Comprehensive Review on Visual Tracking Systems

Ghassan Ahmad Ismaeel

College of Pharmacy, University of Mosul,
Mosul, Iraq

Date of Receiving: 10 September 2022, Date of Acceptance: 29 September 2022, Date of Publication: 01 October 2022

ABSTRACT

Visual tracking is an emerging field of study in the field of computer vision applications. In order to achieve performance perfection, researchers have recently developed a number of unique tracking methods. In this study, a number of contemporary visual tracking techniques will be analyzed and classified into four different groups: Discriminative Trackers, Generative Trackers, Correlation Filter-Based Trackers, and Combined Trackers. In addition, this paper analyzes and tabulates the methodology used by each newly presented visual tracking approach. This study is to offer the reader a thorough understanding of the many features of tracking techniques and the future direction of tracking research.

I INTRODUCTION

The eyes, which are one of the human sight sensors, is one of the crucial information trackers that we use to make focus that seizes our attention on the surrounding area. According to the literature, most of the information that the human sense is via the human eye vision, such as shapes, colors, moving, distance, and dimensions. Similar to humans, computer vision is a technology to mimics the human eye to identify and track an object. As we use images and video every day, Visual tracking become a hot topic recently as it is used for different applications in our life such as in the medical field, robotic field, agriculture, camera surveillance, and now self-driving cars, etc.

As it is a hot topic many researchers worked on the same topic and added new ideas for different video lengths and different durations and then they have been evaluated accordingly, as shown in Table 1. Where ST states for short-term, and LT states for long-term, Te states for the Test set, Tr states for the training set.

Table 1 the visual tracking works in recent years [2].

Dataset	Year	Videos	Duration
PETS [2]	2004	28	ST
VIVID [3]	2005	9	LT
OTB-50 [7]	2013	50	ST
PTB [15]	2013	100	ST
ALOV++ [16]	2013	314	ST
VOT [12]	2014-2018	25, 60, 60^1, 60^2, 60^3	ST
TC-128 [17]	2015	129	ST
UAV-123 [19]	2016	123S + 20L	ST & LT
NfS [20]	2017	100	ST
DTB-70 [21]	2017	70	ST
AMP [22]	2017	100	ST
TLP [23]	2017	50	LT
YTBB [5]	2017	380,000 Tr	ST & LT
VOT-LT [24]	2018	35	LT
TrackingNet [6]	2018	30,132 Tr + 511 Te	ST & LT
LTWB [25]	2018	366	LT

2. VISUAL TRACKING ALGORITHM AND ARCHITECTURES

Recently, new technologies for visual tracking have been investigated [1-16]. First, as an initialization process, where, the recent frame is chosen to be under the attention of the tracking system first. Later, the following posteriority of tracking frames is estimated using the appropriate model. Later on, in time, the tracking focus is getting updated using the current frame and the past ones. Figure 1, shows the process of intended visual tracking, which illustrates the input frames, motion frames, to be input to the feature observation, hence, the feature extraction. Lastly, using ensemble to synthesize the final frame prediction.

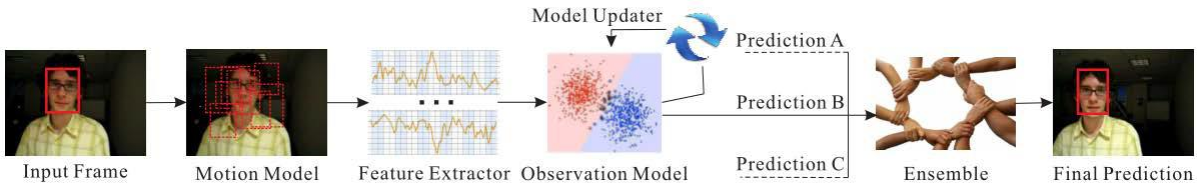

Figure 1. the schematic of the visual tracking scheme [13].

