Object Tracking Methods:A Review

Zahra Soleimanitaleb
M.S student
Malek-Ashtar University of Technology
z_soleimanitaleb@mut.ac.ir

Mohammad Ali Keyvanrad
Assistant Professor
Malek-Ashtar University of Technology
keyvanrad@mut.ac.ir

Ali Jafari
Assistant Professor
Malek-Ashtar University of Technology
iustuser@mut.ac.ir

***Abstract*— Object tracking is one of the most important tasks in computer vision that has many practical applications such as traffic monitoring, robotics, autonomous vehicle tracking, and so on. Different researches have been done in recent years, but because of different challenges such as occlusion, illumination variations, fast motion, etc. researches in this area continues. In this paper, various methods of tracking objects are examined and a comprehensive classification is presented that classified tracking methods into four main categories of feature-based, segmentation-based, estimation-based, and learning-based methods that each of which has its own sub-categories. The main focus of this paper is on learning-based methods, which are classified into three categories of generative methods, discriminative methods, and reinforcement learning. One of the sub-categories of the discriminative model is deep learning. Because of high-performance, deep learning has recently been very much considered.**

Keywords—Object tracking, Generative learning, Discriminative Learning, Reinforcement Learning, Deep Learning

I. Introduction

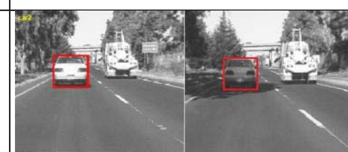
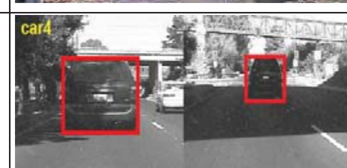
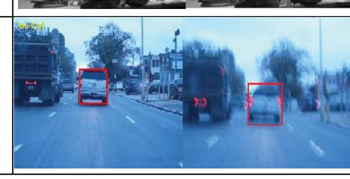
Object tracking is one of the important tasks in computer vision that tries to detect and track objects in image sequences. In object tracking, the target specifies in the first frame and must be detected and tracked in the next frames of the video.

Object tracking has different applications. Object tracking applicable in areas such as traffic monitoring (e.g. monitoring of traffic flow [1] and detection of traffic accidents [2]), robotics (e.g. ASIMO humanoid robot [3]), autonomous vehicle tracking (e.g. path-tracking [4]–[6]), medical diagnosis systems (e.g. tracking of ventricular wall [7] and medical instruments control [8]), and activity recognition (e.g. learning activity patterns [9] and human activity recognition [10]).

There are many challenges in object tracking that have led to ongoing research in this area. Some of these challenges in the OTB[1] dataset [11] are presented in TABLE I.

There are two OTB models, OTB100 and OTB50, where 100 and 50 represent the number of sequences available. The sequences in these datasets manually tag with 9 attributes, which represents the challenging aspects in visual tracking.

TABLE I
SOME OF THE OBJECT TRACKING CHALLENGES[12]

challenge	describe	example
Illumination Variation	the illumination in the target region is significantly changed	
Background Clutters	the background near the target has a similar color or texture as the target	
Low Resolution	the number of pixels inside the ground-truth bounding box is low	
Scale Variation	the ratio of the bounding boxes of the first frame and the current frame is out of the range	
Occlusion	the target is partially or fully occluded	
Change the target position	During the movement, the target may be rotated, deformed, and so on.	
Fast Motion	the motion of the ground truth is large	

[1] Object Tracking Benchmark

The aspects of tracking have serious challenges and there's some expectations from tracking system, so when designing each tracking algorithm, should try to address those expectations as much as possible. Some of these features are listed below.

Robustness: Robustness means that the tracking system can track the target even in complicated conditions such as background clutters, occlusion and illumination variation.

Adaptability: In addition to the environment changes, the target is to changes, such as the complex and sudden movement of the target. In order to solve this problem, the tracking system must be able to detect and track the current apparent characteristics of the target.

Real-time processing of information: A system that deals with image sequences should have high processing speeds. So there is a need to implement a high-performance algorithm.

II. CLASSIFICATION OF THE OBJECT TRACKING METHOD

Object tracking methods have different categories, for example, Fiaz et al. have a comprehensive study of tracking methods that categorize tracking methods into two groups of methods based on the correlation filter and the noncorrelation filter [13]. Li et al. have reviewed and compared deep learning-based tracking methods [14], Verma has reviewed methods of object detecting and tracking and categorized tracking methods into five categories of feature-based, segmentation-based, estimation-based, appearance-based and learning-based methods [15].

