IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 2, Ver. I (Mar.-Apr. 2017), PP 26-32
www.iosrjournals.org

Human Activity Recognition Based on Object Detection

Aarsh Joshi, Hetal Parmar, Khyati Jain,Chandni Shah ,Prof. Vaishali R Patel
joshi.aarsh21@gmail.com ,
chandnishah7040@gmail.com,khyatu.jain@gmail.com,hetalparmar13@gmail.com,vaishalirajpatel@gmail.com
Department of Computer Science and Engineering, Shri S'ad Vidya Mandal Institute of Technology Bharuch

Abstract: *Human activity recognition and object detection are techniques for detecting a human and objects from an image respectively. Which can be greatly useful for medical, security and industrial fields. This paper presents a new approach by combining both the techniques for extra usability and accurate detection of activities. This paper presents the HOG features for detecting the human in an image and SURF features for extraction of matched points from the image. The accuracy of results are higher due to combination of both the techniques by merging the results of features obtained from an image. The algorithm helps to perform better in low-light conditions which helps to perform better feature extraction. This technique determines the distance of the object from a human and provides the results of activities based on the distance.*

Keywords: *Human detection; Object detection; Activity detection; HOG and SURF fetaures.*

I. Introduction

Human activity recognition either with object or without any external object is aimed to retrieve the precise result about the activities being performed in the input image. The system has various stages for detection which include human detection, object detection and human activity recognition like running walking etc. The algorithm used for human detection is HOG features extraction which is more efficient to detect the features of a human from a still image.

The features include the different postures of human in different activities. The object detection is used to detect different object with the SURF detector for better efficiency to detect the correct object. There are several algorithms for the same but the test results indicate that surf is more efficient detector. For the ease of system there are only ten objects supported for detection. The images which have both object and human in it are to be detected with the result if the activity is being performed by using the object or not. For obtaining such result the system calculates the distance between the human and object and if the distance is less or they both collide with each other in the image then the result is shown with the activity performed using the object. So the distance between the object and human contributes to obtain more efficient result of the activity detection.

In this paper the different methods and algorithms are compared for detection like HOG, SURF, and SIFT, MSER, SVM, and HARRIS etc. with the parameters of time and number of features. The comparison is made to obtain the best method for both kind of detection process as human and object.

II. Related Works

A. HOG$_{[1]}$

Histogram of Oriented Gradients (HOG) was first introduced by Dalal and Triggs for pedestrian detection in 2005 and since then has been widely used to detect many different objects. In their framework each image window is first gamma and color normalized then is divided into small regions called Cells. In order to make the descriptor robust to contrast, a bigger region called Block is defined. Blocks consist of a number of adjacent cells and have overlaps with each other. After calculating gradients for all pixels in each cell of the image, they will be quantized into a user defined number of bins (suppose b). So each cell will be represented by a b-bin Histogram of Oriented Gradients. By concatenating these vectors for all cells inside each block of image, HOG feature vector is formed.

B. SURF & SVM$_{[1]}$

The work of Gavrila and Philomin compare edge images to an exemplar dataset using the chamfer distance. Extended their Haar-like wavelets to handle space-time information for moving-human detection. The "Integral Image" allows very Fast evaluation of Harr-wavelet type features, known as rectangular filters. This led to a real-time face detection system that was later extended to a human detection system using rectangular filters both in space and time. But the main drawback of SVM method is we can detect human only from within the dataset there are different techniques to extract features and images for human detection. But out of some like harr-wavelets, lbp(local binary patterns), hog. We will be using hog (Histogram of oriented gradient) for its better feature. Out of all time is little consuming but accuracy is more in hog.

C. HOG/SIFT [1]

The HOG/SIFT representation has several advantages. It captures edge or gradient structure that is very characteristic of local shape, and it does so in a local representation with an easily controllable degree of invariance to local geometric and photometric transformations: translations or rotations make little difference if they are much smaller that the local spatial or orientation bin size. For human detection, rather coarse spatial sampling, fine orientation sampling and strong local photometric normalization turns out to be the best strategy, presumably because it permits limbs and body segments to change appearance and move from side to side quite a lot .Provided that they maintain a roughly upright orientation.

