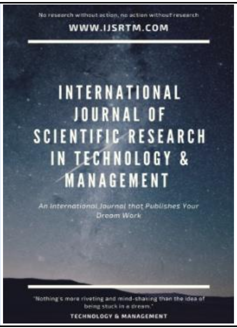
https://www.ijsrtm.com
Vol.4 Issue 1 Mar 2024: 17-20
Published online 11 March 2024

E-ISSN: 2583-7141

International Journal of Scientific Research in Technology & Management



A Review of Visual Trackers & its Applications over the Real World

Amit Saxena
Dept. of Computer Science & Engineering
Rabindranath Tagore University
Bhopal, Madhya Pradesh, India
amit.saxena78@gmail.com

Sitesh Kumar Sinha
Dept. of Computer Science & Engineering
Rabindranath Tagore University
Bhopal, Madhya Pradesh, India
siteshkumarsinha@gmail.com

Sanjeev Kumar Gupta
Dept. of Computer Science & Engineering
Rabindranath Tagore University
Bhopal, Madhya Pradesh, India
sanjeevgupta73@yahoo.com

***Abstract*— Visual tracking is one of the hot topics and is widely emerging research in the field of computer vision applications. Over the years, the application of visual tracking has been gradually expanded with the continuous development of technology. Researchers have developed various novel tracking methodologies to improve the performance. Although various approaches has been proposed robust visual tracking remains a great challenge.In this paper various visual tracking methods are surveyed and classified into major classifications and also future trends are identified.**

***Keywords*— Moving Object Detection, Object Tracking, CNN, R-CNN, SSD, YOLO, Pattern Recognition.**

I. Introduction

Visual tracking is a visual processing skill that occurs when the eyes focus on an object as it moves across the field of vision. Visual tracking occurs with movement of the eyes to follow a moving object and not movement of the head. The eyes have the ability to track an object in the vertical and horizontal, diagonal, and circular planes. Visual tracking is an important task within the field of computer vision. The proliferation of high-end computers, the availability of high quality video cameras, and the increasing need for automated video analysis have generated a great deal of interest in visual tracking algorithms. Visual tracking, in general, is a very challenging problem due to the loss of information caused by the projection of the 3D world on a 2D image, noise in images, cluttered-background, complex object motion, partial or full occlusions, illumination changes as well as real-time processing requirements, etc. In the early years, almost all visual tracking methods assumed that the object motion was smooth and no abrupt appearance change. However, tremendous progress has been made in recent years. Some algorithms can deal with the problems of abrupt appearance change, leaving out from scenes and drifting, etc.

II. Related Works

There are different approaches proposed by various researchers that having certain flaws. The motive is to deep analyze the researches related to the moving object detection along with object tracking. In the discipline of computer vision, object detection is a fundamental study area. For more difficult computer vision tasks like target tracking, pattern recognition, and semantic comprehension, it serves as a crucial precursor. It seeks to properly identify the category, locate the target of interest inside the image, and provide the bounding box of each target. Small data size, low portability, lack of pertinence, high temporal complexity, and lack of robustness only in certain basic situations are the primary drawbacks of this technology.