In this paper, the last grouping is used and as shown in Figure 1, a comprehensive classification of object tracking methods is provided and details of each one is given below.

A. Feature-Based Methods

This method is one of the simple ways of object tracking. To track objects, features, such as color, texture, optical flow, etc., are extracted first. These extracted features must be unique so that objects can be easily distinguishable in the feature space. Once the features are extracted, then the next step is to find the most similar object in the next frame, using those features, by exploiting some similarity criterion.

One of the problems with these methods is at the extraction step because the unique, exact and reliable features of the object should be extracted so that it can distinguish the target from other objects. Here are some of the features that are used for object tracking.

1) Color

Color feature can show appearance of the object. This feature can be used in different ways, one of the most common methods to use this feature is color histogram.

The color histogram shows the distribution of colors in an image, in fact, shows the different colors and the number of pixels of each color in the image. The disadvantage of color histograms is that the representation just depends on color of object and ignore shape and texture of object , so two different objects may have same histogram.

There are some papers that use color histogram for tracking such as[16]–[18].

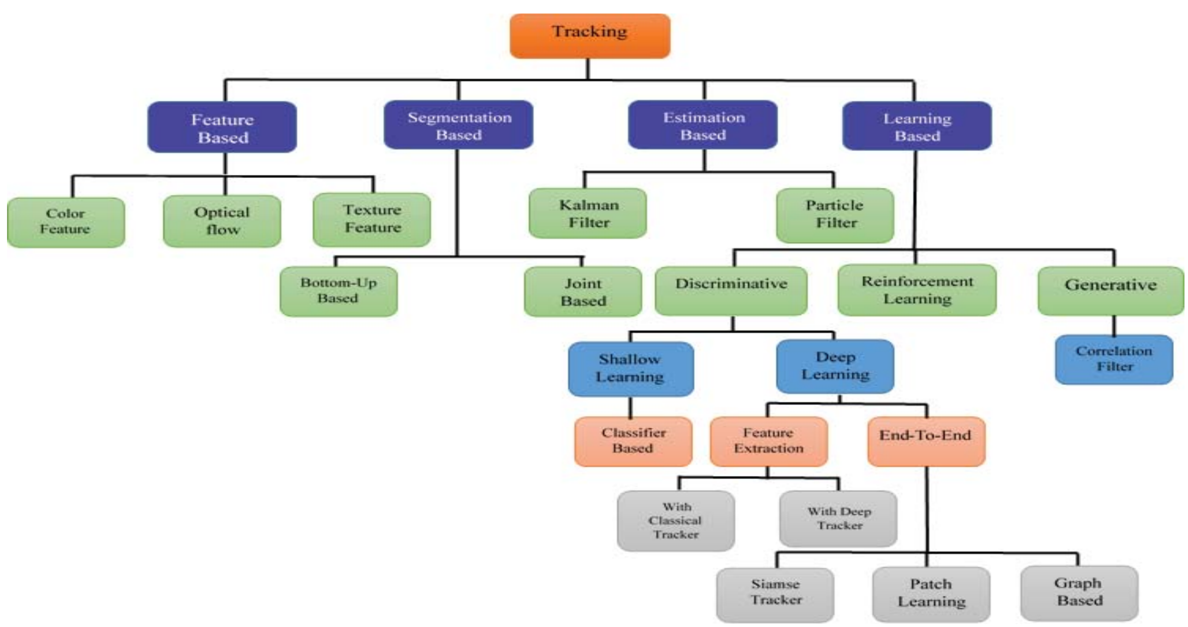


Figure 1: Classification of object tracking methods

2) Texture

Texture is a repeated pattern of information or arrangement of the structure with regular intervals. Texture features are not obtained directly. These features are generated by image preprocessing techniques.

The texture feature is an important feature in the image, it can be used along with the color feature to describe the contents of an image or a region of the image. Because the color feature is not sufficient to identify the similar objects and sometimes it can be seen that different images have the same histogram.

