



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 9, September 2021

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 7.569

# Review on Object Detection and Tracking in Videos using Deep Learning

Parul Pulaiya<sup>1</sup>, Mr. Siddharth Singh Parihar<sup>2</sup>

M. Tech. Scholar, Dept. of Electronics and Communication Engineering, SISTec, Gandhinagar Bhopal, India<sup>1</sup>

Assistant Professor, Dept. of Electronics and Communication Engineering, SISTec, Gandhinagar Bhopal, India<sup>2</sup>

**ABSTRACT:** The growth in amount of video content has led to development of number of applications like video summarization, extraction of useful content from videos, development of automatic video surveillance systems etc. One common use analysis for such applications is detection and tracking of moving objects from videos. There are some smart techniques such as background detection, frame difference, Gaussian mixture modeling, optical flow based techniques, that have been proposed by researchers in this context. But the problem is still evergreen because of number of challenges that complicate detection and tracking process like camera jitter, rapid illumination changes, dynamic backgrounds, moving cameras and shadow detection. These problems cannot be tackled by conventional algorithms because of complexities, impreciseness and improbable factors introduced at various intermediate steps. To cater these problems, soft computing based studies had begun to take center stage and have demonstrated a sound progress in overcoming these issues which are difficult to be handled by traditional approaches.

**KEYWORDS:** - Dilation, Morphology, Erosion

## I. INTRODUCTION

Many computer vision applications require object detection and tracking in video sequences as a vital step for target detection, target monitoring and activity analysis of target in videos. Since video based systems are relatively inexpensive and can capture a huge amount of desired information, there is an immediate need for automated video understanding methods in which object detection and tracking can be done without involvement of human operators. This growth has led to great demand on real time video processing. Object detection and tracking objects are among the most challenging task to accomplish for determining meaningful events and suspicious activities from video sequence. The general framework for object detection and tracking in videos is shown in Figure 1.1. The framework is initialized by feeding video imagery. The process of object detection begins with object segmentation followed by extraction of object from video sequence. Detection of object is followed by labeling and tracking of that object. Labeling is important to differentiate and track an object over series of frames. Preprocessing operations can be applied before or after an object is detected. Preprocessing before object extraction will target to enhance quality of video sequences by suppressing of noise or redundant information to enable more accurate and precise detections. It is different from other perspective where it is targeted to enhance boundaries or shapes of detected object with help of morphological operations.

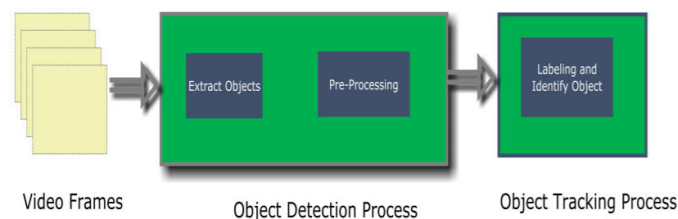


Figure 1: Framework of object detection and tracking in videos

## Object Detection

The interpretation and analysis of video streams is an active research field and one of the important tasks in this field is object detection. Object detection is the process of locating the occurrence of object using number of techniques like background subtraction, feature extraction, statistical methods etc. Figure 1.2 shows some example of object detection in videos. Colored boxes on the figures are used to highlight the detected object on to frame. Applications like video

event detection need to detect and track objects firstly, followed by recognizing what is going on around those tracked objects.



Figure 2: Examples of object detection in videos

## II. LITERATURE REVIEW

**Ganbayar Batchuluun et al. [1]**, the studies discussed above are focused towards discussion of related studies on specific applications like video surveillance, background subtraction or human detection as center of compilation. Although compilation of studies targeting certain suitable applications is important but contributions of soft computing techniques in developing these applications shall not be ignored. The exploits made by soft computing approaches is the major reason for success of these studies. Such contributions shall be acknowledged by someone and it is an attempt in this direction. As the work in these applications prospered together with challenges thrown by real time environmental conditions like background cluttering, occlusions, shape, appearance changes of object, studies based on soft computing techniques had really evolved over the past few years.

