International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 347-360 www.ijaem.net ISSN: 2395-5252

Quality Enhancement of Degraded Video and Object Tracking With Local Binary Pattern Approach

Mr. Suyash G. Deshmukh, Dr. S. R. Gupta,
M.E. Scolar, P. R. M.I.T. R., Badnera
Professor, P. R. M.I.T. R., Badnera

Submitted: 01-08-2021 Revised: 07-08-2021 Accepted: 10-08-2021

I. CHAPTER 1

INTRODUCTION

Video surveillance has an objective to monitor a given environment and report the information about the observed activity that is of significant interest. In this respect, video usually utilizes electro-optical sensors that is video cameras to collect information from the environment. Moving object detection and tracking of a video image signals, by using visible light image sensor a thermal infrared, low light level imaging sensor uptake of the moving target. After the corresponding digital image processing, detection and extraction of moving targets in video file is performed [1]. The detection and tracking of moving targets both are the closely related processes. Detection is the basis of tracking, and tracking is to obtain the target motion parameters, such as position, velocity and trajectory, for the subsequent motion analysis, understanding the motion of the target behavior and to provide reliable data source to complete higher level mission and provide help for moving target detection.

Digital cameras, and in particular binocular stereo rigs, at the moment do not reach the geometric accuracy of range sensors such as LIDAR, but offer the advantage that in addition to the scene geometry they deliver rich appearance information, which is more amenable to semantic interpretation. Recent work has shown that with modern computer vision tools, visual environment modelling for robot navigation is becoming possible [2]. A key component of these approaches is that they strongly rely on semantic object category detection—in the context of road traffic especially detection and tracking of pedestrians and cars.

To support dynamic path planning, it is not sufficient to detect those scene objects; one also has to track them i.e. estimate their trajectories over time to be able to predict their future locations. As the two tasks of detection and tracking are closely related: several of the most successful tracking methods at present follow the tracking-by-detection paradigm, in which the output of (appearance-based) object detectors serves as observation for tracking. The task of multi-object tracking then amounts to linking the right detections across time to form object trajectories [3]. The approach presented here extends the tracking-by-detection framework to better cope with difficult scenarios with many moving objects close to each other.

In a typical surveillance system, these video cameras are mounted in fixed positions or on pan-tilt devices and transmit video streams to a certain location, called monitoring room [2]. Then, the received video streams are monitored on displays and traced by human operators. However, the human operators might face many issues, while they are monitoring these sensors. One pro fact that the operator must navigate through the cameras, as the suspicious object moves between the limited field of view of cameras and should not miss any other object while taking it. Thus, monitoring becomes more and more challenging, as the number of sensors in such a surveillance network increases. Therefore, surveillance systems must be automated to improve

1.1 Motivation

The rapid improvement in technology makes video acquisition sensors or devices better in compatible cost. This is the cause of increasing the applications that can more effectively utilize digital videos. So now, more information is present in the video about the object and background that are changing with respect to time. The area of video tracking is currently of immense interest due to its implication in different functional areas. Therefore

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 347-360 www.ijaem.net ISSN: 2395-5252

it is seen that there is a wide range of research possibilities are open in relation to video tracking. Along with this, detecting and tracking of objects in a particular video sequence or any surveillance camera is really a challenging task in computer vision application. Video processing is really time consuming due to huge number of data is present in the video sequence. But as the scope is growing in normally all application areas. It is necessary to develop methods for proper and efficient object detection and tracking.

1.2 Aim

This system aims to perform object detection and tracking as an important challenging task within the area of Computer Vision that try to detect, recognize and track objects over a sequence of images called video. It helps to understand and describe object behavior instead of monitoring computer by human operators. Here the system aims to detect moving objects from the video file or surveillance camera. It will try to improve the invention of high quality of the imaging sensor, quality of the images and resolution of the images with proper and efficient algorithms.

1.3 Objectives

The current dissertation work is dedicated to achieve some of the following objectives:

- Enhancement of low quality degraded video to quality video with higher frame quality.
- To improve the speed and accuracy of object detection and tracking technique used for finding target object.
- To increase quality of frame that works well in blur image, camera motion, illumination and scale conditions.
- To find target object and match with each frames in video by using object detection and object tracking methodology.

1.4 Scope

With the decrease in costs of hardware for sensing and computing, and increase in the processor speeds, this system aims to provide robust surveillance at an affordable price. There is wide scope of this system as the surveillance systems have become commercially available, and they are now applied to different number of applications, such as traffic monitoring, airport and bank security etc. With the current advance techniques like Haar Wavelet decomposition, the video quality gets improved, it becomes useful for different video processing applications. With quality frames and using Template matching methodology, object detection and tracking makes the surveillance task more accurate and easy to handle. This makes the system more useful in all its application areas.

1.5 Organization of report

- Chapter 1, covers the introduction Video surveillance, data warehousing, issue is motion analysis techniques, motivation, objective and scope of the dissertation.
- Chapter 2, covers the various techniques used for data processing also as multimedia data processing.
- Chapter 3, during this section the architecture of proposed system, its working methodology is discussed intimately.
- Chapter 4, this section covers the implementation of audio data processing with its result and graphs.
- Chapter 5, this section covers the various Applications and advantages.
- Chapter 6, this section covers the conclusion and future scope of the project.

