A Robust Object Tracking Method for Surveillance Applications to Handle Occlusion

Madah-Ul-Mustafa and Zhu Liang Yu

***Abstract*—The work proposed in this paper, addresses the issue of robust tracking scheme by further studying the problem occlusion that causes tracker to drift. The proposed work addresses the problem within context of real-time tracking for surveillance applications. Firstly, we studied the occlusion and drift problems and how it is linked to the visual object tracking framework. Secondly we proposed a robust tracking scheme that can handle occlusion and drift problems as well as other visual object tracking challenges to predict the target object position when occlusion is occurred. The proposed scheme adopts an efficient integration of motion modeling via particle-kalman-filter (PKF) into the kernelized correlation filter (KCF) tracking framework to achieve an efficient and robust tracking scheme that mitigate the problem of tracker drift. In the proposed tracking scheme KCF acts as our basic tracker due to its better performance and high efficiency but the tracker lags behind other state-of-the-art when there are problems like occlusion and illumination variation causing it to drift. When the occlusion occurs, the PKF will be used to predict the target object location and will use the available position and state of the target object before occlusion is occurred. An experimental result on publicly available dataset demonstrates that the proposed scheme achieves a competitive performance as compared with other state-of-the-art trackers.**

***Index Terms*—Kernelized correlation filter (KCF), occlusion, tracking-by-detection, particle kalman filter (PKF).**

I. Introduction

Vision-based Various machine vision algorithms for intelligent surveillance systems comprises of object detection and object tracking as its fundamental and crucial part, that led the realization of high-level and complex systems. The applications of object tracking are not only limited exclusively to machine vision like biomedical imaging [1], video security systems [2] and vision-based traffic monitoring [3], but it is further extended to more highly integrated systems including such as human computer interaction [4], navigation systems [5], robotic vision [6] and anomaly detection [7]. However to develop a robust and persistent visual object tracking framework is still a contemporary problem due to the challenging factors such as occlusion, in-plane rotation, out-of-plane rotation, background clutters, uneven motion, variation of scale and illumination.

In an effort to address the above challenges, good progress has been made in this regard and tracking methods like data association [8], energy minimizing [9], point tracking [10] and feature descriptors [11]. Some methods proposed object modeling based on the object appearance to 2d shapes [12] or 3d shapes [13]. These various techniques have been proposed in different domains of visual object tracking such as human tracking [14], hand tracking [15] and vehicle tracking [16]. However these proposed methods succeeded high efficient performance as well as significant tracking speed but these methods fail to maintain the efficiency while tracing arbitrary objects due to their massive dependence on the offline target modeling. To cope with such problems discriminative methods has gained attention in visual object tracking. In these methods, the tracker is initialized in the first frame of the video sequence with the location of target object. The goal of the tracker is to estimate the target object location in the later frames of the video sequence. In discriminative methods, object detection is treated as binary problem and mostly rely on classification schemes to realize the feasibility of target object detection schemes. These schemes have shown significant results and some of the remarkable work has been implemented in [17]-[19]. These methods are also referred as tracking-by-detection methods.

Correlation filters made recent advancement in discriminative methods and achieved competitive results, due to its simple structure, high computational efficiency and achieve hundred of frames per second making it capable to for real-time applications. In [20] proposed a novel correlation filter named minimum output sum of squared error (MOSSE), by using adaptive learning scheme that showed promising results. Later numbers of correlation filter based tracker were developed and showed promising results [21]-[23], most of the trackers fail to explicitly handle occlusion; a common flaw in most state-of-the-art trackers. A typical example can be seen in kernelized correlation filter tracker (KCF) [24], which is capable of achieving remarkably fast and efficient tracking results. Despite its speed and efficiency, KCF tracker just like other correlation filter based trackers, fails to handle occlusion that severely hampers the tracker's performance in most of the applications, that leads the tracker to drift and in most scenarios it fails to recover the target object. In real-time applications such as surveillance applications, there is a need of such tracking scheme that is capable of tackling drift, irrespective to the duration of video sequence. Thus it merits further research to robustifying such tracking schemes.