Gabor wavelet [19] is one of the most studied texture features. The most important property of Gabor filters is their invariance to illumination, rotation, scale, and translation which makes it suitable for object tracking. A method to detect the location and body shapes of moving animals using a Gabor filter is presented in [20]. The local binary pattern is another textual descriptor [21]. Zhao et al. have used LBP to describe moving objects and also have used a Kalman filter for target tracking [22].

3) Optical Flow

Optical flow is the apparent motion of brightness patterns in the image. apparent motion can be caused by lighting changes without any actual motion. The optical flow algorithm calculates the displacement of brightness patterns from one frame to another. Algorithms that calculate displacement for all image pixels are known as dense optical streaming algorithms, while light-flow algorithms estimate displacement tension for a selective number of pixels in an image [23].

There are some papers that use optical flow for tracking such as[24], [25].

B. Segmentation Based Methods

Segmenting foreground objects from a video frame is fundamental and the most critical step in visual tracking. Foreground Segmentation is done to separate foreground objects from the background scene. Normally, the foreground objects are the objects that are moving in a scene. To track these objects, they are to be separated from the background scene [15]. In the following, some of the object tracking methods based on the segmentation are examined.

1) Bottom-Up Based Method

In this type of tracking, there must be two separate tasks, first the foreground segmentation and then the object tracking.

The foreground segmentation uses a low-level segmentation to extract regions in all frames, and then some features are extracted from the foreground regions and tracked according to those features[26], as shown in Figure 2.

There are some papers that use a bottom-up based method for tracking such as [27]–[31].

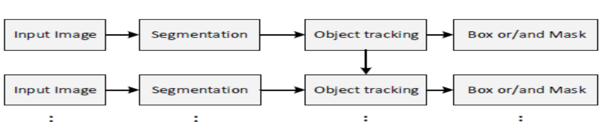


Figure 2: Bottom-Up based framework [26]

2) Joint Based Method

In bottom-up method, foreground segmentation and tracking are two separate tasks; one of the problems with this method is that the segmentation error propagates forward, causing error tracking. To solve this problem, as shown in Figure 3, the researchers merged the foreground segmentation and tracking method, witch improved tracking performance.

There are some papers that use joint based method for tracking such as [32]–[34].

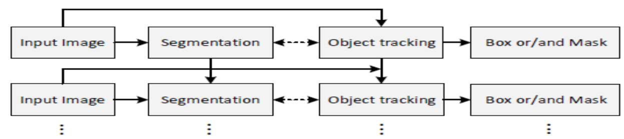


Figure 3: Joint based framework [26]

C. Estimation Based Methods

Estimation methods formulate the tracking problem to an estimation problem in which an object is represented by a state vector. The state vector describes the dynamic behavior of a system, such as the position and velocity of an object. The general framework for the dynamic mode estimation problem is taken from Bayesian methods [15].

The Bayesian filters allow the target to continuously update its position on the coordinate system, based on the latest sensor data. This algorithm is recursive and consists of two steps: prediction and updating.

The prediction step estimates the new position of the target in the next step using the state model, while the updating step uses the current observation to update the target position using the observation model. The prediction and updating steps are performed on each frame of the video. Here are some examples of this method.

1) Kalman Filter

To use the Kalman filter [35] in object tracking, should design a dynamic model of the target movement. The Kalman filter is used to estimate the position of a linear system assumed that the errors are Gaussian. In many cases, dynamic models are nonlinear, so in this case, the Kalman filter is not used and other suitable algorithms are used. One of these algorithms is the extended Kalman Filter [36]. The framework for using the Kalman filter is shown in Figure 4.

There are some papers that use kalman filter for tracking such as[22], [37].

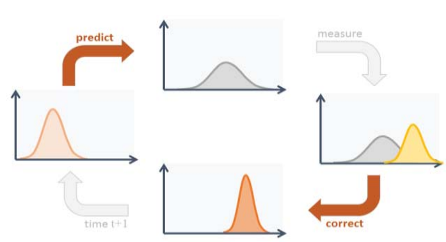


Figure 4: Kalman filter framework [38]

2) Particle Filter

Most tracking issues are non-linear. Therefore, particle filter has been considered for solving such problems. The particle filter is a recursive Monte Carlo statistical calculation method that is often used for non-gaussian noise measurement models. The main idea of the particle filter shows the distribution of a set of particles. Each particle has a probability weight, which represents the probability of sampling that particle from the probability density function. One of the problems with this method is that the particles that have more probability to be selected several times, and resampling is used to solve this problem. The framework for using particle filters is shown in Figure 5.