**B. Oswald-Tranta et al. [2]**, Soft computing approaches are tolerant of uncertainty, partial truth, and approximation and imprecision and have sound history in offering good solutions in field of object detection and tracking. But studies on these solutions are scattered around and there is need to acknowledge contributions of these studies and group them together under one umbrella. A survey based on contribution of soft computing based studies for object detection and tracking will widely contribute in this direction and is likely to be appreciated by fellow researchers. The chapter is an attempt in that direction and will focus to track development, use and progress of soft computing based methods in the development of some good applications. The primary constituents of this chapter are studies on soft computing (SC) have been studied (1) Neural Network Based approaches (2) Fuzzy Logic Based Approaches (3) Evolutionary Algorithm Based approaches (4) Hybrid approaches. The chapter will focus on soft computing based object detection and tracking approaches implemented by various applications and discusses object detection and object tracking separately.

**G. Batchuluun et al. [3]**, a double difference approach was introduced three frame difference approach based on mathematical morphological operation such as erosion and dilation for object detection. The proposed approach computes frame difference between three consecutive images and then applies AND operator on difference. Mathematical morphological operations such as dilation and erosion are applied to determine moving object on resultant difference.

**G. Batchuluun et al. [4]**, has utilized adaptive frame differencing, moving object localization and automatic thresholding technique for simultaneous detection of slow and fast moving objects in videos. Some researchers embarked upon that combination of frame difference methods with other conventional techniques to improve its

performance in field of object detection in videos. The third category of conventional object detection approaches for object detection in videos is Statistical approach. They make use of properties of pixels or group of pixels to construct dynamically updated background model. In this approach foreground pixels are identified by comparing statistics of group of pixels in question with the statistics of the background model.

**S. Farsiu et al. [5]**, this method use a single gaussian distribution method per pixel to compensate uncorrelated noise. To overcome limitation of single gaussian distribution approach, Gaussian mixture model is effective in terms of handling low illumination variation but shadows and rapid illumination variation is still problematic. The noisy images from UAVs were stabilized using improved Fourier Mellin Variant method and de-noised using video block matching 3D filtering. Gaussian mixture based background/foreground segmentation algorithm was applied to achieve object detection and Kalman filter was applied for tracking of human beings in the processed video.

**Y. Matsushita et al. [6]**, proposed an update framework for learning adaptive mixture models of background scene for the real-time tracking of moving objects. The proposed approach serves both as an adaptive model for background subtraction and as classifier of pixels for foreground as well as for background. As a result, efficiency was increased and there was reduction in processing time. The optical flow models are one of the oldest and still active domains of object detection in videos. This model is usually able to deal with sudden illumination variation and camera jitter but some problems such as variation in object size, still remains an active area using optical flow based approach.

**H. Linyang et al. [7]**, used region-based mixture of Gaussian (RMoG) to model dynamic backgrounds by taking neighboring pixels. It proposed a methodology to detect and track moving objects acquired by Unmanned Aerial Vehicles (UAVs). Some authors have utilized various other statistical approaches such as support vector models and subspace learning techniques for modeling background. The statistical methods can model less dynamic scenes, noise of the camera sensor and provides robust performance in case of camera jitter. Performance of statistical methods degrades unfortunately from vacillating backgrounds (e.g. swaying trees), moving background elements and slow environmental light changes. Another conventional approach that can be used by various researchers for object detection in videos is Optical flow based model. In this approach flow vectors of moving objects are calculated over time to determine moving objects in videos.

**H. Park et al. [8]**, these models for object detection in videos were presented in literature presented an algorithm based on optic flow based approach for traffic surveillance application. In the proposed approach median filter was employed to remove noise and thresholding operation was utilized to achieve accuracy in detection of moving objects. In this method, original optical flow field was estimated by Horn–Schunck optical flow algorithm and followed by enhancement of object using local mean algorithm, finally the noise was filtered out using self-adaptive thresholding function.

**D. Bianli et al. [9]**, put forward a combined kirsch operator and optical flow based approach for moving object detection. Kirsch operation was applied to determine contours of objects and motion vector field for video sequence was established using optical flow based approach. Neural networks are family of models inspired by working of human brain and nervous system. Neural network have the ability to learn from historic data and neural based algorithms are powerful tools already proven for providing useful and efficient solutions for classification problems. As a model to simulate the learning function of human brains, this method is able to perform classification of moving objects accurately when network is trained with sufficient training data. With the advent of techniques like deep learning, deep convolution networks the success gained by these studies is also quite remarkable.

