

A Review on Visual Object Tracking Methods & Impact for Benchmarks

Amit Saxena

Research Scholar, Computer
Science & Engineering,
Rabindranath Tagore University
Bhopal, Madhya Pradesh, India
amit.saxena78@gmail.com

Sitesh Kumar Sinha

Professor, Computer Science
& Engineering,
Rabindranath Tagore University
Bhopal, Madhya Pradesh, India
siteshkumarsinha@gmail.com

Manish Manoria

Professor, Computer Science
& Engineering,
Truba Institute of Engineering &
Information Technology,
Bhopal, Madhya Pradesh, India
manishmanoria@gmail.com

Sanjeev Kumar Gupta

Dean Academics,
Rabindranath Tagore University
Bhopal, Madhya Pradesh, India
sanjeevgupta73@yahoo.com

Abstract—Visual Object Tracking is a significant exploration point in Computer vision pattern recognition. The aim of Visual Object Tracking is to automatically acquire the environment of the object in the ensuing video outlines. Visual tracking is now very useful for tracking various moving objects like football in a match and many more for efficiently tracking the target for better decision making. In Artificial Intelligence; visual tracking is more challengeable because of instability of object in the frames. Conventional techniques are not efficient to deal with this kind of challenges. For tracking the object more efficiently; it has been targeted to achieve it through machine learning techniques. This paper intends to review various implemented researches that accepted the challenges for visual tracking and compare to the ground-truth. There are so many researches have been done in this field but somewhere somehow missing the target in certain benchmarks. The most promising technology which has been used by various researchers is Convolutional Neural Network. A network should be trained with all occlusions and acquire better level of precision with minimal overflows.

Keywords—Visual Tracking, Computer Vision, CNN, ANN, OpenCV, Machine Learning, OTB.

I. INTRODUCTION

Visual tracking is a trending part of moving object recognition by using pattern recognition. In the very initial stage; it focuses on object area and its size as it is considered as in the first frame then system keep tracking the particular area and size as per the pattern recognized in the first frame till the last frame. Many researchers use distinct methodologies for tracking object in a video frames with various challenge factors such as Deformations (DEF), Illumination Variations (IV), Background Clutter (BC), In-Plane Rotations (IPR), Fast Motions (FM), Occlusions (OCC), Out of Plane Rotations (OPR), Motion Blurs (MB), Scale Variations (SV), Out of Views (OV) and Low Resolutions (LR) [1]. Visual tracking is the method involved

with finding a moving object over the long haul by utilizing a camera. It has an assortment of uses, some of which are H-C interface, security and reconnaissance and expanded reality, traffic control, clinical imaging and video altering. Video tracking can be a tedious interaction because of the measure of information that is contained in video. Intricacy is the conceivable need to utilize object recognition methods for tracking. The goal of video tracking is to relate target objects in back to back video frames. The network can be particularly complex when the items are moving quick comparative with the frame rate. Another circumstance that expands the intricacy of the issue is the point at which the tracked object changes the direction over the long run. For these circumstances video tracking frameworks normally utilize a movement model which portrays how the picture of the objective may change for various potential movements. To perform video tracking a calculation examines consecutive video frames and yields the development of focuses between the frames. There are an assortment of calculations, each having qualities and weakness. Considering the expected use is important while picking which methodology is to utilize. There are two significant parts of a visual tracking framework: target presentation and localization [2].

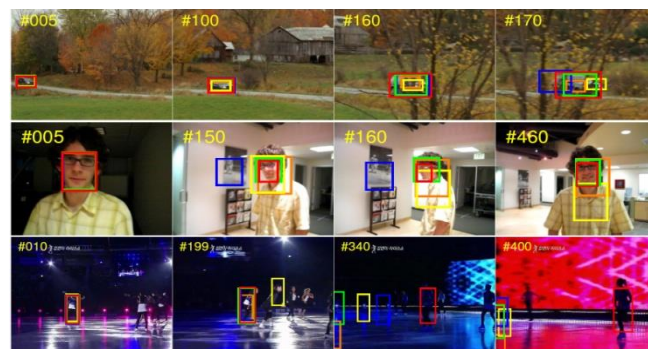


Fig. 1. Visual Tracking OTB50 [3]