II. CHAPTER 2

LITERATURE REVIEW

2.1 Background History

The development of video databases has impelled research for structuring multimedia content. Traditionally, low-level descriptions are provided by image and video segmentation techniques. The best segmentation is achieved by the human eye, performing simultaneously segmentation and recognition of the object thanks to a strong prior knowledge about the objects' structures. To generate similar high-level descriptions, a knowledge representation should be used in computer based systems. One of the challenges is to map efficiently the low-level descriptions with the knowledge representation to improve both segmentation and interpretation of the scene [13].

There are three key steps in video analysis: detection of interesting moving objects, tracking of such objects from frame to frame, and analysis of object tracks to recognize their behavior [13]. In its simplest form, Segmentation of moving objects in image sequences is one of the key issues in computer vision, since it lies at the base of virtually any scene analysis problem. In particular, segmentation of moving objects is a crucial factor in content- based applications such as interactive TV, content-based scalability for video coding, content-based indexing and retrieval, etc. Obviously, such applications require an accurate and stable partition of an image sequence to semantically meaningful objects.

Here, only the representative video surveillance systems are discussed for better understanding of the fundamental concept. Tracking is the process of object of interest within a sequence of frames, from its first appearance to its last. The type of object and its description within the system depends on the application. During the time that it is present in the scene it may be occluded by other objects of interest or fixed obstacles within the scene. A tracking system should be able to predict the position of any occluded objects. Object tracking systems are typically geared towards surveillance application where it is desired to monitor people or vehicles moving about an area [14].

The basic framework of moving object detection for video surveillance is shown in figure below.

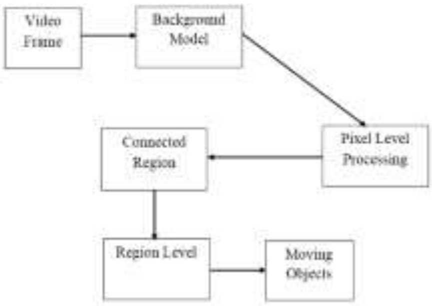


Figure 2.1.1: Basic framework for Video Object Detection System

In computer vision and video processing areas, moving object detection is a very important research topic. The process of moving object detection in video consists of two steps - background extraction and moving object detection. The preliminary idea is to capture a series of video pictures at regular intervals; the video is divided into n number of frames to describe the vector information of the region. This is the basic framework for all types of video stream or file as shown in the figure 2.1.1. To the output of this framework, different techniques needs to be applied for proper extraction of required objects. The proposed system here uses proper techniques and algorithms to extract useful object from the above framework.

2.2 Related Work

There are several number of methods and techniques are performed in this area for detecting object from video frames. Out of which some important related work done by different authors are given below. The authors in [1], presented a novel approach for multi-object tracking, that couples object detection and trajectory estimation in a combined model selection framework. This approach does not rely on a Markov assumption, but can integrate information over long time periods to revise its decision and recover from mistakes in the light of new evidence. As this approach is based on continuous detection, it can operate with both static and moving objects.

2.3Summary & Discussions

Sr. No.	Paper Title	Author	Year	Proposed Method	Limitations
1	Coupled Object Detection and Tracking from Static Cameras and Moving Vehicles	Bastian Leibe, Konrad Schindler, Nico Cornelis, and Luc Van Gool	2008	Multi-object tracking that couples object detection and trajectory estimation in a combined model selection framework	the system as a whole is not yet capable of real-time performance

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 347-360 www.ijaem.net ISSN: 2395-5252

2	Video Object Tracking Mechanism	R. Bharathi	2014	LSK Object tracking and Bayesian filtering framework	The tracking area gets smaller resulting, after 160 frames
3	Improved Multiple Object Detection and Tracking Using KF-OF Method	Akshay S, Sajin Thomas, Ram Prashanth A	2016	combination of Optical flow and Kalman filter	segmentation and tracking algorithms needs to be improved & building higher-level intelligence applications based on motion tracks needs to be extended
4	Robust Object Tracking via Sparse Collaborative Appearance Model	W. Zhongz	2014	Sparse collaborative model that exploits Both holistic templates and local representations	Runs at 3 frames per second
5	A Spatial Mean and Median Filter For Noise Removal in Digital Images	N. Kumar	2015	Mean and Median image filtering algorithms	Mean filter total noise not reduced

Table 2.3.1: Summary of literature study

2.4 Methodologies Used

The main methods used in this work are as follows:

1. Haar Wavelet Transform

Wavelet:

A wave is a fluctuating function of time or space and is periodic. In contrast, wavelets are localized waves. Wavelet means a “small waves”. Wavelets are mathematical tools for stratified decomposing functions. Wavelets are mathematical functions which help in representing the original image into an image in frequency domain, which can else be divided into sub band images of different frequency components.