D. SURF Features

Speed up robust features it approximation and performs previously proposed schemes with the respect to repeated. [7]Hessian matrix based measure for the detector and a distribution based descriptor [8]. SURF is a comparison to state of art and also faster to compute [7]. Interest point detector is used in Harris corner detector [9]. To make the SURF feature faster. This allows to detect interest points in an image, each with their own characteristic scale. Experimented with both the determinant of the Hessian matrix as well as the Laplacian [8]. Hessian-based detectors are more stable and repeat-able than their Harris-based counterparts. Using the determinant of the Hessian matrix rather than its trace (the Laplacian) seems advantageous, as it fires less on elongated, ill-localized structures. Also approximations like the DoG can bring speed at a low cost in terms of lost accuracy.[8] SHFIT is a good performance to compare to other descriptor. SURF contain s the same similarities of SHFIT properties.[7]

Working:

The first step consists of fixing a reproducible orientation based on information from a circular region around the interest point. Then, we construct a square region aligned to the selected orientation, and ex-tract the SURF descriptor from it.[8]

SURF descriptor outperforms the other descriptors in a systematic and significant way, with sometimes more than 10% improvement in recall for the same level of precision [7].The main difficulty is in object tracking to choose the suitable feature .SURF algorithm is for continuous image recognition and also in video. It reduce the search space of interest points. It also add a lot of features to make it faster in every iteration. Interest points are reputably but noise free. SURF use to track object using interest point matching and the updating. It is difficult to recognize the object when the object is moving fast in the frame (video).SURF efficiency is improved using harris corner detector.[8]

Steps:

1) Interest Points: The detection of this points include a very small amount of iterations. When more than iteration points are detected the we need to choose the best matched
2) HAAR Wavelet Filtering: A Haar-like feature uses adjacent rectangular regions at a specific location in a detection window and sums up the pixel intensities in each region and the difference between these sums is calculated. For feature description, SURF uses Wavelet responses in horizontal and vertical direction around the point of interest. The position of these rectangles acts like a bounding box to the target object.

It then continuously extracts feature for recognition. For feature extraction SURF uses the sum of the Haar wavelet response around the interest point. The main advantage of Haar wavelet transform is its calculation speed. SURF (Speeded Up Robust Features) algorithm is used here for continuous image recognition and tracking in video. The SURF feature descriptor operates by reducing the search space of possible interest points inside of the scale space image pyramid. SURF adds a lot of features to improve the speed in every step. The resulting tracked interest points are more repeatable and noise free. SURF is good at handling images with blurring and rotation. SURF descriptor is three times as fast as SIFT feature descriptor[8].

E. Harris Corner Detector

It is good for the blurred and rotated images corner detector are used to locate the interest points in the image.(harris corner detector). Corner detection is good for obtaining image features for object tracking and recognition. Interest points in an image are located using corner detector. By using harris corner detection algorithm along SURF feature descriptor, tracking efficiency is improved. Harris corner detection is used for its computation speed. Harris corner detector is rotation and scale invariant. Corner detection is an approach used to extract certain kinds of features and image contents. Corner detection is used in motion detection, image registration, video tracking, panorama stitching, and object recognition etc. The Corner Detection block finds corners in an image using the Harris corner detection. Harris detector considers the differential of the corner score with respect to direction directly, instead of using shifted patches.[8]

F. MSER

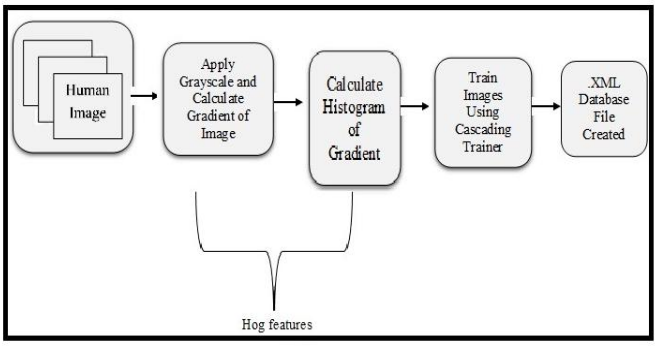
The MSER detector incrementally steps through the intensity range of the input image to detect stable regions. The Threshold Delta parameter determines the number of increments the detector tests for stability. You can think of the threshold delta value as the size of a cup to fill a bucket with water. The smaller the cup, the more number of increments it takes to fill up the bucket. The bucket can be thought of as the intensity profile of the region. The MSER object checks the variation of the region area size between different intensity thresholds. The variation must be less than the value of the MaxAreaVariation parameter to be considered stable. At a high level, MSER can be explained, by thinking of the intensity profile of an image representing a series of buckets. Imagine the tops of the buckets flush with the ground, and a hose turned on at one of the buckets. As the water fills into the bucket, it overflows and the next bucket starts filling up. Smaller regions of water join and become bigger bodies of water, and finally the whole area gets filled. As water is filling up into a bucket, it is checked against the MSER stability criterion. Regions appear, grow and merge at different intensity thresholds[10].