In an order to address this problem, we proposed an occlusion robustification scheme by exploiting the advantages of basic KCF tracker and further extended it to object motion model estimation to predict the position of the target. This motion modeling is proven to be sufficient in tackling occlusion without trading-off the speed of the KCF tracking framework. This paper focuses on the efficient

Manuscript received July 29, 2020; revised October 17, 2020.

Madah-Ul-Mustafa and Zhu Liang Yu are with School of Automation Science and Engineering, South China University of Technology, Guangzhou 510641, China (e-mail: aumadah_mustafa@mail.scut.edu.cn, zlyu@scut.edu.cn).

DOI: 10.7763/IJCTE.2021.V13.1283

object motion modeling problem within the proposed kernelized correlation filters tracker, by treating the tracking task as an object motion modeling problem. Experimental results attained on various challenging scenarios demonstrate the robustness of the proposed scheme to occlusion, leading to drift-alleviation while significantly maintaining the tracking speed characteristic of kernelized correlation filter tracking framework.

II. Related Work

Visual object tracking has been studied intensively and have numerous applications in the real world scenarios. Generally speaking visual object tacking method consists of two important components i.e. a motion model and an observation model. A motion model depicts the states of an object over the period of time and predicts its likely state. Most commonly used motion models are Kalman filters [25], [26] and particle filters [27], [28]. While, an observation model describes the appearance information of the target object and verifies its prediction in each frame [29]. The modern day tracking methods are mostly based on observation model due to its better performance as compared to motion model [30].

In an Observation model, the tracking methods are further categorized into either generative methods [31], [32] or discriminative methods [17], [18]. Generative methods deal the problem by searching the most similar regions that are close to the tracked object. These methods includes template-based [33], incremental subspace learning [31], sparse representation [34] and density estimation [35] based methods. While discriminative methods also known as tracking-by-detection methods treat tracking as binary problem and classifier is trained to distinguishes target from background. Various traditional machine learning techniques have applied to address the tracking problem but not limited to K-means [36], boosting [37], support vector machine (SVM) [17], ranking SVM [38], structured learning [39], naive bayes [40], multiple instance learning [41], P-N learning [42], random forest [43] and metric learning [44].

In the past few years correlation filters have gained attention in the field of visual object tracking due to its fast computational efficiency and impressive results [45]. In [20] proposed a simple adaptive correlation filer based tracker with a minimum output sum of squared error (MOSSE) filter that can achieve hundreds of frames per second speed. In [19] proposed circulant structure kernel (CSK) that performs classification on non-linear feature spaces by using kernel trick can achieve good speed. CSK was further efficiently improved by using HOG features and the kernel method to strengthen multiple channels in kernelized correlation filters (KCF) [24]. In addition to KCF several other methods also introduced kernels into correlation filters such as work presented in [46], [47]. Later different variants of KCF have been explored to improve tracker's efficiency such as to handle the scale variations of the target object, discriminative scale space tracker (DSST) [45] were presented that learns adaptive filters using Histogram Oriented Gradients (HOG) features for translation and scaling separately. Some methods integrated motion based modeling such as presented in kalman-based kernelized correlation filter (KKCF) tracker [48], where KCF tracker is used for basic tracking and kalman Filter is used to predict target object position when there is occlusion. To improve the robustness several methods introduced local patch based correlation trackers [49], [50] or either parts based correlation filter trackers [51], [52].

In the domain of surveillance applications there is a need of such tracking schemes that is capable to tackle the problems like occlusion. Despite the great success of correlation filter based tracking methods but these methods are not robust enough to tackle occlusion that is due its dependence on the maximum response value and trackers may lose the target object when the maximum response value is uncertain or not reliable. To cope with this problem the motivation of the work in this chapter is developed. We integrated the particle-kalman-filter (PKF) into KCF tracking framework to predict the target object position when the KCF tracker's result is not reliable. This motion modeling is proven to be sufficient to handle occlusion and experimental results achieved on numerous challenging scenarios shows the robustness of the proposed tracking scheme to occlusion that leads tracker to drift while maintaining sufficient tracking speed of kernelized correlation filter tracker.

