A Review on Comparative Study of Different Algorithms for Object Detection

ABHISHEK DAS[1] ,KESHAV KISHORE[2] ,NISHA GAUTAM[3]
AP Goyal Shimla University

Abstract- Object detection and tracking are playing an important role in many computer vision and pattern recognition applications such as video classification, vehicle navigation, surveillance and autonomous robot routing.It is a challenging field in computer visualization and pattern analysis research area. There are many techniques which have been proposed and developed. In this paper we present different approaches of detecting objects using different methods. The objective of this paper is to focus on the main techniques and algorithms on object detection.

Keywords- *Object Detection, Computer vision, Algorithms, SLAM, SIFT, SURF, HOG, DPM, Occlusion.*

I. INTRODUCTION

Object tracking and detection is a classical research area in the field of computer vision from decades. Numerous kinds of applications are dependent on the area of object detection, such as advance driving assistance system, traffic surveillance, scene understanding, autonomous navigation etc.[1]The core challenge and the basic step in tracking is to accurately detect the object in different environments, but due to complex backgrounds, weather conditions, cast shadows and occlusions it becomes difficult to track an object. In this paper we will be focusing on different object detection algorithms. Many computer vision algorithms suffer due to the presence of occluded objects in a scene. The region, which is occluded though, depends on the camera viewpoint. In some scenarios angle of the camera can define which part is occluded and which one is not, hence minimization approach, temporal selection, graph cut method and sum of squared distance are followed for handling the same problem of occlusion.[2] Object detection is the process of finding instances of real-world objects such as faces, bicycles, and buildings in images or videos. Object detection algorithms typically use extracted features and learning algorithms to recognize instances of an object category. It is commonly used in applications such as image retrieval, security, surveillance, and automated vehicle parking systems.[3]

You can detect objects using a variety of models, including

A. Feature-based object detection

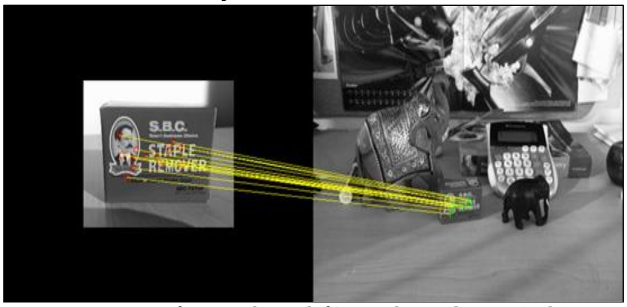


Fig 1: Detecting a reference object (left) in a cluttered scene (right) using feature extraction and matching. RANSAC is used to estimate the location of the object in the test image.

B. Viola-Jones object detection

Fig 2:Face detection (left) and stop sign detection (right) using the Viola-Jones Object Detector.

C. SVM classification with histograms of oriented gradients (HOG) features

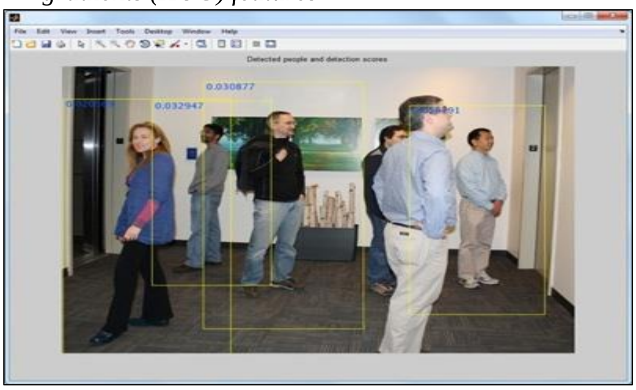


Fig 3: Human detection using pertained SVM with HOG features.