Haar Wavelet Transform:

The Haar wavelet is a sequence of rescaled "square-shaped" functions which together form a wavelet family or basis. The Haar sequence was proposed in 1909 by AlfrédHaar. Haar used these functions to give an example of an orthonormal system for the space of square-integrable functions on the unit interval [0, 1].

One such type wavelet transform used here is Haar Wavelet Transformation. Haar wavelet enumerate a wavelet transform to represent image. It is the basic transformation from space to a local frequency domain. A HWT disintegrate each signal into two components, one is called average (approximation) or trend and the other is known as difference (detail) or fluctuation. This process is repeated repeatedly upto desired number levels by taking consideration of size of image /frame in the video.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 347-360 www.ijaem.net ISSN: 2395-5252

Properties of Haar Transform:

- Haar Transform is real and orthogonal.
- The basis vectors of the Haar matrix are consecutively organized.
- Orthogonally: The original signal is split into a low semifinal matrix (T) whose rows and columns have a high frequency part and filters enabling the diverging without replicating information are said to orthogonal.
- Linear Phase: To obtain linear phase, symmetric filters would have to be used.
- Perfect reconstruction: If the input signal is transformed and inversely modified using a set of weighted basis functions and the reproduced sample values are equivalent to those of the input signal, the transform is said to have the perfect reconstruction property

2. Template Matching Methodology

Template matching is a powerful technique in digital image processing for finding small parts of an image which match a template image. This can also be used for classifying objects. Template matching techniques compare portions of images against one another. Sample image may be used to recognize similar objects in source image. Templates are most often used to identify printed characters, numbers, and other small, simple objects [17].

In various fields, there is a necessity to detect the target object and also track them effectively while handling occlusions and other included complexities. Many researchers (Almeida and Guting 2004, Hsiao-Ping Tsai 2011, Nicolas Papadakis and Aure lie Bugeau2010) attempted for various approaches in object tracking. The nature of the techniques largely depends on the application domain. Some of the research works which made the evolution to proposed work in the field of object tracking are depicted as follows.

OBJECT DETECTION

Object detection is an important task, yet challenging vision task. It is a critical part of many applications such as image search, image auto-annotation and scene understanding, object tracking. Moving object tracking of video image sequences was one of the most important subjects in computer vision. It had already been applied in many computer vision fields, such as smart video surveillance (ArunHampapur 2005), artificial intelligence, military guidance, safety detection and robot navigation, medical and biological application. In recent years, a number of successful single-object tracking system appeared, but in the presence of several objects, object detection becomes difficult and when objects are fully or partially occluded, they are obtruded from the human vision which further increases the problem of detection. Decreasing illumination and acquisition angle. The proposed MLP based object tracking system is made robust by an optimum selection of uniquefeaturesandalsobyimplementingtheAdaboost strongclassificationmethod.

Existing Methods:

2.1 ResNet

To train the network model in a more effective manner, we herein adopt the same strategy as that usedfor DSSD(theperformance oftheresidual network is betterthanthat oftheVGGnetwork). The goal is to improve accuracy. However, the first implemented for the modification was the replacement of the VGG network which is used in the original SSD with ResNet. We will also add a series of convolution feature layers at the end of the underlying network. These feature layers will gradually be reduced in size that allowed prediction of the detection results on multiple scales. When the input size is given as 300 and 320, although the ResNet–101 layer is deeper than the VGG–16 layer, it is experimentally known that it replaces the SSD's underlying convolution network with a residualnetwork, anditdoesnotimproveitsaccuracybutratherdecreasesit.

2.2 R-CNN

To circumvent the problem of selecting a huge number of regions, Ross Girshick et al. proposed a method where we use the selective search for extract just 2000 regions from the image and he called them region proposals. Therefore, instead of trying to classify the huge number of regions, you can just work with 2000 regions. These 2000 region proposals are generated by using the selective search algorithmwhichiswrittenbelow.

SelectiveSearch:

1. Generatetheinitialsub-segmentation, wegeneratemanycandidateregions
2. Use thegreedyalgorithmtorecursivelycombinesimilarregionsintolargerones
3. Usegeneratedregionstoproducethefinalcandidateregionproposals

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 347-360 www.ijaem.net ISSN: 2395-5252

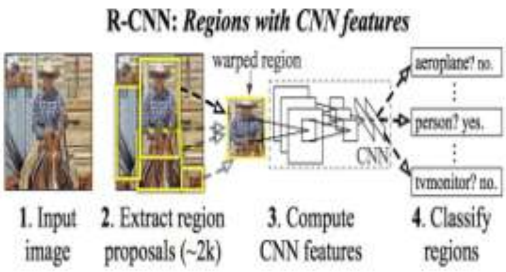


Figure 2.2.1: R-CNN Regions with CNN Features

These 2000 candidate regions which are proposals are warped into a square and fed into a convolutional neural network that produces a 4096-dimensional feature vector as output. The CNN plays a role of feature extractor and the output dense layer consists of the features extracted from the image and the extracted features are fed intoan SVM for the classify the presenceofthe object within that candidate region proposal. In addition to predicting the presence of an object within the region proposals, the algorithm also predicts four values which are offset values for increasing the precision of the bounding box. For example, given the region proposal, the algorithm might have predicted the presence ofapersonbutthefaceofthatpersonwithinthatregionproposalcouldhavebeencutinhalf. Therefore, theoffsetvalueswhichisgivenhelpinadjustingtheboundingboxoftheregionproposal.

