International Journal of Emerging Technologies in Engineering Research (IJETER)
Volume 6, Issue 7, July (2018) www.ijeter.everscience.org

A Comparative Analysis on Visual Object Tracking System using Discriminative Correlation Filter and Extended Correlation Filter

V.Ramalakshmi @ Kanthimathi [1], Dr. M. Germanus Alex [2]
[1]Research Scholar, Bharathiar University
[1]Assistant Professor, Kamarajar Govt.Arts College, Surandai, Tamilnadu, India
[2]Research Guide, Bharathiar University
[2]Assistant Professor, Kamarajar Govt.Arts College, Surandai, Tamilnadu, India

Abstract – Tracker is the term used often in worldwide web. Visual object tracking in video is very common. It is the process of locating and tracking the motion and orientation of one or multiple moving objects over time in the given video sequence.The variations in the appearance of a tracked object include changes in geometry/photometry, camera viewpoint, illumination, or partial occlusion pose a major challenge to object tracking. In this paper, we present the design and implementation of a visual object tracking system using video scenes captured from a single surveillance camera. The proposed system uses a discriminative correlation filter based model which is robust and computationally efficient for real time tracking. The proposed system is developed in Matlab. Experiments are conducted on real world publicly available benchmarking video sequences. Our experimental results show that the proposed trackers perform better than most of the recently designed trackers.

Index Terms – Object tracking, Visual object tracking, Correlation Filter, Extended Correlation Filter.

1. INTRODUCTION

Visual object tracking is one of the challenging problems in the field of computer vision. It has several related applications such as automated video surveillance, traffic monitoring, security and robotics [1]. The objective of a visual object tracker is to estimate the location of a target in all the frames of a video sequence based on the given initial location (or a bounding rectangle) of the target. Object tracking problem has been studied by the computer vision community for several decades. However, it remains a challenging task to design an efficient and robust visual object tracking system for all the practical real- world applications. Further, there are several factors affect the performance of the object tracker such as illumination variations, scale variations, occlusions, deformations, motion blur, rotations, and low resolutions [2].

Over the years, several visual object trackers have been proposed to address some of the challenges mentioned above [3] [4] [5] [6] [7]. Broadly there are two categories of trackers found in the literature: 1) generative models [4] [5] and 2) discriminative models [6] [7]. Generative models learn an appearance model of the target object and formulate the tracking problem as finding the best matching appearance window (or the bounding box) in the subsequent frames. They are also referred as template or subspace matching models. In general, generative models do not consider the background information which might be useful for differentiating the target from the background. Whereas, the discriminative models formulate the target tracking task as a classification problem which differentiates the target objects from backgrounds of the video sequences. It is shown that discriminative models outperform generative models [2].

In the recent years, correlation filters [8] are in widespread usage for developing many object detection and recognition applications. They are commonly referred as "tracking-by-detection" methods. They gained much attention to visual object tracking tasks recently as they are transferred into Fourier domain for element-wise multiplication which is computational efficient [9] [10]. In this paper, we present the design and development of a visual object tracking system using discriminative correlation filter and Extended Correlation Filter. The proposed system was developed in Matlab. We evaluated the experiments on real world publicly available benchmarking video sequences. Our experimental results show that the proposed trackers perform better than most of the recently designed trackers.

The rest of the paper is organized as follows. In Section II, we present the related works. In Section III, we present the design of the proposed visual object trackers followed by performance metrics in Section IV. In Section V, we present the experimental setup and results using performance metrics. In Section VI, we conclude the paper.

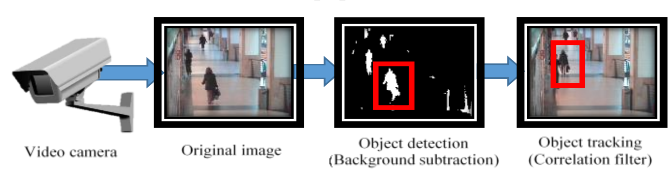

Fig. 1: Proposed visual object tracking system using DCF

2. RELATED WORK

Visual object tracking has been studied extensively with lots of applications. In this section, we introduce the approaches closely related to our work. In this section, we review the existing literature which is broadly classified into two broader categories: 1) Generative models and 2) Discriminative models.

On the discriminative model side, several kernel-based methods are proposed. For example, Hare et al. [3] presented an adaptive visual object tracking framework based on structured output prediction. They used a kernel-based structured support vector machine for performing online learning and adaptive tracking. Wang et al. [6] proposed a metric learning framework for object tracking. Both visual tracking and appearance modeling are performed online in their framework. Further, their formulations can handle multiple objects and occlusions. In contrast with all these methods, we used a discriminative correlation filter and its extension which are robust in detecting objects under several challenging scenarios such as occlusions, scale variation, among others. The following table shows the details of recent tracking methods with its motion model used for tracking objects in the video frames.

