International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 20 (2018) pp. 14505-14509

R-CNN Based Object Detection and Classification Methods for Complex Sceneries

Jaswinder Singh [#1], Dr. B.K. Sharma [*2]

[#] *Research scholar, Dr. A.P.J Abdul Kalam Technical University, Lucknow, India.*
[*] *Principal Scientific Officer, NITRA Technical Campus, Ghaziabad, India.*

Abstract

The computer vision system based application growths are very many needs to gather the information in the unusual environment. Surveillance systems are using in many applications. To get better surveillance system, object detection and recognition techniques need to discuss. In this field, object recognition has gained lots of attraction for real-time application. Various techniques have proposed for recognizing and detecting the object from the picture. In the proposed article, R-CNN techniques are presenting for object recognition in the unusual environment. This technique gets reduce the complexity to detect the object from complex scenarios.

Keyword: Image classification, object detection, convolution neural network, deep-learning.

INTRODUCTION

Computer Vision Techniques have proven as a promising technology for real-time applications such as the surveillance system, medical imaging. In real-time applications like video surveillance system, various occlusions and objects affect the do in surveillance system. The technique of object detection is a promising and challenging technique to apply in the field. Object detection technique is to find the objects in any real time from natural scene images such as the face, dog and building. These are presenting then these objects are detecting by using object detection scheme. Object detection techniques perform the desired operation with the help of feature extraction technique and feature learning scheme. Each object belongs to the different object class which known as object classification and carried out with the help of classification techniques. Classification is a process to find the category of the detected object in any image. Each object or each class of the object has its own features and characteristics responsible to make be difference in each class. This technique helps to recognize and classify the objects from any input image. Object recognition techniques are using in various real-time applications such as image retrieval, surveillance, security, vehicle tracking and parking systems. Various challenges are presenting in this field of object recognition. The challenging issue is considering about the robustness of the object recognition system. This technique is using for various real-time applications where guaranteed performance is demand, this performance is degrade due to image scale variation, the different viewpoint, deformations and imaging conditions of image acquisition. Another issue is cause due to the huge amount of data in the every object class which causes conflict resulting in performance degradation.

LITERATURE REVIEW

Various techniques have been proposed for object detection recently such as frame difference based method [1][2], optical flow computation [3], point detectors [4], temporal difference method [5], background subtraction techniques [6].

Frame differencing method [2] is applied to video sequence to detect the objects. This method computes the difference between two consecutive frames or images and detects an object. This method shows good performance even if computation environment is dynamic in nature. This method suffers from the issue of a complete outline of the moving object which results in performance degradation of object detection. Optical flow [3] based computation method which depends on the computation of optical flow and image clustering. This method is capable to provide information regarding movement of each object present in the image. Compare to background subtraction [6] based method, this method provides 85% better accuracy for object detection. The typical challenges of background subtraction in the context of video surveillance have such as, Illumination Changes, Dynamic Background, Occlusion, Clutter, Camouflage, Presence of Shadows, Motion of the Camera, Bootstrapping, Video Noise, and Speed of the Moving Objects, Intermittent Object Motion and Challenging Weather. Point detector is also used for object detection in computer vision system based applications. This is a technique which extracts the key points from the image considering texture and location information of the image. This approach provides better results where illumination and camera viewpoints are varying frequently. In this field, Harris detector [7], KLT detector method [8] and SIFT detector [9]. The temporal difference method for detecting the object and Working process of this method is similar to the frame differencing method.

The pixel-wise difference of two-or-three frames is computed resulting in the extraction of moving region from input image sequence. If moving object [10][11] consists of uniform texture pattern then this approach results in better adaptability considering dynamic scene change for object detection. The key challenge in this technique includes various aspects such as computational complexity and noise-sensitivity which makes it more complex to implement for real-time applications. Object recognition technique based on the

feature extraction method using linear SVM technique [12]. A new approach for object recognition using deep convolution neural network, named as ConvNet or CNN to obtain higher and faster rate of object detection [13]. Initially, image integral is obtained which helps to improve the computation time [14].

Object detection performance [16] depends upon the training process of the feature set and feature extraction technique. Variation in domain shift factors, spatial location performance and image qualities are the main components which are responsible for object detection performance in video or image datasets.