III. Related Work

Viola and Jones [1] have described a machine learning approach for visual object detection which is capable of processing images very rapidly and achieving high detection rates. Their proposed method is distinguished by the introduction of a new image representation called the “Integral linage” which is nothing but the features and a learning algorithm, based on AdaBoost that selects a small number of critical visual features from a larger set and yields extremely efficient classifier. This trained model distinguishes the interested object from background. Yao et al., [2] have modelled the unknown parts using latent variables. In doing so they extend the online algorithm named pegasos to the structured prediction case (i.e., predicting the location of the bounding boxes) with latent part variables. Nam and Han have proposed the method which pre-trains an algorithm with CNN using a large set of videos with tracking ground truths to obtain a generic target representation [3]. The proposed method is composed of shared layers and multiple branches of domain specific layers, where domains correspond to individual training sequences and each branch is responsible for binary classification to identify target in each domain. A simple yet efficient Dual Linear Structured Support Vector Machine (DLSSVM) algorithm was proposed by Ning et al., [4] to enable fast learning and execution during tracking. By analysingthe dual variables, they have proposed a primal classifier update formula where the learning step size is computed in closed form. A framework for adaptive visual object tracking based on structured output prediction was proposed by Hare et al., [5] by explicitly allowing the output space to express the needs of the tracker, which is able to avoid the need for an intermediate classification step. Liu et al., [6] have proposed an efficient L1 tracker with minimum error bound and occlusion detection which is named as Bounded Particle Resampling (BPR)-L1 tracker. This method, initially calculates the minimum error bound from a linear least square equation, and serves as a guide for particle resampling in a particle filter framework. Secondly, it performs the occlusion detection by investigating the trivial coefficients in the minimization. Xue et al., [7] have proposed a two-stage sparse optimization, which jointly maximize the discriminative power and minimize the target reconstruction error.. As the target template and discriminative features usually have temporal and spatial relationship, dynamic group sparsity (DGS) is utilized in their algorithm.Correlation filters for long-term visual object tracking was proposed by Montero et al., [8] with great interest, in this they have presented a fast scalable solution based on the Kernalized Correlation Filter (KCF) framework. They have introduced an adjustable Gaussian window function and a key-point based model for scale estimation to deal with the fixed size limitation in the Kernalized Correlation Filter. Furthermore, they have integrated the fast HoG descriptors and Intel's Complex Conjugate Symmetric (CCS) packed format to boost the achievable frame rates. Elman et al., [9] have proposed hidden unit patterns which can be fed back to themselves; the internal representations which develop thus reflect task demands in the context of prior internal states. The networks are able to learn interesting internal representations which incorporate task demands with memory demands: indeed, in their approach the notion of memory is inextricably bound up with task processing. Tokola et al., [10] have introduced a new method for jointly learning an ensemble of correlation filters that collectively captures as much variation in object appearance as possible. During training, the filters adapts the needs of the training data with no restrictions on size or scope. A novel method named as sequential binary code selection (SBC) [11] has been proposed to learn a set of compact binary codes for image patch representation. Using the sparse projections, the high dimensional feature can be rapidly embedded into the compact binary codes with preserving both the label information and geometrical distance. A tracking algorithm [12] that consists of a deep convolution neural network (DCNN)-based detection module and a probabilistic-model-based tracking module is implemented for the efficient and robust astronaut visual tracking. The authors have improved the DCNN in the detection module through optimizations of light weight network architecture design, parameters model compression and inference accelerate.

A. Categories of tracker

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.

B. Discriminative tracker

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 [13]. After finding the target feature, machine learning trains the classifier to characterize the target from the background image. 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.

C. Generative Tracker

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: $(y|x)$ = $(x,)$ (x) ..(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.

D. Correlation Filter based tracker

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, [14], tested variations in lighting, scale, pose, and non-rigid deformations while operating at 669 frames per second. 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) (τ)= $\int f*(t)\ g(t-\tau)dt\ \infty\ -\infty$ …(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 $O(nlogn)$.

E. *Combined trackers*

The combination of multiple object trackers is a further investigation path to improve the tracking efficiency in all sorts of scenarios. Recently, deep learning architectures are customized with layers or objective functions for tracking feature discrimination. A correlation filter can be combined with deep architecture to incorporate the cross domain advantages into a single specific objective. Zhong et al., [15] have proposed a robust object tracking algorithm using a collaborative model to develop a Sparsity-based Discriminative Classifier (SDC) and a Sparsity-based Generative Model (SGM). In the S-DC module, the authors have introduced an effective method to compute the confidence value that assigns more weights to the foreground than the background. Zhang et al., [16] have proposed a multi-expert restoration scheme to address the model drift problem in online tracking

F. *Trackers for Deep learning*

Recently, many researchers started to use the technique of neural networks and convolutional neural networks (CNN)[17], which provide tracking features for tracking. Later deep learning (DL) has been introduced to achieve better performance[18]. 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.