### III. VIDEO RECONSTRUCTION

Essentially, the true aim of salient object detection is to find objects that are distinctive from the image background. It needs to calculate the contrast between the objects and the image background and then select those with high contrast as salient objects. However, the local and global contrast based methods do not identify which regions form the image background. They blindly assume the neighboring regions or the entire image to be the background and then calculate the contrast between each location and the assumed background. As their assumed background may not be the real one, the determined contrast also becomes incorrect, which in turn reduces the performance of saliency detection. To overcome these problems, a few emerging methods [7], [8] using background priors were proposed based on the idea of modeling the property of background first and thereby separating salient objects from the background. Specially, Wei

et al. [9] exploited the boundary and connectivity priors about the background in natural images and detected saliency based on the geodesic distance. Considering that the salient object may be partially cropped on the boundary, this paper adopts an existing saliency detection method [10] to compute the saliency of boundary patches and generates weights for the virtual background nodes. However, in some challenging images, where [11] could not calculate the saliency of boundary patches precisely, the method of [12] is difficult to obtain satisfactory results. Yang et al. [13] modeled saliency detection as a manifold ranking problem and proposed a two-stage scheme for graph labeling. They represent the image as a close-loop graph with superpixels as nodes. In saliency detection, they first use the nodes on the image boundary as background seeds to rank other nodes in the graph. Then, in the second stage, they select the salient nodes from the detection results of the first stage and use them to refine the saliency of other nodes in the graph. On the assumption that the image boundary is mostly background, these methods result in a background template. As a result, the contrast between salient object and background can be precisely obtained. By incorporating background priors into traditional contrast-based methods, they show improved results in saliency detection. However, existing background prior-based methods still have certain limitations. Typically, there are four scenarios where performing background prior-based saliency detection as summarized in the following.

- 1) The entire image boundary is a large and smoothly connected region (see the first row of Fig. 4).
- 2) The regions defined within the image boundary look different, whereas they may share certain latent pattern (see the second row of Fig. 4).
- 3) The background is complex (for example, containing small-scale high-contrast patterns) and the regions of image boundary are different, as shown in the third row of Fig. 4.
- 4) Salient objects significantly touch the image boundary and parts of them are wrongly considered as the background, as shown in the fourth row of Fig. 1. As can be seen in Fig. 4, existing background prior based approaches [12] are effective for the first scenario and moderately effective for the second scenario. However, unsatisfactory results are produced in dealing with the last two scenarios. In this paper, we propose a novel background prior based saliency detection framework using a stacked denoising auto encoder (SDAE) with deep learning architectures.

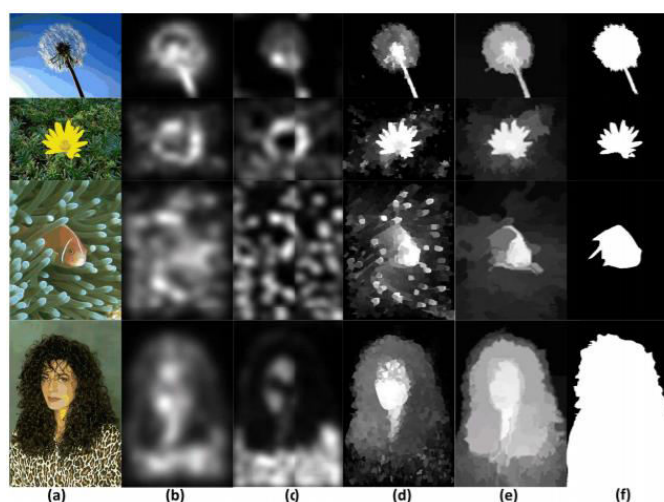


Fig. 3: Some examples of saliency detection for Video Collection

In this paper, the SDAE is used to model image background. Rather than adopting hand-designed features as used in previous works [13], [14], the deep-structured SDAE is employed to learn more powerful representation directly from the raw image data in an unsupervised way, which also enables capturing the latent pattern of the input data hierarchically. It thus helps to deal with the second scenario (shown in the second row of Fig. 4) where the background regions share latent patterns. Then, the measure of contrast between the salient objects and the background is formulated as the reconstruction residuals in the deep-structured SDAE. Unlike the previous works [18], [19], which mainly focused on the way to calculate the similarity or distinctiveness between a certain image patch and the image boundary, this paper pays more attention to modeling the background regions.

