DETECTION OF ANIMAL HUNTERS IN FOREST USING REGIONAL CONVOLUTIONAL NEURAL NETWORK ALGORITHM

Abstract

Detection of objects moving in a video is regarded as a vital problem in image processing applications i.e. the detection during the movement of camera is considered significant in video processing. In this paper, the movable objects for instance humans moving in forest environment is detected using a deep learning algorithm called regional convolutional neural network (RCNN). In this classification model, the classifier segregate the moving human in a dynamic environment. The region of motion is initially detected via background compensation method. The RCNN is utilised then to locate accurate the person moving in that region of motion. The RCNN model is a lightweight classifier that is designed specifically for the smaller objects. The results of both motion and object detection is fused together to obtain the moving objects.If a moving information is missed in the current frame, it is then recalled in the subsequent frame as per the spatial or temporal data. The simulation is conducted on 30 videos, where 24 is used for training the classifier and remaining 6 is taken for testing purposes. The result of simulation shows that the proposed RCNN model obtains improved accuracy, specificity, sensitivity and f-measure.

Keywords

Animal Hunters, Forest, Convolutional Neural Network, Object Detection

1. Introduction

Moving detection and recognition of objects is a major subject of research for computer vision, in which intelligent video monitoring, robot vision browsing, virtual reality and medical diagnostics are key. The rise of the Internet of Things (IoT), in particular, has heightened interest in video sequence detection of moving objects in recent years [1] [2]. IoTs have advanced imaging abilities where the camera can function with various movement degrees and autonomy, but the motion blur and dynamic background cause certain challenges. Moreover, due to the moving object appearance due to light, occlusion and shadows in outdoor situations, it is very vital to build a robust motion detection and recognition approach to improve the exactness of moving object detection.

When using moving cameras, the background of an image appears to rotate, translate, and enlarge, as opposed to fixed-camera moving object identification techniques [3]-[5]. The backdrop or moving objects based on optical flow are so difficult to model. Therefore, we are proposing a movement detection approach for objects to tackle this problem by combining background adjustment with deep education [6]-[11]. The background compensation determines the movement of data in a frame. Firstly, according to the coordinate relation of feature points in consecutive frames, the motion parameter is measured and then the binary mask of moving areas is generated using the interframe difference approach. We restrict the feature points by separate and geographic distribution the outer points depending on the outcome of the interframe. Object matching makes the match between the feature points more accurate. The outside points are feature marks for moving objects, which have a severe impact on the recording of the image.

The deep learning method namely convolutional neural network (CNN) seeks to find a faster and higher sensing performance for the moving goals of objects, a model for object detection, more precisely. By fusing the moving objects, the results of these two procedures can be correctly identified. However, the backdrop compensation approach is frequently not used to detect movements, due to the effects of illumination variation, the appearance of moving items, sudden motion, occlusion, and shadows, etc., resulting in motion loss. We employ time and space information in consecutive frames to reminisce about the loss of objectives to address this problem [12].

Because of the significant development in technology, the IoT industry is becoming increasingly popular and can do a large number of tasks that human beings consider harmful and inaccessible. Missions such as forest monitoring might be conducted efficiently with IoT, as the installed IoT camera ensures a large amount of coverage. Deforestation has become a popular human activity in recent decades, transforming wooded areas into non-forest regions for use, such as metropolitan areas, logging regions, and agricultural land.

But for a vast forest area, a large-scale inspection is a labour-intensive, time-consuming and inefficient. This paper presents a human object algorithm that uses IoT vision as an effective and efficient method for forest monitoring of the presence of illegal human entry and preventing illegal logging activity, in order to reduce labor requirements and increase the surveillance process. Acknowledging that the deep learning concept has recently gained considerable interest in various fields of applications, this paper models a deep learning-based algorithm which can find characteristics by optimising certain loss functionalities, in contrast to traditional machine learning approaches which extract characteristics from image data.

This research explores in-depth learning technology which incorporates IoT-captured visual information that allows forests to be detected by Regional Convolutional Neural Network (RCNN) [19]-[22]. The focus of this research is the creation of an algorithm for the detection of human objects with deep knowledge. Practically, this research is designed to construct an IoT vision for forest monitoring human object detection algorithm. This project represents a significant investment in long-term forest management and redevelopment for the benefit of IoT monitoring.