Particle filters are also used in [39]–[41].

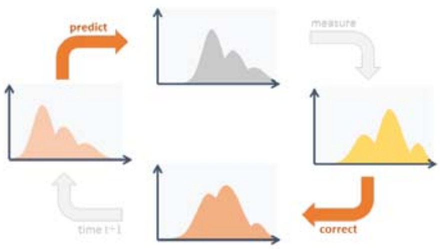


Figure 5: Particle filter framework [42]

D. Learning-Based Methods

In learning based methods, the features and appearance of different targets and their prediction are learned in next frames, and then in the test time, based on what they learned, they can detect the object and track it in next frames. Learning based methods are often divided into three types of generative, discriminative and reinforcements learning.

1) Discriminative Methods

Discriminative trackers usually consider tracking as a classification problem that discriminates the target from the background. Discriminative learning is divided into two categories contains Shallow learning and Deep learning.

a) Shallow Learning

Object tracking can be considered as a decision-making process where classification is trained to learn the discrimination between the target and the background. After the training, at the test time, it decides that the object is target or no. Features are extracted from different objects and the various methods such as support vector machines [43]–[46] can be used for classification.

b) Deep Learning

Shallow learning with fewer layers predicts the model, but deep learning has too many layers. Another difference is that shallow learning requires important and discriminatory features extracted by professionals, but deep learning itself extracts these important features. Deep learning [47] has made impressive developments in various areas, including computer vision, in recent years. One of the areas where deep learning has affected it and increased its accuracy is object tracking. Due to this increasing accuracy, we will focus on deep learning methods which is explained below. In this paper, deep learning methods are classified into feature extraction based and end-to-end methods.

i. Feature extraction Based Methods

Wang et al. [48] show that feature extraction is also a very important issue in the design of a robust tracker. From the research experience of classical tracking algorithms, it can be concluded that any development in the methods of the feature extracting or machine learning techniques may lead to the development of tracking. Therefore, since deep learning techniques have shown great abilities in feature extraction and object classification, it can be concluded that the use of deep learning can improve tracking performance. Due to the success of the deep features in image classification, some of the deep-network tracking methods are used to extract features that are known as a feature extraction network.

These methods separate the detection and tracking sections. Detection uses deep learning methods that can extract deep features, but for tracking, these methods can be classified into two categories of tracking with classical methods and tracking with Deep methods.

Yang et al have used Faster RCNN, KCF[2] and Kalman filter trackers [49]. Chahyati et al. have used Faster RCNN and the Siamese network [50]. Agarwal and Suryavanshi have used Faster RCNN and GOTURN [51]. Ghaemmaghami has used SSD, YOLO, and LSTM, and an example of the architecture of this method is shown in Figure 6.

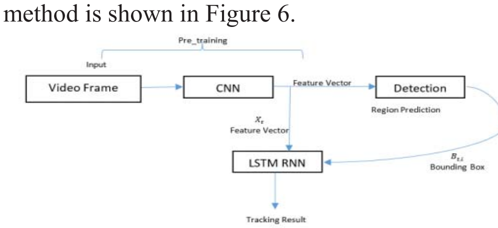


Figure 6: Feature extraction based tracking example [52]

ii. End-To-End Methods

End-to-end methods train a network to conduct both feature extraction and candidate evaluation. In this paper, end-to-end methods are classified into three categories of Siamese trackers, patch learning and graph-based trackers.

- *Siamese tracker*

Siamese networks have two inputs and produce one output, it captures two inputs and measures the similarity between the two images to determine whether or not the same objects exist in the two images.

These types of networks are capable of learning similarities and common features. There are some papers that use the siamese tracker method for tracking such as [53]–[56]. An example of tracker architecture is shown in Figure 7.

[2] Kernelized Correlation Filter

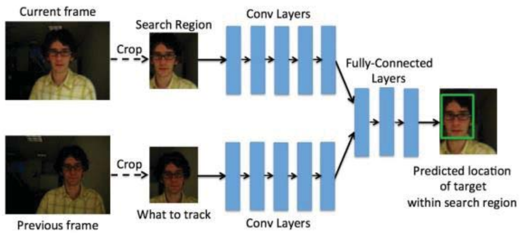


Figure 7: Siamese tracker example [53]

❖ *Patch learning tracker*

In patch learning method positive and negative samples are extracted, then train the model on these samples. The model is tested on the number of samples, and the maximum response indicates the target position. There are some papers that use patch learning method for tracking such as [57]–[59]. An example of this method shown in Figure 8.