III. Proposed Tracking Algorithm for Real-Time Surveillance Applications

The proposed methodology is briefly explained in this section. In the first part basic kernelized correlation filter tracking is explained and in the second part the proposed PKF prediction system method.

A. Kernelized Correlation Filter Based Tracking

The simple kernel structure and its combination with correlation filters have already been proven. Thus encourages its selection for illustration of the work proposed in this paper. The kernelized correlation filter (KCF) tracker performs a simple element wise operation in the Fourier domain instead of using the time consuming convolution to reduce the memory and computational costs. The tracker predicts the location of target object by training a correlation filter using all possible cyclic shifts on an image patch x of size $M\ X\ N$ pixels of HOG features that is centered around the target. All the cyclically shifted variants of, $x_{m,n}(m,n) \in \{0,\dots,M-1\} \times \{0,\dots,N-1\}$ are generated as training samples to train a filter for the classifier. These samples are labeled with a Gaussian function label $y_{m,n}$, so that that $y_{m,n}$ is the label for $x_{m,n}$. The filter w can be trained be calculated by minimizing the error between the training sample $x_{m,n}$ and their labels $y_{m,n}$.

$$w = \arg\min_{w} \sum_{m,n} \left|\langle \emptyset(x_{m,n}), w\rangle - y(m,n)\right|^2 \ \lambda||w||^2 \quad (1)$$

While $\emptyset$ is the mapping to the Hilbert space induced by the kernel k and $\lambda \geq 0$ is the regularization parameter. By using kernel trick presented in [19], the objective function can identically be expressed as a linear combination of the inputs:

$$w = \sum_{m,n} \alpha(m,n)\emptyset(x_{m,n}) \quad (2)$$

where α is the matrix, whose elements are denoted by $\alpha_{m,n}$. The matrix α can be calculated in Eq. (3) based on the circulant structure and the Fourier transform.

$$\alpha = F^{-1}\left(\frac{F(y)}{F(k^x)+\lambda}\right) \tag{3}$$

where F and F^{-1} represents the Fourier Transform (FFT) and inverse Fourier transform (IFFT) respectively. k^x is the kernel correlation of x with itself i.e. $k^x = k\,(x_{m,n}, x)$. During the object tracking process, a patch z with the same size as x is cropped out from the new frame. The location of target is estimated by calculating the of maximum response score:

$$f(z) = F^{-1}\big(F(k^z) \odot F(\alpha)\big) \tag{4}$$

where $\odot$ is the element-wise product, $k^z = k\,(z_{m,n}, \hat{x})$ and $\hat{x}$ is the learned appearance of the target object. The filter is updated in Eq. (5):

