Asian Journal of Computer Science and Technology
ISSN: 2249-0701 Vol.7 No.S1, 2018, pp. 112-114

RetinaNet Based Environment Classification

R. Balamurugan[1], R. Arunkumar[2] and S. Mohan[3]
[1]Research Scholar, [2&3]Assistant Professor
[1,2&3]Department of Computer Science and Engineering, Annamalai University, Tamil Nadu, India
E-Mail: r.balamuruganitau@gmail.com, arunkumar_an@yahoo.com, mohancseau@gmail.com

***Abstract* - Environmental classification is very useful for visually impaired persons and Robotic applications. The main objective of this work is to detect and recognize the objects present in a scene and identify the environment based on the occurrence probability of the objects in the scene. Objects from the real-time images are detected and recognized by means of RetinaNet. Occurrence probabilities of the recognized objects are used to identify the environment.**
***Keywords:* Object Detection, Object Recognition, RetinaNet**

I. INTRODUCTION

Environment classification will be more useful for visually impaired person and Robotic applications. Number of different approaches has been proposed to detect and recognize the object. Object detection and recognition plays a vital role in robotic applications. The objects appearing in an image may have a large range of variation due to: viewpoint changes, shape changes (e.g., non-rigid objects), photometric effects and scene clutter. Deep learning classifier RetinaNet has been used to detect and recognize objects in the scene.

A. Real -time Scenes: Real-time scenes consist of a variety of natural and manmade objects. In which every object may present in a specific environment or in very few environment only, which lead us to identify the environment based on the recognized objects in the image or scene.

Fig. 1 Computer Lab image

The above Fig.1, Fig.2 & Fig.3 shows the computer lab, hall and bed room, real time images with different objects. Each image contains a different set of objects and also few common object different images.

Fig. 2 Hall image

Fig. 3 Bed room image

B. Outline of Work: Environmental classification consists of different modules they were described in the upcoming sections. Pre-processing is described in section 2. Detection and Recognition of objects is described in section 3. Environmental classification is described in section 4. Experimental results are described in section 5. Conclusion is provided in section 6. Position estimation of the objects present in the recognized scene will be described as a future enhancement process.

II. PRE-PROCESSING

A. Log Polar: Image Log-polar pictures [1, 4] are specified devote a high sight within the centre of the sphere of read, so lost objects will be perceived with tremendous quality, as a result of the resolution decreases exponentially with eccentricity, the dimensions of the log-polar image is tiny.

B. Resize: Input images are resized [3] to a standard form 320 by 240 dimension, to reduce the Storage and the computational complexity.

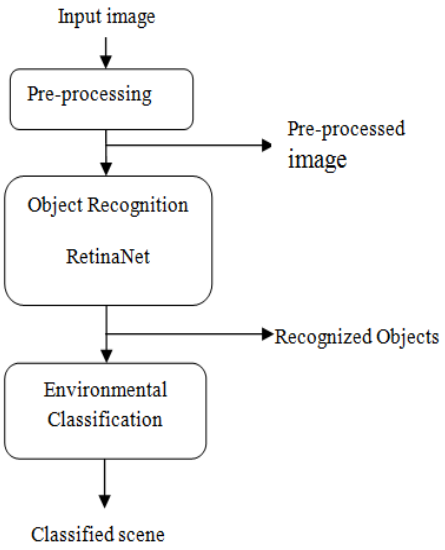

Fig. 3 Overall block diagram of the work

III. OBJECT DETECTION AND RECOGNITION OF OBJECTS

Object detection [4, 6, 7, 15] is the process of finding instance of real word object in a digital image or video. Object recognition, [3, 5, 13] identifies a specific object in an image or video. Object detection and recognition plays a major role in modern world applications, including surveillance monitoring, robotic applications, automatic identification in real world games, medical imaging etc. Large numbers of object detection and recognition approaches have been proposed, Feature-based object, Template-based object, Classifier Based Object and, Motion- based Object Detection. In our proposed work RetinaNet- deep learning model have been used for object recognition.