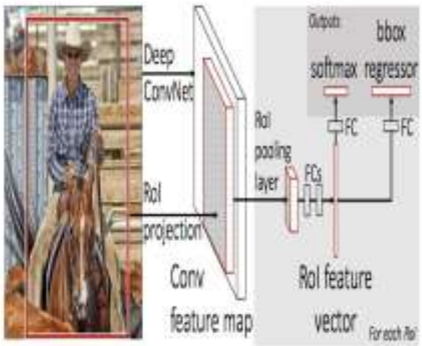


Figure2.2.2 :R-CNN

2.2.1 Problems with R-CNN

- It still takes a huge amount of time to train the network as you would have to classify 2000 region proposalsperimage.
- Itcannotbeimplementedrealtimeasittakesaround 47secondsforeachtestimage.
- The selective search algorithm is a fixed algorithm. Therefore, no learning is happening at that stage.Thiscouldleadtothegenerationofbadcandidateregionproposals.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 347-360 www.ijaem.net ISSN: 2395-5252

2.3 Fast R-CNN

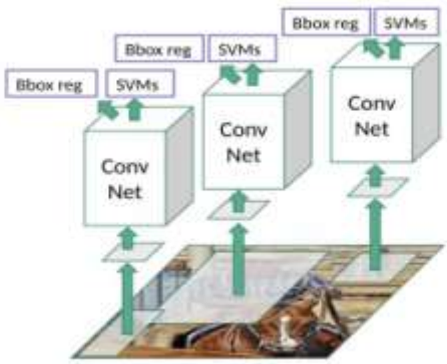


Figure2.3.1 :FastR-CNN

The same author of the previous paper(R-CNN) solved some of the drawbacks of R-CNN to build a fasterobject detection algorithmand itwas calledFast R-CNN.Theapproachis similartothe R-CNN algorithm. But, instead of feeding the region proposals to the CNN, we feed the input image to the CNNtogenerateaconvolutionalfeaturemap.Fromthe convolutionalfeaturemap, wecanidentifythe region of the proposals and warp them into the squares and by using anRoI pooling layer we reshape them into the fixed size so that it can be fed into a fully connected layer. From the RoI feature vector, wecanuseasoftmaxlayertopredicttheclassoftheproposedregionandalsotheoffsetvaluesforthe boundingbox.

The reason “Fast R-CNN” is faster than R-CNN is because you don’t have to feed 2000 region proposals to the convolutional neural network every time. Instead, the convolution operation isalways doneonlyonceperimageandafeaturemapisgeneratedfromit.

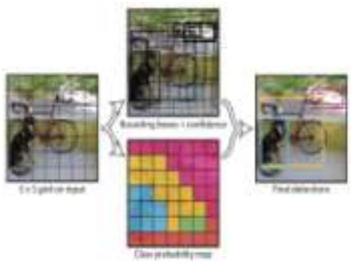


Figure2.3.2 : Comparison of object detection algorithms

2.4 Faster R-CNN

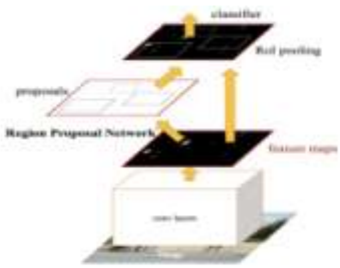


Figure 2.4.1: Faster R-CNN

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 347-360 www.ijaem.net ISSN: 2395-5252

Both of the above algorithms(R-CNN & Fast R-CNN) uses selective search to find out the region proposals. Selective search is the slow and time-consuming process which affect the performance of thenetwork.

Similarto Fast R-CNN,theimageisprovidedasaninputtoaconvolutionalnetworkwhichprovidesa convolutional feature map. Instead of using the selective search algorithm for the feature map to identify the region proposals, a separatenetwork is used to predict the region proposals. Thepredicted the region which is proposals are then reshaped using an RoI pooling layer which is used to classify theimagewithintheproposedregionandpredicttheoffsetvaluesfortheboundingboxes.

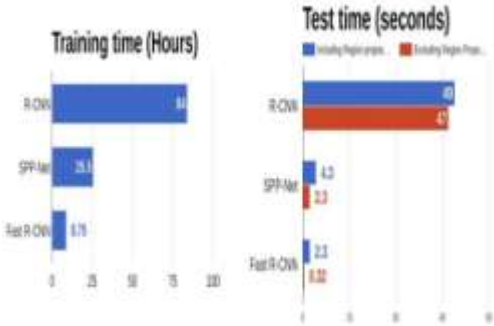


Figure2.4.2 : Comparison of test-time speed of object detection algorithms

Fromtheabovegraph, youcanseethatFasterR-CNNismuchfasterthanit'spredecessors.Therefore, itcanevenbeusedforreal-timeobjectdetection.

