Real Time Object Identification Using fast Figure-Ground Separation

Priyadarsini.S
Dept. of Computer Engineering
College of Engineering, Chengannur
Kerala, India
E-mail: spriyadarsini12@gmail.com

Dhanya Raveendran
Dept. of Computer Engineering
College of Engineering, Chengannur
Kerala, India
E-mail: dhanyaishwar@gmail.com

Abstract— Computer vision is a field which includes some methods for acquiring, processing, analyzing, and scene understanding in machines .Object Identification is the ultimate aim of scene understanding, which is a major problem of computer vision. In this paper, saliency detection is done after acquiring video at real time by a robot which ultimately reduces the amount of data for further processing. Object detection is done in the resultant saliency map mainly using thresholding and bounding box technique. Object Identification is the final step in which recognition of objects detected is done after comparing with a trained data set and classify to appropriate class.

Keywords- Computer vision, scene understanding, saliency map, object detection, bounding box, object identification, knn classifier.

I. INTRODUCTION

The main aim behind the development of the computer vision is to duplicate the abilities of human vision and scene understanding by electronically capturing and interpreting an image. An affluent stream of visual data enters into the eyes of robots every second. Processing this data in real-time is needed before object identification which is an extremely overwhelming task. Also the complexity and speed of these processes depends on data submitted to the machine as the captured image. Thus the robots need to reduce the amount of data to be processed so as to work at real-time. This paper presents a method to reduce the amount of data to be processed by extracting the foreground of the images taken and then give it as input to the machine for further processing. The next step is object detection which is the search to find the objects present in the captured scene, which can be limited to the salient region which is found earlier using the fast foreground detection. The objects found can be called as candidate objects for identification or recognition step. Identification can be done by comparing the detected objects with some class of already trained objects. The

knn classifier can be used to classify the detected objects to different categories.

The rest of this paper is organized as follows : Section 2 surveys related work to put this paper in the right context. Section 3 describes our proposed system of real-time Object identification. Section 4 provides extensive experimental results and Section 5 draws conclusions and discusses future work.

II. RELATED WORKS

A lot of research is being going on in the field of robotics. X.hou et.al introduced a simple image descriptor referred to as the image Signature in their paper[1]. They presented, within the theoretical framework of sparse signal mixing, that the quantity spatially approximates the foreground of an image. They experimentally investigated whether that approximated foreground overlapped with visually conspicuous image locations by developing a saliency algorithm based on the image signature.

The ability of human visual system to detect visual saliency is extraordinarily fast and reliable. However, computational modelling of this basic intelligent behavior is a challenge. Bruce and Tsotsos presents a simple method for the visual saliency detection in their paper [2]. The model is independent of features, categories, or other forms of prior knowledge of the objects. By analyzing the log-spectrum of an input image, the spectral residual of an image in spectral domain is extracted , and proposed a fast method to construct the corresponding saliency map in spatial domain.

A more relevant study came from Hou and Zhang [3]. They found that the residual Fourier amplitude spectrum, the difference between the original Fourier amplitude spectrum and its smoothed copy, could be used to form a saliency map. The residual retains more high-frequency information than low, where the smoothed copy is similar to the original.

Holistic image processing short-circuits the need for segmentation, key-point matching, and other local operations. Bolstered by growing general interest in large-scale image retrieval systems, holistic image descriptors have become a topic of intense study in the computer vision literature. GIST [4] is an excellent example of such an algorithm in the field.

Another related work by Harel,Koch and Perona [5] attempted to ultimately highlight a handful of significant locations where the image is informative according to some criterion A new bottom-up visual saliency model, Graph-Based Visual Saliency (GBVS),was proposed .This model powerfully predicted human fixations but it was not fast enough to work at real- time.

III. PROPOSED SYSTEM

The previous works in the field attempted to find the saliency map of images in database .The proposed system was designed to take the image in real time as machines or robots really do , generate the saliency map in very less time and then identification process to be done on that map. The overall design of the project was organized as four processes. The first process was the image capturing which was supposed as the robot was taking image of real world a scene and preprocessing was done. The second process was the saliency map generation of the captured image. After obtaining the saliency map, the third step object detection was done by finding possible foreground boundary. Finally object identification was done after training the system.