The state-of-the-art based visual trackers have particular qualities like reconciliation of different models, learning techniques, training datasets, network destinations, network yields, sorts of took advantage of profound provisions, CPU/GPU executions, programming dialects and frameworks, speed, etc. Therefore, this work gives a similar investigation of visual trackers, benchmark datasets, and assessment measurements to research proposed trackers exhaustively and work with creating progressed trackers. Visual objective trackers can be generally arranged into two fundamental classes, prior and then afterward the insurgency of DL in computer vision. The principal classification is essentially explored by [4], which incorporate traditional trackers dependent on old style appearance and movement models, and afterward inspect their upsides and downsides efficiently, tentatively, or both. These trackers utilize physically planned elements for target displaying to lighten appearance varieties and to give productive computational intricacy. For example, all these trackers are appropriate to execute on the flying robots because of the limitations of utilizing progressed GPUs, they need more fervor to deal with the difficulties of in-thewild videos. Ordinarily, these trackers attempt to outfit different components to develop a corresponding arrangement of visual signals. Considering visual trackers' critical advancement as of now, they looked into techniques by the referenced works [5].

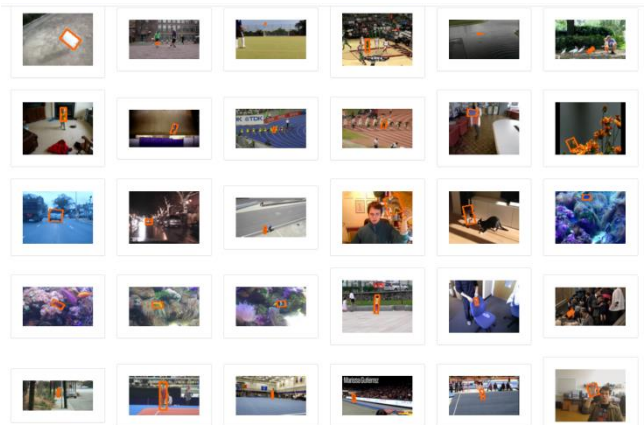


Fig. 2. VOT Benchmarks [1]

Fig. 2 shows the VOT benchmark where challenges have been given and researchers have to test their systems by accepting these challenges and pertain their precision accordingly.

II. RELATED WORKS

A. Related Works

Haojie Li et al. [6] proposed a network that is intended to track the object with MA-Dual technique that is following spatial transient approach for tracking the patterns in each and every frames of a dataset. This paper is based on 3d convolutional approach where system extracts the features on the basis of their structures and follow the same in entire frame. System is bit inefficient to obtain the object location in certain datasets because some challenges are bit difficult because of various prospects such as motion blur, low

resolutions and many more. System integrity may get differ accordingly because data may get highlighted distinctly in different luminance. System has been tested with various datasets such as UAV123, OTB benchmarks, VOT and TC128 too. The investigation results show that the proposed strategy accomplishes an exceptionally encouraging tracking execution, and is particularly acceptable at taking care of testing conditions, like disfigurement, scale variety, enlightenment changes, and so forth. Linyu Zheng et al. [7] proposed a Gaussian Process Regression based tracker (GPRT) which is a reasonably normal tracking approach. Contrasted with all the current CF trackers, the limit impact is wiped out completely and the part stunt can be utilized in our GPRT. Also, Authors present two productive and successful update techniques for our GPRT. Analyses are performed on two public datasets: OTB-2013 and OTB-2015. Without extravagant accessories, on these two datasets, our GPRT acquires 84.1% and 79.2% in mean cross-over exactness, individually, outflanking every one of the current trackers with hand-created highlights. An original tracking framework, GPRT which applying the Gaussian Regression Processes to visual tracking, has been introduced in this paper. Contrasted with all the current CF trackers, our GPRT not exclusively doesn't exist the limit impact, however al so can exploit the bit stunt simultaneously. Expansion, Authors propose two distinct proficient and compelling up date techniques for our GPRT. Authors perform extensive tests on two benchmark datasets: OTB-2013 and OTB 2015. Benchmarks pertain more than 100 videos with thousands of frames in all challenges like birds, bolt, box, cars, bikers, blur bodies, football, human, dudek, david, crowds and many more.