2.5 YOLO—YouOnlyLookOnce

All the previous object detection algorithms have used regions to localize the object within the image. The network does not look at the complete image. Instead, parts of the image which has high probabilities ofcontaining the object. YOLO or You Only Look Once isan object detection algorithm muchisdifferentfromtheregionbasedalgorithmswhichseenabove.In YOLOasingle convolutional networkpredictstheboundingboxesandtheclassprobabilitiesfortheseboxes.

Figure 2.4.3: YOLO

YOLO works by taking an image and split it into an SxS grid, within each of the grid we take m bounding boxes. For each of the bounding box, the network gives an output a class probability and offset values for the bounding box. The bounding boxes have the class probability above a threshold valueisselectedandusedtolocatetheobjectwithintheimage.

YOLOisorders ofmagnitudefaster (45 frames persecond) than anyotherobjectdetectionalgorithms. The limitation of YOLO algorithm is that it struggles with the small objects within the image, for example, it might have difficulties in identifying a flock of birds. This is due to the spatial constraints ofthealgorithm.

III. CHAPTER 3
PROPOSED SYSTEM ANALYSES AND DESIGN

3.1 Existing System:

3.1.1ResNet

To train the network model in a more effective manner, we herein adopt the same strategy as that usedfor DSSD (theperformance oftheresidual network is betterthanthat oftheVGGnetwork). The goal is to improve accuracy. However, the first implemented for the modification was the replacement of the VGG network which is used in the original SSD with ResNet. We will also add a series of convolution feature layers at the end of the underlying network. These feature layers will gradually be reduced in size that allowed prediction of the detection results on multiple scales. When the input size is given as 300 and 320, although the ResNet–101 layer is deeper than the VGG–16 layer, it is experimentally known that it replaces the SSD's underlying convolution network with a residualnetwork, anditdoesnotimproveitsaccuracybutratherdecreasesit.

3.1.2 R-CNN

To circumvent the problem of selecting a huge number of regions, Ross Girshick et al. proposed a method where we use the selective search for extract just 2000 regions from the image and he called them region proposals. Therefore, instead of trying to classify the huge number of regions, you can just work with 2000 regions. These 2000 region proposals are generated by using the selective search algorithmwhichiswrittenbelow.

SelectiveSearch:

1. Generatetheinitialsub-segmentation, wegeneratemanycandidateregions
2. Use thegreedyalgorithmtorecursivelycombinesimilarregionsintolargerones
3.Usegeneratedregionstoproducethefinalcandidateregionproposals

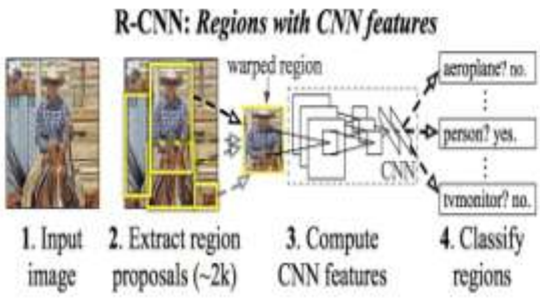


Figure 3.1.2.1: R-CNN Regions with CNN Features

3.2 Existing Technology

3.2.1 Object Detection from Video:

In a video there are primarily two sources of information that can be used for detectionand tracking of objects: visual features (e.g. color, texture and shape) andmotion information. Robust approaches have been suggested by combining the statisticalanalysis of visual features and temporal analysis of motion information [4]. Atypical strategy may first segment a frame into a number of regions based on visualfeatures like color and texture, subsequently merging of regions with similar motionvectors can be performed subject to certain constraints such as spatial neighborhoodof the pixels.

A large number of methodologies have been proposed by a number of researchersfocusing on the object detection from a video sequence. Most of them make useof multiple techniques and there are combinations and intersections among differentmethodologies. All these make it very difficult to have a uniform classification ofexisting approaches.

The different approaches available for moving object detection from video are:

1. Background Subtraction
2. Temporal Differencing
3. Statistical Approaches
4. Optical Flow

3.2.2 Object Tracking:

Object detection in videos involves verifying the presence of an object in a sequenceof image frames. A very closely related topic in video processing is possiblythe locating of objects for recognition – known as object tracking [5].There are a wide variety of applications of object detecting and tracking in computervision—video surveillance, vision-based control, video compression, humancomputerinterfaces, robotics etc. In addition, it provides input to higher level visiontasks, such as 3D reconstruction and representation. It also plays an important rolein video databases such as content-based indexing and retrieval.Popular methods of object tracking are as follows:

1. Mean-shift
2. Kanade–Lucas–Tomasi (KLT)
3. Condensation
4. TLD
5. Tacking Based on Boundary of the Object

3.2.3 Challenges of Object Detection and Tracking:

Object tracking fundamentally entails estimating the location of a particularregion in successive frames in a video sequence. Properly detecting objects can be aparticularly challenging task, especially since objects can have rather complicatedstructures and may change in shape, size, location and orientation over subsequentvideo frames [6]. Various algorithms and schemes have been introduced in the fewdecades, that can track objects in a particular video sequence, and each algorithmhas their own advantages and drawbacks. Any object tracking algorithm willcontain errors which will eventually cause a drift from the object of interest. Thebetter algorithms should be able to minimize this drift such that the tracker isaccurate over the time frame of the application.In object tracking the important challenge that has to consider while theoperating a video tracker are when the background is appear which is similar tointerested object or another object which are present in the scene [7]. This phenomenais known as clutter.The other challenges except from cluttering may difficulty to detect interestedobject by the appearance of the that object itself in the frame plane due to factors which are described as follows:

Object poses in the video frame: In a video file, since the object is moving so the appearance of an interested object may vary its projection on a video frame plane.

