

# **A Robust Object Detection and Recognition using Deep Features based Appearance-Model**

Ranjana Kapila  
Chandigarh Engineering  
College, Landran, India

Heena Wadhwa  
Assistant Professor  
Chandigarh Engineering,  
College, Landran, India

## **ABSTRACT**

The object detection models are very important parts of the vehicular monitoring, traffic surveillance or traffic planning in the urban or crowded areas of the cities. The satellites images are the primary sources of the urban surveillance; hence the proposed model has been proposed to work with the satellite imagery data. The object detection and localization plays the vital role in the various surveillance applications, vision based applications, augmented reality and robotics. The object detection and recognition becomes very important to localize the specific objects for augmented reality and robotics applications. In this thesis, the work would be carried forward on the mechanism using the Deep Features based Appearance-Model for the object detection and recognition in the target images.

## **General Terms**

Machine Learning, Artificial intelligence, Image recognition, Algorithm

## **Keywords**

Deep learning, Neural network, Appearance model, Pattern recognition, deep features.

## **1. INTRODUCTION**

The object oriented classification is a new classification method which is based on image segmentation [1]. This is opposed to per pixel image segmentation method. In this kind of image segmentation objects are the basic units for classification which are composed of a group of pixels representing real world objects (that the user intends to identify) rather than per pixel as the basic unit of classification. These objects are created by image segmentation [2] process mainly the region based segmentation process. The purpose of image segmentation is to subdivide an image into different segments/polygons that corresponds to objects in the terrain. Image segmentation techniques make use of both spectral information (feature vector of the pixels) and spatial information [3] (size, shape, texture, contextual information and adjacency to other pixels) to identify and delineate suitable segments within an image. Most approaches to shape matching, find a transformation that gives a factor of dissimilarity between two shapes. The extent of this dissimilarity decides whether we have a match. People have used various properties like curvature, contour to describe shape boundaries [4]. Shape matching has also been implemented using various feature-based and brightness based methods. Feature based methods [4] has been used for silhouette images. The feature being Characteristic of the object is used in content based matching approaches. In eigenvector or modal matching based approaches, sample points are transformed into a mass-spring damper equation

and the vibrations are used to estimate matching. PCA [5] was successfully used for faces to create a dimensionally reduced mean shape that was used for matching shapes. The region based segmentation process is performed by different methods. Some of the methods are clustering method, histogram based method, region growing method, split and merge method, graph partitioning method, and model based segmentation method, multiscale segmentation method and semi automatic segmentation method. For the present lab region growing segmentation method [6] will be us. Some of the disadvantages of this kind of classification method are tedious while creating the segments, unavailability of software packages which is actually limitation of this lab rather than disadvantage of this method. Also the classification method is overly dependent on the accuracy of the segmentation method used [7]. In the last few years [4], the problem of learning object class models and localizing previously unseen instances in novel images has received a lot of attention. While many methods use local image patches as basic features, recently several approaches based on contour features have been proposed. These are better suited to represent objects defined by their shape, such as mugs and horses. Most of the methods that train without annotated object segmentations [5] can localize objects in test images only up to a bounding-box, rather than delineating their outlines. It is believed that the main reason lies in the nature of the proposed models, and in the difficulty of learning them from real images, as opposed to hand-segmented shapes. The models are typically composed of rather sparse collections of contour fragments with a loose layer of spatial organization on to. A few authors even go to the extreme end of using individual edgels as modeling units. In contrast, an explicit shape model formed by continuous connected curves completely covering the object outlines is more desirable, as it would naturally support boundary-level localization in test images. In order to achieve this goal, an approach is proposed [11] which bridge the gap between shape matching and object detection. Classic non-rigid shape matchers produce point-to-point correspondences, but need clean pre-segmented shapes as models. In contrast, a method that can learn complete shape models directly from images is proposed. Moreover, it can automatically match [8] the learned model to cluttered test. In this research, the problem of a robot grasping 3D objects of known 3D shape from their projections in single images of apple trees. In the context of object grasping and manipulation, object recognition has always been defined as simultaneous detection and segmentation [9] in the 2D image and 3D localization. 3D object recognition has experienced a revived interest in both the robotics and computer vision communities with [10] RGB-D sensors having simplified the foreground-background segmentation problems. Nevertheless, difficulties remain as such sensors cannot generally be used in outdoor environments yet [9]. The goal of this paper is to