D. Image segmentation and blob analysis

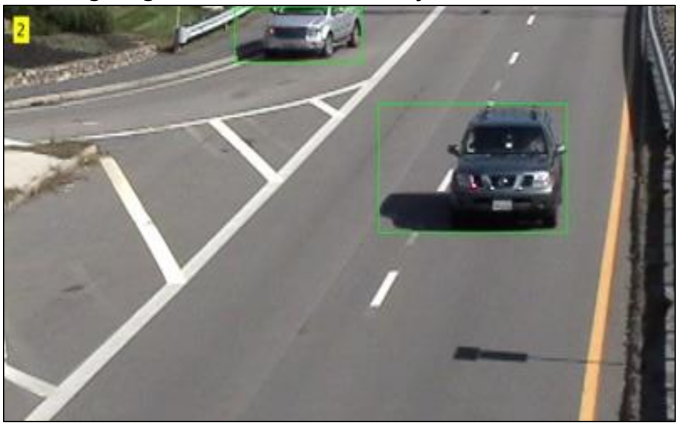


Fig 4: Moving cars are detected using blob analysis.

Other methods for detecting objects with computer vision include using gradient-based, derivative-based, and template matching approaches.[3]

II LITERATURE REVIEW

During literature review we came across a wide range of techniques and algorithms but here we will be focusing on the impactful algorithms given in the below image.

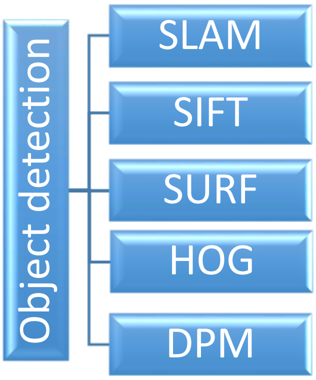


Fig 5:Important and reviewed techniques / algorithms

A. SLAM – Simultaneous Localizationand Mapping[4]

SLAM can be implemented in many ways. First of all there is a huge amount of different hardware that can be used. Secondly SLAM is more like a concept than a single algorithm. There are many steps involved in SLAM and these different steps can be implemented using a number of different algorithms. SLAM consists of multiple parts; Landmark extraction, data association, state estimation, state update and landmark update. There are many ways to solve each of the smaller parts. It is a process by a mobile device or a robot builds the map of an environment and uses this map at the same time to deduce its location in it. The SLAM is commonly used for the robot navigation systems in an unknown environment, it's more like a concept than a single algorithm. The first step in the SLAM process is to obtain data about the surroundings of the robot. Extended Kalman filter, a traditional approach is also quite often used for the estimation in robotics. While in SLAM during robot navigation in an environment the robot localize itself using maps. It is basically concerned with a problem of building a map of an unknown environment by a mobile robot while at the same time navigating the environment using the map. SLAM consists of multiple parts; Landmark extraction, data association, state estimation, state update and landmark update. The SLAM is based on Extended Kalman filter which utilizes the a priori map of the locations. The objective of the SLAM problem is to estimate the position and orientation of the robot together with the locations of all the features. SLAM is frequently used in Autonomous underwater vehicles, unmanned aerial vehicles and autonomous ground vehicles.

B. SIFT – Scale Invariant Feature Transform[5]

SIFT, in computer vision is used to detect and describe local features in images. In any particular image for object detection, interesting points can be figured out to describe a set of features. The particular description taken out from a training image in terms of numbers can be further used on a testing image to identify a particular object. For the best recognition results the features should be detected even in noisy and illuminated scenes or images. SIFT works on the principle of Euclidean distance of the feature vectors, firstly the key points of SIFT are extracted from a reference image and further stored in a database, after this an object is recognized in a new image by comparing the features of both the images. Scale space filtering is used to detect the larger corners with larger windows. In SIFT, Difference of Gaussian is used which is an approximation of LOG. The initial step of SIFT is to create internal representations of the original image to ensure scale invariance which is done by generating a scale space. The Laplacian of Gaussian (LOG) is great for finding interesting points which is the nucleus of this algorithm. At last with scaling and rotation invariance it's easy to identify the last set of unique features.

C. SURF – Speeded Up Robust Features[6]

Speeded up robust features (SURF), is inspired by Scale Invariant feature transform (SIFT), it is a local feature detector and descriptor that can be used for tasks such as object recognition or 3D reconstruction. The basic version of SURF is several times faster than SIFT and is much more robust. The algorithm works on interest point detection, local neighbourhood description and matching. SURF uses wavelet responses in horizontal and vertical direction for a neighbourhood pixels. The surf algorithm is based on two basic steps feature detection and description. The detection of features in SURF relies on a scale-space representation, combined with first and second order differential operators. The originality of the SURF algorithm (Speeded up Robust Features) is that these operations are speeded up by the use of box filters techniques.