The system of visual tracking can be divided into two groups based on the method of feature extraction. The first group is productive (generative) and the other one is discriminative. Both groups are used to predict the projection of the intended frame on the surveillance scheme. We will discuss both tracking groups in detail in this section explaining with examples the difference between both.

Figure 2. visual tracking examples [14].

2.1 Discriminative Trackers

This is used with machine learning to extract features from the intended image.
In this method, the direct learning decision function is used as a prediction model in the process of learning. The discriminative method also known as a posterior probability function is established and predicts the model directly [24]. After finding the target feature, machine learning trains the classifier to characterize the target from the background image. The main architecture of the discriminant method tracking is illustrated in Figure 3.

The discriminative tracking focus on the difference between signal not on the way the data has been generated as explained in the next section. The discriminative method tracks the signal via classifying the binary signal and then putting them in categories in different groups. That is why the tracking is happening at the base of the frame-by-frame detection problem.

The main method to achieve discriminative tracking is via tracking-by-detection [25] also called dynamic detection. This method usually has two kinds: correlation filtering, which is using regression as a filter to train the input features as a target Gaussian distribution to find the maximum value that represents the location of the target. The second kind is deep learning, which is working via updating the weights forward and backward classifier filters. That can provide a strong ability to distinguish the target from the surrounding backgrounds. Other machine learning techniques that are used in discriminative tracking are listed in Table 2.

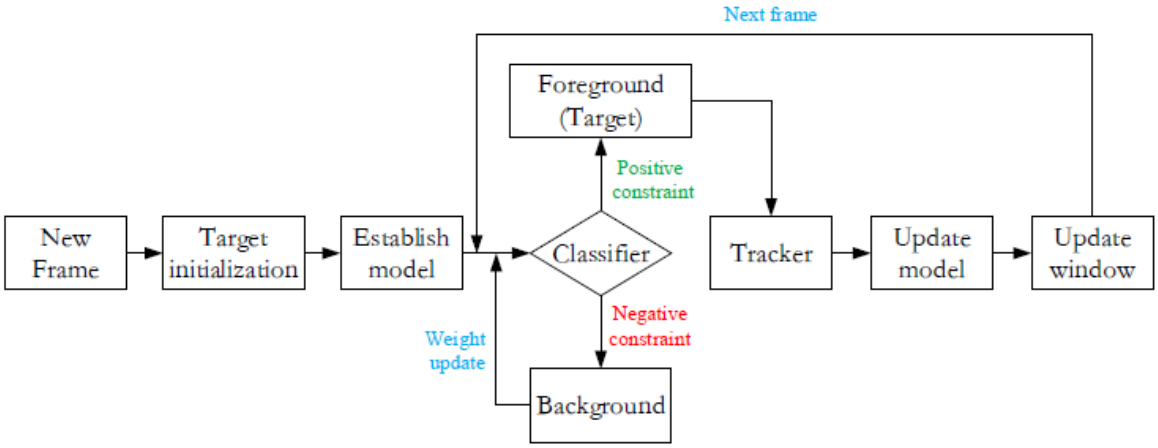

Figure 3. the discriminative tracking diagram [24].

Table 2. different methods of machine learning used in discriminative tracking.

Method	
Integrated Learning	Stacking [2],Bosting[4], Xgboost[9],decision tree[8]
Random learning	SVM[10], KNN[11], Multiline learning[8]
Deep learning	CNN[15], SAE[17], RNN[19]
Bayes Classifier	Naieve Bayes [11],Gbn [12]
Regression techniques	Linear Regression [13],Logistic Regression [14]

2.2 Generative Trackers

In the process of obtaining conditional probability distribution in visual tracking, learning is called the generative tracking method. The conditional probability distribution can be shown as:

$$P(y|x) = \frac{P(x,y)}{P(x)} \quad ..(1)$$

And it is called the probability of **y** given **x** equals the joint probability of **x** and **y** divided by the probability of **y**.