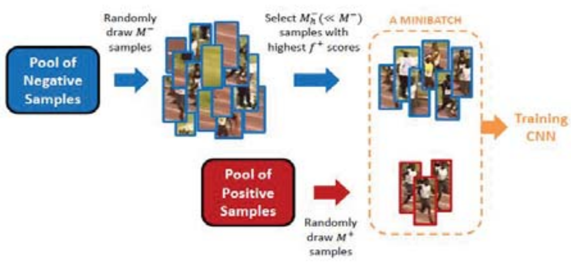


Figure 8: Learning patch tracker example [60]

❖ *Graph-based tracker*

In computer vision, graph theory has been successfully applied to solve many tasks. Graphs offer a rich and compact representation of the spatial relations between objects in an image and they can be used to model many types of relations and processes.

There are some papers that use graph-based methods for tracking such as [61]–[63]. An example of this method shown in Figure 9.

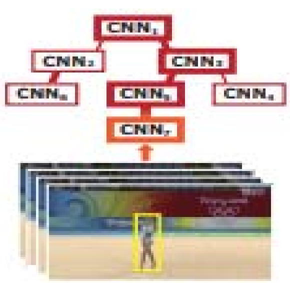


Figure 9: Graph-based tracker example[61]

2) Generative Methods

These methods focus on searching in areas that are more similar to object. Correlation Filter based methods are one of the examples of these trackers.

The main idea of the correlation filter is to estimate an optimal image filter so that this filter produces optimal output on the input image.

In the first frame, the target is identified by a bounding box and the correlation filter is trained on it, then at each time step, the patch is cropping to its predicted position for tracking. Then, as shown in Figure 10, different features can be extracted from the input data. A cosine window is usually used to smooth boundry effects.

Then, the correlation between the current input and the frequency filter learned in the frequency domain is made based on the Convolution theorem. After the correlation, a spatial confidence map is obtained by Inverse FFT (IFFT), The maximum value of that can be predicted as the new position of target. Then, appearance at the newly estimated position is extracted for training and updating the correlation filter with a output.

There are some papers that use the correlation filter method for tracking such as [64]–[66].

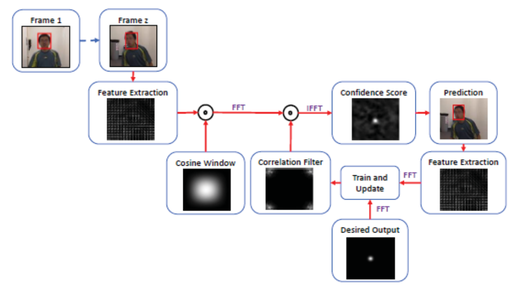


Figure 10: Correlation filter tracker framework [67]

3) Reinforcement Learning

In a reinforcement learning[68] problem, we encounter a agent that interacts with the environment through trial and error and learns to select the optimal action to achieve the goal.

There are some papers that use a reinforcement learning method for tracking such as [69]–[71]. An example of this method shown in Figure 11.

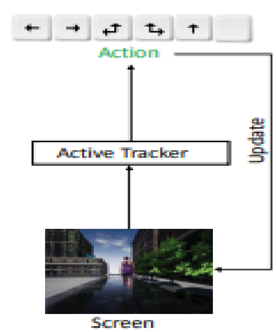


Figure 11: Reinforcement learning tracker example [72]

III. Conclusion

In this paper, a comprehensive classification of object tracking algorithms is presented. In this category, tracking algorithms are divided into feature-based, segmentation-based, estimation-based and learning-based categories. This paper focuses on learning-based approaches. Learning-based tracking algorithms, and especially deep learning, have recently received much attention. Deep learning is a new and exciting field in various fields, especially computer vision, and in many fields it is very accurate and has made a lot of progress.

Deep learning networks have many layers and extract different features of the object in each layer so it can have good accuracy in object tracking as well because it has many layers and model complexity is slower than shallow networks. However, it can't be said that deep learning works best in all cases and always should be used, but by knowing the advantages and disadvantages of all methods one can find out which method can work best in problem solving.