2. Related works

Adel Hafiane et al. [13] have taken off a corner feature block to accommodate the dynamic background. In order of extracting the relevant features, Setyawan et al. [14] employed the Harris corner detector and this approach proved to be more rugged and less sensitive than the other approaches. The classification approach requires the selection of an appropriate tracker for the long-term path of the character points before the use of the clustering approach to identify the path that is in the background from other trajectories.

Yin et al. [15] used optical flow to offset camera movement before utilising a watershed transformation to cluster and then conducted primary analysis for the reduction of unusual pathways.

Brox et al. [16] similarly used the optical flow approach to create point paths in video sequences but to categorise the foreground and background pathways by applying spectral clustering with spatial regularity. The purpose of this method is considered different, although tracking objects can also be considered to be moving objects.

The CNN is a new approach to object representation. Danelljan [17] has launched the Discriminative Correlation Filter factor convolution operator to limit model parameters. Overfitting problems and Computational complexity is fixed, thus improving tracking speed and performance.

Deep Sort [18] is an algorithm for multi-target tracking. First, the method identifies the aims of each frame per object detection and forecasts the motion path via a Kalman filter and utilises the weighted method to match boxes, which ensures that pedestrian multi-objective tracking has a favourable effect.

3. Proposed Method

In this section, we offer a moving object detection approach based on a combination of object detection and background compensation, consisting of three modules that detect, detect objects and match objects. Figure 1 shows the proposed strategy and the method proposed is described below.

Step 1: **Pre-processing**: With the background compensation method, the motion detecting module acquires a raw motion.

Step 2: **Detection**: RCNN is used for recognising the location of objects, it is introduced in the object detection module - RCNN.

Step 3: **Validation**: In the object matching, the detection of moving object is carried out via motion and detection results while the missing detections are recalled using temporal and spatial information types from subsequent frames.

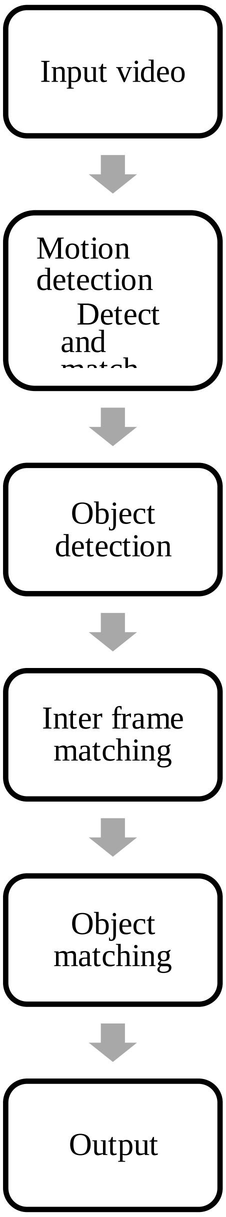

Figure 1: Proposed Architecture of motion detection

a. Pre-processing the video using Motion Detection

The background compensation method is based on the co-ordinate relationships of the corresponding feature points. The performance of multiple detection point techniques was tested and the SIFT technique was selected. Experiments indicated that the more feature points with distribution can be extracted using the SIFT technique, and the speed also met the requirements.

For the individual functional points, a 128-dimensional vector is displayed with the SIFT algorithm, and two neighbouring frames are used for the matching of SIFT functional points with a minimal Euclidean distance. The following considerations must be taken into account in the feature selection: If the feature intensity is higher, they will not be lost. There may be excessive estimations when features that are matched with an excessive number of feature points. Excessive feature point concentration can result in substantial mistakes in motion parameters. Therefore, an average and broad distribution of the function points and not external points should be provided.

The spatial distribution of the feature points is constrained. To make the distribution of feature points as uniform as feasible, but with adequate feature points, the image is into $S \times S$, where $S=10$ has been picked. The strongest characteristic points are preserved and other characteristic points are filtered away. Finally, the results produced after object matching across frames show that feature points are filtered out of the grids where the object bounding boxes are placed.

The selected feature points contain a small number of outpoints, and the random consensus sample approach is used to determine the transformation perspective parameters in order to estimate the high accuracy of the outside data. The transformation of perspective allow translation, scaling, rotation, and shearing. In comparison with the affinity, the transformation is flexible and can precisely define the form of motion for the image background, as in the following equation:

$$[x', y', w'] = [\mu, v, \omega] \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = [\mu, v, \omega] \times T$$

where

μ, v– original image coordinates,

x_0, y_0– transformed image coordinates, and

T – Transformation matrix of order 3×3.