Ambient illumination: In a video, it is possible to change in intensity, direction and color of ambient light in appearance of interested objects in a video frame plane.

Noise: In the acquisitions process of video, it may possible to introduce a certain amount of noise in the image or video signal. The amount of noise depends upon sensor qualities which are used in acquitting the video.

Occlusions: In a video file, moving object may fall behind some other object which are present in the current scene. In that case tracker may not observe the interested object. This is known as occlusion.

3.2.4Implementation of Existing System

Currently, capturing images with high quality and good size is so easy becauseof rapid improvement in quality of capturing device with less costly but superiortechnology. The video can provide more information about the object when scenariosare changing with respect to time. Therefore, manually handling videos are quiteimpossible. So it needs an automated devise to process these videos. In this system, one such attempt has been made to track objects in videos. Many algorithms andtechnology have been developed to automate monitoring the object in a video file.

Simple object detection compares a static background frame at the pixel levelwith the current frame of video. The existing method in this domain first triesto detect the interest object in video frames. One of the main difficulties inobject tracking among many others is to choose suitable features and models forrecognizing and tracking the interested object from a video. Some common choiceto choose suitable feature to categories, visual objects are intensity, shape, colorand feature points.

Here, Haar Wavelet decomposition technique will be used for enhancement or improving the quality of low degraded video frames in video. After that template matching methodology will be used for object detection and tracking of object in video. Preliminary results fromexperiments have shown that the adopted method is able to track targets withtranslation, rotation, partial occlusion and deformation.

3.3 Hardware and Software Requirement:

Hardware Requirements:

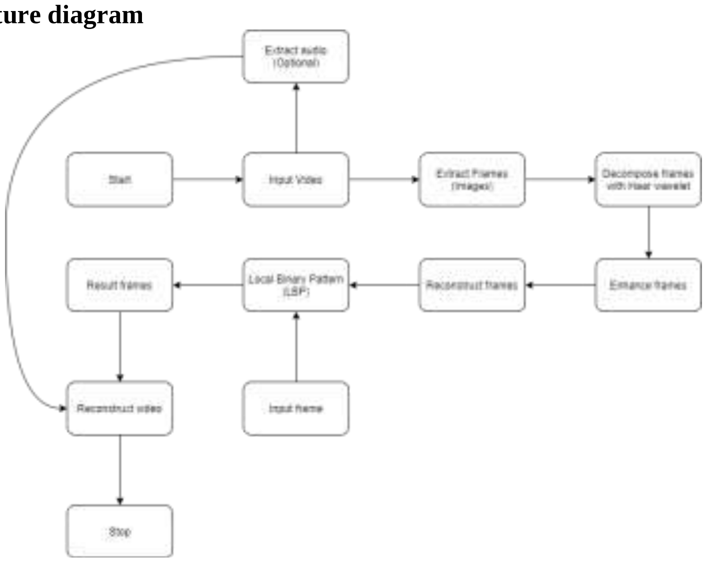
- Processor: Intel Core 2.0 GHz or more
- RAM: 1 GB or More
- Hard disk: 50GB or more
- Monitor: 15” CRT or LCD monitor
- Keyboard: Normal or Multimedia
- Mouse: Compatible mouse

Software Requirements:

- Operating system : Windows XP/07/10
- Development Tool : Matlab
- Backend Structure : System Directory
- Technologies used : .net framework, image processing

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 347-360 www.ijaem.net ISSN: 2395-5252

3.4 Top view architecture diagram



3.5 Algorithm

3.5.1 Frame's extraction

i. Start
ii. Input video (v)
iii. Foreach frames (f) in Video (v)
If (Format (f) == "image type")
Add to frame directory
End
iv. Save
v. Stop

3.5.2 Audio Extraction

i. Start
ii. Input video (v)
iii. Foreach audio frame (f) in Video (v)
If (Format (f) == "audio type")
Add to audio directory
End
iv. Save
v. Stop

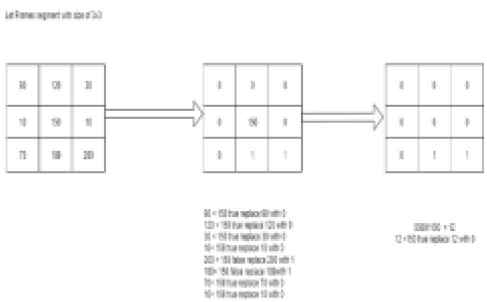
3.5.2.1 Haar wavelet

vi. Start
vii. Input Haar level L
viii. Foreach frames (f) in frame directory
For i=0; i<L; i++
F = f(height/2 and width/2)
End
ix. Save Frames
x. Stop