The idea of Generative tracking is to determine the way that the data has been generated.

Figure 4 shows the generative tracking method. It works via selecting the target in the required video frame as an initialization. Extract the feature of the target in the current frame. Then the generative tracking will track the probability distribution of the target and search in the next frame to find the matching region and hence the target. At the end, the target will be put in a box to keep eye on.

To achieve this tracking many researchers utilized different methods such as Gaussian mixed model [17], incremental learning [16], linear subspace [18], Bayesian network [19], Kernal trick[14], hidden Markov model [25], sparse representation [15]. As a confidence metric, the similarity function is used to reflect the reliability of each tracking accuracy.

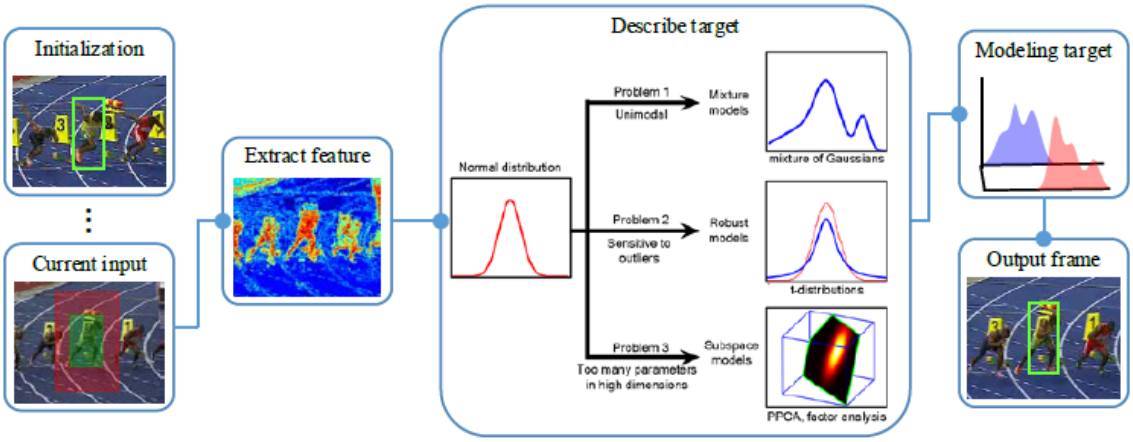

Figure 4 Generative tracking diagram [16]

2.3 Correlation Filter-Based Trackers

The correlation-based tracker correlates the windows via filtering the video with subsequence frames. Increasing the value of correlation means getting closer to the target location in the image directory. The adaptive correlation tracker introduced by Bolmein 2010 via Minimum Output Sum of Squared Error (MOSSE) filter, [7], tested variations in lighting, scale, pose, and non-rigid deformations while operating at 669 frames per second. In [7] Figure 5, shows a video tracking a face via finding a low point in MOSSE.

The principle of the Correlation filter can be described via the convolution response of correlated signal *f* and correlated signal *g* to be greater than a unit. If the signal f^* is a complex conjugate of *f:*

$$(f * g)(\tau) = \int_{-\infty}^{\infty} f^*(t)\, g(t-\tau)dt \qquad \ldots(2)$$

To reduce the computationally, the circular matrix and the fast Fourier transform (FFT) are introduced so the speed of tracking increases and the complexity of computationally decreased from $O(n^2)$ to $O(n\ log\ n)$.

Blome et al. also used synthetic exact tracking (AST) [15], to correlate over filter-based tracking (CFT). The article in [18] used MOSSE with signal channel gray feature and illustrates the high speed of 615 frames per second FPS. Later, CSK in [17] extends the MOSSE with the multichannel feature, as a version of the Kernal Correlation filter KCF. Then [15] extend that to a color frame by increasing the channel frame from 615FPS to 292FPS. After that, the speed of the tracks started to increase rapidly.