The three-frame differentiation method is used to obtain a binarization mask in the moving region after background adjustment. The formula is the following for the three-frame approach.

$$d_k(x, y) = \begin{cases} 1 & I_{k-1}(x, y) > T \,\&\&\, I_{k+1}(x, y) > T \\ 0 & otherwise \end{cases}$$

where

$I_k(x,y)$ - pixel in k^{th}frame at a point (x,y),

$d_k(x,y)$ - gray value at a point (x,y),

A – moving pixel, and

T- threshold difference.

The moving target in frames will not change when the movement speed of an object is very sluggish. This is why we have a three-frame interval in place. The inter-frame differential approach produces a binary mask for moving objects that contains noises because of an unknown threshold, background errors and changes in irregular items Morphological procedures and the features of interconnected regions can be analysed to limit the effects of noise, but only interframe differences can be detected reliably and completely.

b. Object Detection using RCNN

We use the RCNN object detection for the detection of potentially moving items in the image. First, considering the possibility of using an IoT and a monitoring camera, the size of the model must be reduced further and the speed of the model improved. Secondly, a huge number of small moving goals are involved in the moving detection. We have counted the number of pixels in the photos for the moving targets and have determined the targets pixels are the majority, which means that for smaller targets, the RCNN network needs optimization.

The RCNNs have semantic information in the deep layers, whereas the shallow levels carry information about the position. Smaller object characteristics can be lost due to the deepening of Network Layers; therefore, CNNs use upsampling to fuse the multiple layers of feature maps to avoid this problem. In RCNNs, we employ the same strategy for enhanced detection of small goals. The 4x down sample layer has been chosen as the final network output layer; the 8x down sample layer has been linked 4x layer, the 16x down sample layer to the 8x4 down sampler layer and the 32x down sampling layer has been linked to the 16x down sample layer and 8x Down sampling layers. In order to fuse shallow and deep information, a branch fusion structure is created, thus enhancing the quantity of location and statistical information for small subjects. To prevent the gradient from disappearing during training and to increase reuse of features, it changes the overall output layer into a convolutional layer and two residual units.

c. Object Matching

A moving object with motion attributes is defined in this section, where the motion detection and object detection results must overlap for the image. We use intersection over union (IoU) to unify the detection results. The bounding box for motion detection is assumed to be *M* and the bounding box is *D*, so the formula for calculating IoU is as follows.

$$IoU = \frac{A(M) \cap A(D)}{A(M) \cup A(D)}$$

The motion detection result is usually inaccurate and only the approximate area containing the moving object is provided. The object detection detects the target but it lacks motion information. Thus, we use IoU to add motion information for object detection, where the object detection box is used as the final output when IoU >threshold. The moving object is hence determined as below:

Usually, the results of motion detection are incorrect and the moving item is only approximately contained. The object detection method can detect the target completely, but it does not have any movement information. Therefore, we used IoU to add movement information to the object detection when the IoU value is greater than the threshold when the object detection box is the final output. Thus, the moving item is as follows:

$$MO = \begin{cases} D_i & \max\left(IoU\left(D_i, M_j\right)\right) > 0.2 \\ 0 & otherwise \end{cases}$$

where

D_i– results of object detection, and

M_j– results of motion detection.

The object detection using RCNN can detect objects accurately, because of the lack of motion information, it can never be treated as a moving object. Therefore, the motion detection module and the object detection module are combined, but the low movement detection reminder module limits the joint algorithm's performance. Thus, we introduce object matching across frames to remember the missed detections. In the continuous image sequence, the moving objects demonstrate temporal and spatial continuity, i.e. in a short period, the moving objects do not suddenly vanish and their positions do not change at once. The bounding boxes of the same target are so highly overlapping across the adjacent frames, and it is therefore easy to establish, by calculating IoT. To detect missing detections, the RCNN bounding box in the current frame and the bounding box in the previous frame are calculated. The RCNN bounding box is employed for detection purposes if the IoT value is greater than 0.5. Furthermore, when a moving item stops behind the screen, we count the number of frames in the detection related to motion that have lost detection, and when a specific threshold is met, we determine that the target has stopped moving.