#### IV. METHODOLOGY

Research design is a strategy, a plan and a structure for conducting research. It is comprehensive master plan of research study to be undertaken, giving a general statement of methods to be used. Selection of research design is to ensure that requisite data is in accordance with the problem at hand and is collected accurately and economically. It is frame work or blueprint for research study that guides collection and analysis of data. Research design depending upon needs of researcher may be a detailed statement or only furnishing of minimum information required for planning research project. Every design has its positive and negative sides and best design depends upon research question as well as orientation of researcher. Various modes of research design are: Theoretical research design, applied research design, exploratory design, descriptive design, quasi-experimental design and experimental design.

##### Artificial Neural Network

Neural networks are the data handling frameworks, which are built and executed to show the human cerebrum. The main object of the neural network analysis is to develop a process of computing device for modeling the brain to execute various process of computing tasks at a faster rate than the traditional systems. Artificial Neural Networks execute various tasks such pattern matching and classification, optimization function and data clustering. These errands are exceptionally troublesome for conventional PCs, which are quicker in algorithmic procedure of registering undertakings and exact number juggling tasks. ANNs gangs substantial number of exceedingly interconnected preparing components called hubs or unit or neuron, which more often than not work in parallel and are arranged in standard models [18]. Every neuron is associated with the other by an association interface. Every association interface is related with weights, which contain data about the info flag. This data is required by neuron net to take care of a specific issue. ANN, s aggregate conduct is portrayed by their capacity to learn, review and sum up preparing examples or information like that of human mind.

##### Multilayer Feed-forward Neural Network

The back propagation algorithm performs learning on multilayer feed-forward neural network. An example of such network is shown in figure 3.1. The inputs correspond to the attributes measured for each training sample. The information sources are sustained at the same time into a layer of units making up the info layer. The weighted yields of these units are thus, encouraged all the while to a moment layer of "neuron like" units, known as concealed layer. The shrouded layer's weighted yields can be contribution to another concealed layer, et cetera. The quantity of shrouded layer is discretionary, in spite of the fact that practically speaking, normally just a single is utilized. The weighted yields of the last shrouded layer are contribution to units making up the yield layer, which emanates the system's expectation for given examples [18].

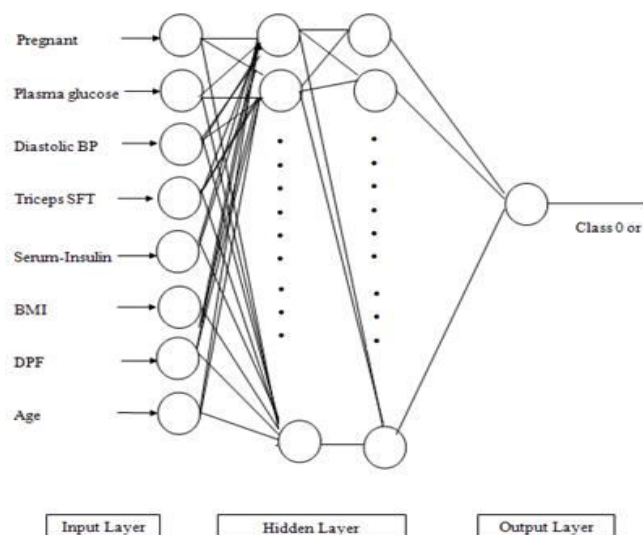


Fig. 4: Feed forward NN model for diabetes diagnosis

## V. CONCLUSION

Object detection and tracking is one of the challenging research tasks of computer vision aimed at detecting moving objects from video sequence. It is followed by predicting the path of moving object for the duration of its presence in video frame sequences. This study has provided a comprehensive review of the state-of-the-art methods on object detection and tracking with focus on soft computing based approaches. Research articles utilizing classic theories of soft computing have been explored and categorized suitably. The study has also explored effectiveness of recently proposed techniques of soft computing like deep learning and hybrid approaches. It has presented an insight into failures as well as success of the existing soft computing based approaches to help new researchers to have a wider view on the entire field. Availability of datasets, research trends and challenging issues in detection and tracking has also been highlighted. Analysis of various soft computing techniques on different datasets is also tabulated for interested readers.

