Journal of Engineering Sciences
Vol 11, Issue 1, JAN/2020
ISSN NO:0377-9254

Tracking of Foreground objects using FCNN classifier

[1]P Jayanthi, [2]Shaik Taj Mahaboob, [3]Varna Kumar Reddy
[1]M.Tech Student, [2]Assistant Professor, [3]Assistant Professor (Adhoc)
[1]palagijayanthi853@gmail.com
[1]Electronics and Communication Engineering,
[1]JNTUA College of Engineering Pulivendula, Pulivendula, India.

Abstract: Video tracking means locating a moving object or effectively focusing on a specific object over time by using a camera. Due to the amount of data carries in that video, it is a time-absorbing task. So, to reduce the difficulty, we use object recognition techniques for tracking. Target appearance with a single spatial grid layout without the involvement of various grids can affect their performance and the target is affected by rapid motion, heavy occlusions, large variations in deformation, background clutter. This paper introduces a process that the object can be properly tracked in a video by using similarity learning method followed by FCNN classifier, and features are extracted using Harris corner features to track objects in video. FCNN is used to improve speed and eliminate noisy feature maps to improve tracking accuracy. This approach can be used in various video applications like interfaces with human computers, monitoring, traffic control, analysis of motion.

Key words: FCNN classifier, Harris corner detectors, nonlocal similarity learning, visual tracking.

I.INTRODUCTION

Visual tracking is the ability to control eye movements through the use of the oculomotor system (visual muscles and eyes working together). There are two types of tracking: maintaining your attention on a moving object and shifting your focus between two objects as well as being One of the key components in a wide variety of computer vision and video processing applications, including various practical applications for example surveillance, robots, traffic control, navigation, augmented reality, interaction with human computers and medical imaging.

In view of the initialized state (e.g. position and size) of the target object in the video frame, the aim of the tracking is to estimate the target status in the subsequent frames. Although tracking of object has been studied for several decades, a lot has been achieved in recent years [3]. It's still a difficult issue. Many factors affect tracking algorithm efficiency, such as variations in illumination conditions, occlusion, and background clutter, and there is no single tracking solution capable of managing all situations effectively.

In [4] U. Shalit, and S. Bengio design an OASIS (Online Algorithm for Scalable Image Similarity) learning which learns to measure sparse representations that have a bilinear similarity. OASIS is an interactive dual solution that uses a passive-aggressive learning algorithm with a high margin requirement and a cost-effective failure of the hinge. [5] Zhang et al. defines the model multi-task learning (MTL), tasks sharing function dependencies or learning specifications are mutually managed to draw the implicit relationships between them. In this area, several studies showed that the MTL could be extended with modern problems like object labeling and classification of image.

In [6] the merits of both subspace and incomplete visual tracking representations, D. Wang and H. Lu introduces $l_1$regularization when restoring the PCA develops the novel algorithm for an object with a few templates and the explicitly description for the data and noise. Objects are represented as monitoring

online updates were learned from the sparse prototypes.

In [7] Luca Bertinetto, Jack Valmadre is equipping a simple tracking procedure along with a new, fully-convoluted Siamese network trained end-to-end on the ILSVRC15 video object detection dataset. The tracker runs outside in real time at frame rates and despite its extreme simplicity. However, the development to robust and efficient tracer is still difficult task due to partial occlusion problems, variability in lighting, backdrop clutter, motion blur, and point of view change, and so on. Most of the existing methods cannot handle these issues well and are failed to distinguish the forepart from the backdrop with well-defined co-existing objects.

This paper proposes a method that can track targeted object or person accurately in the video. Input video is divided into number of frames. We divide object and also backdrop representatives within the set of nonoverlapping frames. Each frame is represented with Harris corner detector and hsv features. Harris features is used to detect the corners of foreground objects and hsv is used to detect an object with certain color to reduce the influence of light intensity from outside. To boost tracking accuracy, the FCNN classifier is used to improve speed and eliminate noisy feature maps.

II. RELATED WORK

In recent years, due to their success in detecting objects, we own more interest in these discriminatory classifiers in tracking methods. [8] R. Caseiro and P. Martins describes the circulating tracking structure through kernel detection. They are trained online by collecting samples while they are being tracked. There is a possible huge amount of samples with statistical significance that conflicts explicity with actual time criteria. On the other side, restriction of samples can compromise efficiency. By adding more number of samples, the difficulty is getting a circulating framework. Using the well-established Circulant Matrix Theory, we have a reference to the Fourier Analysis, which extends the probability of fast analyzing and identification of Fast Fourier Transform.