III. Human Detection

Human Detection from an image is a challenging task with respect to their appearance and wide range of poses. For detecting the human we need to use HOG feature, as HOG provide the fast and efficient result as compared to sum.

Firstly the color image is taken and it is converted into gray scale. Then the gradient of the image is found. With the help of some positive and negative images training cascading is perform on the images using histogram. Then after finishing the training the image, the human is extracted.

Trainer for Human Detection


Human Image
Apply Grayscale and Calculate Gradient of Image
Calculate Histogram of Gradient
Train Images Using Cascading Trainer
.XML Database File Created
Hog features

Human Images:-

For Detecting Human we need the database of positive and negative images. Here by, comparing the positive and negative images human extraction is possible.

Gradient Image:-

The color images are taken as an input, then grayscale is applied on the colored image. For that rgb2gray function is used.Rgb2gray is used to convert the true-color image RGB to the grayscale intensity image I. rgb2gray converts RGB images to grayscale by eliminating the hue and saturation information while retaining the luminance.

After finding the grayscale of the image we need to find the gradient of the image. There are various methods for finding gradient of the image. So using the best method out of it we need to calculate the magnitude and the direction of the grayscale image.

Histogram of the image:-

Histogram of the image computes the frequency distribution of the elements in the input for calculating the histogram of the image imhist is used. imhist(I) calculates the histogram for the intensity image I and displays a plot of the histogram. The number of bins in the histogram is determined by the image type.

If I is a grayscale image, imhist uses a default value of 256 bins.

If I is a binary image, imhist uses two bins.

Train the cascading detector using positive and negative images:-

After finding the histogram of the image, we needs to train the detector. For training, the cascade detector is used by positive image which would contain images with people in it and negative images which would contain anything but people in it.

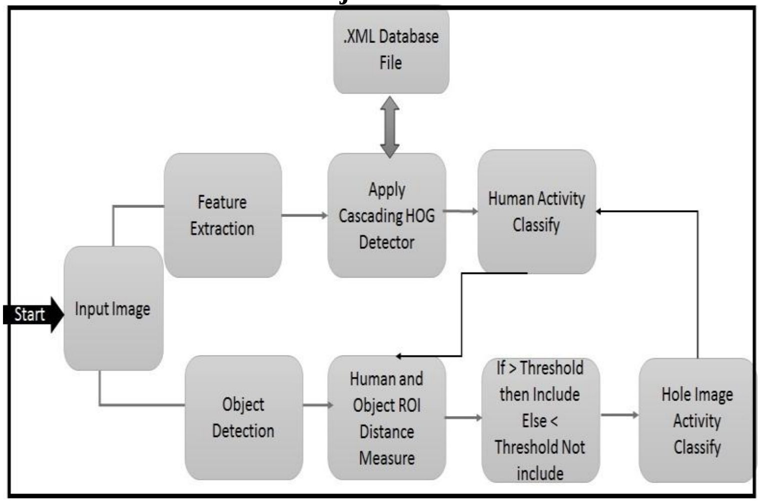
XML file:-
After training the images one xml file is generated which contains the information about the images.
For detecting Human firstly a dataset of positive and the negative images are taken. After that the image is converted into grey colored image using grayscale and the gradient of the image is calculated.
Histogram of gradient is calculated using HOG features. The positive and negative images are then trained using the cascading detector. After training an .xml file is created.
For detecting Human here four layers are used.
People Detector: The people detector detects people from an input image using the Histogram of Oriented Gradient (HOG) features and a trained Support Vector Machine (SVM) classifier. The detector detects concluded people in an upright position.

Face Detector: The cascade detector uses the Viola-Jones algorithm to detect people's faces, noses, eyes, mouth, or upper body. You can also use the train Cascade Object Detector function to train a custom classifier to use with this System object.

Upper body Detector: The cascade detector uses the Viola-Jones algorithm to detect people's faces, noses, eyes, mouth, or upper body. You can also use the train Cascade Object Detector function to train a custom classifier to use with this System object.

Custom Detector: This detector detects humans using the positive and negative trained images. Using all these layers a human is detected.