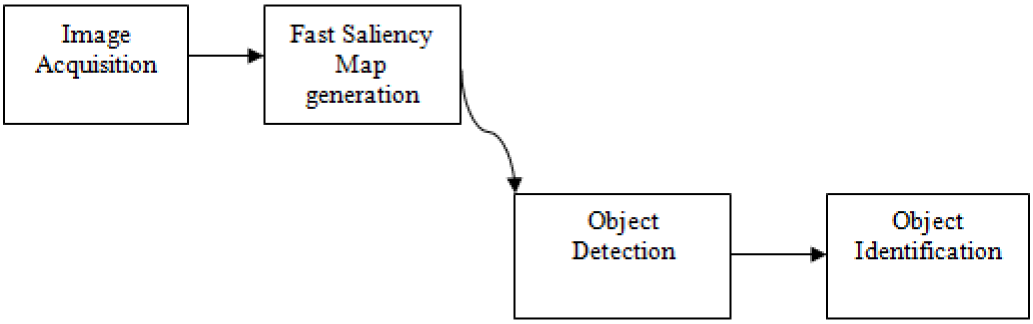

Figure 1 . Overall design of the work done.

A. *Image acquisition*

The machines are provided with camera in the place of eyes in humans. The first stage of the paper was to capture the path as a dynamic video, and then sampled frames or images are extracted from the video. These extracted frames were inputs to the next stages which was the saliency map generation.

B. *Fast Saliency Map Generation*

The paper assumed that an image foreground as visually conspicuous relative to its background. The image from camera was taken as input. 2- D Discrete Cosine Transform (DCT2) was done on that image and the changed image was stored in variable called X. Sign (X) was taken where Sign function was called Signum function and it was stored in X1. i.e., X1=sign(X)

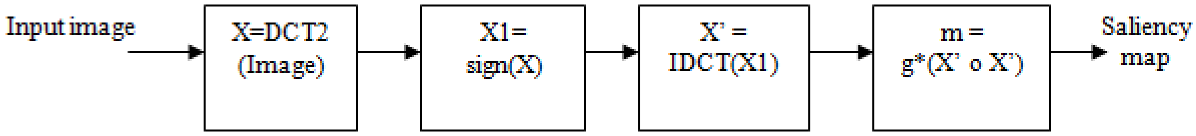

Figure 2. Steps for saliency map generation from real-time captured scene .

Then we took the 2- D Inverse Cosine Transform (IDCT2) of the above coefficient X1 and were stored in X. The last step used the Hadamard product operator to multiply the coefficient X and that result was multiplied by g, where g is a Gaussian Kernel. That is $m = g*(X' \circ X')$. The Output was the saliency map of the image.

The approximate foreground location highlighted by the above method was found to be remarkably consistent with the locations of human eye movement fixations .In short during the saliency map detection step the coordinates of the image that correspond to information rich regions are extracted. The map generation was done in two different color spaces for a comparative study.

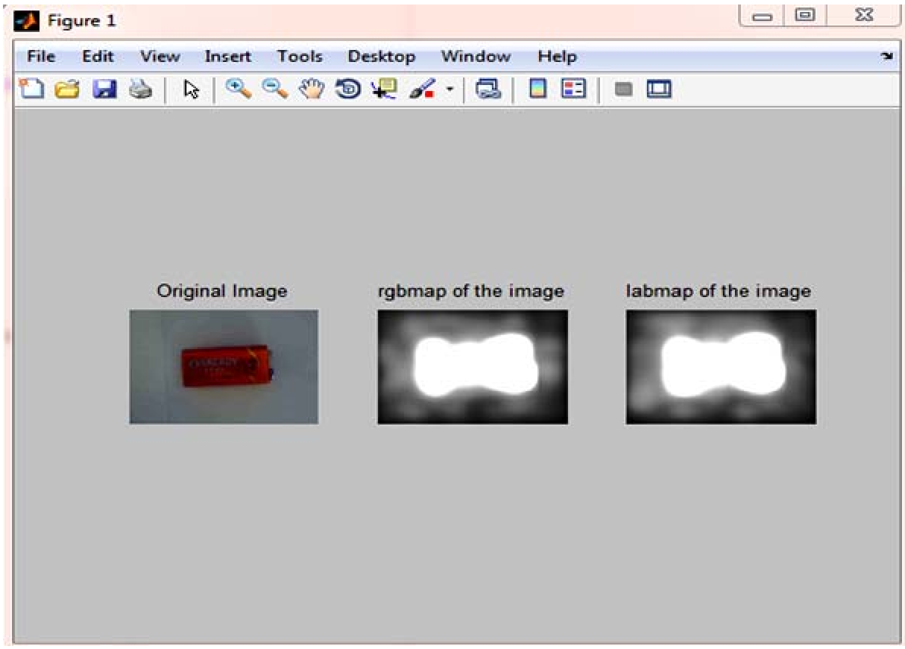

Figure 3. Saliency map generation in RGB and Lab color spaces

C. Object Detection

The Object detection is the process of finding instances of real-world objects such as faces, vehicles, buildings, humans etc in images or videos. Object detection algorithms typically use extracted features and learning algorithms to recognize instances of an object category. The saliency map obtained was the input of detection stage .