Object detection [17] by considering background subtraction as a benchmark technique for fixed camera view scenarios. However, various components are present which affect the performance of object recognition or detection. These components [15] are illumination variation, shadow casting, background movement and camera vibration etc. studies regarding background subtraction technique to improve the accuracy of object detection model.

The method [18] mitigates the performance issue of existing saliency detection method. Salient region extraction is performed by taking existing saliency detection technique along with distance weighting, binarization and morphological methods. Later, superpixel segmentation model is applied where the Bayesian model is used for saliency refinement resulting accurate saliency mapping. An iterative optimization method is designed to obtain better saliency results by exploiting the characteristics of the output saliency map each time. Through the iterative optimization process, the rough saliency map is updated step by step with better and better performance until an optimal saliency map is obtained.

Object detection [19] by extending the hierarchical model resulting in sequential frame updating considering dynamic scenes. The method utilizes dictionary learning based scheme developed using two-layer model. The model is developed using three-stage modelling for improving the sparse representation resulting in significant improvement in object detection and classification.

Feature extraction based model [20] also utilizes saliency technique as a base model. Hybrid feature modelling used to improve the detection performance. Since images are diverse nature which shows that images have different context significant nature. In order to deal with diversity, here a feature detector model is developed by introducing feature quality measurement. This approach shows significant performance as tested with benchmark datasets. The saliency-based technique is developed [21] which is named as Deep Saliency. Object detection performance also depends on the semantic properties of salient objects. By taking this into account, authors developed a saliency model using convolution network where inputs are given as raw images and output is obtained in the form of saliency map.

Confidence relationship model [22] used to detect the object from moving object detection technique. When dealing with the salient object that contains several regions with different appearances, salient object detection can be a difficult task. Often only parts of the salient object are highlighted and consistency between the salient regions is poor. This study tackles this problem by introducing object to assist the salient object detection. Rather than treating the object in the same manner as other low-level cues (e.g. uniqueness, location etc.) for the determination of regional saliency values, the authors emphasise that object should also play a significant role in tuning the consistency between salient regions.

The authors integrate object, uniqueness and centre bias to find potential salient regions and then enforce consistency between these regions using a full-connected Gaussian Markov random field with the weights determined by the object score [23]. In this field of learning based object detection, [24] a new approach and experimented on PASCAL VOC dataset. These methods are complex to implement and uses both low-level and high-level feature to produce higher-level feature vector.

Finally, this method is combined with region proposals and the proposed method is described as R-CNN. The key contribution of this work is denoted by its computation speed, hence it is known as the Fast R-CNN technique for object detection.

PROPOSED OBJECT RECOGNITION AND DETECTION TECHNIQUES

R-CNN is an object detection framework based on deep convolutional networks, which includes a Region Proposal Network (RPN) and an Object Detection Network. Both networks are trained for sharing convolutional layers for fast testing. In this work, we have combined shape feature extraction with the existing model of feature extraction. This combination of features improves detection and classification accuracy. This model is tested with publically available dataset VOC 2007 which contains various categories of objects. Our main aim is to detect each category in the image even occlusion are present in the image.

Multi task loss for R-CNN

$$\mathcal{L}(p,u,t^u,v) = \mathcal{L}_{cls}(p,u) + \lambda[\, u >= 1 \,]\, \mathcal{L}_{loc}(t^u,v)$$

u is the ground-truth class

v is ground-truth bounding-box regression target

Where $\mathcal{L}_{cls}(p, u) = -\log p_u$ is the log loss for true class u

For Feature Extraction, Deep Neural Network is implemented, this technique based on Krizhevsky's (Image Net Classification with Deep Convolutional Neural Networks). In Caffe, Deep Neural Networks contain eight learned layers – five convolutional and three fully connected. When the system is complex sometimes it gives random noise which has the undefined relation. It is called over-fitting. To reduce that it uses ReLU and Dropout layers. Following sequence shows the sequence of layers used and how image passes through each layer.