IV. Object Detection


.XML Database File
Feature Extraction
Apply Cascading HOG Detector
Human Activity Classify
Start
Input Image
Object Detection
Human and Object ROI Distance Measure
If > Threshold then Include Else < Threshold Not include
Hole Image Activity Classify

Input Image:-
Detecting the human activity and objects is a very typical task from the color image. So, here in our work, explaining that from the color image it is possible to extract the image. Here the dataset of the colored images is taken for the different poses of the images along with the colored images of different objects. Using this dataset from any images human activity and the objects can be detected.

Feature Extraction:-
After the dataset of the image is generated, we need to extract the features of the image. For using the HOG features, firstly we need to convert the color image into gray scale image. The gray scale image coverts the RGB image into grayscale using rgb2gray. Then after finding grayscale image we need to apply the HOG feature on the image.

HOG feature is extracted from the grayscale image applied as an input. These features are returned in 1-by-N vector, where N is the HOG feature length. The returned features encode local shape information from regions within an image. This information can be used for the tracking, detecting and classification purposes.

XML file:-
After training the image using the cascade detector, one XML database file will be generated which contains each and every information of the image. The xml file extension is .xml file.

Human Activity Classify:-
As database is created and training is applied on the images now, we need to find whether the human can be detected or not by comparing the negative and positive images and the image which we provide as an input. After detecting the human we need to classify the different basic activities of the human like running, walking, sitting etc.

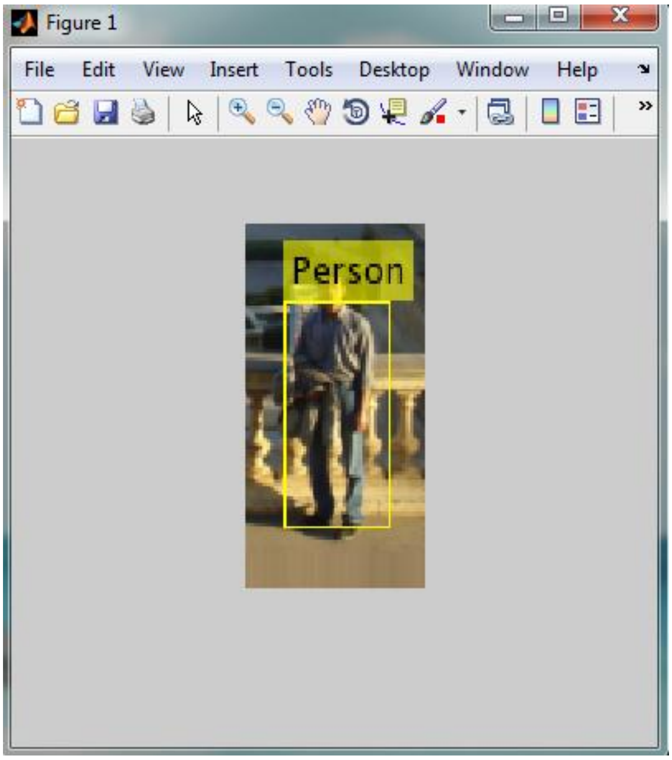
Object Detection:-
Object detection is the identification of an object in an image or video.

Human and Object ROI Distance Measure:-
When we have detected the human and the object after that we need to find the ROI distance between the human and the object. A region of interest (ROI) is a portion of an image that you want to filter or perform some other operation on. You define an ROI by creating a binary mask, which is a binary image that is the same size as the image you want to process with pixels that define the ROI set to 1 and all other pixels set to 0.More than one ROI can be defined in an image.
Measuring the distance between object and human.-
After detecting the human we need to find the distance between the human and object.

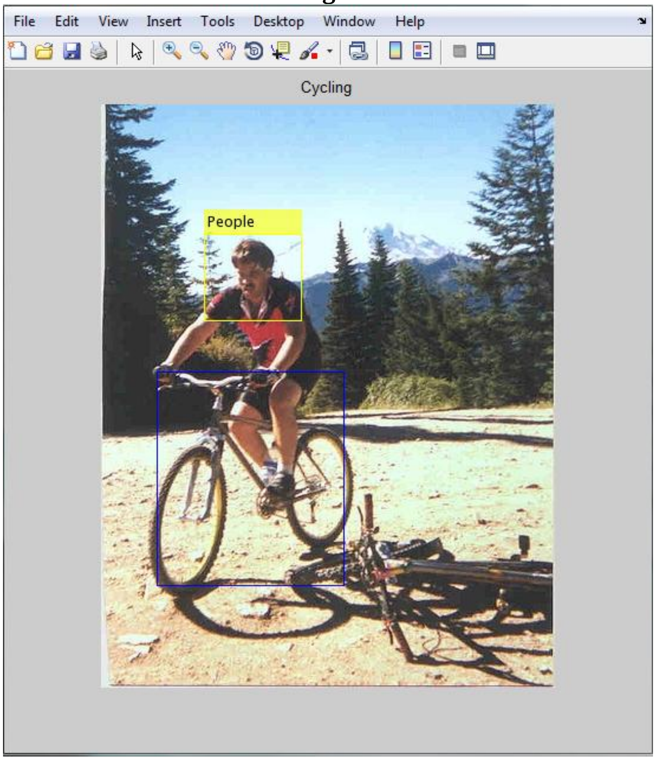
Threshold:-
The Human Activity and the Human and Object ROI distance are compared if the distance is greater than the threshold then the image was included else the image it will not be included.