D. DPM – Deformable Part Based Model[7]

Deformable part based model is based on an object category that represents the appearance of the parts and how they relate to each other. The part is any element of an object or scene that can be reliably detected using only local image evidence. In part based model, each part represents local visual properties. The basic idea behind deformable parts is to represent an object model using a lower resolution root template and set a spatially flexible high resolution part templates. Deformable part based model is the next revolutionary idea after the Histogram of orientation gradient in object detection. Threshold employed in the deformable part based model in the non-maximum suppression filter is the key root of this algorithm. Lower the threshold, higher the number of detection.

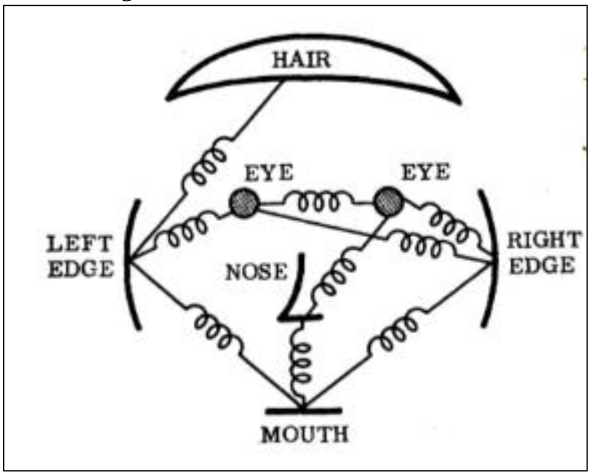


Fig 5: Basic idea behind DPM

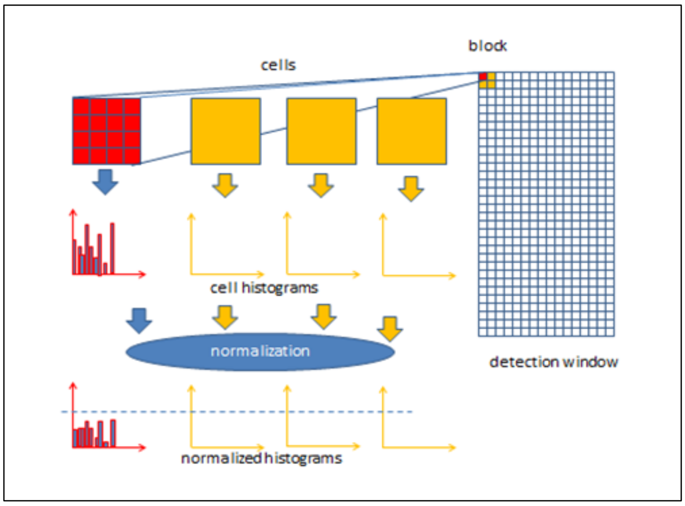
E. HOG-Histogram of Oriented Gradients [8]

Histogram of oriented gradients (HOG) is a feature descriptor used to detect objects in computer vision and image processing. The HOG descriptor technique counts occurrences of gradient orientation in localized portions of an image - detection window, or region of interest (ROI).

Implementation of the HOG descriptor algorithm is as follows:

1. Divide the image into small connected regions called cells, and for each cell compute a histogram of gradient directions or edge orientations for the pixels within the cell.
2. Discretize each cell into angular bins according to the gradient orientation.
3. Each cell's pixel contributes weighted gradient to its corresponding angular bin.
4. Groups of adjacent cells are considered as spatial regions called blocks. The grouping of cells into a block is the basis for grouping and normalization of histograms.
5. Normalized group of histograms represents the block histogram. The set of these block histograms represents the descriptor.