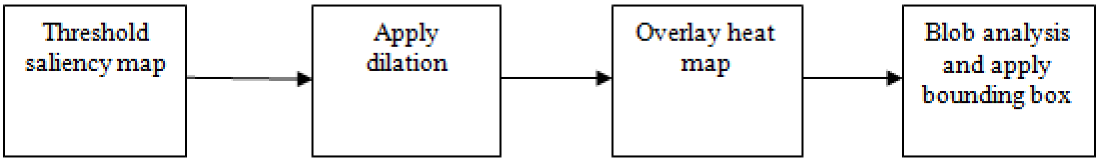

Figure 4. Steps for Object detection from the saliency map .

The saliency map obtained from the Saliency map generation phase was a gray scale image with blurred edges. Dilation causes objects to dilate or grow in size. The amount and the way that they grow or shrink depend upon the choice of the structuring element. The basic effect of the operator on a binary image is to gradually enlarge the boundaries of regions of foreground pixels. Thresholding is used to split an image into smaller segments, or junks, using at least two colors. Here the two colors used are black and white. So the resultant image was a black and white image instead of the gray level input image. For a binary image, white pixels are normally taken to represent foreground regions, while black pixels denote background.

Heatmap is a false color 2D image of the data values in a matrix. HeatMap displays the matrix data values as colors in a two-dimensional map. Blob detection was usually done after colour detection and noise reduction to find the required object from the image. A blob is defined as a region of connected pixels. Blob analysis is the identification and study of these regions in an image. Those regions could signal the presence of objects or parts of objects in the image domain with application to object recognition and/or object tracking. The method detected points and/or regions in the image that differ in properties like brightness or color compared to the surrounding. The algorithm discerned pixels by their value and place them in one of two categories: the foreground or the background . Bounding Box is a rectangle which containing an object and which sides touch the object. A minimal bounding box was preferred over the detected blob.

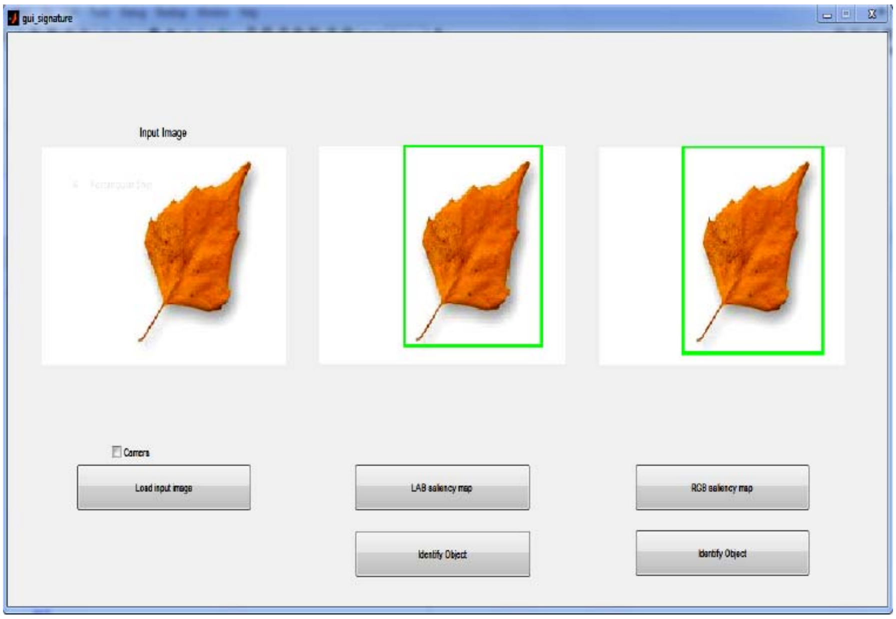

Figure 5. Application of bounding box in Object Detection phase

D.Object Identification

In Object identification phase we attempted to interpret the captured scene to find which all objects are present. It was done through a series of steps. To identify an object the previous knowledge about that object was necessary for machines. For that we gave a training to the computer .There was the classification module which had the main task to classify the extracted regions of interest presented to its input into the category they belonged to. Further the identification can be performed in the bounding box and avoid the rest of the image at the time of processing.