4. Results and Discussions

This section validates the moving targets from the input video frame(in images) are used for training the classifier (Dataset: Conservation Drones (http://lila.science/datasets/conservationdrones)), where the motion and object detection is validated. For the purpose of testing, the proposed method uses drone to capture the hunters or poachers, where the video data is utilised for testing the classifier.There are 48 real aerial TIR videos and 124 synthetic aerial TIR videos (generated with AirSim), for a total of 62k and 100k images, respectively.

The performance of each model is validated in terms of four different metrics that includes accuracy, specificity, sensitivity and f-measure. It is further tested for mean average percentage error to check the errors present while detecting the motion in a video frame. The proposed method is compared with existing motion object detection methods that includes CNN-LSTM, CNN, RNN, 3 layered ANN and BPNN. The entire experiments are conducted in Core i9 generation computing with a graphic acceleration unit. The simulation is conducted in python, where the RCNN for object detection is built in Tensorflow.

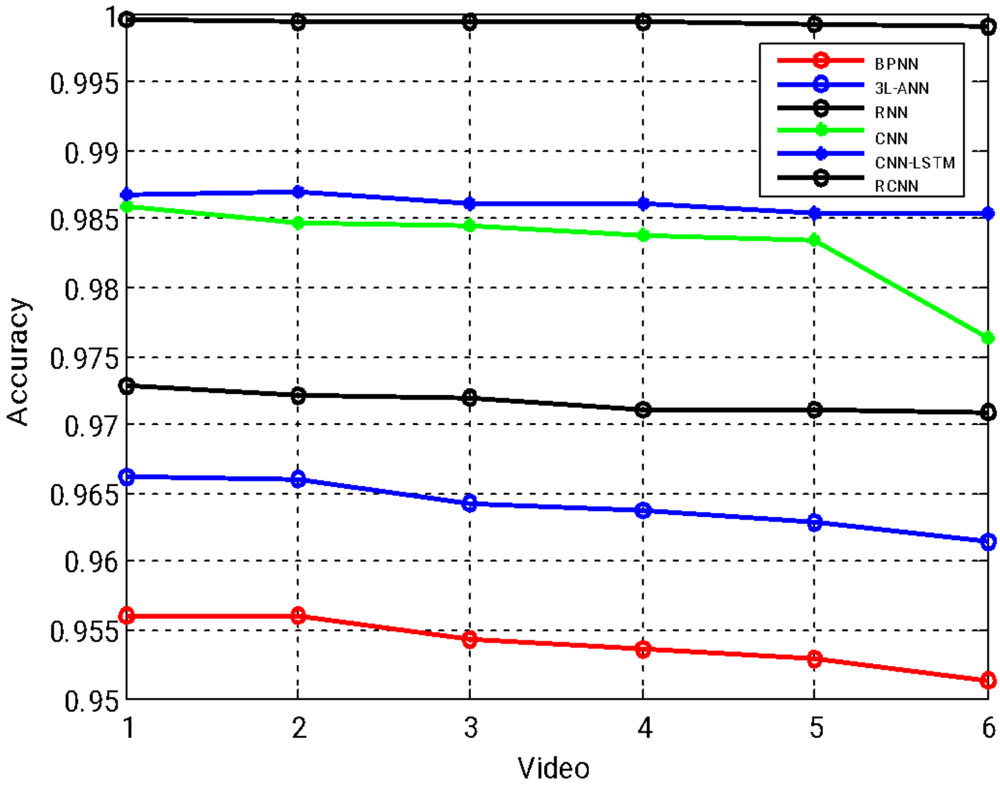

Figure 2: Accuracy

Figure 2 shows the illustration of accuracy of RCNN with existing motion object detection methods that includes CNN-LSTM, CNN, RNN, 3 layered ANN and BPNN. The results of simulation shows that for each video, the accuracy varies with respect the video frames and richness of colour in the videos. The comparative results show that the RCNN obtains improved accuracy than CNN-LSTM, CNN, RNN, 3 layered ANN and BPNN for all the videos.