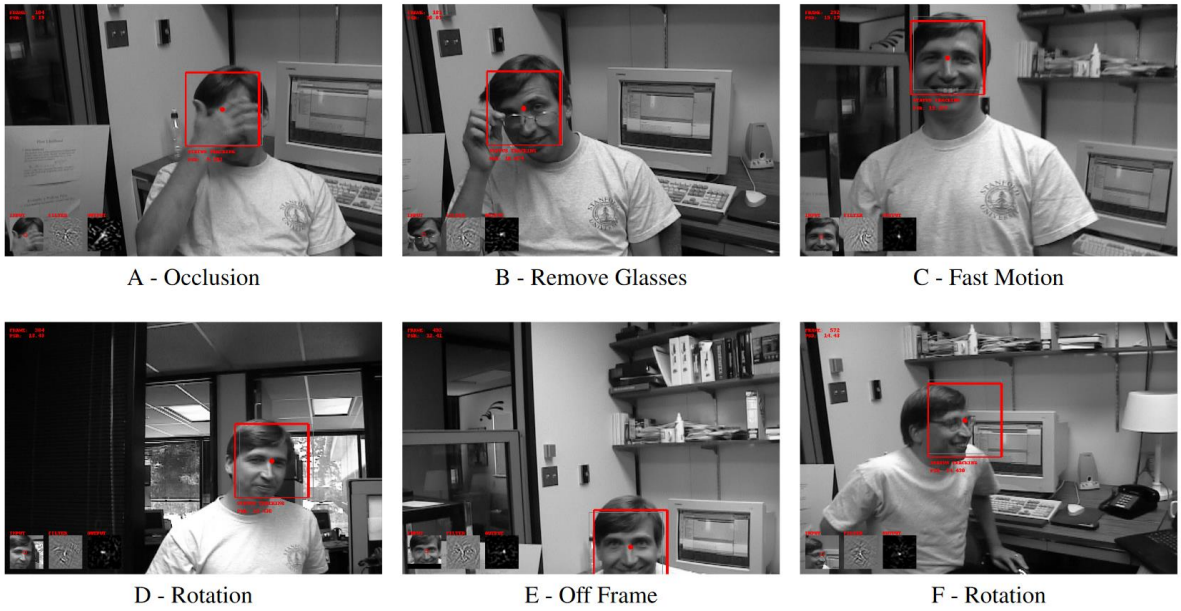

Figure 5, shows a video tracking a face via finding the low point in MOSSE [7].

Table 2 shows the comparison of frame speed of tracking

	Speed FPS	Accuracy %
MOSSE [16][7]	615	43.1
KCF [17]	172	73.2
DCF [17]	292	72.8

2.4 Trackers for Deep Learning

Recently, many researchers started to use the technique of neural networks and convolutional neural networks (CNN)[15], which provide tracking features for tracking. Later deep learning (DL) has been introduced to achieve better performance[11].

In 2015 a Korean team introduced a combination of CNN and the supportive vector machine (SVM) classifiers to observe objects by a sequential Bayesian filter.

CONCLUSION

In this review paper, we investigate different articles that study the visual tracking system using different methods of machine learning techniques. We also discuss the structural and main challenges of these methods. Four groups of visual tracking have been presented generative, discriminative, Correlation Filter-Based Trackers, and Combined Trackers. We provide explicit explanations assisted by examples. These two methods were then evaluated using

accuracy tables and hence connected to surveillance schemes. Many articles discuss that and utilized it in modern applications in our life such as automobiles and video cameras.

ACKNOWLEDGMENT

Ghassan Ahmad Ismaeel would like to thank the University of Mosul for its support.