IV. References

[1] B. Tian, Q. Yao, Y. Gu, K. Wang, and Y. Li, "Video processing techniques for traffic flow monitoring: A survey," in *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2011, pp. 1103–1108.

[2] J.-C. Tai, S.-T. Tseng, C.-P. Lin, and K.-T. Song, "Real-time image tracking for automatic traffic monitoring and enforcement applications," *Image and Vision Computing*, vol. 22, no. 6, pp. 485–501, 2004.

[3] Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki, and K. Fujimura, "The intelligent ASIMO: System overview and integration," in *IEEE/RSJ international conference on intelligent robots and systems*, 2002, vol. 3, pp. 2478–2483.

[4] M. Brown, J. Funke, S. Erlien, and J. C. Gerdes, "Safe driving envelopes for path tracking in autonomous vehicles," *Control Engineering Practice*, vol. 61, pp. 307–316, 2017.

[5] M. Menze and A. Geiger, "Object scene flow for autonomous vehicles," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3061–3070.

[6] V. A. Laurense, J. Y. Goh, and J. C. Gerdes, "Path-tracking for autonomous vehicles at the limit of friction," in *2017 American Control Conference (ACC)*, 2017, pp. 5586–5591.

[7] A. C. F. Colchester and D. J. Hawkes, Eds., *Information Processing in Medical Imaging: 12th International Conference, IPMI '91, Wye, UK, July 7-12, 1991. Proceedings*. Berlin Heidelberg: Springer-Verlag, 1991.

[8] S. Walker *et al.*, "Systems and methods for localizing, tracking and/or controlling medical instruments," Apr-2017.

[9] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747–757, Aug. 2000.

[10] R. Bodor, B. Jackson, and N. Papanikolopoulos, "Vision-based human tracking and activity recognition," in *Proc. of the 11th Mediterranean Conf. on Control and Automation*, 2003, vol. 1.

[11] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 2411–2418.

[12] "Visual Tracker Benchmark." [Online]. Available: http://cvlab.hanyang.ac.kr/tracker_benchmark/datasets.html. [Accessed: 11-Jun-2019].

[13] M. Fiaz, A. Mahmood, and S. K. Jung, "Tracking Noisy Targets: A Review of Recent Object Tracking Approaches," *arXiv:1802.03098 [cs]*, Feb. 2018.

[14] P. Li, D. Wang, L. Wang, and H. Lu, "Deep visual tracking: Review and experimental comparison," *Pattern Recognition*, vol. 76, pp. 323–338, 2018.

[15] R. Verma, "A Review of Object Detection and Tracking Methods," *International Journal of Advance Engineering and Research Development*, vol. 4, pp. 569–578, Oct. 2017.

[16] X. Li and N. Zheng, "Adaptive target color model updating for visual tracking using particle filter," in *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*, 2004, vol. 4, pp. 3105–3109.

[17] M. Fotouhi, A. R. Gholami, and S. Kasaei, "Particle filter-based object tracking using adaptive histogram," in *2011 7th Iranian Conference on Machine Vision and Image Processing*, 2011, pp. 1–5.

[18] A. Zhao, "Robust histogram-based object tracking in image sequences," in *9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA 2007)*, 2007, pp. 45–52.

[19] D. Barina, "Gabor Wavelets in Image Processing," *arXiv:1602.03308 [cs]*, Feb. 2016.

[20] D. A. Wagenaar and W. B. Kristan, "Automated video analysis of animal movements using Gabor orientation filters," *Neuroinformatics*, vol. 8, no. 1, pp. 33–42, 2010.

[21] T. Ojala, M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions," in *Proceedings of 12th International Conference on Pattern Recognition*, 1994, vol. 1, pp. 582–585.

[22] Z. Zhao, S. Yu, X. Wu, C. Wang, and Y. Xu, "A multi-target tracking algorithm using texture for real-time surveillance," in *2008 IEEE International Conference on Robotics and Biomimetics*, 2009, pp. 2150–2155.

[23] J. L. Barron, D. J. Fleet, S. S. Beauchemin, and T. A. Burkitt, "Performance of optical flow techniques," in *Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1992, pp. 236–242.

[24] J. Hariyono, V.-D. Hoang, and K.-H. Jo, "Moving object localization using optical flow for pedestrian detection from a moving vehicle," *The Scientific World Journal*, vol. 2014, 2014.