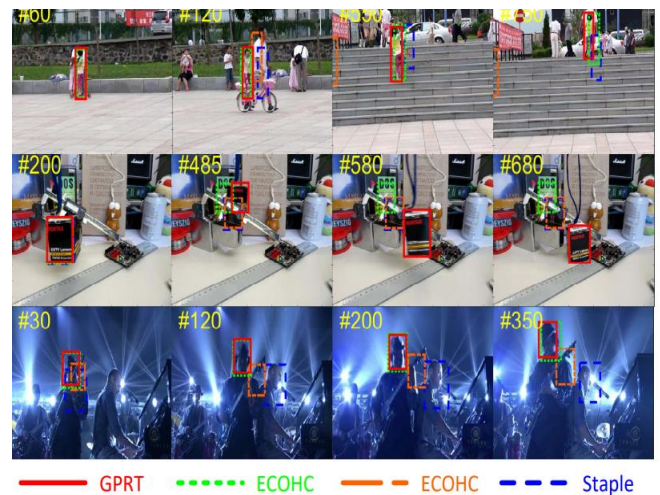


Fig. 3. GPRT Visual Tracking[7]

Martin Danelljan et al. [8] proposed a probabilistic regression detailing and apply it to tracking. System's network predicts the restrictive likelihood thickness of the objective state given an info image. Significantly, system's plan is equipped for demonstrating name commotion originating from in precise explanations and ambiguities in the assignment. The regression network is prepared by limiting the Kullback Leibler dissimilarity. When applied for tracking, system's definition not just permits a

probabilistic portrayal of the yield, yet additionally generously works on the presentation. System's tracker sets another best in class on six datasets, accomplishing 59.8% AUC on LaSOT and 75.8% Accomplishment on TrackingNet. Kai Chen et al. [9] proposed a regression method that follows the convolutional network for tracking the moving object. Here the system is based on edge regression that also extract the edges for tracking the object by its patterns and textures. System is also based on back propagation model that follows the rendering technique with different layers of convolutional model. In the DCF model, each layer has been designed or trained with different perspective that follows the various integration features that object pertains and on the basis of these parameters system tracks the object with different proportions and challenges with various iterations and back propagations. It is one more approach to manage come out as comfortable with the relapse model for visual following single convolutional layer. Maybe than learning the immediate relapse model in a shut construction, creators endeavor to deal with the relapse issue by propelling a one-channel-yield convolution layer with GD. In particular, the piece size of the convolution layer is set to the size of the item. Rather than DCF, it is attainable to intertwine all "certified" models cut from the whole picture. An essential issue of the GD approach is that most of the convolutional tests are negative and the responsibility of positive models will be covered. To determine this issue, creators propose a cunning target ability to clear out straightforward negatives and update up-sides. To accelerate the preparation stage, authors additionally propose a worked on objective capacity to kill simple negatives and improve positives. The outcomes show that the proposed calculation accomplishes extraordinary execution and beats the majority of the current DCF-based calculations. Anton Millan et al. [10] proposed a system which is based on recurrent neural network. Authors present a distinct way to deal with multi-target tracking dependent on RNNs.

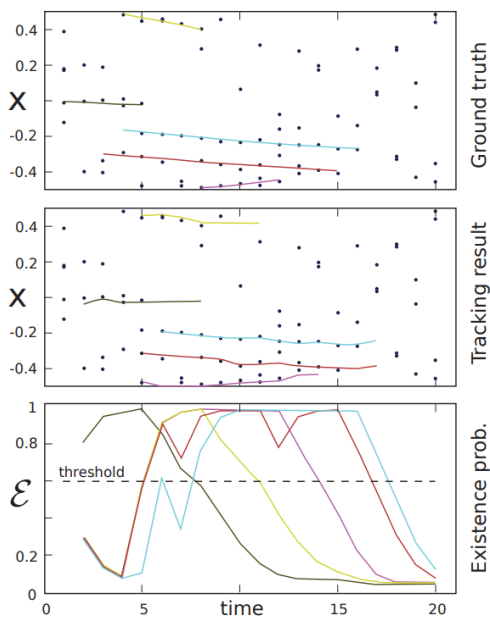


Fig. 4. Result Comparison with Ground-truth [10]

