Test Case ID	1.3	Test Case Engineer	Sumit
Product Module:	Classification of Moving objects from HEVC Videos	Test Date:	22nd march
Product Version:	1	Testing Cycle:	1
Revision History:		Status:	Pass
Purpose:	To check if the all the algorithms are working fine		
Assumptions:	MATLAB is running		
Pre-conditions:	Database of surveillance video		
Steps to reproduce:	1. Input the video. 2. Check algorithm one by one		
Expected Results:	Moving object extraction		
Actual Outcome:	Moving object extraction takes place		

Table 3: Test Case C

`

Test Case ID	1.4	Test Case Engineer	Sumit
Product Module:	Classification of Moving objects from HEVC Videos	Test Date:	22nd march
Product Version:	1	Testing Cycle:	1
Revision History:		Status:	Pass
Purpose:	To check if the all the algorithms are working fine		
Assumptions:	MATLAB is running		
Pre-conditions:	Database of surveillance video		
Steps to reproduce:	1. Input the video. 2. Check algorithm one by one		
Expected Results:	Classification		
Actual Outcome:	Classification of moving object takes place		

Table 4: Test Case D

`

Test Case ID	2.1	Test Case Engineer	Sumit
Product Module:	Complete system on 1st dataset accuracy	Test Date:	23rd march
Product Version:	1	Testing Cycle:	1
Revision History:		Status:	Pass
Purpose:	To find the working of the complete system after classification		
Assumptions:	MATLAB and program is running successfully.		
Pre-conditions:	Database of surveillance video. User should know which datset should be selected.		
Steps to reproduce:	1. Input the video. 2. Check the output video after all the processing takes place 3.Check the classified results.		
Expected Results:	95%		
Actual Outcome:	90%		

Table 5: Test Case E

`

Test Case ID	2.2	Test Case Engineer	Sumit
Product Module:	Complete system on 2nd dataset accuracy	Test Date:	23rd march
Product Version:	1	Testing Cycle:	1
Revision History:		Status:	Pass
Purpose:	To find the working of the complete system after classification		
Assumptions:	MATLAB and program is running successfully.		
Pre-conditions:	Database of surveillance video. User should know which datset should be selected.		
Steps to reproduce:	1. Input the video. 2. Check the output video after all the processing takes place 3.Check the classified results.		
Expected Results:	Accuracy should be 95%		
Actual Outcome:	92%		

Table 6: Test Case F

`

5.2 Summary of the Results

This picture shows the GUI of MATLAB as well as what is the view after the code has run and we have got our output saved to output1 video.