RetinaNet [1] is a deep learning model developed by Facebook which works very well on various object detections. It is a single, unified network; it is composed of a backbone network and two task specific subnetworks, Classification subnet and Box Regression Subnet. Backbone network is responsible for computing a convolutional feature map over an entire input image and is an off-the-self convolutional network. The Classification subnet performs convolutional object classification on the backbone's output; the Box Regression Subnet performs convolutional bounding box regression.

A. Backbone Network: Retina Net [1, 6, 18] network architecture uses a Feature Pyramid Network (FPN) as a backbone network. FPN enlarges the convolutional network with a top-down pathway and lateral connections to enlarge the network efficiently. Every levels of the pyramid have been used for object detection at a different scale. FPN is constructed on top of the ResNet architecture, with levels P_3 through P_7, here pyramid level is indicated by L (P_L has resolution 2^L lower than the input), C = 256 channels corresponding to every pyramid levels.

We employed translation-invariant anchor boxes that have areas of 32^2 to 512^2 on pyramid levels P_3 to P_7, respectively. Three aspect ratios {1:2, 1:1, 2:1} were used for anchors at each pyramid level. For denser scale coverage, at each level we added anchors of sizes $\{2^0, 2^{1/3}, 2^{2/3}\}$ of the original set of 3 aspect ratio. These improve AP in our setting. Totally there are A= 9 anchors per level and across levels they cover the scale range 32 - 813 pixels with respect to the network's input image.

B. Classification subnet: The probability of object present in every spatial position for all objects and anchors are predicted by the classification subnet. input feature map with C channels are taken from a pyramid level, this subnet apply four 3x3 convolution layers, each with C filters and followed by ReLU activations. Finally sigmoid activations are attached to the outputs.

C. Box Regression Subnet: Box regression subnet [1, 8] estimates the location of the objects with respect to anchor box. Both subnet shares the general structure with different parameters.

IV. ENVIRONMENTAL CLASSIFICATION

Real-time scenes consist of a number of different natural and manmade objects. In which every object may present in a specific environment or in very few environment only. In a real word environment, bed is only available on the bed room, cylinder, mixi, and grinders are specifically available on the kitchen, similarly tv is mostly available on Hall and bed room, likewise every environment has some specific objects which are only available on that environment this provides discriminate characteristics among different scenes and also has some common objects among different environment. In our proposed work, obprob model have been developed which gave high weight value to the specific object to an environment and low weight value to the common objects.

Recognized objects and its co-occurrence pattern and occurrence probabilities (obprob) are used to identify the environment. We proposed the obprob model, in which co-occurrence patterns of objects in a particular environment and its occurrence probabilities are used to assign weight value for each object in all environments. After objects are recognized from an image, weight value for each object, from the obprob model is multiplied with the current number of occurrence and summed. The environment which have above 80% resulted sum value will be chosen as a background environment.

V. EXPERIMENTAL RESULTS

Performance efficiency of the Environmental Classification based on RetinaNet was measured for variety of scenes.

Four categories of indoor objects Environments are considered such as Bed Room, Hall, kitchen, and lab. The performance of Environmental classification merely depends on the object detection performance. It can be able to produce 100% environmental classification accuracy.

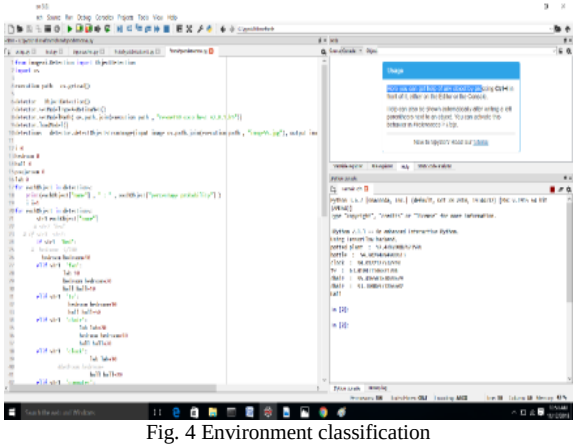

Fig. 4 Environment classification

Fig. 5 Detect objects in the hall and categorize

Fig. 6 Detect objects in the computer lab and categorize

Fig. 7 Detect objects in the Bed Room and categorize

VI. CONCLUSION

The environments of real-time image are identified. The objects in an image are detected and recognized by RetinaNet deep learning model. RetinaNet performs accurate object recognition. Obprob model have been developed and applied to the recognized object to identify the environment. In future work, the obprob model will be developed for all real time environments to identify all real time environments and location of each object in an image will be estimated, which will be very useful for robotic applications and visually impaired persons.

