Object Detection System Using K-Means Clustering

S. Srihari[1], Prof. Sridhar Ranganathan[2]

[1]SCOPE-School of computer science Tamil Nadu Vellore Institute of Technology, Chennai, Tamil Nadu
[2]Vellore Institute of Technology, Chennai Tamil Nadu

ABSTRACT: Object Detection is one of the most popular applications in the branch of computer vision. While accuracy has always been the focus, focus has gradually also shifted to lightweight models. In this paper we propose a light weight system where object classification is not required but only object detection using clustering methods.

INDEX TERMS: Object detection, Computer vision, Elbow method, K-means clustering

I. INTRODUCTION

Computer vision is the branch of Artificial Intelligence that deals with visual world. Object detection is one of the most popular applications of computer vision. Object detection applications generally deal with Object detection and Object classification.

Various models have been developed for Object detection. Over the years. Object Detection models have been generally based deep learning based neural networks. These models generally have a lot of size and generally classification module is part of it. But in applications where there is no need for object classification, we can use a lighter alternative. In this paper, an method based on simple k means clustering has been proposed for object detection. Object Detection is the process of detecting the location/coordinates of objects present in a frame. A very popular application, a k means clustering based approach wherein the number of objects is decided by elbow method is discussed in this paper. Detailed explanations of the methodolgy proposed followed by a graphical explanation of our project has been discussed. The discussed method can be extremely efficient especially in cases where object classification is not required

II. RELATED WORK

Kim J et al., [1] discusses an unsupervised moving object segmentation cum recognition system for ITS(Intelligent Transportation Systems). The first step is object detection. It is done by the combination of variation detection and adaptive thresholding. Moving objects can be detected by change in intensity. So a combination of difference between two successive frames and difference between frame and background image.

Adaptive thresholding as said before is used to set the threshold. A binary motion mask is obtained, where changing pixels are set to one and others are set to one. Clustering is sued for object segmentation. There may be a chance of over-segmentation. Using temporal data, region merging can be done, to solve this problem. The segmented region can be fed into neural networks which will give the object classification.

Nguyen, H. T., Lee et al.,[2] suggests a method for multiple object detection cum classification using deep learning methods and clustering. Underwater Sonar images and three dimensional point cloud LiDAR data are used for this research. Fully Convolutional Network, a variation and rather an improved version of CNN. The output size is reduced while keeping the input size intact. This steps helps in object segmentation out of background using which input data is prepared for clustering algorithms. K- Means Clustering and DBSCAN algorithms were used for clustering. Clustering helps in figuring out relevant objects and filtering out irrelevant objects. The results from both the clustering methods including both elbow and silhouette method were compared and DBSCAN clustering showed higher accuracy result. Redmon, J., & Farhadi, A suggests an approach called YOLO(You Only Look Once) in their paper [3] which is a very popular means of object detection. In this system the input image is divided into S x S grids. For every grid as the input to the CNN, anchor boxes are predicted per grid. The output shape will be in the form of (width * height *((no_of_anchor_boxes)* (BOX_INFO))). For every anchor boxes, coordinates data (based on which actual coordinates will be obtained) , object class probabilities and objectiveness score is predicted. Here objectiveness score is a meaure of whether an object is there in a predicted box. After filtering out the low quality boxes final object detection and classification results are obtained.

Girshick, R. [4] proposes an Object Detection Sysrem based on Fast Region-based Convolutional Network method (Fast R-CNN). Based on previous works, Fast R-CNN builds to efficiently classification of object proposals using deep convolutional networks. When compared with previous works, Fast R-CNN uses newer techniques for improving training and testing speed while also increasing accuracy of detection. Fast R-CNN trains the very deep VGG16 network 9 times faster than that of R-CNN, is 213 times faster at test-time, and achieves a higher mAP on PASCAL VOC 2012. Compared to

SPPnet, Fast R-CNN trains and tests faster VGG16, and is more accurate.

III. PROPOSED SYSTEM

In the proposed methodology for object detection, the background image containing the background scene(for eg – a wall , a screen etc) and a foreground image with objects to be detected are given as inputs. After both the inputs are checked to be of the same size, both the images are converted from the RGB format to Grayscale format. After converting the input images to grayscale, a new image is saved with value as the difference between the foreground and background image. The difference between the images helps to segregate the objects to be detected in the foreground from that of the background.