## REFERENCES

- [1] Ganbayar Batchuluun, Jin Kyu Kang, Dat Tien Nguyen, Tuyen Danh Pham, Muhammad Arsalan And Kang Ryoung Park, "Deep Learning-based Thermal Image Reconstruction and Object Detection", IEEE Access, 2020.
- [2] B. Oswald-Tranta, "Temperature Reconstruction of Infrared Images with Motion Deblurring," J. Sens. Sens. Syst., vol. 7, pp. 13–20, 2018.
- [3] G. Batchuluun, Y. W. Lee, D. T. Nguyen, T. D. Pham, and K. R. Park, "Thermal Image Reconstruction Using Deep Learning", IEEE Access, vol. 8, pp. 126839–126858, 2020.
- [4] G. Batchuluun, H. S. Yoon, D. T. Nguyen, T. D. Pham, and K. R. Park, "A Study on the Elimination of Thermal Reflections," IEEE Access, vol. 7, pp. 174597–174611, 2019.
- [5] S. Farsiu, M.D. Robinson, M. Elad, and P. Milanfar, "Fast and Robust Multiframe Super Resolution," IEEE Trans. Image Process., vol. 13, pp. 1327–1344, 2004.
- [6] Y. Matsushita, H. Kawasaki, S. Ono, and K. Ikeuchi, "Simultaneous Deblur and Super-Resolution Technique for Video Sequence Captured by Hand-Held Video Camera," in Proc. IEEE International Conference on Image Processing, Paris, France, 27–30 Oct. 2014, pp. 4562–4566.
- [7] H. Linyang, L. Gang, and L. Jinghong, "Joint Motion Deblurring and Superresolution from Single Blurry Image," Math. Probl. Eng., vol. 2015, pp. 1–10, 2015.
- [8] H. Park and K.M. Lee, "Joint Estimation of Camera Pose, Depth, Deblurring, and Super-Resolution from a Blurred Image Sequence," Sep. 2017, arXiv: 1709.05745v1. [Online]. Available: <https://arxiv.org/abs/1709.05745v1>.
- [9] D. Bianli, R. Xiaokang, and R. Jie, "CNN-based Image Super-Resolution and Deblurring," in Proc. Association for Computing Machinery, Wuhan, China, 29–31 Oct. 2019, pp. 70–74.
- [10] X. Zhang, F. Wang, H. Dong, and Y. Guo, "A Deep Encoder-Decoder Networks for Joint Deblurring and Super-Resolution," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, Calgary, Canada, 15–20 Apr. 2018, pp. 1448–1452.
- [11] X. Zhang, H. Dong, Z. Hu, W.-S. Lai, F. Wang, and M.-H. Yang, "Gated Fusion Network for Joint Image Deblurring and Super-Resolution," Jul. 2018, arXiv: 1807.10806v1. [Online]. Available: <https://arxiv.org/abs/1807.10806v1>.
- [12] F. Albluwi, V.A. Krylov, and R. Dahyot, "Image Deblurring and Super-Resolution Using Deep Convolutional Networks," in Proc. IEEE 28th International Workshop on Machine Learning for Signal Processing, Aalborg, Denmark, 17–20 Sept. 2018, pp.1–6.
- [13] J.U. Yun and I.K. Park, "Joint Face Super-Resolution and Deblurring Using a Generative Adversarial Network," Dec. 2019, arXiv: 1912.10427v1. [Online]. Available: <https://arxiv.org/abs/1912.10427v1>.
- [14] F. Vankawala, A. Ganatra, and A. Patel, "A Survey on Different Image Deblurring Techniques," Int. J. Comput. Appl., vol. 116, pp. 15–18, 2015.
- [15] S.F. Hamood, M.S.M. Rahim, O. Farook, and D. Kasmuni, "A Survey on Various Image Deblurring Methods," J. Eng. Appl. Sci., vol. 3, pp. 561–569, 2016.



**Impact Factor:  
7.569**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details