[25] D.-S. Kim and J. Kwon, "Moving object detection on a vehicle mounted back-up camera," *Sensors*, vol. 16, no. 1, p. 23, 2016.

[26] R. Yao, G. Lin, S. Xia, J. Zhao, and Y. Zhou, "Video Object Segmentation and Tracking: A Survey," *arXiv preprint arXiv:1904.09172*, 2019.

[27] J. Son, I. Jung, K. Park, and B. Han, "Tracking-by-segmentation with online gradient boosting decision tree," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3056–3064.

[28] A. Adam, E. Rivlin, and I. Shimshoni, "Robust fragments-based tracking using the integral histogram," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2006, vol. 1, pp. 798–805.

[29] T. Zhou and D. Tao, "Shifted subspaces tracking on sparse outlier for motion segmentation," in *Twenty-Third International Joint Conference on Artificial Intelligence*, 2013.

[30] Z. Cai, L. Wen, Z. Lei, N. Vasconcelos, and S. Z. Li, "Robust deformable and occluded object tracking with dynamic graph," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5497–5509, 2014.

[31] C. Gu and M.-C. Lee, "Semiautomatic segmentation and tracking of semantic video objects," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 5, pp. 572–584, 1998.

[32] C. Bibby and I. Reid, "Robust real-time visual tracking using pixel-wise posteriors," in *European Conference on Computer Vision*, 2008, pp. 831–844.

[33] A. Milan, L. Leal-Taixé, K. Schindler, and I. Reid, "Joint tracking and segmentation of multiple targets," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5397–5406.

[34] F. Schubert, D. Casaburo, D. Dickmanns, and V. Belagiannis, "Revisiting robust visual tracking using pixel-wise posteriors," in *International Conference on Computer Vision Systems*, 2015, pp. 275–288.

[35] N. Najafzadeh, M. Fotouhi, and S. Kasaei, "Object tracking using Kalman filter with adaptive sampled histogram," in *2015 23rd Iranian Conference on Electrical Engineering*, 2015, pp. 781–786.

[36] "A Constrained Extended Kalman Filter for Target Tracking - Google Search." [Online]. Available: https://www.google.com/search?q=A+Constrained+Extended+Kalman+Filter+for+Target+Tracking&oq=A+Constrained+Extended+Kalman+Filter+for+Target+Tracking&aqs=chrome..69i57.380j0j7&sourceid=chrome&ie=UTF-8. [Accessed: 13-Jun-2019].

[37] P. R. Gunjal, B. R. Gunjal, H. A. Shinde, S. M. Vanam, and S. S. Aher, "Moving Object Tracking Using Kalman Filter," in *2018 International Conference On Advances in Communication and Computing Technology (ICACCT)*, 2018, pp. 544–547.

[38] "Object Tracking: Kalman Filter with Ease." [Online]. Available: /Articles/865935/Object-Tracking-Kalman-Filter-with-Ease. [Accessed: 14-Jun-2019].

[39] L. Mihaylova, P. Brasnett, N. Canagarajah, and D. Bull, "Object tracking by particle filtering techniques in video sequences," *Advances and challenges in multisensor data and information processing*, vol. 8, pp. 260–268, 2007.

[40] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet, "Color-based probabilistic tracking," in *European Conference on Computer Vision*, 2002, pp. 661–675.

[41] Y. Li, H. Ai, T. Yamashita, S. Lao, and M. Kawade, "Tracking in low frame rate video: A cascade particle filter with discriminative observers of different life spans," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1728–1740, 2008.

[42] "Object Tracking: Particle Filter with Ease." [Online]. Available: /Articles/865934/Object-Tracking-Particle-Filter-with-Ease. [Accessed: 14-Jun-2019].

[43] S. Avidan, "Support vector tracking," *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 8, pp. 1064–1072, 2004.

[44] M. Tian, W. Zhang, and F. Liu, "On-line ensemble SVM for robust object tracking," in *Asian conference on computer vision*, 2007, pp. 355–364.

[45] T. Malisiewicz, A. Gupta, and A. A. Efros, "Ensemble of exemplar-SVMs for object detection and beyond.," in *Iccv*, 2011, vol. 1, p. 6.