The step to be followed next is to do binary thresholding of the difference image. This step is really significant considering the following steps. By binary thresholding, values of grayscale over a particular value is set to max value of 255 and others are set to zero. While the difference images gives a visual description of objects to be detected, the thresholded image gives an more accurate view.

The thresholded image is further processed to remove any noise which is present. This is done using a Gaussian noise filter. From the filtered images, coordinates should be collected where the grayscale value is equal to max value of 255. The points collected signify the points of objects to be detected. From the points collected clustering is to be done for object detection is to be done. The clustering method that is used in the present paper is k means clustering.

To explain in short the method of k means clustering, a number of centroids are defined initially. Over various iterations distances of all points from the defined centroid and each point is assigned to the centroid which is close. After every iteration a new centroid is choosen based on the formed clusters. When there is no change in centroids in successive iterations, the process is stopped and centroid and clusters are fixed, K means clustering of the points collected are done for various values for k from 1 to max_no_of_objects . The elbow method is used to reason the optimal number of k i.e the number of objects in the image.

Elbow method is a very common method in determining the optimum number of clusters. In this method distortion i.e the sum of square of distances from the centroids are plotted against number of clusters, the point at which elbow or knee occurs is choosen for number of clusters

Using the clusters formed , the point's in the clusters are analysed and the leftmost point(x_l, y_l) , the topmost points (x_t, y_t) , the rightmost point(x_r, y_r) and the bottommost point(x_b, y_b) for each cluster is collected and bounding boxes for object detection are drawn with the following coordinates :
$(x_l - margin, y_t - margin)$, $(x_r + margin , y_t - margin)$, $(x_l - margin , y_b + margin)$, $(x_r + margin , y_b + margin)$

Fig 1. Flow diagram of proposed methodology
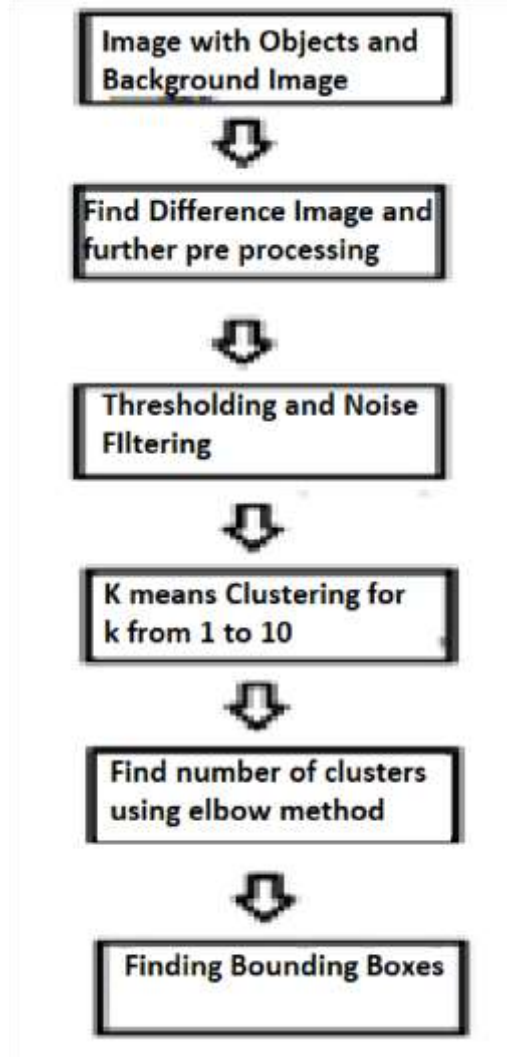
IV. EXPERIMENTAL SETUP

In our proposed setup, videos and photos are taken in a car. The objects are placed in a car and then photo's and videos are taken. Some of the objects placed include a laptop, tablet and a metal flask.

**Fig 2.Example of a photo of a object in backseat**

- Various combinations of these objects are placed and totally 18 photos and 4 videos were taken. The next section details our findings

V. RESULTS AND DISCUSSION

In our discussion of results let us see various scenarios and how our object detection system works for those scenarios. First let us look at the general scenario. In this example, there are two objects which are at a distance with each other. After the subtraction of background and subsequent processes, we do k means clustering with k from 1 to 10. After plotting the graph, we obtain the following.