Image monitoring does not protect time-changing stationary video sources. While most current algorithms are capable of tracking good targets in controlled conditions, normally fails in the existence of major variations within the shape of the target and its adjacent radiance conditions. [9] J. Jim and M.H. Yang implements a tracking system that slowly learns how to interpret a less proportional subspace, that adapts online efficiently to change the aspects of objects. Progressive process for main evaluation of features are included with two essential factors: a procedure for updating the test mean correctly and a variable for ensuring, the low modeling energy which is used to match previous considerations.

B.Babenko and M. Hsuan Yang[10] suggests Multiple Instance Learning (MIL) method for addressing issue of video object tracking. Tracking by detection method trains an interactive discriminating classifier to isolate the item from the background and use the current tracker status to delete both positive and neutral samples from the present frame. Consequently, minor defects within the tracker leads to improperly marked training that reduces the classifier and leads to drift. MIL prevents these issues and leads to a more stable tracker with less parameter changes, which was more strengthened as [11] - [13]. Hare et al. [14] propose visual tracking as functional information problem and also, we use Haar-like components as the target representations.

In [15] Mei and Ling use target arrangements and inessential frequency templates to create vocabulary, and retrieve a minimum re-establish error patch by resolving a l_1 depreciation problem.

In [16] Qingshan Liu and Jiaqing Fan describes about a method logistic regression classifier and

histogram of oriented gradients. LRC is used for classification and hog is used for feature extraction purpose, these methods have drawbacks like improper tracking and less recognition rate. So, it is necessary to improve the recognition rate and tracking accuracy.

III. METHODOLOGY

In this section our proposed method is explained. We know that video is separated into sum total number of frames and foreground target is separated from background in each frame. The proposed method is given in a flowchart as shown in fig(1), In the first step input video is divided into frames using a nonlocal similarity learning function that considers the interactions of the grid characteristics, not only from the same spatial locations, but also from different ones, in order to essentially control the major variations in the appearances. Then we represent each frame with Harris corner detector and hsv for feature extraction purpose, classification of frames to track foreground objects is done by using FCNN classifier in a video.

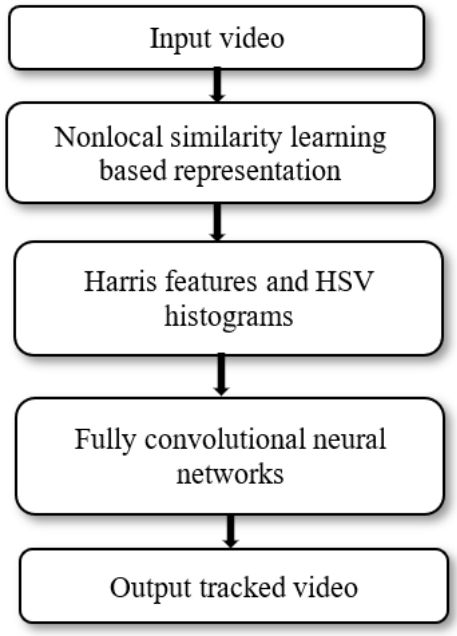


Fig (1): Flowchart for proposed method

Feature extraction by using Harris corner detector: Corner Detector of Harris is used for the observation of corner performance, and it is also widely used for identifying edges that infer object features in computer vision algorithms. It takes into account the difference in a corner point with direct orientation reference, to the direction, rather than using the moving patches for each 45-degree angle, the distinguishing between edges and corners was shown to be more precise. The corner is an end where the localized region is located as both presiding and distinct directions of the edge. In other words, a corner can be comprehended as a two-edged convergence whereas the edge is an abrupt change in radiance of an image.

Process of corner detection for Harris:
The procedure of Harris detector is shown in fig (2).

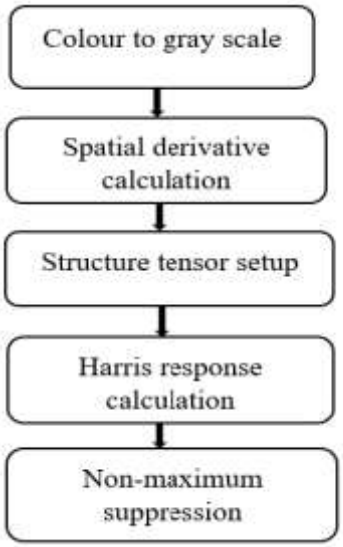


Fig (2): Process of Harris corner detector

Color to grayscale: In this first step we are taking corner detection of Harris in a color image, to convert it into a grayscale image, that improves the speed of processing.