IV. CONCLUSION

Visual tracking as an important component of Computer Vision with many applications which makes it a highly attractive research problem. In this paper various visual tracking methods are classified into major classifications. The history of development and current research status of the tracking algorithm are introduced in time sequence, and the advantages and disadvantages are analyzed. Although the generative method framework has the advantages of good real-time performance and less adjustment parameters, its modeling complexity limits its further development. With the development of correlation filter and deep learning, discriminative method algorithm based on Tracking-by-Detection architecture has become the mainstream. Finally, the use of specific environmental information is also an important research direction. Such as vehicle tracking, cars should be kept on the road, not on the sky or on the wall. This kind of semantic or environmental information is also very useful for the development of trackers.

REFERENCES

[1]. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features, in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. I– 511–I–518.

[2]. Yao R, Shi Q, Shen C, Zhang Y, van den Hengel A (2013) Part-based visual tracking with online latent structural learning, in IEEE Conference on Computer Vision and Pattern Recognition, pp. 2363–2370

[3]. H. Nam and B. Han (2016) Learning multi-domain convolutional neural networks for visual tracking, in The IEEE Conference on CVPR.

[4]. Ning J, Yang J, Jiang S, Zhang L, Yang MH (2016) Object tracking via dual linear structured svm and explicit feature map, in The IEEE Conference on Computer Vision and Pattern Recognition –CVPR

[5]. Hare S, Saffari A, Torr P (2011) Struck: Structured output tracking with kernels, in ICCV, pp. 263–270

[6]. Liu B, Yang L, Huang J, Meer P, Gong L, Kulikowski C (2010) Robust and fast collaborative tracking with two stage sparse optimization, in IEEE ECCV, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, vol. 6314, pp. 624–637.

[7]. Mei X, Ling H, Wu Y, Blasch E, Bai L (2011) Minimum error bounded efficient l1 tracker with occlusion detection. CVPR 2011:1257–1264

[8]. Montero AS, Lang J, Laganiere R (2015) Scalable kernel correlation filter with sparse feature integration.in IEEE International Conference on Computer Vision Workshop (ICCVW), pp. 587–594

[9]. Elman JL (1990) Finding structure in time. Cognitive Science 14(2):179–211 http://dblp.uni-trier.de/db/journals/cogsci/cogsci14.html#Elman90

[10]. Tokola R, Bolme D (2015) Ensembles of correlation filters for object detection, in IEEE Conference on WACV, pp. 935–942.

[11]. Guo X, Xiao N, Zhang L (2019) Sequential binary code selection for robust object tracking. Multimedia Tools and Applications, pp:1–13

[12]. Unlu HU, Niehaus PS, Chirita D, Evangeliou N, Tzes A (2019, October) Deep learning-based visual tracking of UAVs using a PTZ camera system. In IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society 1:638–644

[13]. Wang Q, Gao J, Xing J, Zhang M, Hu W (2017) Dcfnet: Discriminant correlation filters network for visual tracking, CoRR, vol. abs/1704.04057. https://arxiv.org/abs/1704. 04057

[14]. . Wang F, Liu G, Zhang H, Hao Z (2018) Robust long-term correlation tracking with multiple models. https://doi.org/10.1049/iet-ipr.2018.6209 Multimedia Tools and Applications

[15]. Zhong Wei LH, Ming-Hsuan Y (2012) Robust object tracking via sparsity-based collaborative model, in IEEE Conference on CVPR, ser. CVPR '12. Washington, DC, USA: IEEEComputerSociety, pp. 1838– 1845

[16]. Zhang J, Ma S, Sclaroff S (2014) MEEM: robust tracking via multiple experts using entropy minimization, in Proc. of the European Conference on Computer Vision (ECCV).