Input Image→ First Layer [Conv Layer → Maxpooling]→ Second Layer [Conv Layer→ Maxpooling] → Third Layer [Conv Layer] →Forth Layer[Conv Layer]→Fifth Layer[Conv Layer → Maxpooling]→ Six Layer [FC Layer] → Seventh Layer [FC Layer] → Eight Layer [FC Layer]

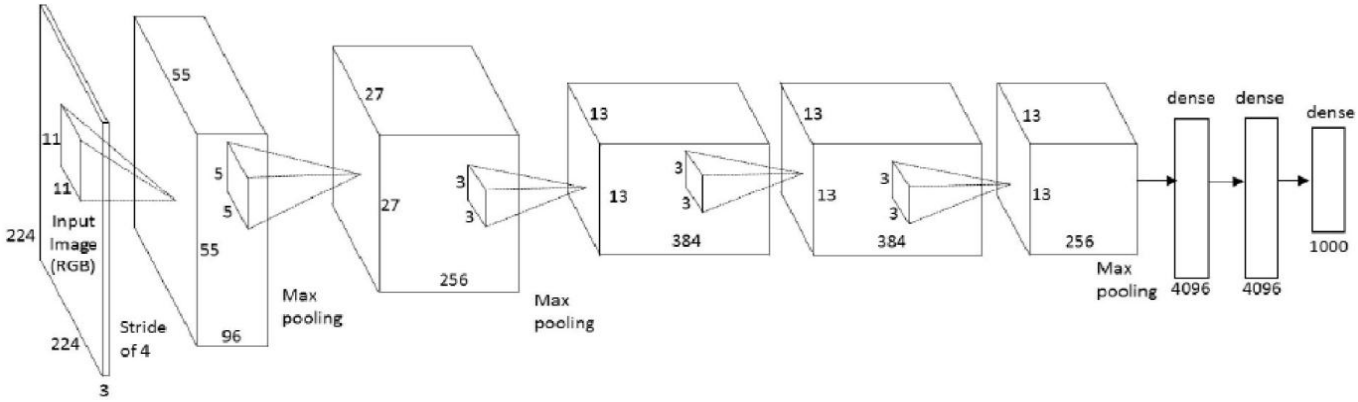

Figure 1. R-CNN Networks

Past research implementations have that multi-scale representation and it's providing the effective combination in much recent deep learning assignment. The sequence of the experimental work of Region based Convolutional Neural Network shown in figure-1.

RESULT AND DISCUSSIONS

For the object recognition, final result images are obtained from complex sceneries as shown in figure-2 like water, boating, ground, hill, bike riding, animal, bus, and traffic. R-CNN results when trained and tested using various region proposal methods. These results use the ConvNet for Selective Search (SS). From this result, proposed method (combined shape feature extraction with the existing model of feature extraction) gives good result in complex sceneries.

Figure 2: final result of proposed technique

From the detection results shown in Table 1 two methods are discussed and placed the comparison values for region-based convolutional methods. The experimental results are comparatively fulfilled for mostly in complex sceneries

Table 1. Result of the comparison between PS1 and ES method

Method	Box	mAP	Aero	bike	bird	bottles	bus	car	cat	Chair	cow
ES1	300	7	68.5	74.1	77.2	67.7	53.9	51.0	75.1	79.2	78.9
PS	300	78.8	82.0	77.7	68.9	65.7	88.1	88.4	88.9	63.6	86.3

Method	table	dog	Horse	Mbike	person	plant	sheep	sofa	train	tv
ES1	50.7	78	61.1	79.1	71.4	72.2	75.9	62.5	77.4	66.4
PS	70.8	85.9	87.6	80.1	82.3	53.6	80.4	75.8	86.6	78.9

CONCLUSION

Combined shape feature extraction with the existing model of feature extraction technique provides better performance when compared to existing deep convolution network technique. According to this technique, various techniques are using for improving the testing and training process, resulting, and improvement in efficiency of the network.

Proposed object detection system is to run the real-time frame rates. The learned RPN method improves region scheme quality. The overall object detections are finding necessary accuracy.

REFERENCES

[1] Y. Tang; X. Wang; E. Dellandrea; L. Chen, "Weakly Supervised Learning of Deformable Part-Based Models for Object Detection via Region Proposals," in IEEE Transactions on Multimedia , vol.PP, no.99, pp.1-1, 03 October 2016

[2] H. Yin, Y. Chai, S. X. Yang and X. Yang, "Fast-moving target tracking based on mean shift and frame-difference methods," in Journal of Systems Engineering and Electronics, vol. 22, no. 4, pp. 587-592, Aug. 2011.