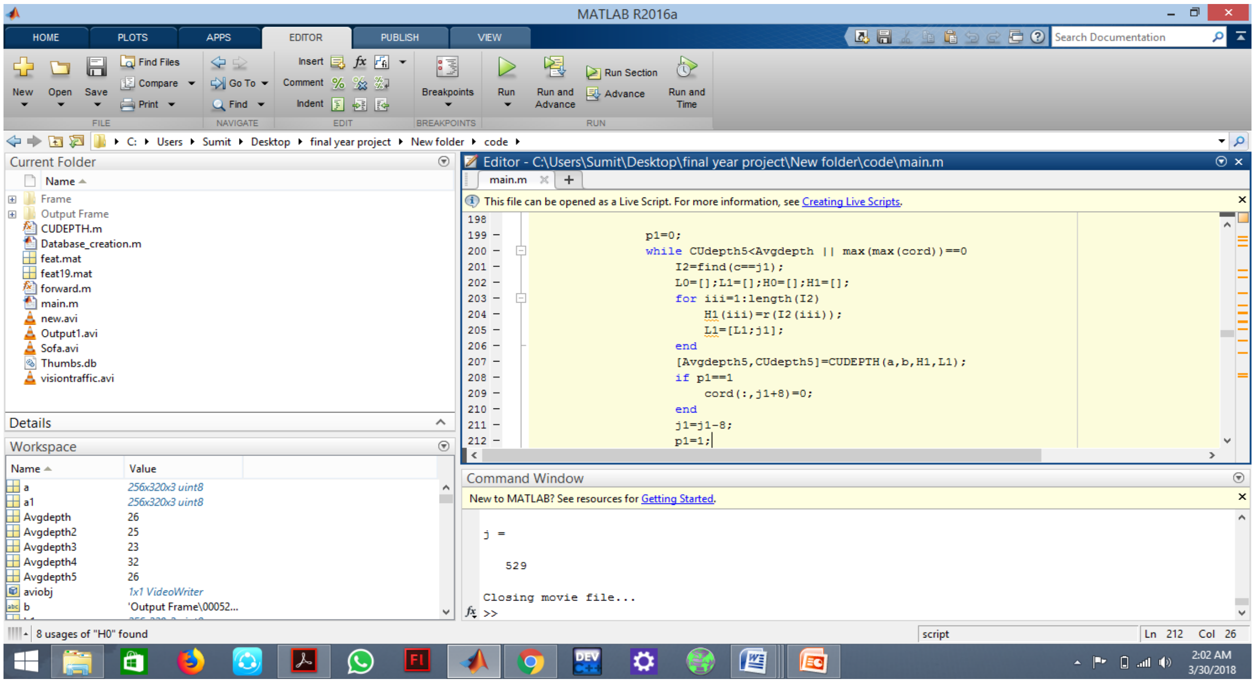

Fig 17: GUI Matlab

1. Preprocessing:

It is performing preprocessing on each frames and this show how the blocks are created around moving objects,

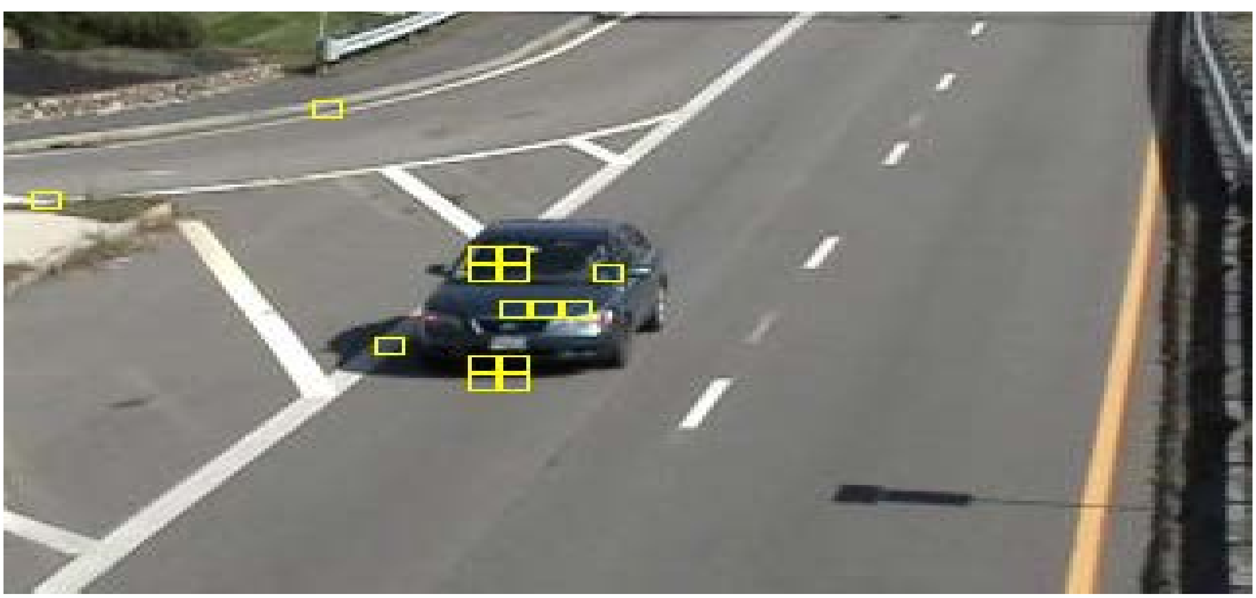

Fig 18: Output of preprocessing

`

2. Object region tracking:

In this step basically tracking of the motion part of the frame takes place and blocks are created around it.