Fig 3. Elbow diagram for determination of K for multiple objects
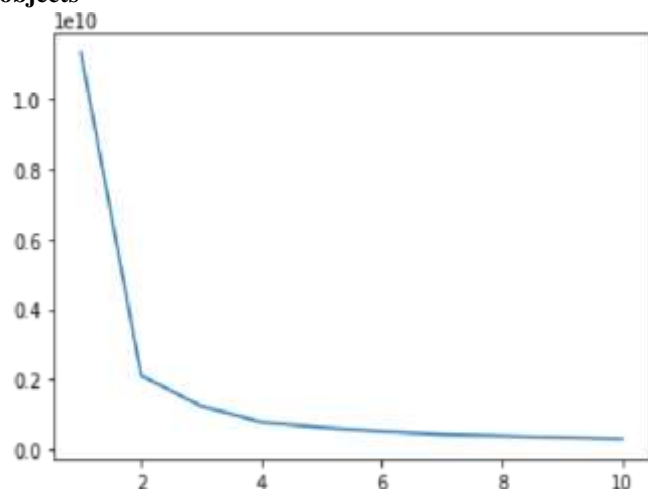
As one can see the elbow is observed at k = 2. This is equal to the number of objects present in the image. Hence the coordinates for bounding boxes are extracted for object detection and the following

Fig 4. Detected Objects

 This is the general scenario as discussed before. But there are few anomalies too using our method.

Single Object problem and k determination anomaly

Consider Fig2. in which a single object is there.
- After doing similar pre-processing and clustering with k = 1 to 10 and then plotting the following graph is obtained

Fig 4. Elbow diagram for a single object figure
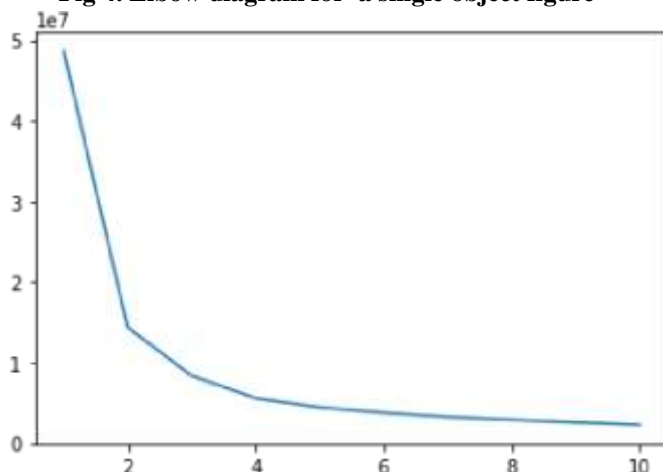
- As one can see the elbow cannot determine the number of objects here because elbow could not be obtained at k = 1. This is a disadvantage with respect to the method discussed in the paper.

- Another area where improvement could be made is when objects are overlapping with each as this might also provide a distorted view with respect to the number of clusters.

- Thus a light weight, and a quite efficient method was proposed for object detection. Results were quite accurate. A few improvements could be made as discussed before for the single object anomaly and the cases where objects overlap with each other.

VI. CONCLUSION

Thus a light weight, and a quite efficient method was proposed for object detection. Results were quite accurate. A few improvements could be made as discussed before for the single object anomaly and the cases where objects overlap with each other

REFERENCES

1. Kim, J. B., Park, H. S., Park, M. H., & Kim, H. J. (2002, May). Unsupervised moving object segmentation and recognition using clustering and a neural network. In *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)* (Vol. 2, pp. 12401245). IEEE.
2. Nguyen, H. T., Lee, E. H., Bae, C. H., & Lee, S. (2020). Multiple Object Detection Based on Clustering and Deep Learning Methods. *Sensors*, *20*(16), 4424.
3. Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7263-7271).
4. Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440-1448).

5. Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*. Vol. 1. Ieee, 2001.

6. Zhao, Zhong-Qiu. "Member, Shou-tao Xu, and Xindong Wu, "Object Detection with Deep Learning: A Review"." *Journal of Latex Class Files* 14.8 (2017).