Spatial derivative calculation: Next, we are calculating eq (1).

$$I_x(x,y), I_y(x,y) \qquad (1)$$

Structure tensor setup:

By using above values, the structure tensor M can be developed.

Harris response calculation: In this step, we are measuring the small number of eigenvalues of the structure tensor with the following estimated eq (2).

$$\lambda_{min} \approx \frac{\lambda_1\lambda_2}{(\lambda_1+\lambda_2)} = \frac{\det(M)}{\text{trace}(M)} \qquad (2)$$

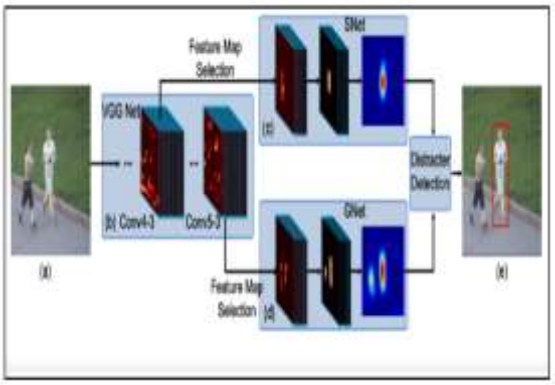
Non-maximum suppression:
We can find the local maximum as corners to get the optimum values to denote corners. Within the frame, which is a 3 by 3 filter.

FCNN Classifier:
This approach can be used to distinguish between target and the background then its quality becomes stronger and more stable slowly it becomes the major method in tracking. The discriminatory technique is also mentioned as Tracking-by-Detection, and this is also involving in deep learning. In order to obtain tracking-by-detection, we can identify the possible objects for all the frames and deep learning is used to identify the targeted object from the candidates. The fundamental network models are of two types: stacked auto-encoders (SAE) and convolutional neural networks (CNN).

- Stacked denoising auto-encoder will increase the ability to express features by adding noise and reconstruction of actual images to the input images.
- In vision of computer and visual tracking, CNN has to enhance the mainstream deep model. CNN can be trained as a tracker as well as a classifier. A fully-convolutional neural network tracker (FCNN) and multi-domain CNN (MD Net) are two representative CNN-based tracking algorithms.

FCNN successfully analyzes and exploits the VGG models feature maps, as the pre-trained ImageNet, and results in the following perceptions:

- CNN feature maps are used for locating an object and also for tracking.
- Most CNN function maps are noisy and irrelevant to distinguishing from the perspective of a particular object.
- Higher layers represent linguistic classification constructs, while lower layers can encode a greater number of critical features to take the intra-class variation**.**



Fig(3):Block diagram of FCNN classifier

In the fig (3) an input ROI region(a) is taken and shows how FCNN develops the feature selection network to specify the most related feature maps against (fig3 (b) and (c)) conv4–3 and conv5–3 layers of the VGG network. So, as to keep away from over fitting with the noisy ones, two additional channels called SNet and GNet are designed for choosing highlight maps from two layers independently, i.e. from (fig3(c) and (d)).

The GNet catches the object's class data while the SNet distinguishes the object with a similar appearance from a background. All networks are initialized in the first frame along with specified boundary box to obtain the object's heat maps, and a region of interest (ROI) is focused at the position of object and the last frame is cropped and generated for new frames. Eventually, the classifier receives two warmth

maps for prediction via SNet and GNet, and also the tracker determines that the heat maps are used to produce last tracking outcomes even if the distractors are present.

IV. RESULTS AND DISCUSSION

The performance of the method that is used for enhancing the video tracking has been accessed by using MATLAB software tool and each second we can run 5 frames on a pc with intel i3. We use CVPR2013 benchmark dataset for tracking. The proposed methodology is evaluated for various videos and the output of the target object is analyzed in each case by using NSL method followed by FCNN classifier.