$$H_t = \frac{y}{\hat{x}_t * \hat{x}_t^* + \lambda} \odot \hat{x}_t \tag{5-a}$$

$$W_t = (1 - \gamma W_{t-1}) + \gamma H_t \tag{5-b}$$

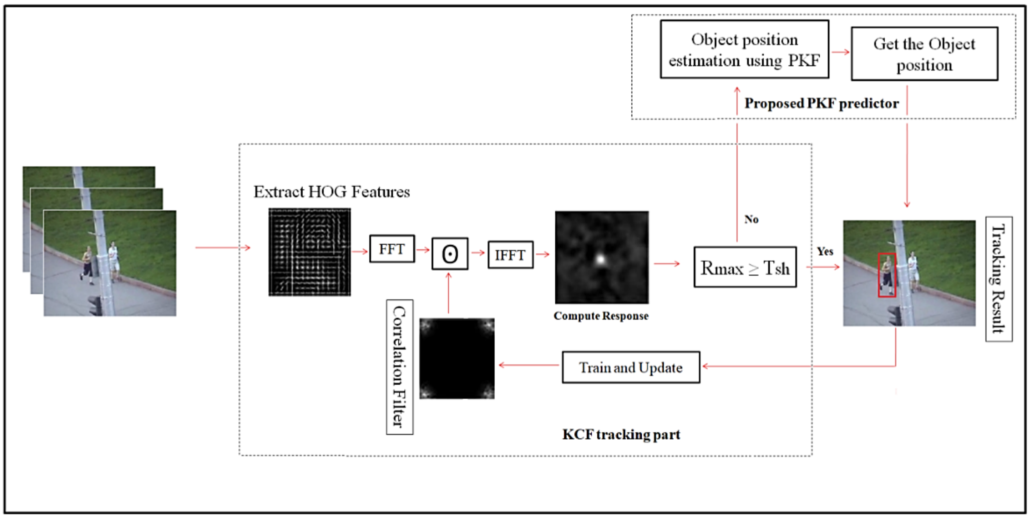


Fig. 1. Flowchart of proposed tracking algorithm illustrated by the Jogging1 sequence.

where H_t is the correlation filter at t-th frame, W_t is the updated correlation filter model and γ is the learning rate used to update the correlation filter model only in the current frame. A general flow-chart of the proposed scheme is shown in Fig. 1.

B. *Occlusion Handling via Motion Modeling*

In the domain of surveillance applications there is a need of such tracking scheme that is capable to tackle the challenging factors like background clutter or occlusion, variation of scale and fast motion that leads the tracker to drift and may lose the target object. Regardless of its impressive results in a recent benchmark [53], KCF fails to tackle the fore-mentioned problem that is due to its excessive dependence on maximum response value. The work presented in this paper proposes a new occlusion handling scheme in the kernelized correlation filter tracking framework to make tracking robust and efficient. Most commonly used motion models are kalman filtering and particle filtering, some notable work is presented in [25], [27], [54]. However, these methods are not robust enough or don't meet the criteria of real-time scenarios. Kalman filter cannot handle the non-linear motion of the target object and fails to predict the location. While particle filters are more robust as compared to kalman filter but due to a large number of particles used to track each object that makes the algorithm complex hence making the particle filter slow. Therefore, particle filters are usually not suitable for real-time applications. In this section we used particle kalman filter (PKF) [55] that takes advantages of both kalman filter and particle filter to meet the real-time application requirements to the high dynamics encountered in the surveillance applications. In PKF particles are calculated in two phases, i.e. prediction phase and the update phase. Each particle set has state, covariance matrix and weight. The k-th predicted particles at frame t is represented by $\hat{s}_t^k = \{\hat{x}_t^k, \hat{p}_t^k, \widehat{w}_t^k\}_{k=1\ldots N}$, where $\hat{x}_t^k$ are the predicted particle state of target that is also treated as the position of target object, $\hat{p}_t^k$ is the covariance matrix of predicted particle set, $\widehat{w}_t^k$ is the weights of particle set calculated by $\widehat{w}_t^k = \frac{1}{N}$, where N is the total number of particles. The dynamic model in the particle filter is used to predict the possible position of the particles in the next video frame. The same dynamic model for each particle is equally applied by random-walk scheme due to the uncertainty of motion shown in Eq. (6):

$$\hat{x}_t^k = \hat{x}_t^k + g_t^k \tag{6}$$

where g_t^k is the zero mean gaussian distribution that corresponds to the searching radius of particles around the mean of particles. To ensure the accuracy half of the particles are employed small noise and half of the particles with bigger noise, in case of full occlusion the particles with bigger noise are used for tracking of target object.

$$\hat{x}_t^{k(\frac{1}{2}*N)} = \hat{x}_t^{k(\frac{1}{2}*N)} + g_{t(small)}^{k(\frac{1}{2}*N)} \tag{7-a}$$

$$\hat{x}_t^{((\frac{1}{2}*N+1)N)} = \hat{x}_t^{((\frac{1}{2}*N+1)N)} + g_{t(big)}^{((\frac{1}{2}*N+1)N)} \tag{7-b}$$