3.5.2.2 Local Binary Pattern

xi. Start
xii. Input frame (f)
xiii. Divide frame into size of 3 x 3
xiv. Foreach divided frames (d) in f
Find centre of d Foreach pixel (pi) in d
If pi >d
Replace pi = 1
Else
Pi =0
Convert all pixels to decimal
if decimal value of pixels > center
Add 1 to center pixel
Else
Add 1 to center pixel
End
xv. Match LBP pattern of input frame with target frame
xvi. Save result

3.6 Local Binary Pattern (Example)



IV. CHAPTER 4

SYSTEM IPLEMENTATION & TESTING

4.1 Setting Environment

To implement this idea smoothly, it must have one among the varied versions of windows OS which may be window or onward and wish to put in the Visual Studio 2012 and above version. the various parameters utilized in this system are as follows.

PARAMETER	TYPE
Operating System	Window 10 and Above
Visual Studio	2012 and above version
Database	Any relational database
Tool	Window voice recognition (inbuit)
RAM	Minimum 2 GB
Processor	1.5 GHz Minimum
Hard Disc Drive	-
Voice Capture	External/Internal Mice (Voice capture device)

Table 4.1: System Parameters

4.2 Implementation Details

To implement this technique we are found out proposed system design with Visual Studio 2012. Visual Studio 2012 provides interactive graphics design tools that creates proposed concept design more attractive. Different packages like speech Recognition system, Threading system,text system.io etc. from Visual Studio are used. As there's a requirement to stay voice samples in database, we prefer non-relational database to store these samples.

4.3 System Execution Details

For execution of proposed system, our first requirement is to update dictionary words. Below screen shot shows an execution of proposed system

V. CHAPTER 5

APPLICATIONS AND ADVANTAGES

5.1 Applications

The developed system is in a position to supply robust surveillance systems at a reasonable price. With the decrease in costs of hardware for sensing and computing, and therefore the increase within the processor speeds. The advanced methodology and techniques used for developing makes the system more accurate and straightforward to handle, that creates this technique more useful altogether its application area as:

• It has very large application in Surveillance systems.
• It are often utilized in manufacturing as a neighborhood of internal control .
• Used to supply how to navigate a mobile robot.
• Used as how to detect edges in images.
• It is employed for signal coding, to represent a discrete signal during a more redundant form often as a preconditioning for data compression.
• Practical applications also can be found in signal processing of accelerations for gait analysis, in digital communications and lots of others.
• And number of various applications, like traffic monitoring, airport and bank security etc.

5.2 Advantages

The proposed and developed system has many advantages a number of which are mention and listed as follows:
• It is conceptually simple and fast.
• It is memory efficient, since it are often calculated in situ without a short lived array.
• It is strictly reversible without the sting effects that are a drag with other wavelet transforms.
• Implementation cost are less costly.

• It provides a promising cost savings conjoining with sending less data over switched telephone network where cost of call is basically usually based upon its continuation.
• It not only reduces vault requirements but also overall execution time.

5.3 Limitations

There are a number of the restrictions of the proposed system that must be lookout of, so as to realize proper advantage of the proposed system. the restrictions are as follows:
• In generating each set of averages for subsequent level and every set of coefficients, the algorithm shifts over by two values and calculates another average and difference on subsequent pair.
• The high frequency coefficient spectrum should reflect all high frequency changes. The Haar window is merely two elements wide. If an enormous change takes place from a good value to an odd value, the change won't be reflected within the high frequency coefficients.

VI. CHAPTER 6

CONCLUSION & FUTURE SCOPE

6.1 Conclusion

By using this thesis and supported experimental results we are ready to detect obeject more precisely andidentify the objects individually with exact location of an obeject within the picture in x,yaxis.This project also provide experimental results on different methods for object detection and identification andcompares each method for his or her efficiencies.

6.2. Future Scope

- Geometric properties of the image are often included within the feature vector for recognition.
- Using unsupervised classifier rather than a supervised classifier for recognition of the thing .
- The proposed visual perception system uses grey-scale image and discards the colour information.
- The colour information within the image are often used for recognition of the thing .Colour based objectrecognition plays vital role in RoboticsAlthough the visual tracking algorithm proposed here is strong in many of the conditions, it can bemade more robust by eliminating a number of the restrictions as listed below:
- within the Single Visual tracking, the dimensions of the template remains fixed for tracking. If the dimensions of theobject reduces with the time, the background becomes more dominant than the thing being tracked. During this case the thing might not be tracked.
- Fully occluded object can't be tracked and thought of as a replacement object within the next frame.
- Foreground object extraction depends on the binary segmentation which is administered by applyingthreshold techniques. So blob extraction and tracking depends on the edge value.
- Splitting and merging can't be handled alright altogether conditions using the only camera thanks tothe loss of data of a 3D object projection in 2D images.
- For already dark visual tracking, night-sight mode should be available as an inbuilt feature within theCCTV camera.