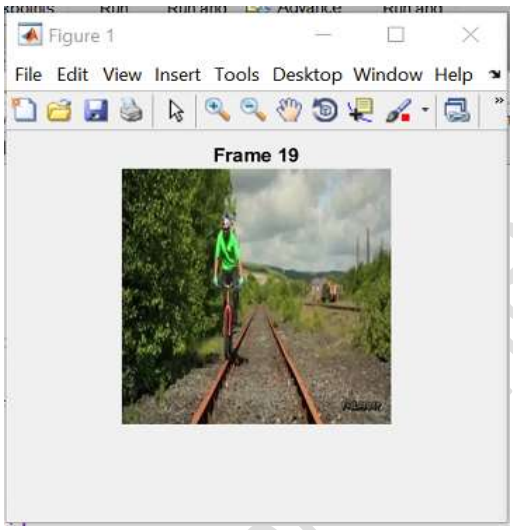


Fig 4(a): Input video

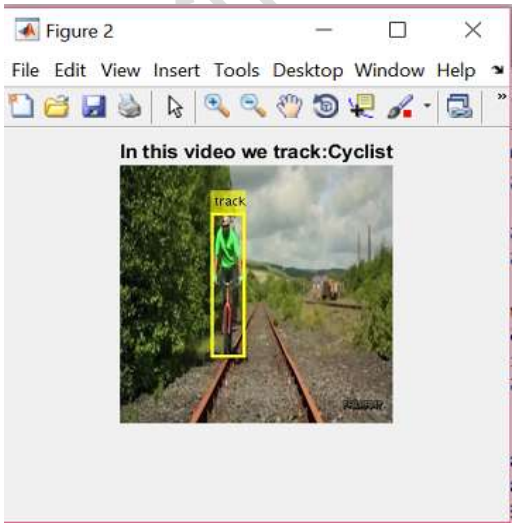


Fig 4(b): Extracted object tracked in video

Fig 4(c): output parameters

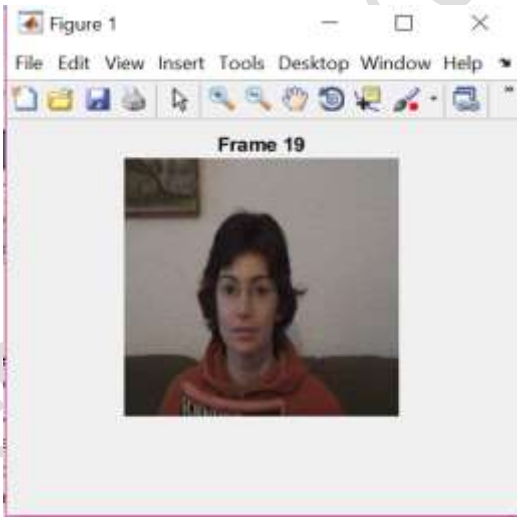


Fig 5(a): Input video

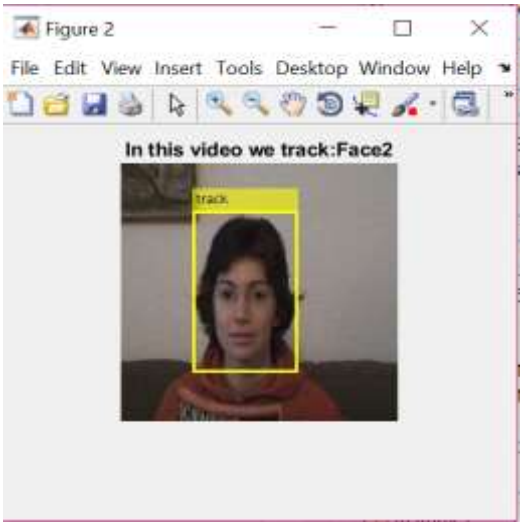


Fig 5(b):Extracted object tracked in video

```
Command Window
  Accuracy--
       0.9875

  Precision--
       0.9786

  Recall--
       0.9975
```

Fig 5(c): output parameters

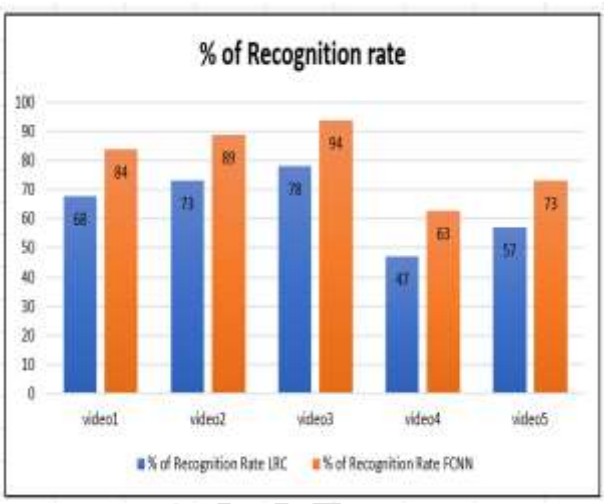
The above results are showing the performance of object tracking for different videos when the target undergoes fast motion and illumination change. Fig 4(a) shows Input video i.e., a person while cycling on the track goes to drift condition which appears on the screen because of rapid motion. fig 4(b) and fig 4(c) shows how our targeted object is tracked and also calculated different parameters for the output video.