REFERENCES

[1] T.Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection" *In CVPR*, 2017.

[2] Janardhana Rao, and O. Venkata Krishna, "The Log Polar Transformation for Rotation Invariant Image Registration of Aerial Images", *IJCTA*, Vol. 4, pp. 833-840, 2013.

[3] R. Arunkumar, M. Balasubramanian, and S. Palanivel, "Indoor Object Recognition System using Combined DCT-DWT under Supervised Classifier", *IJCA*., Vol. 82 – No3, pages: 17- 21, November 2013

[4] J. Dai, Y. Li, K. He, and J. Sun., "R-FCN: Object detection via region-based fully convolutional networks", *In NIPS*., 2016.

[5] A. Wahi, P. Ravi, M. Saranya, "A neural network approach to rotated object recognition based on edge features: Recognition rate and CPU time improvement for rotated object recognition using DWT", *Computing Communication and Networking Technologies (ICCCNT)*, Vol. 29-31, pp. 1 – 6, July 2010.

[6] Junliang Li, Hon-Cheng Wong, Member, IEEE, Sio-Long Lo, and Yuchen Xin, "Multiple Object Detection by a Deformable Part-Based Model and an R-CNN".

[7] O. Barinova, V. Lempitsky, and P. Kholi, "On detection of multiple object instances using hough transforms", *IEEE Trans. Pattern Anal. Mach. Intell*., Vol. 34, No. 9, pp. 1773–1784, Sep. 2012

[8] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", *In CVPR*., 2014

[9] S. R. Bulo, G. Neuhold, and P. Kontschieder, "Loss maxpooling for semantic image segmentation", *In CVPR*, 2017

[10] S. Bell, C. L. Zitnick, K. Bala, and R. Girshick, "Insideoutside net: Detecting objects in context with skip pooling and recurrent neural networks", *In CVPR*., 2016

[11] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He and Piotr Dollar, "Focal Loss for Dense Object Detection", Feb 2018

[12] Jian Wu, Zhiming Cui, Victor S. Sheng, Pengpeng Zhao, Dongliang Su, and Shengrong Gong, "A Comparative Study of SIFT and its Variants , *Measurment Science Review*., Vol 13, No. 3, 2013

[13] Nasser H. Dardas, and Nicolas D. Georganas, "Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques", *IEEE Trans. on Instrumentation and Measurement*., Vol. 60, No. 11, November 2011

[14] Seyyid Ahmed Medjahed, "A Comparative Study of Feature Extraction Methods in Images Classification", *IJIGSP*, Vol. 7, No. 3, pp. 16-23, 2015.

[15] Olga Barinova, Victor Lempitsky , Pushmeet Kholi, "On detection of multiple object instances using hough transforms", *IEEE Trans. Pattern Anal. Mach. Intell*., Vol. 34, No. 9, pp. 1773–1784, Sep. 2012

[16] Yang Liu, Lei Huang, Xianglong Liu, and Bo Lang, "A Novel rotation adaptive object detection method based on pair Hough model", *Neurocomputing*., Vol. 194, pp. 246-259, 2016.

[17] Chunsheng Liu, Faliang Chang, and Chengyun Liu, "Cascaded split-level color Haar-like features for object detection", *Electronics Letters*., Vol. 51, No. 25, pp. 2106–2107, 10th December 2015.

[18] Sanjivani Shantaiya, Keshri Verma, and Kamal Mehta, "A Survey on Approaches of Object Detection", *International Journal of Computer Applications*, Vol. 65, No.18, pp. 0975 – 8887, March 2013.