The covariance matrix is measured by the Eq. (8) after the first iteration of the PKF. The weights of the particle in the predicted particle set are calculated by the likelihood function in Eq. (9):

$$\hat{p}_t^k = \tilde{p}_t^k + Q \tag{8}$$

$$\widehat{w}_t^k = \frac{1}{\sqrt{2\sigma\pi}} exp\left(-\frac{\left[1-\mathrm{D}[q,x_t^k]\right]}{2\sigma^2}\right) \tag{9}$$

where σ is the normalization constant and $\mathrm{D}[\,q, x_t^k]$ is bhattacharya distance of measurement model and the particle. After updating the particle set in the prediction phase, the observation vector z_t is calculated from predicted particle set using Eq. (10):

$$z_t = \sum_{k=1}^{N} \widehat{w}_t^k, \hat{x}_t^k \tag{10}$$

After calculating the observation vector, kalman gain is calculated by Eq. (11):

$$\widehat{K}_t^k = \hat{p}_t^k (\widehat{K}_t^k + R_t^k)^{-1} \tag{11}$$

where R_t^k is the measurement noise and if R_t^k is small then observation vector z_t is accurate. After computing observation vector and kalman gain, the position, covariance matrix and weights of particles are updated to generate a new particle set in Eq. (12):

$$\tilde{x}_k^t = \hat{x}_t^k + \widehat{K}_t^k (z_t - \hat{x}_t^k) \tag{12-a}$$

$$\tilde{p}_t^k = \hat{p}_t^k - \widehat{K}_t^k \hat{p}_t^k \tag{12-b}$$

$$\widetilde{w}_t^k = \frac{1}{\sqrt{2\sigma\pi}} exp\left(-\frac{\left[1-\mathrm{D}[q,x_t^k]\right]}{2\sigma^2}\right) \tag{12-c}$$

The position of the target object is calculated by the particle set.

$$x_t = \sum_{k=1}^{N} \tilde{x}_t^k, \widetilde{w}_t^k \tag{13}$$

As compared to conventional particle filter, the particle resampling is performed on updated particle at the end of every iteration by duplicating the particles with high weights and the particles that have low weights are discarded to prevent particle degeneracy.

IV. Implementation Details

The proposed algorithm is shown in Algorithm 1. The tracker begins object position in the first frame to train a correlation filter. The HOG features are extracted in the search window and convolve it with the correlation filter to calculate the maximum response. We assumed that there is no obvious change occurred in target appearance in the first two frames. The maximum response is calculated from t > 2. This maximum response value is then compared with the predefined threshold to decide which method is applied to predict the target object location. If the maximum response value is greater than the predefined threshold then KCF method is used to predict the target object location and the model is updated. And if maximum response is less than the threshold then PKF method is applied to predict the target object location. When occlusion is occurred at *j*th frame while the target object is being tracked by the KCF tracker and then motion model considers the position of target in the previous frame (*j-1)*th frame before occlusion occurs. The whole process is repeated until the position of the target object in the last frame of the video.

Algorithm 1: Proposed Occlusion Handling Algorithm

1. **Input:** The *t*-th frame, target location BB_t within frame *t*.
2. **Output:** Target location BB_t in each frame.
3. Extract the target HOG features in the initial frame
4. Train the model by using Eq. (3).
5. **if** $t < T$, where *t* is the current frame and *T* is the total number of frames **then**
6. Compute the correlation filter response by using Eq. (4).
7. For the $t > 2$, calculate the maximum response R_{max} and compare it with the predefined threshold T_{sh}.
8. **if** $R_{max} \geq T_{sh}$ **then**
9. Get the target position by using KCF tracker at frame t and the target object size.
10. **end if**
11. Update the correlation filter by using Eq. (5).
12. **if** $R_{max} \leq T_{sh}$ **then**
13. Implement the PKF to predict the target positions using object position before occlusion is occurred.
14. Initialize predicted particle set and compute predicted particle set $\hat{s}_k^t = \{\hat{x}_t^k, \hat{p}_t^k, \widehat{w}_t^k\}_{k=1\ldots N}$
15. Compute observation vector z_t by using Eq. (10).
16. Compute the kalman gain $\widehat{K}_t^k$ by using Eq. (11).
17. Update the particle set $\tilde{s}_k^t = \{\tilde{x}_k^t, \tilde{p}_k^t, \widetilde{w}_k^t\}_{k=1\ldots N}$, by using Eq. (12).
18. Get the position of the target model at frame *t* by using Eq. (13).
19. Resample the particles.
20. **end if**
21. **end if**

