ISSN 2348 – 7968

Backgroung Subtraction Based Detection and Tracking Of People In Video

Sneha S. Joshi [1], J. R. Panchal [2]

P.G. Student, Department of E&TC Engineering, Siddhant College of Engineering, Sudumbare, Pune.

Associate Professor, Department of E&TC Engineering, Siddhant College of Engineering, Sudumbare, Pune.

Abstract: **People Detection in video is generally used in high level multimedia applications like intelligent surveillance systems, augmented reality ,etc. People detection is based on background subtraction. Generally videos are available in compressed form due to which noise will be added in it. Many algorithms are used to detect people and control their rate. 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. The ability of human visual system to detect visual saliency is extraordinarily fast and reliable. However, computational modeling of this basic intelligent behavior still remains a challenge. In this paper we put our attention into background subtraction and Gaussian grouping of pixels for detection of people in low quality video i.e. improve accuracy**.

Keywords: Background subtraction, Gaussian grouping, Kalman filter, Segmentation.

I. INTRODUCTION

The field of computer vision is concerned with problems that involve inter- facing computers with the environment through visual means. The increase of high-powered computers and the availability of high-quality and inexpensive video cameras extend the computer vision's applications to everyday life technology. People tracking is one of the most important tasks in computer vision which can be defined simply as the problem of estimating the trajectory of each person in the image plane as s/he moves around in the scene. In other words, a people tracker system can recognize each person in consecutive frames of a video. Depending on the applications of people tracking, additional information can be also provided by the system during tracking.

Background subtraction technique find the foreground object from video and then classify it into categories like human, animal, vehicle etc., based on shape , color , motion or other features. Most of the multimedia videos are available in compressed format. Usually higher the compression rate, lowers the correct hits and quality of video due to noise added in it. A modern object detection algorithm can be divided into five parts: pre-processing and normalization, local rectification and compensation of small shape variations, computation of descriptor set, machine learning classification, and post-processing to fuse multiple detections.

. In this paper, we focus on detection schemes based on background subtraction because of their widespread use and the possibilities they offer in implementing real-time object detection systems. background subtraction is nothing but foreground detection. Any motion detection system based on background subtraction needs to handle a number of critical situations such as, noise image due to a poor quality image source, variations in the lighting conditions, small movement of non-static objects, shadow regions that are projected by foreground objects, multiple objects moving in the scene both for long and short periods. The main objective of this paper is to develop an algorithm that can detect people. Unfortunately, images and video are usually available in compressed format which makes object detection more difficult because of the additional distortion noise. In this paper we propose a saliency map algorithm and compare it with background method to improve accuracy.

The process algorithm is described as follows:

1. Sequences of Video Frames
2. Frame Separation
3. Image Sequence
4. Separation of Image Sequence in Current frame Image and Background Frame Image.
5. Perform Background Subtraction
6. Detection of Moving Object
7. Perform Background Updating
8. Noise Removal

9. Shape Analysis

Here shape feature is used to determine whether the moving object is a human being or not. For that two criterias are to be considered such as the object area is larger than the set threshold and the aspect ratio of the object region should conform to the set ratio. If these two conditions are met, the moving object is the moving human body or not a human body is conform. Since the difficulty of the background , the discrepancy image obtained contains the motion region as well as large number of noise. These noise might be included in the image due to environmental factors , illumination changes, during transmission of video from the camera to the further processing. Therefore noise needs to remove. To remove this noise, Gaussian filter is used. The flow chart of process algorithm is as shown in fig.1.

Object detection has been studied for about four decades producing a wide range of object detection techniques