Tracking multiple objects in true scenes includes many difficulties, including a) a deduced obscure and time-differing number of targets, b) a persistent state assessment of every current objective, and c) a discrete combinatorial issue. Most past strategies include complex models that require dreary tuning of boundaries. Here, author proposed interestingly, a start to finish learning approach for online multi-target tracking. Existing profound learning strategies are not intended for the above challenges and can't be inconsequentially applied to the assignment. Sangdoo Yun et al. [11] proposed a system which is based on deep reinforcement learning algorithm. In this paper, authors proposed an original activity driven technique utilizing deep convolutional networks for visual tracking. The proposed tracker is constrained by an ADNet, which seeks after the objective object by consecutive activities iteratively. The activity driven tracking methodology makes a critical contribution to the decrease of calculation intricacy in tracking. Likewise, RL makes it conceivable to utilize to some degree named information, which could extraordinarily add to the structure of preparing information with a little exertion. As per the assessment results, the proposed tracker accomplishes the best in class execution in 3 casings/s, which is multiple times quicker than the current profound organization based trackers utilizing a tracking-by-location procedure. Besides, the quick form of the proposed tracker accomplishes an ongoing rate (15 Frames/second) by changing the metaparameters of the ADNet, with a precision that outflanks cutting edge constant trackers. Da Zhang et al. [12] proposed a framework which is based on Deep Reinforcement Learning similarly to previous authors. In this paper; authors acquaint a completely end-to-end approach for visual tracking in videos that figures out how to anticipate the bouncing box areas of an objective object at each edge. A significant knowledge is that the tracking issue can be considered as a successive dynamic cycle and historical semantics encode exceptionally pertinent data for future. In light of this instinct, they detailed out model as a repetitive convolutional neural network that cooperates with a video additional time, and model can be prepared with Region learning (RL) algorithms to learn tracking strategies that focus on continuous outline boxes and amplify tracking performance over the long haul. The proposed tracking calculation accomplishes best in class execution in a current tracking benchmark and works at frame rates quicker.

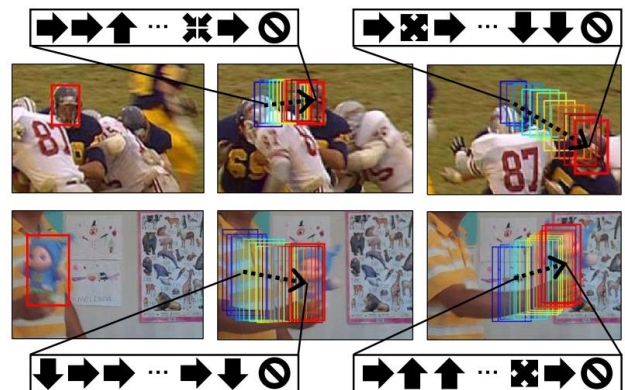


Fig. 5. Sequential Visual Tracking [12]

Joao F. et al. [13] proposed a system which is based on Kernelized Correlation Filters. In this work, authors exhibited that it is feasible to systematically display regular image interpretations, showing that under certain conditions the subsequent information and bit frameworks become circulant. Their diagonalization by the DFT gives an overall diagram to making quick algorithms that arrangement with interpretations. Authors have applied this outline to straight and piece edge relapse, getting best in class trackers that run at many FPS and can be carried out with a couple of lines of code. Expansions of our essential methodology appear liable to be helpful in different issues. Since the primary variant of this work, circulant information has been taken advantage of effectively for different algorithms in location and video event recovery. An intriguing heading for additional work is to loosen up the presumption of occasional limits, which might further develop execution. David S. Bolme et al. [14] proposed a system which is based on Adaptive Correlation Filters. This paper has shown that the visual tracking issue, which is generally utilizing significant bulky classifiers, complex appearance models, and stochastic techniques can be supplanted by proficient and easier MOSSE correlation channels. The outcome is not difficult to execute, that can be comparably exact, and is a lot quicker. In this paper the tracker was kept easy to assess the channel's capacity to follow and adjust to troublesome tracking scenarios. There are various basic ways that tracker can be improved. For instance, if the presence of the objective is moderately steady, floating could be relieved by occasionally re-centering the channel dependent on the underlying casing. The tracker can likewise be stretched out to appraise changes in scale and pivot by separating the log-polar. Debi Prasad Dogra et al. [15] proposed a system which is based on Kalman Filter. This paper presents a novel multi-way investigation based video object tracking calculations. Direction of the moving object is refined by Kalman filter. The proposed calculation has been utilized effectively to dissect one of the mind boggling neurological assessments that frequently alluded to lateral shifting test. This is a significant trial of neurological appraisal cycle, and this test is hard to grade by visual perception. It has been displayed in this paper that the proposed video object tracking calculation can be utilized to examine the videos of quick objects by fusing application-explicit data. For instance, the proposed tracking calculation can be utilized to evaluate sidelong shifting trial of the Mallet smith neurological assessments. Sungdoo Yun et al. [16] proposed a system which is based on Deep Reinforcement Learning. This paper intended a novel tracker which is constrained by successively seeking after activities learned by reinforcement learning. As opposed to the current trackers that utilizing deep neural networks, the proposed tracker is intended to accomplish a light calculation just as palatable tracking accuracy in both area and scale. The deep reinforcement control activities are pre-prepared assessment and tweaked during tracking for online transformation to target and background changes. The training is finished by using deep reinforcement learning just as supervised learning. The utilization of learning even to some degree marked to be effectively used for semi-directed learning.