REFERENCES

[1]. T. Ojala, M. Pietikäinen, and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971–987, 2002.
[2]. T. Ojala, M. Pietikäinen, and D. Harwood, “A comparative study oftexture measures with classification based on featured distribution,”Pattern Recognition, vol. 29, no.1, pp. 51–59, 1996.
[3]. T. Ahonen, A. Hadid, and M. Pietikäinen, “Face recognition with localbinary patterns,” in Proc. Euro. Conf. Computer Vision (ECCV), 2004,pp. 469–481.
[4]. A. Hadid, M. Pietikäinen, and T. Ahonen, “A discriminative feature spacefor detecting and recognizing faces,” in Proc. Int. Conf. Computer Visionand Pattern Recognition (CVPR), 2004, pp. 797–804.
[5]. D. P. Huijsmans and N. Sebe, “Content-based indexing performance: aclass size normalized precision, recall, generality evaluation,” in Proc.International Conference on Image Processing (ICIP), 2003, pp.733-736.
[6]. D. Grangier and S. Bengio, “A discriminative kernel-based approach torank images from text queries,” IEEE Trans. Pattern Analysis and MachineIntelligence, vol. 30, no. 8, pp. 1371-1384, 2008.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 347-360 www.ijaem.net ISSN: 2395-5252

[7]. W. Ali, F. Georgsson, and T. Hellström, "Visual tree detection for autonomousnavigation in forest environment," in Proc. IEEE Intelligent VehiclesSymposium, 2008, pp. 560-565.14
[8]. L. Nanni and A. Lumini, "Ensemble of multiple pedestrian representations,"IEEE Trans. on Intelligent Transportation Systems, vol. 9, no. 2,pp. 365-369, 2008.
[9]. T. Mäenpää, J. Viertola, and M. Pietikäinen, "Optimising colour andtexture features for real-time visual inspection," Pattern Analysis andApplications, vol. 6, no. 3, pp.169-175, 2003.
[10]. M. Turtinen, M. Pietikäinen, and O. Silven, "Visual characterization ofpaper using Isomap and local binary patterns," IEICE Transactions onInformation and System, vol. E89D, no. 7, pp. 2076-2083, 2006.
[11]. M. Heikkilä and M. Pietikäinen, "A texture-based method for modelling the background and detecting moving objects," IEEE Trans. PatternAnalysis and Machine Intelligence, vol. 28, no. 4, pp. 657-662, 2006.
[12]. V. Kellokumpu, G. Zhao, and M. Pietikäinen, "Human activity recognitionusing a dynamic texture based method," in Proc. The British MachineVision Conference (BMVC), 2008.
[13]. A. Oliver, X. Lladó, J. Freixenet, and J. Martí, "False positive reduction inmammographic mass detection using local binary patterns," in Proc.Medical Image Computing and Computer-Assisted Intervention(MICCAI), 2007.
[14]. S. Kluckner, G. Pacher, H. Grabner, H. Bischof, and J. Bauer, "A 3Dteacher for car detection in aerial images," in Proc. IEEE InternationalConference on Computer Vision (ICCV), 2007.
[15]. A. Lucieer, A. Stein, and P. Fisher, "Multivariate texture-based segmentationof remotely sensed imagery for extraction of objects and their uncertainty,"International Journal of Remote Sensing, vol. 26, no. 14,pp. 2917-2936, 2005.
[16]. http://www.ee.oulu.fi/mvg/page/lbp_bibliography. The availability of thelink was last checked on 15 Nov., 2010.
[17]. H. Jin, Q. Liu, H. Lu, and X. Tong, "Face detection using improved LBPunder Bayesian framework," in Proc Int. Conf. Image and Graphics(ICIG), 2004, pp. 306–309.
[18]. L. Zhang, R. Chu, S. Xiang, and S. Z. Li, "Face detection based on Multi-Block LBP representation," in Proc. Int. Conf. Biometrics (ICB), 2007,pp. 11-18.
[19]. H. Zhang and D. Zhao, "Spatial histogram features for face detection incolor images," in Proc. Advances in Multimedia Information Processing:Pacific Rim Conference on Multimedia, 2004, pp. I: 377-384.
[20]. T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with localbinary patterns: application to face recognition", IEEE Transactions onPattern Analysis and Machine Intelligence (PAMI), vol. 28, no. 12, pp.2037–2041, 2006.
[21]. C. Chan, J. Kittler, and K. Messer, "Multi-scale local binary patternhistograms for face recognition," in Proc. Int. Conf. Biometrics (ICB),2007, pp. 809-818.
[22]. X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognitionunder difficult lighting conditions," in Proc. Analysis andModeling of Faces and Gestures (AMFG), 2007, pp. 168-182.
[23]. S. Liao and A. C. S. Chung, "Face recognition by using elongated localbinary patterns with average maximum distance gradient magnitude," inProc. Asian Conf. Computer Vision (ACCV), 2007, pp. 672-679.