Similarly, fig 5(a) shows input video i.e., a person moves her head with changing in the light conditions then the targeted object goes through a sudden illumining change. Fig 5(b) and fig 5(c) shows tracking results of output video and different parameters are calculated for the tracked object in that video which can be done by using NSL followed by FCNN classifier the targeted objects are tracked stably.

Table: Tracking results of logistic regression classifier and fully convolutional neural networks for different videos.

Input Frames	Correctly Detected		Wrongly Detected		Percentage of Recognition Rate	
	LRC	FCNN	LRC	FCNN	LRC	FCNN
Video 1 19 Frames	13	16	6	3	68	84
Video 2 19 Frames	14	17	5	2	73	89
Video 3 19 Frames	15	18	4	1	78	94
Video 4 19 Frames	9	12	10	7	47	63
Video 5 19 Frames	11	14	8	5	57	73

Graph: Graphical representation of different videos for LRC and FCNN classifier.



V. CONCLUSION

This paper recommends a technique for foreground tracking of objects in a video and Features are extracted using local features using Harris corner features for tracking the objects in video. Most of the existing methods identify a target with a single spatial framework format without taking into account the correspondence between various matrix, thereby affecting their presentation when the objective appearance experiences from rapid motion, illumination conditions, occlusion, pose variations. The tracking features are updated for FCNN classifier to improve speed and remove noisy feature maps to make the tracking object more accurate.

REFERENCES:

[1] A. Yilmaz, O. Javed, and M. Shah. Object Tracking: A Survey. ACM Computing Surveys, 38(4):1–45, 2006.

[2] K. Cannons. A Review of Visual Tracking. Technical Report CSE2008-07, York University, Canada, 2008.

[3] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-Based Object Tracking. PAMI, 25(5):564–577, 2003.

[4] G. Chechik, V. Sharma, U. Shalit, and S. Bengio, "Large scale online learning of image similarity through ranking," Journal of Machine Learning Research, vol. 11, no. Mar, pp. 1109–1135, 2010.

[5] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, "Robust visual tracking via multi-task sparse learning," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 2042–2049, 2012.

[6] D. Wang, H. Lu, and M.-H. Yang, "Online object tracking with sparse prototypes," IEEE Transactions on Image Processing, vol. 22, no. 1, pp. 314–325, 2013.

[7] Luca Bertinetto, Jack Valmadre, Joo F. Henriques, Andrea Vedaldi, and Philip H. S. Torr, "Fully-convolutional siamese networks for object tracking," in European Conference on Computer Vision, 2016, pp. 850–865.

[8] Joo F. Henriques, Caseiro Rui, Pedro Martins, and Jorge Batista, "Exploiting the circulant structure of tracking by-detection with kernels," in European Conference on Computer Vision, 2012, pp. 702715.

[9] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," International Journal of Computer Vision, vol. 77, no. 1-3, pp. 125–141, 2008.

[10] B Babenko, M. H. Yang, and S Belongie, "Visual tracking with online multiple instance learning.," in Computer Vision and Pattern Recognition, 2009, pp. 983–990.

[11] H. Song, "Robust visual tracking via online informative feature selection," Electronics Letters, vol. 50, no. 25, pp. 1931–1933, 2014.

[12] K. Zhang, L. Zhang, and M.-H. Yang, "Fast compressive tracking," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 10, pp. 2002–2015, 2014.

[13] K. Zhang, L. Zhang, and M.-H. Yang, "Real-time object tracking via online discriminative feature selection," IEEE Transactions on Image Processing, vol. 22, no. 12, pp. 4664–4677, 2013.

[14] S. Hare, A. Saffari, and P. H. Torr, "Struck: Structured output tracking with kernels," in Proceedings of the IEEE International Conference on Computer Vision, pp. 263–270, 2011.

[15] X. Mei and H. Ling, "Robust visual tracking and vehicle classification via sparse representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 11, pp. 2259–2272, 2011.

[16] Qingshan Liu, Jiaqing Fan, Huihui Song, Wei Chen, Kaihua Zhang, "Visual tracking via nonlocal similarity learning," *IEEE* transactions on circuits and systems for video technology, 1051-8215, 2016.