Through assessment of the OTB dataset, the proposed tracker is validated to accomplish a competitive execution that is multiple times faster than current trackers.

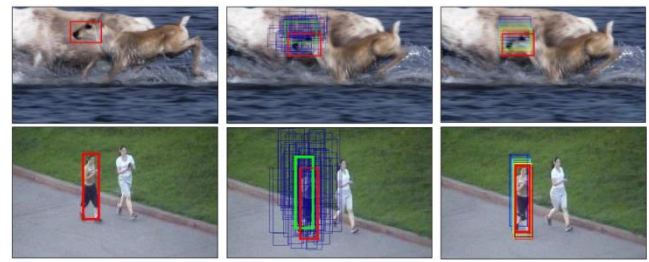


Fig. 6. Deep Reinforcement Visual Tracking [16]

Linyu Zheng et al. [17] proposed a system which is based on Robust Gaussian algorithm. An original tracking structure, GPRT applying the Gaussian Regression Processes to visual tracking, that has been introduced in this paper. Contrasted with all the current CF trackers, the GPRT is not exclusively limit impacts, however it can exploit the part simultaneously. Authors proposed two distinct productive and powerful techniques for GPRT. They performed far reaching system with two benchmark datasets: OTB-2013 and OTB 2015. The GPRT beats every one of the current trackers with this components on both two datasets. Bo Li et al. [18] proposed a system which is based on Siamese RPN++. In this paper, authors have introduced a bound together system, alluded as SiamRPN++, to end to end train a Siamese network for visual tracking. They showed theoretical and experimental proof; that how to prepare a regional network for Siamese tracker. This network is made out of a multi-facet module that gathers the progression assessment with total various degrees of portrayal and a profundity insightful connective layers which permits the network to lessen calculation and excess boundaries while additionally prompting. Utilizing SiamRPN++, authors acquired state-of-the-art results on the VOT2018 continuously, showing the viability of SiamRPN++. SiamRPN++ likewise achieved state-of-the-art results on huge datasets like LaSOT and TrackingNet showing its generalizability. Xin Li et al. [19] proposed a system which is based on Dual Regression model. In this paper, creator proposed a dual-regression tracking outline including a discriminative convolutional neural network and CF module. The proposed tracker acquires a discriminative capacity by learning the difference between the objective and background, and decreases the computational expense by taking advantage of the completely convolutional network. As the completely convolutional network with deeper provisions holds less spatial subtleties, a CF module running on the shallow layer highlights with higher spatial goal to refine the target position. The two streams structure are required; one forward CNN pass to deduce the objective position, which gives a proficient design to deep trackers. The another led on three public datasets show the adequacy and productivity of the proposed calculation. Bingfei Zhang et al. [20] proposed a system which is based on Support Vector Regression. This paper proposed another technique to control the uncalibrated visual servoing for 3D movement tracking. Initially, PI control based development is utilized in image plane. Then, support vector regression (SVR) is

utilized to develop the visual planning model. At long last, a level and three-dimensional space movement tracking is accomplished by means of utilizing constant mapping model. Contrasted and the 3D movement visual tracking with the customary BP neural network technique, the test results exhibited that the SVR had a phenomenal approximating ability under the state of little example learning. Tingting Wang et al. [21] proposed a system which is based on Robust Model Predictive Control. In this paper, a visual-based control law for a two-rotational levels of opportunity video tracking framework with strong model prescient control strategy was proposed.

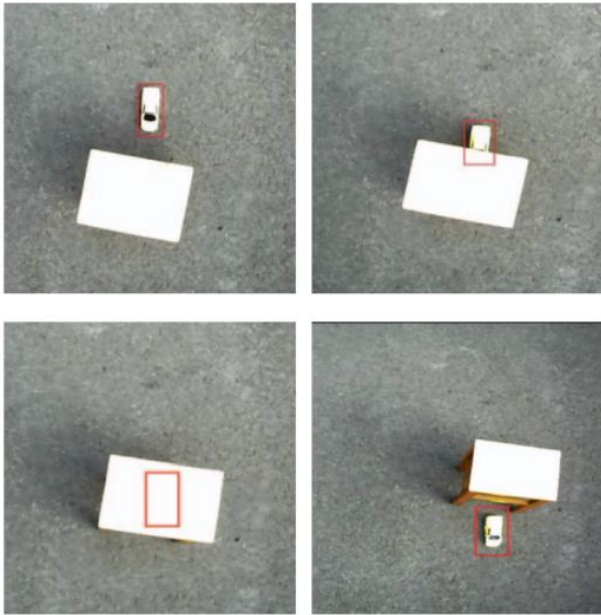


Fig. 7. Tracking Result in Frame Sequence[21]