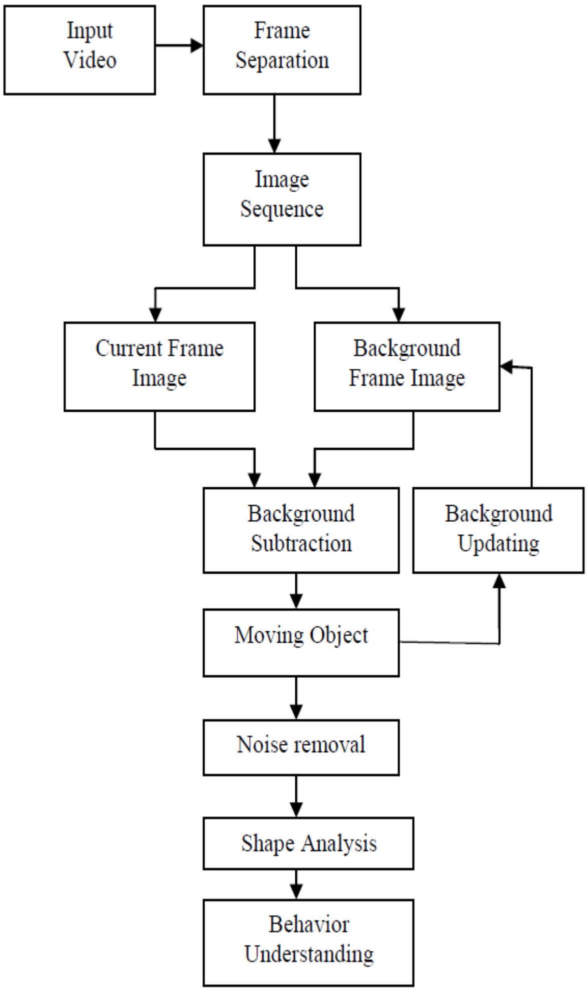


Figure1: Process Algorithm

. This research effort has been fostered by the many application scenarios where object detection can be employed (e.g., active video surveillance, assisted or autonomous drive, database search, data classification) and by the need of increasing the robustness of existing algorithms to differentlighting conditions, poses, scales, locations, and geometries .

II. LITERATURE SURVEY

A important stream of research within computer vision which has gained a lot of importance in the last few years is the understanding of human activity from a video. Understanding human activity has applications in various fields, the most important

t of which is surveillance. Other applications include character animation for games and movies, avatars for teleconferencing, advanced intelligent user interfaces, biomechanical analysis of actions for sports and medicine, etc. Before the complexity of human activity can be understood, we first need automatic methods for finding humans in an image or a video. Once the human is detected, depending on the application, the system can do further processing to go into the details of understanding the human activity. This paper selects a representative sample of papers from the broad literature on full-body human detection, and presents a review and classification of the various methods. It is not intended to be comprehensive, and does not deal with specialized domains such as detection of faces, gestures or characterizing human activity, each of which possess an extensive literature of their own. Our problem is to find people in a given video (or an image). The relevant literature can be divided into techniques which require background subtraction or segmentation and techniques which can detect humans directly from the input without such pre-processing.

Background subtraction techniques usually find the foreground object from the video and then classify it into categories like human, animal, vehicle etc., based on shape, color, or motion or other features. Here, we review the following techniques which perform human detection after background subtraction. Direct techniques operate on (features extracted from) image or video patches and classify them as human or non-human. We can also classify techniques based on the features which are used to classify a given input as human or not. These features include shape (in the form of contours or other descriptors), color (skin color detection), motion, or combinations of these.

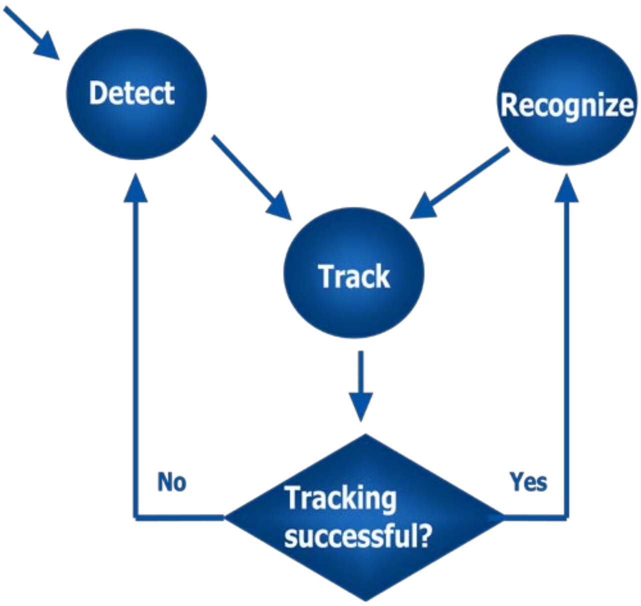
Many detection algorithm algorithms have been proposed in order to evaluate performance of object detection for video surveillance [3]. A survey of techniques for human detection [6]. Most of the presented algorithms depend on the the statistics of the orientation of edges (Histograms of Oriented Gradients or HOGs) and color histograms[9]. Background subtraction algorithm is used to improve the effectiveness of human motion detection[8]. In this paper ,we improve the correct hits and PSNR as compare to saliency based algorithm[1].