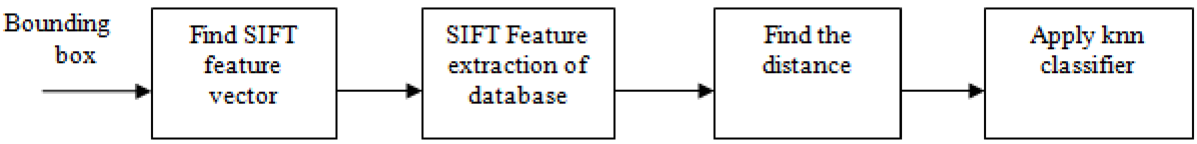

Figure 6. Object identification steps

The features set extracted the relevant information from the bounding box to represent the candidate object for identification. Corner detection was performed in the bounding box .The non maximal suppression and thresholding were performed. We extracted the local maxima by performing a grey scale morphological dilation and then found points in the corner strength image that matched the dilated image and greater than the threshold. The result was a binary image marking row coordinates and column coordinates of corner points. The sift vector around the corner was found and stored that vector in to a matrix. For training we found the the sift properties of the images of different class of objects in the database and formed a histogram of similar features. The next step was to find the similarity by finding the Euclidian distance between the object in captured image and the database. Apply knn classifier to classify the objects to best matching class.

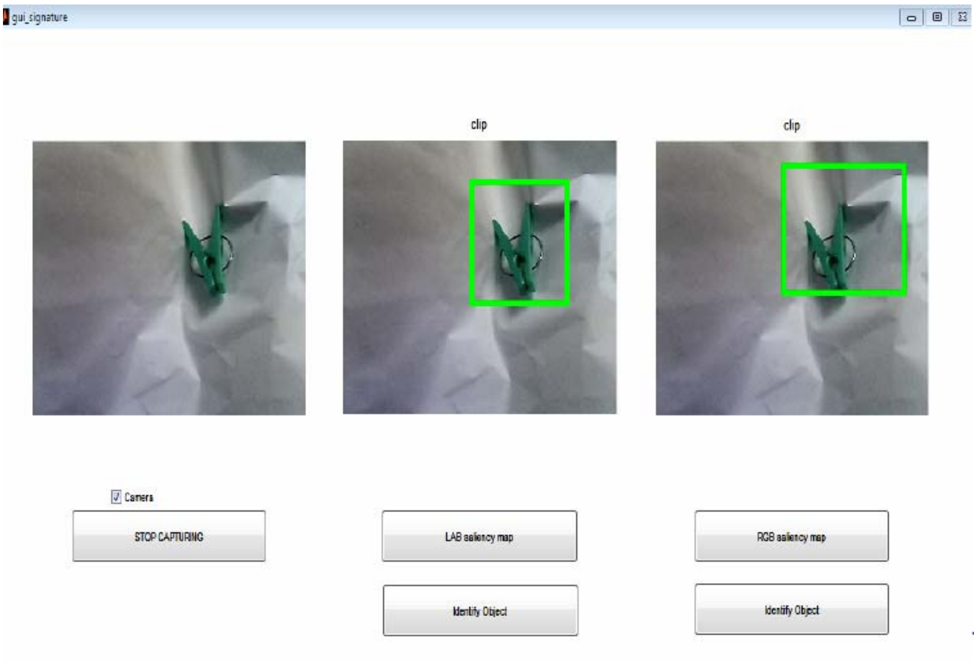

Figure 7. Object Identification in the Bounding box

IV. EXPERIMENTAL RESULTS

The idea in this paper was designed for robots which can be used for real time object detection and identification. The image frames were extracted from video input from the camera of robot. Then for the ease of execution and for time saving, saliency map of that scene was created. The machine identified whether the trained object was present in the scene and labels it. it. The two different color spaces, RGB and Lab, were used when finding the saliency map and found to be almost similar. The identification was also found to be same for the two color models. The time taken to identify an object was found to be 2 seconds on an average.

V. CONCLUSION

The work in the paper suffers a little delay in identification due to the processing speed of MATLAB. This can be improved by using other computer vision dedicated softwares such as OpenCV. The robots which are meant to work in certain environment can be trained with the objects which are supposed to present in that place.

REFERENCES

[1] X Hou, J Harel ,C.Koch “,Image Signature: Highlighting Sparse Salient Regions” IEEE Transactions on pattern analysis and machine intelligence, vol 34, No1, pp. 194-201 January 2012

[2] N. Bruce and J. Tsotsos, “Saliency, Attention, and Visual Search: An Information Theoretic Approach “J. Vision, vol. 9, no. 3, pp. 1-24, 2009.

[3] Xiaodi Hou1,2 and Liqing Zhang1,” Dynamic Visual Attention: Searching for coding length increments” Proc. Advances in Pattern Analysis and Machine Intelligence, pp.1-8 ,2007

[4] A. Oliva and A. Torralba, “Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope” Computer Vision, vol. 42,no. 3, pp. 145-175, 2001.

[5] J. Harel, C. Koch, and P. Perona, “Graph-Based Visual Saliency”, Proc. Advances in Neural Information Processing Systems, pp. 681-688, 2007.