REFERENCES

1. SM, Jainul Rinosha, and Gethsiyal Augasta. "Review of recent advances in visual tracking techniques." *Multimedia Tools and Applications* 80.16 (2021): 24185-24203.
2. Javed, Sajid, et al. "Visual object tracking with discriminative filters and Siamese networks: A survey and outlook." *arXiv preprint arXiv:2112.02838* (2021).
3. SM, Jainul Rinosha, and Gethsiyal Augasta. "Review of recent advances in visual tracking techniques." *Multimedia Tools and Applications* 80.16 (2021): 24185-24203.
4. Babenko, Boris, Ming-Hsuan Yang, and Serge Belongie. "Visual tracking with online multiple instance learning." *2009 IEEE Conference on computer vision and Pattern Recognition*. IEEE, 2009.
5. Bao, Chenglong, et al. "Real time robust l1 tracker using accelerated proximal gradient approach." *2012 ieee conference on computer vision and pattern recognition*. IEEE, 2012.
6. Bolme, David S., et al. "Visual object tracking using adaptive correlation filters." *2010 IEEE computer society conference on computer vision and pattern recognition*. IEEE, 2010.
7. Cui, Zhen, et al. "Recurrently target-attending tracking." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
8. Danelljan, Martin, et al. "Beyond correlation filters: Learning continuous convolution operators for visual tracking." *European conference on computer vision*. Springer, Cham, 2016.
9. Dendorfer, Patrick, et al. "Mot20: A benchmark for multi object tracking in crowded scenes." *arXiv preprint arXiv:2003.09003* (2020).
10. Fan, Heng, and Haibin Ling. "Sanet: Structure-aware network for visual tracking." *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2017.
11. Gao, Yuefang, et al. "Unifying temporal context and multi-feature with update-pacing framework for visual tracking." *IEEE Transactions on Circuits and Systems for Video Technology* 30.4 (2019): 1078-1091.
12. Wang, Naiyan, et al. "Understanding and diagnosing visual tracking systems." *Proceedings of the IEEE international conference on computer vision*. 2015.
13. You, Shaoze, et al. "A review of visual trackers and analysis of its application to mobile robot." *arXiv preprint arXiv:1910.09761* (2019).
14. Huang12, Dafei, et al. "Enable scale and aspect ratio adaptability in visual tracking with detection proposals." (2015).
15. Zhu, W. Q., et al. "Survey on object tracking method base on generative model." *Microprocessors* 38.1 (2017): 41-47.
16. Kristan, Matej, et al. "Hager, and et al. The visual object tracking vot2016 challenge results." *ECCV workshop*. Vol. 2. No. 6. 2016.
17. Kumar, Ashish, Gurjit Singh Walia, and Kapil Sharma. "Real-time visual tracking via multi-cue based adaptive particle filter framework." *Multimedia Tools and Applications* 79.29 (2020): 20639-20663.
18. Li, Yang, Jianke Zhu, and Steven CH Hoi. "Reliable patch trackers: Robust visual tracking by exploiting reliable patches." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.
19. Li, Feng, et al. "Learning spatial-temporal regularized correlation filters for visual tracking." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

20. Liu, Fang, and Anzhe Yang. "Application of gcForest to visual tracking using UAV image sequences." *Multimedia Tools and Applications* 78.19 (2019): 27933-27956.
21. Mei, Xue, et al. "Minimum error bounded efficient ℓ 1 tracker with occlusion detection." *CVPR 2011*. IEEE, 2011.
22. Ojala, Timo, Matti Pietikäinen, and Topi Mäenpää. "Gray scale and rotation invariant texture classification with local binary patterns." *European conference on computer vision*. Springer, Berlin, Heidelberg, 2000.
23. Tang, Fuhui, et al. "Robust visual tracking based on spatial context pyramid." *Multimedia Tools and Applications* 78.15 (2019): 21065-21084.
24. Unlu, Halil Utku, et al. "Deep learning-based visual tracking of UAVs using a PTZ camera system." *IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society*. Vol. 1. IEEE, 2019.
25. Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*. Vol. 1. Ieee, 2001.
26. Wang, Qiang, et al. "Dcfnet: Discriminant correlation filters network for visual tracking." *arXiv preprint arXiv:1704.04057* (2017).