III. TECHNIQUE

Background subtraction method is widely used to detect moving human in videos. The concept in this approach is that of detecting the moving objects from the difference between the existing frame and a reference frame. The background image must be a representation of the scene with no moving objects .The effectiveness of this method will affect the accuracy of test results.

Recent research in computer vision has increasingly focused on building systems for observing humans and understanding their appearance, movements, and activities, providing advanced interfaces for interacting with humans, and creating realistic models of humans for various purposes. In order for any of these systems to function, they require methods for detecting people from a given input image or a video. The techniques are classified with respect to the need for pre-processing (background subtraction or direct detection), features used to describe human appearance (shape, color, motion), use of explicit body models, learning techniques. The proposed framework compares the output of the algorithm with the ground truth and measures the differences according to objective metrics.

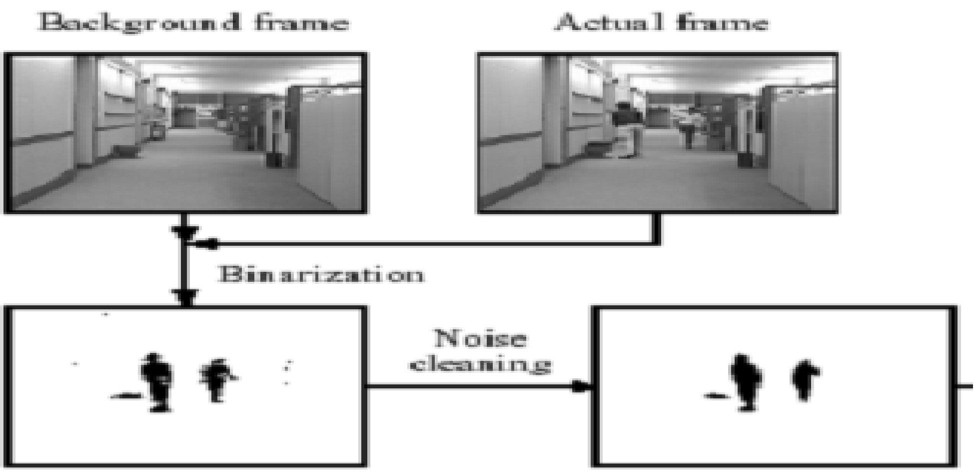
A .IMPLIMENTATION SCHEME



B. BACKGROUND SUBTRACTION

There are so many methods to obtain the initial background subtraction. It computes the absolute difference between the current image and a static background image and compares each pixel to a threshold. Background subtraction is very adaptive to stable environments but is extremely sensitive to dynamic scene changes due to lighting and other conditions.

Our approach is divided into three main parts: (i) building and maintaining the background model., (ii) performing background subtraction and (iii) delineating the foreground. It is performed by subtracting the color channels and edge channels separately and then combining the results.



The number of cameras available worldwide has increased dramatically over the last decade. But this growth has resulted in a huge augmentation of data, meaning that the data are impossible either to store or to handle manually. In order to detect, segment, and track objects automatically in videos, several approaches are possible. Simple motion detection algorithms compare a static background frame with the current frame of a video scene, pixel by pixel. This is the basic principle of background subtraction, which can be formulated as a technique that builds a model of a background and compares this model with the current frame in order to detect zones where a significant difference occurs. The purpose of a background subtraction algorithm is therefore to distinguish moving objects (hereafter referred to as the *foreground*) from static, or slow moving, parts of the scene (called *background*). Note that when a static object starts moving, a background subtraction algorithm detects the object in motion as well as a hole left behind in the background (referred to as a *ghost*). Clearly a ghost is irrelevant for motion interpretation and has to be discarded. An alternative definition for the background is that it corresponds to a reference frame with values visible most of the time, that is with the highest appearance probability, but this kind of framework is not straightforward to use in practice.