Table 1 Recent Tracking Methods with its Motion Model

Tracker	Year	Target Region	Motion Model
Multiple Instance learning Tracking (MIT)	2012	Multiple boxes	Uniform
Tracking, Learning and Detection (LOT)	2012	Patch super pixels	Gaussian
L1-minimization Tracker (LIT)	2011	Multiple boxes	Gaussian
Tracking on the Affine Group (TAG)	2011	Bounding Box	Gaussian
Locally Order less Tracking (TLD)	2010	Multiple boxes	Optical Flow
Foreground-Background Tracker (FBT)	2010	Bounding Box	Uniform
Extended Kalman Filter (EKF)	2015	Bounding Box	Bayes Rule

3. VISUAL OBJECT TRACKER DESIGN

In this section, we present the proposed visual object tracer design and its components. Figure 1 illustrates the different components of the system for object tracking. Firstly, the system receives the video sequences from the video capturing devices such as CCTV or web camera. The next step is object detection which involves detecting the objects of interest from the given first video frame. This step involves applying background subtraction techniques after performing some preliminary preprocessing methods to remove the noise. The basic idea of background subtraction is subtracting a predefined background model frame from the current from to identify the moving objects. Several advanced background subtraction methods have been proposed in the literature which is insensitive to external environmental conditions such as noise. These include approximate median, running Gaussian, and mixture Gaussian methods. After detecting the objects of interest, the next step is called object tracking which involves finding the location of the target object in the subsequent video frames.

A. Discriminative Correlation Filter (DCF)

Over the years, tracking-by-detection methods have become great interest among the researchers for various visual object tracking tasks, because of their excellent tracking performance. In the object tracking task, the tracker needs to estimate the location of the bounding box of the object within each frame of the video sequence. These methods model the target localization problem as a classification problem.

Initially, the target's bounding box or tracking window is selected either manually or automatically. In the manual case, a supervisor marks the rectangle boundary (left, top, right, and bottom) of the target object of interest. Whereas, in the automatic setup as in the proposed system, the prior object detection method outputs the boundary of the target. We use discriminative correlation filter for tracking the object [11].

The object tracking task is performed using these three steps: 1) after selecting object boundary, a correlation filter is trained using image patch cropped from the first frame; 2) in the subsequent frames, the target is tracked by correlating the trained correlation filter over a search window. The window location which gives maximum correlation output is marked as the new location of the target; 3) finally, based on this new location, an online update of the correlation filter is performed. Steps 2 and three are repeated for all the frames in a video sequence.

B. Extended Correlation Filter (ECF)

Initially, a correlation filter is learned using an image patch cropped from the given target position of the first video frame and it is termed as Extended Correlation Filter. Then, the learned correlation filter is applied to the region of interest in the subsequent frame and a response map is generated. The location with maximum value in the response map is predicted as the new location of the target object. Finally, an online update of the correlation filter is done using the new object location.

4. PERFORMANCE METRICS

Performance measures for evaluating the tracking uses ground truth, considering the target presence and position. This requires the considerable amount of annotation, with the consequence that the amount of videos with ground truth is often limited up to this point. Here we provide the most common measures used in single target tracking. The three basic types of errors in tracking are:

• Deviation : The tracker's location deviated from the ground truth.

• False positive: The tracker identifies a target which is not a target.

• False negative: The tracker misses to identify and locate the target.

A reasonable choice for overlap of target and object is the PASCAL criterion.

$$\frac{|T^i \cap GT^i|}{|T^i \cup GT^i|} \geq 0.5 \qquad \text{-------------- (1)}$$

Where T^i denotes the tracked bounding box in frame *i*, and G^i denotes the ground truth bounding box in frame *i*. When Eq. 2 is met, the track is considered to match with the ground truth. In many works, this PASCAL overlap measure is adopted without threshold. We prefer to use it with threshold as it makes it easier to evaluate large sets of sequences.

For n_{tp}, n_{fp}, n_{fn} denoting the number of true positives, false positives and false negatives in a video, *precision=ntp/(ntp+ nfp)*, and *recall = ntp/(ntp+ nfn)*.

$$F = 2 \cdot \frac{precision \cdot recall}{precision + recall} \text{-----------------(2)}$$

Table 1 shows the F-score obtained by different trackers.

5. EXPERIMENTAL RESULTS

The total number of frames in ALOV++ is 89364. It consists of 14 challenge subsets, totally 315 sequences and focuses on systematically and experimentally evaluating trackers' robustness in a large variety of situations including light changes, low contrast, occlusion, etc. The data in ALOV++ are interpreted by a rectangular bounding box along the main axes of flexible size on every fifth frame. In rare cases, when motion is rapid, the annotation is more frequent. The ground truth has been acquired for the intermediate frames by linear interpolation. The ground truth bounding box in the first frame is specified to the trackers. It is the only source of target-specific information available to the trackers. ALOV++ is made available in [24].