The video tracking framework is inferred as a visual servoing model whose boundaries variety in the image Jacobian grid can influence both steadiness and execution. By changing this framework into a raised mix of direct time-invariant vertices structure with the tensor-item (TP) model change technique, the video tracking framework is addressed as a polytopic straight boundary shifting (LPV) framework. The control signal is determined online via doing the arched enhancement including direct lattice disparities (LMIs) in MPC to ensure the shut circle solidness and framework imperatives. The reenactment results show that the proposed regulator can viably follow the moving object with obscure speed and distance data. N. Djelal et al. [22] proposed a system which is based on SURF method. In this paper authors present a blend between a speed up robust features (SURF) and the image based visual servoing control law; to accomplish a powerful tracking mission. The endorsement of the proposed approach guaranteed by the control of turret Dish Slant framework that utilizing the visual components gained by a webcam, and these visual provisions are removed by the SURF calculation; in the imperatives of the variety of the scale and the pivot of the image; with the goal that the utilization of the visual elements in the image based visual servoing control law to register the kinematics screw of the camera in the image outline.

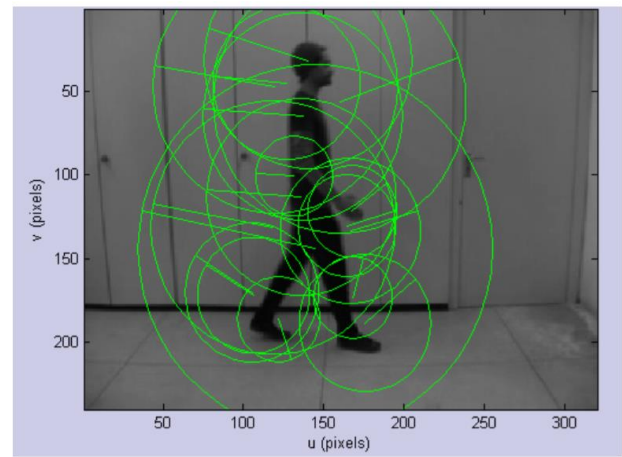


Fig. 8. Support region (scale) orientation (radial line)[22]

C.H. Li et al. [23] proposed a system which is based on Contour method. This review fosters an exceptionally continuous visual tracking framework equipped for tracking a moving objective in jumbled climate. The proposed framework design can be partitioned into three principle parts – the objective indicator, the objective data extractor and the objective tracker. The initial segment distinguishes any moving object in the image, utilizing the moving edges strategy. Then, by coordinating the data from the objective finder and the attractable snake model, the subsequent part consequently extricates the objective's form, brightening and shading data to empower the objective to be followed. At long last, by taking on another mixture coordinating with strategy, the third part empowers the framework to constantly follow the objective with flexibility. Furthermore, the visual tracking framework created in this review is set up on a minimal expense business camera stage, which can be constrained by a PC through a RS232 sequential port. At long last, a few investigations, including object tracking and human head tracking, are led to confirm the power of the tracking framework. Quig Guo et al. [24] proposed a system which is based on Blur and De-blurring technique. In this paper, authors proposed the Blurred Video Tracking (BVT) benchmark to investigate what movement obscure means for visual object tracking and whether state-of-the-art deblurring techniques can help state-of-the-art trackers. The proposed BVT benchmark contains 500 videos for 100 scenes, every one of which has 5 videos with various degrees of movement obscures. As per the assessment aftereffects of 25 trackers on the BVT benchmark, they track down that light movement obscure may effectually affect visual tracking, while serious foggy spots unquestionably compromise the presentation of most trackers. Authors see that current deblurring algorithms can further develop tracking execution on seriously obscured videos while hurting the precision of videos with light movement obscure. Likewise, they proposed an overall haze powerful tracking plan that takes on a tweaked discriminator of DeblurGAN as an assessor to adaptively decide if the current casing be deblurred. This strategy effectively works on the precision of 6 state-of-the-art trackers. We trust our perceptions could move the investigation of the haze heartiness trackers while empowering the advancement of