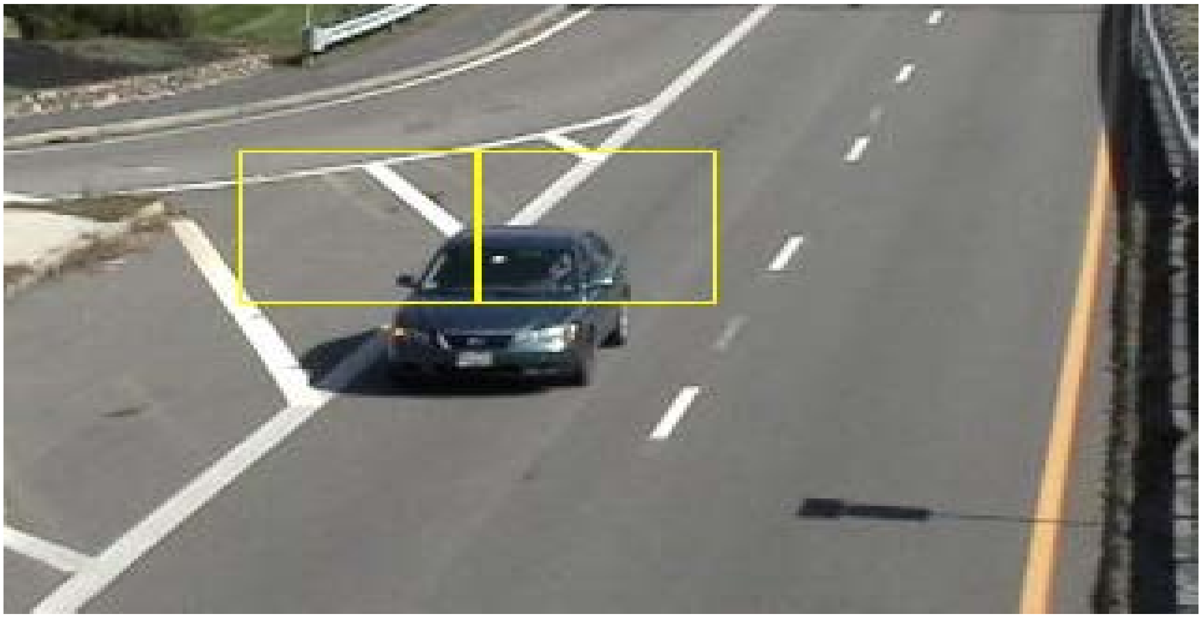

Fig 19: Object region Tracking

3. Object Boundary Refinement:

In this step the object which we found out in previous step the boundary of that object is refined.