The following figure demonstrates the algorithm implementation scheme:

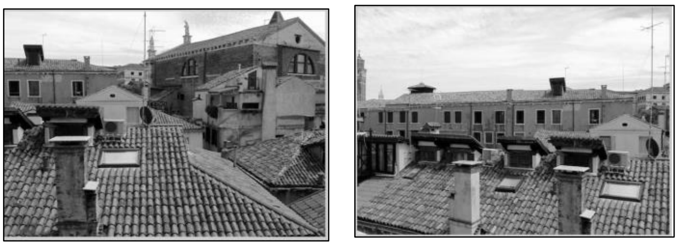


Computation of the HOG descriptor requires the following basic configuration parameters:

- Masks to compute derivatives and gradients
- Geometry of splitting an image into cells and grouping cells into a block
- Block overlapping
- Normalization parameters

III. RELATED PAPER SUMMARY

1. P M Panchal, S R Panchal, S K Shah ," A Comparison of SIFT and SURF", they compared the SIFT and SURF algorithm.[9]Two images are taken and Features are detected in both images using SIFT and SURF algorithm.



(a) Original image1 *(b) Original image*

Comparisons of resultsof SIFT andSURFalgorithm:

This paper has evaluated two feature detection methods for image registration. Based on the experimental results, it is found that the SIFT has detected more number of features compared to SURF but it is suffered with speed. [9]

Process	Percentage accuracy(%)
Only SIFT	56.66
Only HOG	76.66
Only SURF	81.66
SIFT with HOG	73.33
SIFT with SURF	83.33
SURF with HOG	85.00
SIFT,HOG and SURF	96.66

Algo	Detectedfeature Points		Matching feature point	Feature matchingTime(s)
	Image1	*Image2*		
SIFT	892	934	41	1.543
SURF	281	245	28	0.546

So overall accuracy for all classes using the Manhattan distance is 96.66% . It has been checked that different feature extraction process and there combination gives different accuracy for all classes .But in all cases classifier is Manhattan distance.[10]

IV. CONCLUSION

In this paper all the main techniques/algorithms of object detection have been addressed. Feature extraction of objects is one of the most important part of an object detection system. Advance study may open the paths to find efficient algorithms to reduce computational cost and to decrease the time required for detecting the object.

REFERENCES

[1]. Himanshu Chandel and Sonia Vatta," Occlusion Detection and Handling: A Review", International Journal of Computer Applications (0975 – 8887) Volume 120 – No.10, June 2015.
[2]. Sing Bing Kang, Szeliski R. ,Jinxiang Chai, "Handling occlusions in dense multi-view stereo," Computer Vision and Pattern Recognition, 2001.
[3]. https://in.mathworks.com/discovery/object-detection.html
[4]. SLAM for dummies http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-412j-cognitive-robotics-spring2005/projects/1aslam_blas_repo.pdf
[5]. David G. Lowe,"Visual Tracking of Hand Posture with Occlusion Handling,"15th International Conference on Pattern Recognition," Proc. of the International Conference on Computer Vision, 1999.
[6]. Herbert Bay, Tinne Tuytelaars, Luc Van Gool ,"SURF: Speeded Up robust features ,"15th International Conference on Pattern Recognition," 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I , pp. 404-417
[7]. P. Felzenszwalb, D. McAllester, D. Ramanan," A discriminatively trained, multiscale, deformable part model," IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008. pp. 1 – 8
[8]. https://software.intel.com/en-us/node/529070
[9]. P M Panchal, S R Panchal, S K Shah ," A Comparison of SIFT and SURF", International Journal of Innovative Research in Computer and Communication Engineering Vol. 1, Issue 2, April 2013
[10]. Chinmoy Biswas and Joydeep Mukherjee,"Logo Recognition Technique using Sift Descriptor, Surf Descriptor and Hog Descriptor" International Journal of Computer Applications (0975 – 8887) Volume 117 – No. 22, May 2015

Abhishek Das is currently the Mtech Research Scholar of the Department of Computer Science & Engineering at A P Goyal Shimla University, Shimla(H.P.) India.

Keshav Kishore is currently a Professor of the Department of Computer Science & Engineering A P Goyal Shimla University, Shimla(H.P.) India.

Nisha Gautam is currently a Professor of the Department of Computer Science & Engineering A P Goyal Shimla University, Shimla (H.P.) India.