While a static background model might be appropriate for analyzing short video sequences in a constrained indoor environment, the model is ineffective for most practical situations; a more sophisticated model is therefore required. Moreover, the detection of motion is often only a first step in the process of understanding the scene. For example, zones where motion is detected might be filtered and characterized for the detection of unattended bags, gait recognition, face detection, people counting, traffic surveillance, etc. The diversity of scene backgrounds and applications explains why countless papers discuss issues related to background subtraction.

C.GAUSSIAN DISTRIBUTION

Among the high complexity methods , mixture of Gaussian is more beneficial, as it can handle multi-modal distribution. In this, the background is not a frame of values. Rather ,the background model is parametric. Each pixel location is represented by a number of Gaussian functions that sum together to form a probability distribution function. A certain pixel value x, at time t by means of a mixture of Gaussians that sum together to form a probability distribution function as

$$P(x_t)= \Sigma^{k}_{t=1}\ \omega_{i,t}.\eta\ (x_t-\mu_{i,t},\Sigma_{i,t})\ldots\ldots\ldots[11]$$

Where,

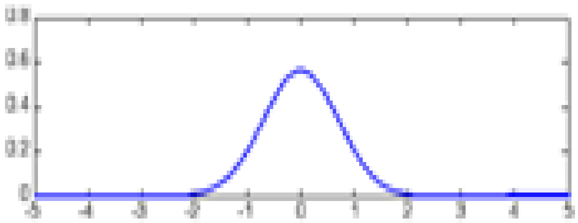
K= Gaussian distributions to describe one of the observable background or foreground objects.

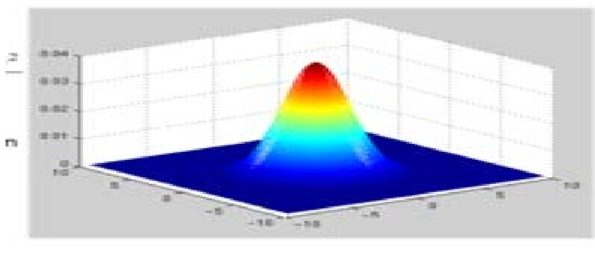
μ= Mean of Gaussian function.

$\omega_{i,t}$= weight of the i^{th} Gaussian in the mixture.

$\mu_{i,t}$ = Mean value of the i^{th} Gaussian in the mixture.

The Gaussian distribution for 1dimentional and 2dimentional image is as follows:





This function is simple and of low computational complexity. However, the object is hard to be precisely detected when both of the background and the foreground are complicated.

D. TRACKING

Kalman filter is used to track the people in video. It is also known as linear quadratic estimation.This filter operates recursively on streams of noisy input data and produces estimates of the current state variables. This filter is widely used for guidance,navigation and control of vehicles, aircraft and spacecraft.

The Kalman filter model assumes that the state of a system at a time *t* evolved from the prior state at time *t*-1 according to the equation

x*t* = F*t*x*t*-1 + B*t*u*t* + w*t* , ……….[12]

where

x*t* = state vector containing the terms of interest for the system (e.g., position, velocity, heading) at time *t*

u*t* = vector containing any control inputs (steering angle, throttle setting, braking force)

F*t* = state transition matrix which applies the effect of each system state parameter at time *t*-1 on the system state at time *t* (e.g., the position and velocity at time *t*-1 both affect the position at time *t*)

B*t* = control input matrix which applies the effect of each

control input parameter in the vector **u***t* on the state vector(e.g., applies the effect of the throttle setting on the system velocity and position)

w*t* = vector containing the process noise terms for each parameter in the state vector. The process noise is assumed to be drawn from a zero mean multivariate normal distribution with covariance given by the covariance matrix Q*t*.

E. COMPRESSION RATE[1]

Most of the object detection algorithms are based on analyzing the distribution of gradients and colors along object borders. Having this in mind, it is possible to design a saliency metric that combines different features into a normalized value. In particular, the input image is divided into blocks and the included pixels are processed computing three different metrics. In fact, it is possible to observe that object detection strategies are sensitive to edge strength, the stationarity of edge directions, and color contrast along the main orientations of an image.