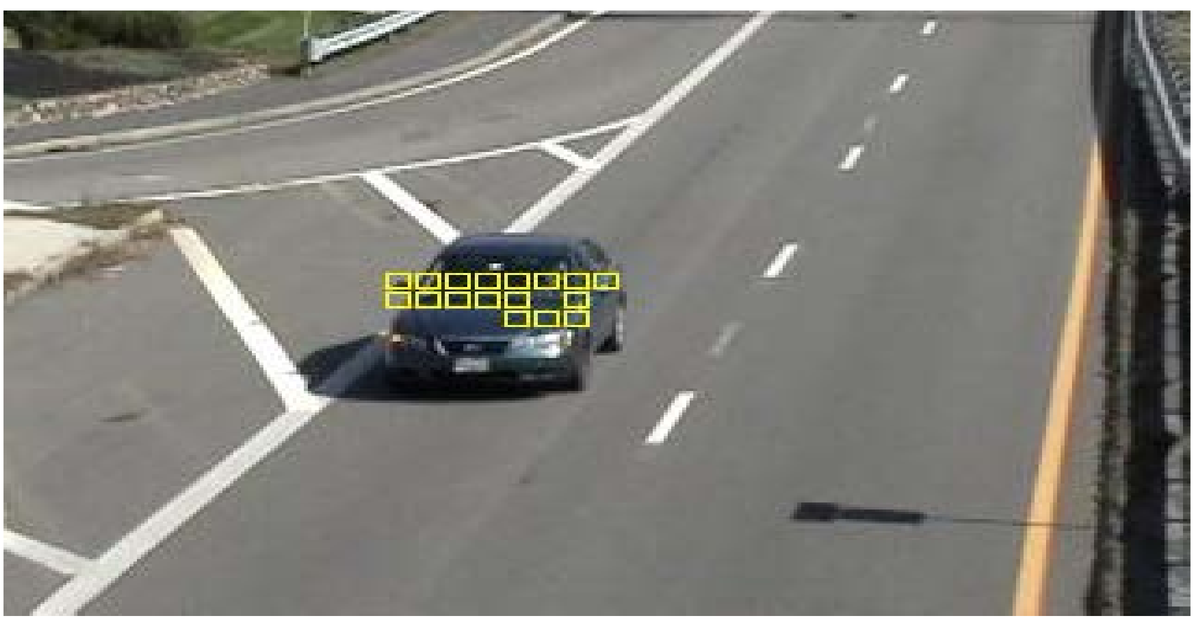

Fig 20: Object Boundary Refinement

`

4. Object Segmentation:

In this step the segmentation of the object finally takes place.

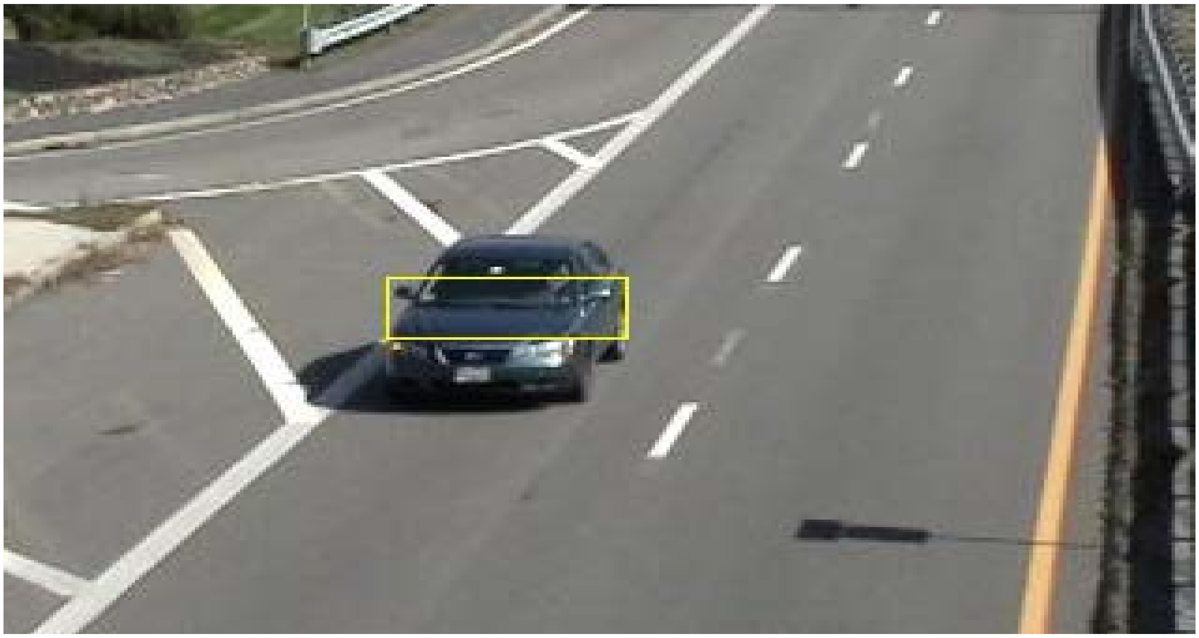

Fig 21: Segmentation

5. Output:

The final step showing the output as vehicle.