Table 2 Overall Performance of the trackers

Trackers/ Video	LOT	TAG	L1T	FBT	MIT	TLD	EKF	DCF	ECF
50	1.00	0.88	1.00	1.00	0.99	0.97	1.00	0.97	1.00
100	0.78	0.49	0.87	1.00	0.90	0.95	1.00	0.90	1.00
150	0.51	0.25	0.61	0.77	0.59	0.72	0.80	0.77	0.82
200	0.32	0.16	0.36	0.49	0.35	0.43	0.50	0.44	0.51
250	0.16	0.08	0.18	0.23	0.19	0.25	0.25	0.24	0.26
300	0.04	0.03	0.05	0.06	0.06	0.08	0.06	0.06	0.07

6. FUTURE WORK AND CONCLUSION

In this paper, we proposed a novel object tracking methods for video sequences. The proposed methods are capable of tracking multiple objects under different challenging scenarios such as illumination variation, occlusion, and low resolution. We evaluated the proposed methods using benchmarking video sequences. The performance of the system was evaluated using

F Score. Based on these metrics, the performance of the proposed object trackers are quantitatively compared with recently designed object trackers. The experimental results clearly show that the proposed object tracking methods performs better than most of the trackers. Multiple object tracking using proposed system is left as future work of this paper.

REFERENCES

[1] A. Yilmaz, O. Javed, and M. Shah. Object Tracking: A Survey. ACM Computing Surveys, 38(4):1–45, 2006.
[2] Wu, Yi, Jongwoo Lim, and Ming-Hsuan Yang. "Online object tracking: A benchmark." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2411-2418. 2013.
[3] S. Hare, A. Saffari, and P. H. S. Torr. Struck: Structured Output Tracking with Kernels. In ICCV, 2011.
[4] J. Kwon and K. M. Lee, “Tracking by sampling trackers,” in Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011, pp. 1195–1202.
[5] V. Belagiannis, F. Schubert, N. Navab, and S. Ilic, “Segmentation based particle filtering for real-time 2d object tracking,” in Computer Vision–ECCV 2012. Springer, 2012, pp. 842–855.
[6] X. Wang, G. Hua, and T. X. Han, “Discriminative tracking by metric learning,” in Computer Vision–ECCV 2010. Springer, 2010, pp. 200–214.
[7] Q. Wang, F. Chen, W. Xu, and M.-H. Yang, “Object tracking with joint optimiza-tion of representation and classification,” Circuits and Systems for Video Technolo-gy, IEEE Transactions on, vol. 25, no. 4, pp. 638–650, 2015.
[8] B. V. K. V. Kumar, A. Mahalanobis, and R. D. Juday, Correlation Pattern Recog-nition, Cambridge University Press, 2005.
[9] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, “Visual object tracking using adaptive correlation filters,” in Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010, pp. 2544–2550.
[10] Zhang, Lei, Yanjie Wang, Honghai Sun, Zhijun Yao, and Shuwen He. "Robust Visual Correlation Tracking." Mathematical Problems in Engineering 2015 (2015).
[11] M. Danelljan, G. H¨ager, F. S. Khan, and M. Felsberg, “Accurate scale estimation for robust visual tracking,” in Proceedings of the British Machine Vision Conference BMVC, 2014.
[12] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, “Robust visual tracking via multi-task sparse learning,” in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012, pp. 2042–2049.
[13] V. Belagiannis, F. Schubert, N. Navab, and S. Ilic, “Segmentation based particle filtering for real-time 2d object tracking,” in Computer Vision ECCV 2012. Springer, 2012, pp. 842–855.
[14] Babenko, B., Yang, M.H., Belongie, S.: Robust object tracking with online multiple instance learning. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(8) (2011)
[15] Zhong, W., Lu, H., Yang, M.H.: Robust object tracking via sparse collaborative appearance model. IEEE Transactions on Image Processing 23(5), 2356–2368 (2014)
[16] Hare, S., Saffari, A., Torr, P.H.S.: Struck: Structured output tracking with kernels. In: Proceedings of IEEE International Conference on Computer Vision (2011)
[17] Henriques, J.F., Caseiro, R., Martins, P., Batista, J.: Exploiting the circulant structure of tracking-by-detection with kernels. In: Proceedings of the European Conference on Computer Vision (2012)
[18] Zhang, K., Zhang, L., Liu, Q., Zhang, D., Yang, M.H.: Fast visual tracking via dense spatio-temporal context learning. In: Proceedings of the European Conference on Computer Vision (2014)
[19] Danelljan, M., Hager, G., Khan, F.S., Felsberg, M.: Accurate scale estimation for robust visual tracking. In: Proceedings of British Machine Vision Conference (2014)
[20] Henriques, J.F., Caseiro, R., Martins, P., Batista, J.: High-speed tracking with kernelized correlation filters. IEEE Transactions on Pattern Analysis and Machine Intelligence 37(3), 583–596 (2015)
[21] Gao, J., Ling, H., Hu, W., Xing, J.: Transfer learning based visual tracking with gaussian processes regression. In: Proceedings of the European Conference on Computer Vision (2014)
[22] Zhang, J., Ma, S., Sclaroff, S.: MEEM: Robust tracking via multiple experts using entropy minimization. In: Proceedings of the European Conference on Computer Vision (2014)
[23] Hong, Z., Chen, Z., Wang, C., Mei, X., Prokhorov, D., Tao, D.: MUlti-Store Tracker (MUSTer): A cognitive psychology inspired approach to object tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (2015)
[24] http://www.alov300.org/