- Edge strength

In order to characterize the significance of edges in different regions of the image I(x; y), it is possible to compute horizontal and vertical Sobelian gradients (named Sx(x; y) and Sy(x; y), respectively) on the whole image using a 3 * 3 Sobel operator. Then, these two gradient maps are merged into a common measure of the gradient strength named

$$G(x,y) = round\left(\frac{(|S_x(x,y)| + |S_y(x,y)|)}{16}\right)$$

- Regularity of edge strength

Re-using the previous results, the orientation of borders A(x; y) can be computed as $A(x,y) = \tan^{-1}\left(S_y(x,y)/S_y(x,y)\right)$. From these where the rounding operation is used to smooth the smallest variations and the normalization factor has been decided after a set of experimental tests. Then, values G(x; y) for the current image are stored in a histogram and the 85-percentile Tg is computed. By selecting those pixel positions (x; y) such that G(x; y) > Tg, it is possible to consider only those positions that present relevant gradient information.

$$R_A(x,y) = \frac{|A(x,y) - A(x+1,y)| + |A(x,y) - A(x,y+1)|}{2} \dots\dots\dots\dots (2)$$

For each image block (referenced with the index b), it is possible to compute the variance σR;b of RA(x; y).

For each image block (referenced with the index b), it is possible to compute to variance.

- Contrast along edges

For every pixel position (x; y) such that G(x; y) is greater than Tq, the algorithm evaluates the color differences along the edge orientation, i.e., it computes

$$C(x,y) = \frac{|R(x+\delta_x, y+\delta_y) - R(x-\delta_x, y-\delta_y)|}{3} + \frac{|G(x+\delta_x, y+\delta_y) - G(x-\delta_x, y-\delta_y)|}{3} + \frac{|B(x+\delta_x, y+\delta_y) - B(x-\delta_x, y-\delta_y)|}{3}$$

where $(\delta_x, \delta_y) \epsilon \{-1,0,1\}^2$ is a displacement array aligned along the normal direction to the edge in (x; y). This parameter proves to be extremely important in identifying those objects that can not be easily detected from the background since a low value of C(x; y) denotes a limited contrast around the edge.

For the current pixel block, the proposed saliency metric evaluates the value Cb, which averages the values C(x; y) for those pixel locations (x; y) in the b-th block such that G(x; y) > Tg. These three metrics are then combined into a single saliency value that is assigned to each block of the image.

- Final saliency metric

The input (uncompressed) image is divided into pixel blocks (indexedwith the variable b), and for each of these thealgorithm computesσR;b and Cb.Starting from this, the saliency value assigned to the current block is

$$S_b = 1 - \begin{cases} \frac{C_b}{k} & if\ 5 < \sigma_{R,b} < 500 \end{cases} \ldots\ldots\ldots\ldots\ldots\ldots (4)$$

where the normalization constant K has been computed in order to make Sb vary in the range [0; 1]. Parameter K is computed after a set of experimental trials on a set of test sequences. In case Sb > 1, its value is clipped as Sb = 1, and viceversa, if Sb < 0, Sb = 0. The value of Sb depends on σR;b since for low and high variance signals, the value of contrast long the edge must be emphasized to fit values within a proper range.

In this work, we implemented this strategy considering the emerging video coding standard HEVC. This saliency algorithm is then compared with pixel based Gaussian distribution.

IV. RESULT

The proposed approach has been tested using the object detector in [1] for people tracking on different video sequences the whole algorithm inherits the key idea of the HOG-based approach [8] but extends object models introducing improvement of people detection by using Gaussian distribution and kalman filter. Graphs shows the percentage of true hits vs target bit rate and rate distortion performance.