continuous deblurring techniques for visual object tracking. Hui Li et al. [25] proposed a system which is based on Visual Saliency and Feature Points Matching. In this paper, propelled by human visual framework, authors presented another intrigued object tracking calculation dependent on visual saliency and element focuses coordinating. The vital thought behind the proposed calculation is that it can copy some intellectual elements of human visual framework like particular consideration. It joins the visual consideration with multi-scale highlight theory to acknowledge intrigued object discovery and tracking, and take care of the issues of conventional strategies need select object physically and the intrigued object tracking isn't sufficient stable successfully. The meanings of the model parts are portrayed exhaustively. The proposed model can be utilized for recognition, area and tracking, which are significant application angles in data processing. In this paper, author just apply the model for tracking a moving object with saliency. Exploratory outcomes show that the proposed strategy can distinguish the intrigued object viably and track the object vigorously.

III. PROBLEM IDENTIFICATION

Kai Chen et al. [9] introduced a system which is based on convolutional regression that is conventional CNN. System tracking the object by using a trained network using CNN, but this approach is bit conventional for visual tracking because it uses back-propagation method and back-propagation is a strategy to discover the contribution of each weight in the errors after a group of information is inclined and the majority of good improvement algorithms (SGD, ADAM) utilizes back-propagation to discover the angles, back-propagation has been doing as such great task but somewhat it is certainly not a productive method of learning, since it needs huge dataset. At the point when authors say translational invariance authors imply that a similar object with marginally change of direction or position probably won't start up the neuron that should perceive that object. Pooling layers is a serious mix-up on the grounds that it loses a ton of significant data and it disregards the connection between the part and the entirety. CNN's are magnificent however it has 2 exceptionally risky defects Interpretation invariance and pooling layers, fortunately author can diminish the risk with information increase yet something is coming up (capsule networks). Object detection matters in the field of object tracking because a pattern or feature can effectively analyzed for tracking as compare to the any conventional method. The proposed system uses two different methods and combining them for acquiring better precision.

IV. CONCLUSION & FUTURE SCOPE

The systems which have been proposed till now are not very much feasible for various challenges because those systems are not efficient to extract the pattern and follow that in entire frames. Most of the systems are based on Prior recognition frameworks repurpose the classifiers or localizers to perform feature extraction. It applies the model to an image at various areas even while object scaling. System required proper analysis for attaining the best level of accuracy. Visual tracking is a useful concept for

implementing video regression that may helpful for visual analysis. The aim of this study is to address a solution to one of the major causes of the inefficient accuracy and poorly trained network. Accuracy is very important with respect to the correct visual regression for ideal system. In future, a system can be implemented that pertains the best accuracy with minimal overflow or losses that works effectively. This system can be enhanced in future where accuracy depends by appending various machine learning based approaches that works effectively in computer vision.