[17]. Wang Y, Luo X, Lu D, Wu J, Shan F (2019) Robust visual tracking via a hybrid correlation filter. Multimed Tools Appl 78(22):31633–31648

[18]. Wang X, Zheng Z, He Y, Yan F, Zeng Z, Yang Y (2020) Progressive local filter pruning for image retrieval acceleration. arXiv preprint arXiv:2001.08878

[19]. Wu Y, Shen B, Ling H (2014) Visual tracking via online nonnegative matrix factorization. IEEE Transactions on Circuits and Systems for Video Technology 24(3):374–383

[20]. Wu Y, Lim J, Yang M-H (2015) Object tracking benchmark. IEEE Trans Pattern Anal Mach Intell 37(9): 1834–1848

[21]. Xue X, Li Y (2019) Robust particle tracking via spatio-temporal context learning and multi-task joint local sparse representation. Multimed Tools Appl 78(15):21187–21204
[22]. Nam H, Baek M, Han B (2016) Modeling and propagating cnns in a tree structure forvisualtracking, CoRR, vol.abs/1608.07242, http://arxiv.org/abs/1608.07242
[23]. Unlu HU, Niehaus PS, Chirita D, Evangeliou N, Tzes A (2019, October) Deep learning-based visual tracking of UAVs using a PTZ camera system. In IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society 1:638–644
[24]. Zhang, Attention-based Target Recognition Algorithm and Applications in the Mobile Robot, Chongqing University, 2013.
[25]. R.B. Fisher, The PETS04 surveillance ground-truth data sets, Proc. 6th IEEE international workshop on performance evaluation of tracking and surveillance2004), pp. 1-5.
[26]. R. Collins, X. Zhou, S.K. Teh, An open source tracking testbed and evaluation web site, IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS 2005)2005), pp. 35.
[27]. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, Imagenet large scale visual recognition challenge, International Journal of Computer Vision, 115 (2015) 211-252.
[28]. Tang, Fuhui, et al. "Robust visual tracking based on spatial context pyramid." Multimedia Tools and Applications 78.15 (2019): 21065-21084.
[29]. Unlu, HalilUtku, 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.
[30]. 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.
[31]. Wang, Qiang, et al. "Dcfnet: Discriminant correlation filters network for visual tracking." arXiv preprint arXiv:1704.04057 (2017).
[32]. Tang, Fuhui, et al. "Robust visual tracking based on spatial context pyramid." Multimedia Tools and Applications 78.15 (2019): 21065-21084.
[33]. Unlu, HalilUtku, 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.
[34]. 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.
[35]. Wang, Qiang, et al. "Dcfnet: Discriminant correlation filters network for visual tracking." arXiv preprint arXiv:1704.04057 (2017).
[36]. Babenko B, Yang MH, Belongie S (2009) Visual tracking with online multiple instance learning, in IEEE Conference on CVPR, pp. 983–990.
[37]. Bao C, Wu Y, Ling H, Ji H (2012) Real time robust l1 tracker using accelerated proximal gradient approach, in CVPR, IEEE Conference, pp. 1830–1837.
[38]. Bertinetto L, Valmadre J, Golodetz S, Miksik O, Torr PHS (2016) Staple: Complementary learners for realtime tracking, in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1401–1409.
[39]. Blanco-Filgueira B, Garcia-Lesta D, Fernández-Sanjurjo M, Brea VM, López P (2019) Deep learning-based multiple object visual tracking on embedded system for IOT and mobile edge computing applications. IEEE Internet Things J 6(3):5423–5431
[40]. Bolme D, Beveridge J, Draper B, Lui YM (2010) Visual object tracking using adaptive correlation filters, in IEEE Conference on CVPR, pp. 2544–2550.
[41]. Cannons K (2008) “A review of visual tracking”, technical report CSE2008–07. York University, Canada.