[3] A. S. Rao, J. Gubbi, S. Marusic and M. Palaniswami, "Crowd Event Detection on Optical Flow Manifolds," in IEEE Transactions on Cybernetics, vol. 46, no. 7, pp. 1524-1537, July 2016.

[4] M. Zandi, A. Mahmoudi-Aznaveh and A. Talebpour, "Iterative Copy-Move Forgery Detection Based on a New Interest Point Detector," in IEEE Transactions on Information Forensics and Security, vol. 11, no. 11, pp. 2499-2512, Nov. 2016.

[5] Z. Li et al., "Rotation Feature Extraction for Moving Targets Based on Temporal Differencing and Image Edge Detection," in IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 10, pp. 1512-1516, Oct. 2016.

[6] W. Kim and Y. Kim, "Background Subtraction Using Illumination-Invariant Structural Complexity," in IEEE Signal Processing Letters, vol. 23, no. 5, pp. 634-638, May 2016.

[7] J. b. Ryu and H. h. Park, "Log-log scaled Harris corner detector," in Electronics Letters, vol. 46, no. 24, pp. 1602-1604, November 25 2010.

[8] C. Zhang, J. Xu, A. Beaugendre and S. Goto, "A KLT-based approach for occlusion handling in human tracking," Picture Coding Symposium (PCS), 2012, Krakow, 2012, pp. 337-340.

[9] S. Luo, W. Mou, K. Althoefer and H. Liu, "Novel Tactile-SIFT Descriptor for Object Shape Recognition," in IEEE Sensors Journal, vol. 15, no. 9, pp. 5001-5009, Sept. 2015.

[10] J. Lei, G. Li, J. Zhang, Q. Guo and D. Tu, "Continuous action segmentation and recognition using hybrid convolutional neural network-hidden Markov model model," in IET Computer Vision, vol. 10, no. 6, pp. 537-544, 9 2016.

[11] M. Xu, L. Zhang and B. Du, "An Image-Based Endmember Bundle Extraction Algorithm Using Both Spatial and Spectral Information," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 8, no. 6, pp. 2607-2617, June 2015.

[12] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR. (2005)

[13] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS. (2012)

[14] Viola, P., Jones, M.: Robust real-time object detection. IJCV (2001)

[15] Gregory D. Hager and Peter N. Belhumeur; “Efficient Region Tracking With Parametric Models of Geometry and Illumination”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20 (10), pp. 1025–1039, 1998

[16] V. Kalogeiton, V. Ferrari and C. Schmid, "Analysing Domain Shift Factors between Videos and Images for Object Detection," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 11, pp. 2327-2334, Nov. 1 2016.

[17] S. K. Choudhury; P. K. Sa; S. Bakshi; B. Majhi, "An Evaluation of Background Subtraction for Object Detection Vis-a-Vis Mitigating Challenging Scenarios," in IEEE Access , vol.PP, no.99, pp.1-1, 13 September 2016

[18] J. Lei et al., "A Universal Framework for Salient Object Detection," in IEEE Transactions on Multimedia, vol. 18, no. 9, pp. 1783-1795, Sept. 2016.

[19] S. Zhu, J. Du, N. Ren and M. Liang, "Hierarchical-Based Object Detection with Improved Locality Sparse Coding," in Chinese Journal of Electronics, vol. 25, no. 2, pp. 290-295, 3 2016.

[20] S. S. Naqvi, W. N. Browne and C. Hollitt, "Feature Quality-Based Dynamic Feature Selection for Improving Salient Object Detection," in IEEE Transactions on Image Processing, vol. 25, no. 9, pp. 4298-4313, Sept. 2016.

[21] X. Li et al., "DeepSaliency: Multi-Task Deep Neural Network Model for Salient Object Detection," in IEEE Transactions on Image Processing, vol. 25, no. 8, pp. 3919-3930, Aug. 2016.

[22] Z. Fan and H. Wang, "Robust moving object detection based on spatio-temporal confidence relationship," in Electronics Letters, vol. 52, no. 10, pp. 825-827, 5 12 2016.

[23] X. Sun, A. Su, S. Chen, Q. Yu and X. Liu, "Objectness to assist salient object detection," in IET Image Processing, vol. 10, no. 5, pp. 391-397, 5 2016.

[24] Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: CVPR. (2014).