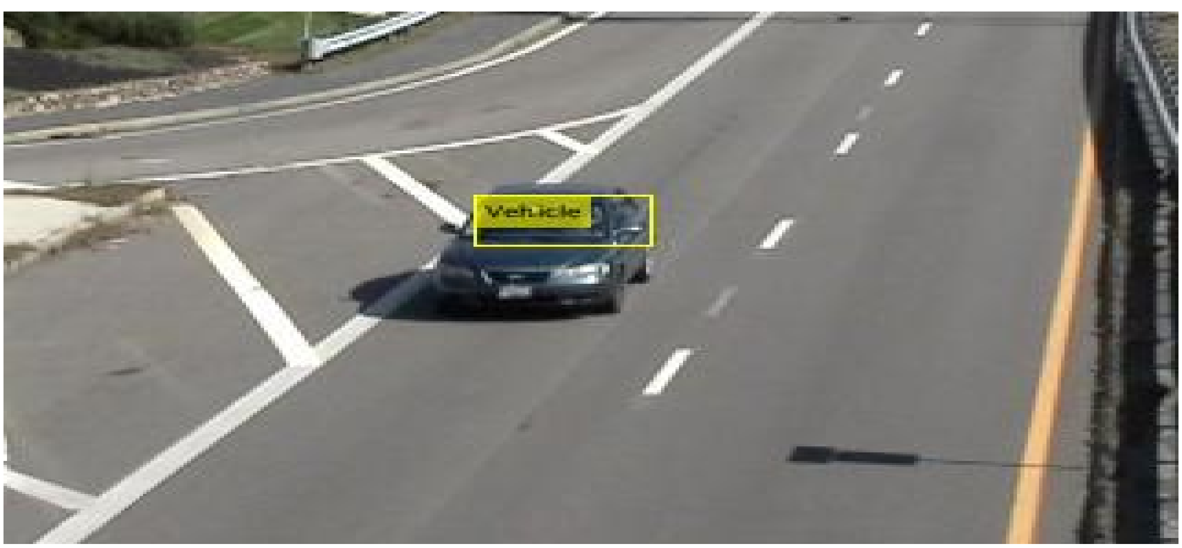

Fig 22: Classification

`

6. Conclusion

This document provides a approach to segment and classify objects in motion of the compressed monitoring HEVC video. In the proposed method only the MVs and the coding modes associated of the compressed stream are used.

Firstly we use interpolation of MV technique for intra-coded PU, for preprocessing we employ MV outlier removal . Then, all the motion vector blocks having value greater than zero are grouped with help of a algorithm called "4 connectivity component labeling" into a connected regions of foreground objects. Finally, we apply object region tracking which has temporal consistency to the connected foreground regions to extract and remove the noise regions. The border/boundary of the moving object region is further refined by using the CU and PU sizes of the blocks. We also used a function to calculate the accuracy of the model and that comes out to be nearly 92%.

I conclude that the approach which I followed worked fine and all the modules were working fine and we got the output and I got a accuracy of around 92% which is pretty god for this model and in future the accuracy could be even better.

Scope for future work

Our project being open source, there is immense scope in adding more and more modules to the system and customizing it as per the needs of specific users.

In our project, we used many algorithms which is used before in a different way. The work we did to build this system is unique in some aspects but many are trying to do best. We used surveillance videos to find the output but we know that they were stored on our system, but we have still to find that how to implement this on live videos. So, in our future work we build the system in which classification can be done on live surveillance videos. This system can also be incorporated with facial detection and can be matched with the "AADHAR" database so that person can be identified at that point only as we know more than 80% people in India have these cards. This system can also be used to detect the owner of the vehicle by scanning the number plate. 2.0 will definitely incorporate all the above suggestions.

`

7. 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 RecommendationH.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. VideoTechnol.*, 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. Ramakrishnan, H. S. Srinivasan, “Video object segmentation: a compression domain approach,” *IEEE Trans. CircuitsSyst. 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, “A joint approach to global motion estimation and 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] 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.

[15] 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.

[16] 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.

[17] M. Laumer, P. Amon, A. Hutter, and A. Kaup, “Compressed domain moving object detection by spatio-temporal analysis of H.264/AVC syntax elements,” in *Proc. IEEE Conf. Picture Coding Symposium (PCS)*, pp.282-286, May 2015.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] M. Moriyama, K. Minemura and K.S. Wong, "Moving object detection in HEVC video by frame sub-sampling," *IEEE Int. Symposium on ISPACS*, pp. 48-53, Nov. 2015.

[30] 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.

[31] 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

[32]Detection & tracking of multiple objects in cluttered backgrounds with occlusion handling by Sukanyathara J and Alphonsa Kuriakos

[33]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.