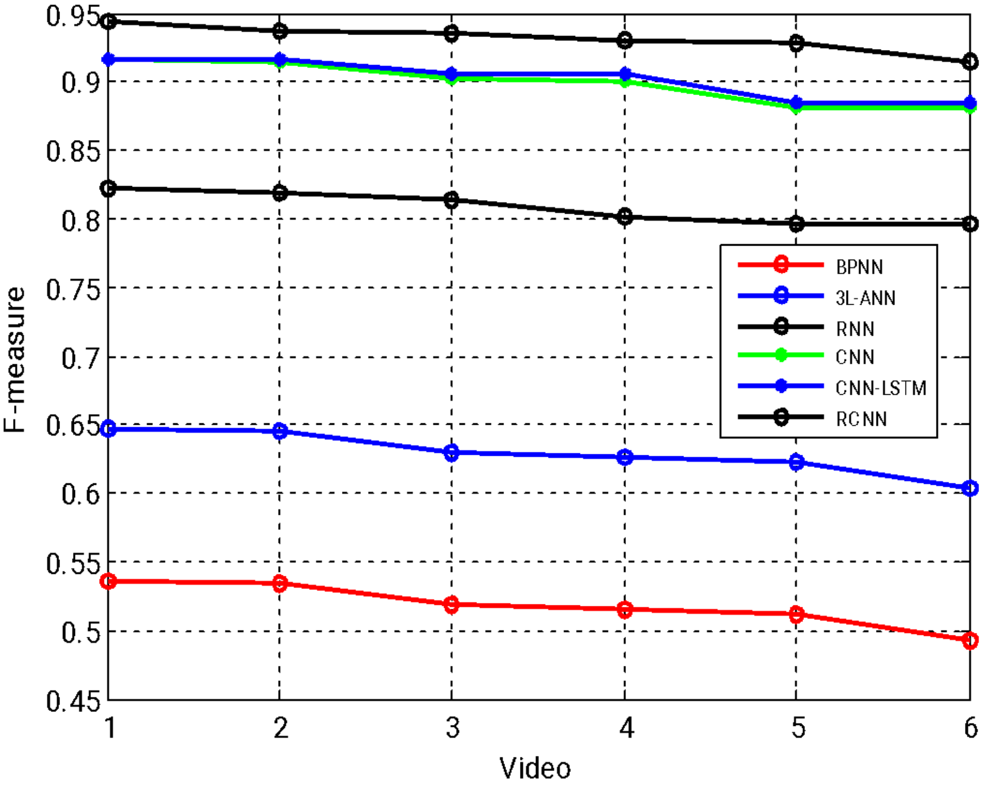

Figure 3: F-measure

Figure 3 shows the illustration of F-measure of RCNN with existing motion object detection methods that includes CNN-LSTM, CNN, RNN, 3 layered ANN and BPNN. The results of simulation shows that for each video, the F-measure varies with respect the video frames and richness of colour in the videos. The comparative results show that the RCNN obtains improved F-measure than CNN-LSTM, CNN, RNN, 3 layered ANN and BPNN for all the videos.