V. Experimental Results

The final results shows the detections in the easy to use GUI here and can detect in three phases like human detection, object detection and activity detection. In first phase, the human features are extracted and a perimeter is created for human. After that the system searches for the object and if found then creates the perimeter to it as well. Then the distance is measured from center to center of the objects and if the perimeters overlap or the distance is minimal then we can safely assume that the human is performing a task by using the object, so if the human and a ball is detected in the image then the system would give the result as ‘the person is playing with a ball’. If there is no object present in the image then the system would try to recognize the pattern of human and detect the activity as walking or sitting etc. so that basically all the aspects can be covered for the detection.



Detecting Human



Detecting Human and Object with its Activity

Compariosn of Features

Methods	Time	No. of Features Extracted	Features
Harris	46.121705 seconds	172	It returns a corner Points object, points. The object contains information about the feature points detected in a 2-D grayscale input image, I. The detectHarrisFeatures function uses the Harris–Stephens algorithm to find these feature points.
MSER	4.206483 seconds	72	It returns an MSERRegions object, regions, containing information about MSER features detected in the 2-D

			grayscale input image, I. This object uses Maximally Stable Extremal Regions (MSER) algorithm to find regions.
SURF	4.457103 seconds	161	It returns a SURF Points object, POINTS, containing information about SURF features detected in the 2-D grayscale input image I. The detectSURFFeatures function implements the Speeded-Up Robust Features (SURF) algorithm to find blob features

VI. Conclusion And Future Scope

This paper is based on the developed system which is capable of detecting the activity of a human from an image, an object from an image and the combined activity of human using the object. With the easy to use GUI all the detections can be accessed easily even by a non-technical person. The algorithms used for human and object detection which are HOG and SURF respectively are efficient enough for higher speed of execution and efficient detection. The system can be implemented for video in future as well as for low light conditions.

References

[1]. Aarsh Joshi,Chandni Shah,Khyati Jain,Hetal Parmar,Prof. Vaishali R Patel A Review on Human Activity Recognition Using HOG Feature IJSRD

[2]. Chi Qin Lai Soo Siang Teoh A Review on Pedestrian Detection Techniques Based on Histogram of Oriented Gradient Feature 978-1-4799-6428-4/14/$31.00 ©2014 IEEE

[3]. Navneet Dalal ,Bill Triggs "Histogram of Oriented Gradients for Human Detection" Proceedings of the 2005 IEEE computer society

[4]. Narges Gheadi ,Maziar Palhang "New Approach for Human Detection in images Using istogram of Oriented Gradient" 2013 IEEE .

[5]. Qiang Zhu ,Shai Avindan ,Mei-chen Yeh ,and Kwang-Ting Cheng "Fast Human Detection Using a Cascade of Histogram of Oriented Gradients" Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR'06)

[6]. P Govardhan ,Umesh C. Pati "NIR Image based Pedestrian Detection in Night Vision with Cascade Classification and validation"

[7]. J.Jasmine AnithaȦ* and S.M.DeepaȦ Tracking and Recognition of Objects using SURF Descriptor and Harris Corner Detection

[8]. Herbert Bay1, Tinne Tuytelaars2, and Luc Van Gool12 SURF: Speeded Up Robust FeaturesEdward Rosten and Tom Drummond

[9]. Machine learning for high-speed corner detection Department of Engineering, Cambridge University, UK{er258, twd20}@cam.ac.uk

[10]. www.mathworks.com/products/**matlab**/

[11]. E.Rosten and T.Drummond, (2005), Fusing points and lines for high performance tracking, *IEEE International Conference on Computer Vision*, volume 2.

[12]. D.Chekhlov, M.Pupilli, W.Mayol-Cuevas, and A. Calway, (2006), Realtime and robust monocular SLAM using predictive multi-resolution descriptors, *2nd International Symposium on Visual Computing*.