[46] S. Zhang, X. Yu, Y. Sui, S. Zhao, and L. Zhang, "Object tracking with multi-view support vector machines," *IEEE Transactions on Multimedia*, vol. 17, no. 3, pp. 265–278, 2015.

[47] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[48] N. Wang, J. Shi, D.-Y. Yeung, and J. Jia, "Understanding and diagnosing visual tracking systems," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3101–3109.

[49] F. Bu, Y. Cai, and Y. Yang, "Multiple Object Tracking Based on Faster-RCNN Detector and KCF Tracker."

[50] D. Chahyati, M. I. Fanany, and A. M. Arymurthy, "Tracking People by Detection Using CNN Features," *Procedia Computer Science*, vol. 124, pp. 167–172, 2017.

[51] A. Agarwal and S. Suryavanshi, "Real-Time* Multiple Object Tracking (MOT) for Autonomous Navigation," Technical report, 2017.

[52] M. P. Ghaemmaghami, "Tracking of Humans in Video Stream Using LSTM Recurrent Neural Net- work," p. 50.

[53] D. Held, S. Thrun, and S. Savarese, "Learning to track at 100 fps with deep regression networks," in *European Conference on Computer Vision*, 2016, pp. 749–765.

[54] R. Tao, E. Gavves, and A. W. Smeulders, "Siamese instance search for tracking," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1420–1429.

[55] X. Wang, C. Li, B. Luo, and J. Tang, "Sint++: robust visual tracking via adversarial positive instance generation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4864–4873.

[56] K. Chen and W. Tao, "Once for all: a two-flow convolutional neural network for visual tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 12, pp. 3377–3386, 2017.

[57] H. Nam and B. Han, "Learning multi-domain convolutional neural networks for visual tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4293–4302.

[58] I. Jung, J. Son, M. Baek, and B. Han, "Real-time mdnet," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 83–98.

[59] B. Han, J. Sim, and H. Adam, "Branchout: Regularization for online ensemble tracking with convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3356–3365.

[60] "Learning Multi-Domain Convolutional Neural Networks for Visual Tracking arXiv : [cs.CV] v1 2015, v Hyeonseob Nam, Bohyung Han Dept. - ppt video online download." [Online]. Available: https://slideplayer.com/slide/11571058/. [Accessed: 19-Jun-2019].

[61] H. Nam, M. Baek, and B. Han, "Modeling and propagating cnns in a tree structure for visual tracking," *arXiv preprint arXiv:1608.07242*, 2016.

[62] T. Wang and H. Ling, "Gracker: A graph-based planar object tracker," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 6, pp. 1494–1501, 2017.

[63] C. Gomila and F. Meyer, "Graph-based object tracking," in *Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429)*, Barcelona, Spain, 2003, vol. 3, pp. II-41–4.

[64] M. Danelljan, A. Robinson, F. S. Khan, and M. Felsberg, "Beyond correlation filters: Learning continuous convolution operators for visual tracking," in *European Conference on Computer Vision*, 2016, pp. 472–488.

[65] M. Danelljan, G. Hager, F. Shahbaz Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 4310–4318.

[66] M. Danelljan, G. Bhat, F. S. Khan, and M. Felsberg, "ECO: Efficient Convolution Operators for Tracking.," in *CVPR*, 2017, vol. 1, p. 3.

[67] Z. Chen, Z. Hong, and D. Tao, "An Experimental Survey on Correlation Filter-based Tracking," Sep. 2015.

[68] M. E. Harmon and S. S. Harmon, "Reinforcement Learning: A Tutorial.," WRIGHT LAB WRIGHT-PATTERSON AFB OH, 1997.

[69] J. Choi, J. Kwon, and K. M. Lee, "Visual tracking by reinforced decision making," *arXiv preprint arXiv:1702.06291*, 2017.

[70] W. Luo, P. Sun, F. Zhong, W. Liu, T. Zhang, and Y. Wang, "End-to-end active object tracking via reinforcement learning," *arXiv preprint arXiv:1705.10561*, 2017.

[71] S. Yun, J. Choi, Y. Yoo, K. Yun, and J. Young Choi, "Action-decision networks for visual tracking with deep reinforcement learning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2711–2720.

[72] D. Zhang, H. Maei, X. Wang, and Y.-F. Wang, "Deep reinforcement learning for visual object tracking in videos," *arXiv preprint arXiv:1701.08936*, 2017.