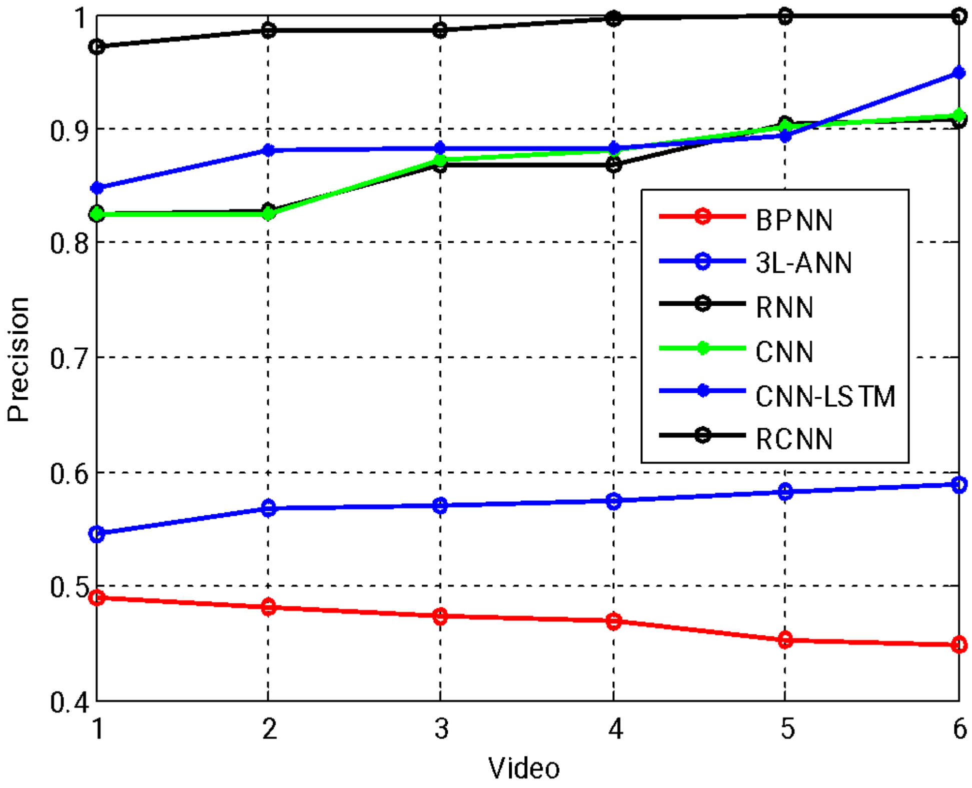

Figure 4: Precision

Figure 4 shows the illustration of precision of RCNN with existing motion object detection methods that includes CNN-LSTM, CNN, RNN, 3 layered ANN and BPNN. The results of simulation shows that for each video, the precision varies with respect the video frames and richness of colour in the videos. The comparative results show that the RCNN obtains improved precision than CNN-LSTM, CNN, RNN, 3 layered ANN and BPNN for all the videos.

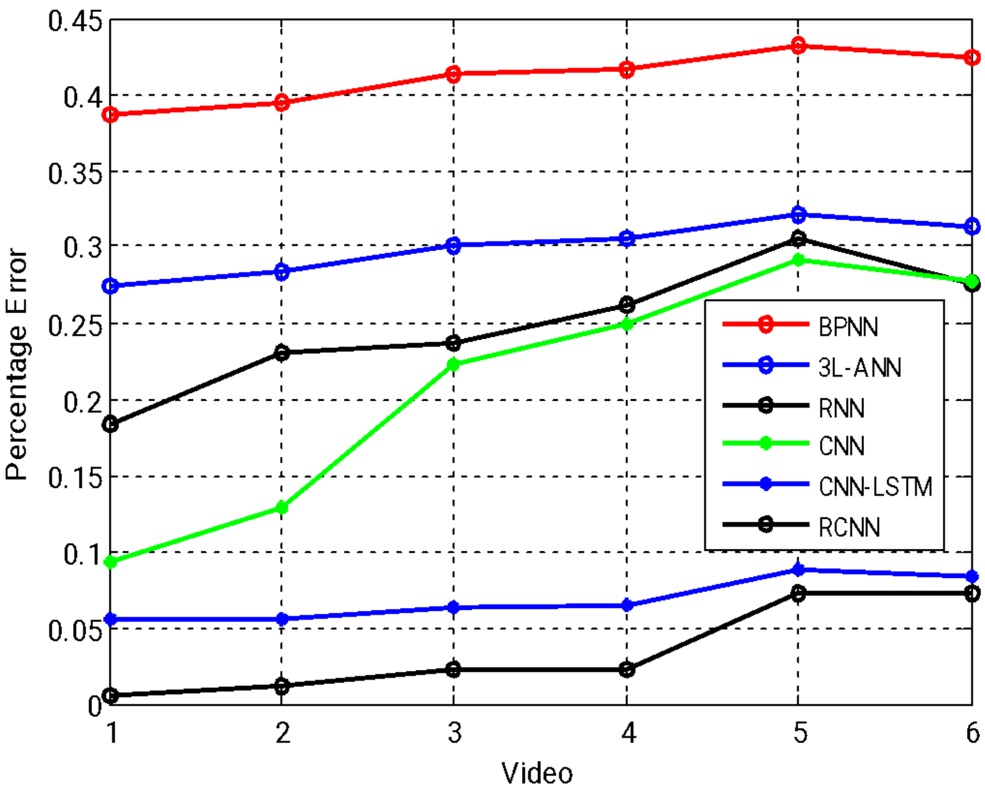

Figure 5: Percentage error

Figure 5 shows the illustration of mean average percentage error of RCNN with existing motion object detection methods that includes CNN-LSTM, CNN, RNN, 3 layered ANN and BPNN. The results of simulation shows that for each video, the mean average percentage error varies with respect the video frames and richness of colour in the videos. The comparative results show that the RCNN obtains reduced mean average percentage error than CNN-LSTM, CNN, RNN, 3 layered ANN and BPNN for all the videos.