REFERENCES

- [1] VOT Challenges, Datasets, 2015. [Online]. Available: <https://www.votchallenge.net/vot2015/dataset.html>, [Accessed: 26-Sept- 2021]
- [2] Wikipedia, Video Tracking, 28-Oct-2014. [Online]. Available: https://en.wikipedia.org/wiki/Video_tracking, [Accessed: 26- Sept- 2021]
- [3] Linkopings Universitet, Discriminative Scale Space Tracker (DSST), 26-Jun-2017. [Online]. Available: <https://www.cvl.isy.liu.se/research/objrec/visualtracking/scalvistrack/index.html>, [Accessed: 26- Sept- 2021]
- [4] A. W. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, "Visual tracking: An experimental survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1442–1468, 2014.
- [5] C. Fu, Z. Huang, Y. Li, R. Duan, and P. Lu, "Boundary effect-aware visual tracking for UAV with online enhanced background learning and multi-frame consensus verification," in *Proc. IROS*, 2019, pp. 4415– 4422.
- [6] H. Li, S. Wu, S. Huang, K. Lam and X. Xing, "Deep Motion-Appearance Convolutions for Robust Visual Tracking," in *IEEE Access*, vol. 7, pp. 180451-180466, 2019, doi: 10.1109/ACCESS.2019.2958405.
- [7] Linyu Zheng, Ming Tang, Jinqiao Wang, " Learning Robust Gaussian Process Regression for Visual Tracking," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence Main track*. Pages 1219-1225. <https://doi.org/10.24963/ijcai.2018/170>.
- [8] Martin Danelljan, Luc Van Gool, Radu Timofte; *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 7183-7192
- [9] K. Chen and W. Tao, "Convolutional Regression for Visual Tracking," in *IEEE Transactions on Image Processing*, vol. 27, no. 7, pp. 3611-3620, July 2018, doi: 10.1109/TIP.2018.2819362.
- [10] Anton Milan, S. Hamid Rezatofighi, Anthony Dick, Ian Reid, and Konrad Schindler. 2017. Online multi-target tracking using recurrent neural networks. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI'17)*. AAAI Press, 4225–4232.
- [11] S. Yun, J. Choi, Y. Yoo, K. Yun and J. Y. Choi, "Action-Driven Visual Object Tracking With Deep Reinforcement Learning," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2239-2252, June 2018, doi: 10.1109/TNNLS.2018.2801826.
- [12] Zhang, Da & Maei, Hamid & Wang, Xin & Wang, Yuan-Fang. (2017). Deep Reinforcement Learning for Visual Object Tracking in Videos.
- [13] J. F. Henriques, R. Caseiro, P. Martins and J. Batista, "High-Speed Tracking with Kernelized Correlation Filters," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583-596, 1 March 2015, doi: 10.1109/TPAMI.2014.2345390.
- [14] D. S. Bolme, J. R. Beveridge, B. A. Draper and Y. M. Lui, "Visual object tracking using adaptive correlation filters," 2010 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2544-2550, doi: 10.1109/CVPR.2010.5539960.
- [15] Dogra, Debi & Badri, Vishal & Majumdar, Arun & Sural, Shamik & Mukherjee, Jayanta & Mukherjee, Suchandra & Singh, Arun. (2014). Video analysis of Hammersmith lateral tilting examination using Kalman filter guided multi-path tracking. *Medical & biological engineering & computing*. 52. 10.1007/s11517-014-1178-2.

- [16] S. Yun, J. Choi, Y. Yoo, K. Yun and J. Y. Choi, "Action-Decision Networks for Visual Tracking with Deep Reinforcement Learning," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1349-1358, doi: 10.1109/CVPR.2017.148.
- [17] Linyu Zheng, Ming Tang, and Jinqiao Wang. 2018. Learning robust Gaussian process regression for visual tracking. In Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI'18). AAAI Press, 1219–1225.
- [18] B. Li, W. Wu, Q. Wang, F. Zhang, J. Xing and J. Yan, "SiamRPN++: Evolution of Siamese Visual Tracking With Very Deep Networks," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 4277-4286, doi: 10.1109/CVPR.2019.00441.
- [19] Xin Li, Qiao Liu, Nana Fan, Zikun Zhou, Zhenyu He, Xiao-yuan Jing, Dual-regression model for visual tracking, Neural Networks, Volume 132, 2020, Pages 364-374, ISSN 0893-6080, <https://doi.org/10.1016/j.neunet.2020.09.011>.
- [20] B. Zhang, X. Zhang and J. Qi, "Support vector regression learning based uncalibrated visual servoing control for 3D motion tracking," 2015 34th Chinese Control Conference (CCC), 2015, pp. 8208-8213, doi: 10.1109/ChiCC.2015.7260942.
- [21] T. Wang and W. Zhang, "The visual-based robust model predictive control for two-DOF video tracking system," 2016 Chinese Control and Decision Conference (CCDC), 2016, pp. 3743-3747, doi: 10.1109/CCDC.2016.7531635.
- [22] Djelal, N.; Saadia, N.; Ramdane-Cherif, A. (2012). [IEEE 2012 2nd International Conference on Communications, Computing and Control Applications (CCCA) - Marseilles, France (2012.12.6-2012.12.8)] CCCA12 - Target tracking based on SURF and image based visual servoing. , (), 1–5. doi:10.1109/ccca.2012.6417913
- [23] C. H. Li and T. I. James Tsay, "Robust Visual Tracking in Cluttered Environment Using an Active Contour Method," 2018 57th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), 2018, pp. 53-58, doi: 10.23919/SICE.2018.8492705.
- [24] Q. Guo, W. Feng, R. Gao, Y. Liu and S. Wang, "Exploring the Effects of Blur and Deblurring to Visual Object Tracking," in IEEE Transactions on Image Processing, vol. 30, pp. 1812-1824, 2021, doi: 10.1109/TIP.2020.3045630.