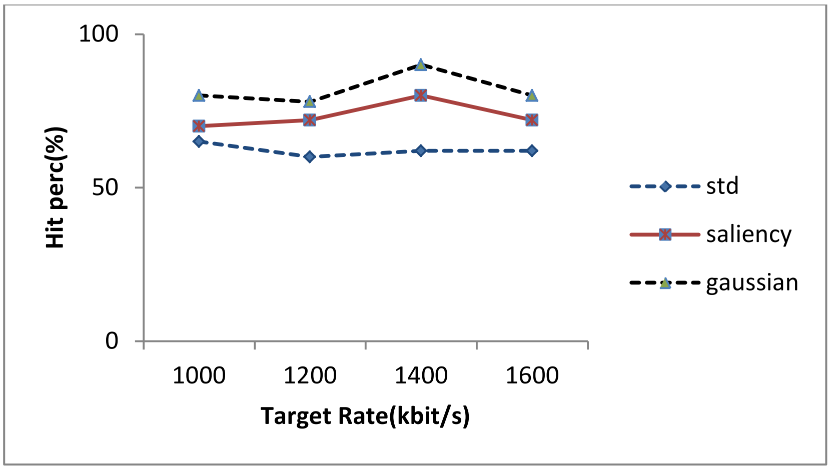


Figure2:Graph of %true hits vs target rate

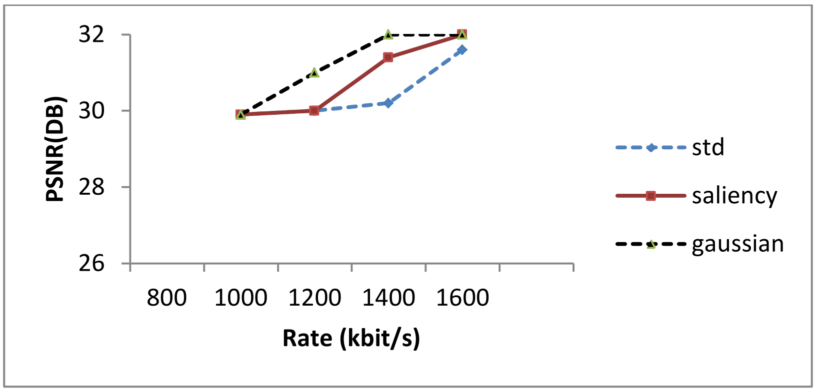


Figure3: Rate distortion performance

IV. CONCLUSION

The saliency map is based on a set of features which give to the encoder an evaluation of the significance of the different regions in a frame. In this paper, we have presented people detection in video by using background subtraction method which effectively improve the accuracy and performance.

REFERENCES

[1] Simone Milani, Riccardo Bernardini, Roberto Rinaldo, “Saliency based rate control for people detection in video”, DIEGM - University of Udine Via Delle Scienze, 208 - 33100 Udine – Italy,IEEE2013

[2] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, and D. Ramanan, “Object detection with discriminatively trained partbased models,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 32, no. 9, pp. 1627 –1645, sept. 2010.

[3] Jacinto Nascimento, Jorge Marques, “Performance evaluation of object detection algorithms for video surveillance”, IST/ISR, Torre Norte, Rovisco Pais, 1049-001, Lisboa Portugal.

[4] David G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[5] M. Reisert and H. Burkhardt, "Equivariant holomorphic filters for contour denoising and rapid object detection," Image Processing, IEEE Transactions on, vol. 17, no. 2, pp. 190 –203, Feb. 2008.

[6] Neeti A. Ogale, "A survey of techniques for human detection from video", Department of computer science ,university of Maryland, college park, MD 20742.

[7] Yongxia Zuo, Guoqiang Wang, and Chuncheng Zuo, "Wavelet packet denoising for pavement surface cracks detection," in Computational Intelligence and Security, 2008. CIS '08.

[8] Rupali S. Rakibe, Bharati D. Patil, "Background subtraction algorithm based human motion detection", International journal of Scientific and research publications, Volume3, May2013.

[9] Navneet Dalal and William Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05), vol. 1, no. 3, pp. 886–893, 2004.

[10] Simone Milani, Riccardo Bernardini, and Roberto Rinaldo,"Adaptive denoising filtering for object detection applications," in Proc of ICIP 2012, Oct. 2012.

[11] Sanjivani Shantaiya, Kesari Verma, Kamal Mehta, "Study and analysis of methods of object detection in video", International journal of soft computing and engineering, volume2, Jan 2013.

[12] Ramsey Faragher, "Understanding the basis of the kalman filter via a simple and intuitive derivation",IEEE signal processing magazine, Sep2012