detect and localize objects in single view RGB images [11] of environments containing arbitrary ambient illumination and substantial clutter for the purpose of autonomous grasping. Objects can be of arbitrary color and interior texture and, thus, we assume knowledge of only their 3D model without any appearance/texture information. Using 3D models [10] makes an object detector immune to intra-class texture variations.

## 2. LITERATURE REVIEW

**Peilin Li et al. (2011)** said that in horticultural industry, conventional harvesting is done by 'handpicking' methods to remove hundreds of fruits such as citrus fruits in random spatial locations on the individual fruit trees. It is well known that harvesting fruits in a large scale is still inefficient and not cost effective. To solve this challenging task, mechanical harvesting systems have been investigated and practiced to enhance profitability and efficiency of horticultural businesses. **B.Sathya Bama et al. (2011)** proposes an efficient computer-aided Plant Image Retrieval method based on plant leaf images using Shape, Color and Texture features intended mainly for medical industry, botanical gardening and cosmetic industry. **Parisa Pouladzadeh et al. (2012)** presented a new recognition algorithm for emerging food classification. The algorithm considering its shape, colour, size, and texture characteristics. Using various combinations of these features, a better classification will be achieved. **Meenu Dadwal et. al (2012)** representing different techniques to detect the rate of ripeness of fruits and vegetables. This paper reports techniques like histogram matching, clustering algorithms based image segmentation and relative value of parameter based segmentation. **Alireza Masoudian et al. (2013)** said that image classification and segmentation are the two main important parts in the 3D vision system of a harvesting robot. Regarding the first part, the vision system aids in the real time identification of contaminated areas of the farm based on the damage identified using the robot's camera. **Fleites, Fausto et. al. (2015)** has worked on the enhancement of the product detection with multicue optimization for TV shopping applications. The proposed model has been designed to realize this use case and detect the products in the content stream must be detected so that the TV system notifies consumers of possibly interesting ones. **Girshick, Ross et. al. (2016)** has worked on the region-based feature extraction for the convolutional networks to achieve the higher accuracy of classification.

### 2.1 Findings Of Literature Review

Image representation is the task exercising the different ways of presentation of the image. The image representation also deals with the image formats and pixel density of the images. The image representation in some ways also handles the conversion of the low resolution images into the high resolution images, especially when there is a need of object or feature extraction from the particular image [12]. The activity of image representing (specifically image reconstruction) explores how images are displayed, based on the pixel as a building block. In particular, the great quantity of data in an image means that we need to use compression to be able to store and transmit it efficiently. The images are represented as the multi-dimensional matrices in the computer systems. In case of the satellite image, [2, 12] the image reconstruction becomes very important. The image reconstruction [12] is the practice which removes the noise and may convert the image into high resolution, which makes it easier to extract the desired feature from the image with high accuracy. The image matrices can be classified as the non-negative matrices [2]

which contain the values between 0 and 255 in general, specifically while kept in the most popular image encoding formats like, double or binary. The following problems are found in the existing schemes:

The use of appearance based models, which utilizes the multiple characteristics for the detection and recognition of the objects in the target images. The appearance based model will be developed with the combination of the contour analysis, followed with non-uniform object centric clustering & color spectrum layout. The improvement will be mainly aim at the removal of the unwanted regions, which are not depicting existence of the objects. The removal of the false positive cases [12] will be carried forward by using the color based features, low level features with speeded up robust features (SURF) and probabilistic classification model.