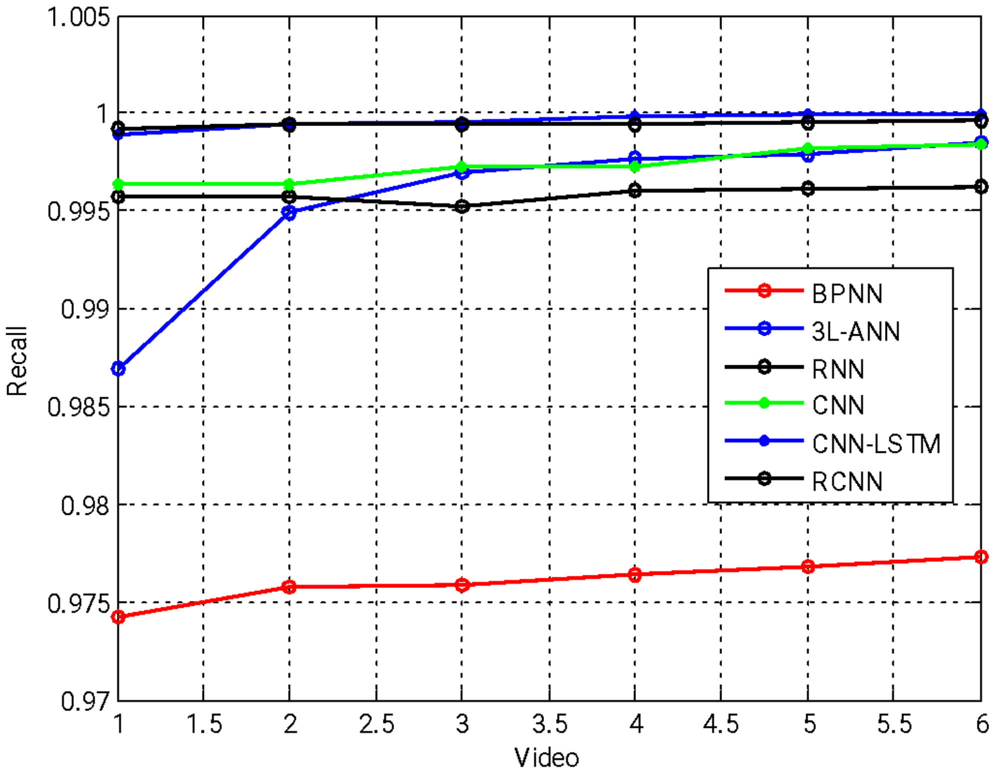

Figure 6: Recall

Figure 6 shows the illustration of recall of RCNN with existing motion object detection methods that includes CNN-LSTM, CNN, RNN, 3 layered ANN and BPNN. The results of simulation shows that for each video, the recall varies with respect the video frames and richness of colour in the videos. The comparative results show that the RCNN obtains improved recall than CNN-LSTM, CNN, RNN, 3 layered ANN and BPNN for all the videos.

5. Conclusions

In this paper, the RCNN is used to classifying the movable humans in forest environment. In RCNN model, the moving humans are effectively separated even in the presence of a rich dynamic background. The background compensation with RCNN helps to effectively find the motion and object detection in dynamic background, where the motions are fused to obtain the resultant objective. The testing on 6 different video input dataset shows that the RCNN model obtains improved accuracy, specificity, sensitivity and f-measure.The improved detection results are due to the recognition of temporal and spatial information from an input data via RCNN. It is further seen that the existing method fails in fusing the moving and objection detection at a time, since none of the methods used background compensation integrated with object detection.

References

[1] Lin, C. Y., Muchtar, K., Lin, W. Y., & Jian, Z. Y. (2019). Moving object detection through image bit-planes representation without thresholding. *IEEE Transactions on Intelligent Transportation Systems*, *21*(4), 1404-1414.