## 3. METHODOLOGY

At the first step, the literature review will be conducted on the concerned neural network techniques for the object detection and image reconstruction. The techniques will be shortlisted for the creation of the hybrid model for the object detection based on the neural network techniques. The problem formulation would be formed after finding the research gaps in the existing object detection models. The proposed model would be framed and finalized using the repeated rounds of algorithm improvement and theoretical debugging. The proposed and existing models would be implemented in the MATLAB simulator and the results would be collected in the form of various performance parameters. The collected results would be analyzed in-depth and the conclusion would be prepared to project the final results of the proposed model.

## 4. CONCLUSION

The proposed method will use a combination of the color and texture features along with robust probabilistic classifier model for the features extraction and positioning for the vehicle detection. The proposed system is aimed to be improved than the existing algorithm for the purpose of object detection and recognition. The proposed model will be using the combination of image segmentation model (Contour analysis), with color (Histogram of oriented gradients) & texture (Speeded up robust features) with robust probabilistic classification algorithm. The contour analysis model is utilized to outline visible objects in the input image using the snake model, where all of the possible objects are detected and outlined using the polygonal outlines. The result of the contour analysis model is further analyzed, and the valuable or descriptive objects are shortlisted for the further analysis. The histogram of oriented gradients (HoG) can be further applied over the extracted data using the contour analysis, which is used to obtain the color based features of the objects and further utilized for the classification purposes. The multiple features will help perform the image reconstruction and representation along with contour analysis, which will improve the overall accuracy of the proposed model on the given image dataset. The comparison of the proposed model will be based upon the various performance parameters like FAR, Precision, Recall, mean accuracy, categorical accuracy.

## 5. REFERENCES

- [1] Plebe A, Grasso G. Localization of spherical fruits for robotic harvesting. *Mach Vis Appl.* 2001;13(2):70-79. doi:10.1007/PL00013271
- [2] Fleites FC, Wang H, Chen SC. Enhancing Product Detection with Multicue Optimization for TV Shopping Applications. *IEEE Trans Emerg Top Comput.*

2015;3(2):161-171.

- [3] Zhang B, Huang W, Wang C, et al. Computer vision recognition of stem and calyx in apples using near-infrared linear-array structured light and 3D reconstruction. *Biosyst Eng.* 2015;139(November 2015):25-34.
- [4] Alfatni MSM, Shariff ARM, Shafri HZM, Saaed OM Ben, Eshanta OM. Oil palm fruit bunch grading system using red, green and blue digital number. *J Appl Sci.* 2008;8(8):1444-1452.
- [5] Pouladzadeh P, Villalobos G, Almaghrabi R, Shirmohammadi S. 2012 IEEE International Conference on Multimedia and Expo Workshops A Novel SVM Based Food Recognition Method for Calorie Measurement Applications. 2012:495-498.
- [6] Ishak W, Ismail W. Hue Optical Properties to Model Oil Palm Fresh Fruit Bunches Maturity Index. :15-17.
- [7] Noorul S, Hassan A, Nadiyah N, Abdul S, Zaw Z, Win SL. VISION BASED ENTOMOLOGY– HOW TO EFFECTIVELY EXPLOIT COLOR AND SHAPE. 2014;4(2):1-12.
- [8] Bama BS. CONTENT BASED LEAF IMAGE RETRIEVAL ( CBLIR ) USING SHAPE , COLOR AND TEXTURE FEATURES. 2011;2(2):202-211.
- [9] Jiménez, A. R., Ceres, R., & Pons, J. L. (2000). A vision system based on a laser range-finder applied to robotic fruit harvesting. *Machine Vision and Applications*, 11(6), 321-329.
- [10] Agriculture F, Haji H, Universiti R, et al. Image based modeling for oil palm fruit maturity prediction. 2015;(April 2010).
- [11] Margaritis D, Thrun S. Learning To Locate An Object in 3D Space From A Sequence Of Camera Images. 1998:332–340.
- [12] Girshick R, Donahue J, Member S, Darrell T, Malik J. Region-based Convolutional Networks for Accurate Object Detection and Segmentation. :1-16.