[2] Cho, J., Jung, Y., Kim, D., Lee, S., & Jung, Y. (2018). Design of moving object detector based on modified GMM algorithm for UAV collision avoidance. *Journal of semiconductor technology and science*, *18*(4), 491-499.

[3] Jarraya, S. K., Hammami, M., & Ben-Abdallah, H. (2010, December). Accurate background modeling for moving object detection in a dynamic scene. In *2010 International Conference on Digital Image Computing: Techniques and Applications* (pp. 52-57). IEEE.

[4] Fan, X., Cheng, Y., & Fu, Q. (2015, April). Moving target detection algorithm based on Susan edge detection and frame difference. In *2015 2nd International Conference on Information Science and Control Engineering* (pp. 323-326). IEEE.

[5] Shin, J., Kim, S., Kang, S., Lee, S. W., Paik, J., Abidi, B., & Abidi, M. (2005). Optical flow-based real-time object tracking using non-prior training active feature model. *Real-Time Imaging*, *11*(3), 204-218.

[6] Zhu, J., Wang, Z., Wang, S., & Chen, S. (2020). Moving Object Detection Based on Background Compensation and Deep Learning. *Symmetry*, *12*(12), 1965.

[7] Wang, P., Yang, J., Mao, Z., Zhang, C., & Zhang, G. (2019). Object Detection Based on Motion Vector Compensation in Dynamic Background. *Ordnance Ind. Autom*, *38*, 6-10.

[8] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).

[9] Suhr, J. K., Jung, H. G., Li, G., Noh, S. I., & Kim, J. (2010). Background compensation for pan-tilt-zoom cameras using 1-D feature matching and outlier rejection. *IEEE transactions on circuits and systems for video technology*, *21*(3), 371-377.

[10] Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, *39*(6), 1137-1149.

[11] Suhr, J. K., Jung, H. G., Li, G., Noh, S. I., & Kim, J. (2010). Background compensation for pan-tilt-zoom cameras using 1-D feature matching and outlier rejection. *IEEE transactions on circuits and systems for video technology*, *21*(3), 371-377.

[12] Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

[13] Hafiane, A., Palaniappan, K., & Seetharaman, G. (2008, July). UAV-video registration using block-based features. In *IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium* (Vol. 2, pp. II-1104). IEEE.

[14] Setyawan, F. A., Tan, J. K., Kim, H., & Ishikawa, S. (2014, September). Detection of moving objects in a video captured by a moving camera using error reduction. In *SICE Annual Conference* (Vol. 2014, pp. 347-352).

[15] Yin, X., Wang, B., Li, W., Liu, Y., & Zhang, M. (2015). Background Subtraction for Moving Cameras based on trajectory-controlled segmentation and Label Inference. *Ksii Transactions on Internet & Information Systems*, *9*(10).

[16] Brox, T., & Malik, J. (2010, September). Object segmentation by long term analysis of point trajectories. In *European conference on computer vision* (pp. 282-295). Springer, Berlin, Heidelberg.

[17] Danelljan, M., Bhat, G., Shahbaz Khan, F., & Felsberg, M. (2017). Eco: Efficient convolution operators for tracking. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6638-6646).

[18] Wojke, N., Bewley, A., & Paulus, D. (2017, September). Simple online and realtime tracking with a deep association metric. In *2017 IEEE international conference on image processing (ICIP)* (pp. 3645-3649). IEEE.

[19] Sarda, E., Deshmukh, P., Bhole, S., & Jadhav, S. (2021). Estimating Food Nutrients Using Region-Based Convolutional Neural Network. In *Proceedings of International Conference on Computational Intelligence and Data Engineering* (pp. 435-444). Springer, Singapore.

[20] Ding, P., Zhang, Y., Deng, W. J., Jia, P., & Kuijper, A. (2018). A light and faster regional convolutional neural network for object detection in optical remote sensing images. *ISPRS journal of photogrammetry and remote sensing*, *141*, 208-218.

[21] Al-masni, M. A., Al-antari, M. A., Park, J. M., Gi, G., Kim, T. Y., Rivera, P., ... & Kim, T. S. (2017, July). Detection and classification of the breast abnormalities in digital mammograms via regional convolutional neural network. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 1230-1233). IEEE.

[22] Rao, T., Li, X., Zhang, H., & Xu, M. (2019). Multi-level region-based convolutional neural network for image emotion classification. *Neurocomputing*, *333*, 429-439.