V. Experimental Results and Discussion

A. Experimental Setup

The details of the experimental platform are presented in Table I. All parameters in the proposed algorithm are fixed for all the experiments for fair comparisons. The value of t_h is set at 0.06 for all experiments and the other parameters for the basic KCF are same used in [24]. The proposed algorithm is designed for real-time surveillance applications, hence the performance is validated on the challenging video scenarios publicly available benchmark dataset and compared with the other state-of-the-art trackers including KCF [24], MIL [41], STRUCK [39], TLD [56] and CT [40].

TABLE I: EXPERIMENTAL PLATFORM FOR PROPOSED ALGORITHM

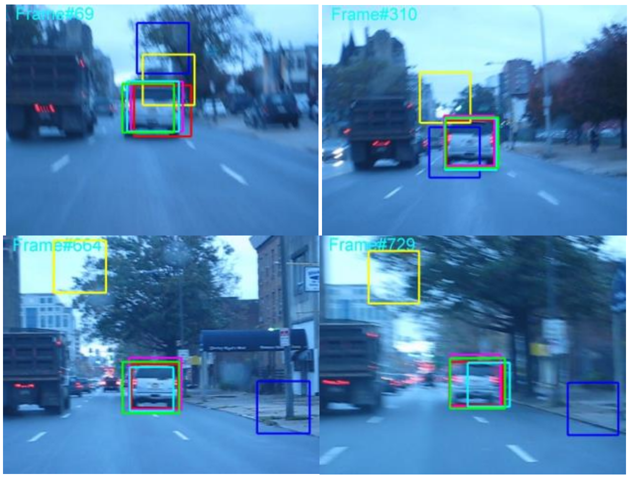
Computing platform	*Windows 7*
Processing Platform	Intel (R) corei7-4702 CPU@2.20GHz 2.20 Hz
RAM	8 GB
Computing Architecture	64 Bits
Implementing platform	Matlab R2015a

B. Experimental Results

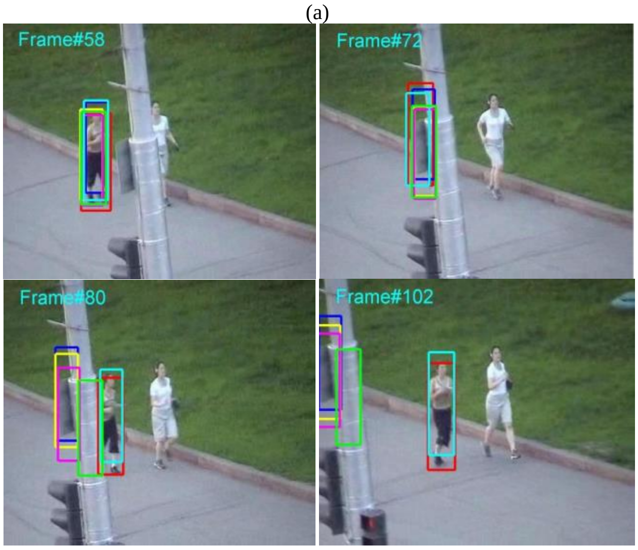
The proposed algorithm is firstly compared with the existing state-of-the-art tracking frameworks on challenging video sequences from publicly available dataset [53]. The video sequences used for evaluation comprises of variety of challenging scenarios like occlusion and other problems mentioned above. This evaluation criterion does not only demonstrate the robustness of the proposed tracking scheme but also emphasizes on its pros and cons as compared to existing state-of-the-art trackers. As the results displayed in the Fig. 2, in the initial frames all the tracking methods perform persistent tracking of target object. Intuitively the performance varies to begin in the later sequences. KCF tracker displays the efficient performance that is due to its capability to exploit kernelized correlation filters using circulant structure. As shown in the experimental results, KCF tracking approach suffers heavily in the presence of occlusion and background clutters, leads the tracker to drift. This drawback is due to KCF tracker inability to tackle the instances of occlusion and some degree of background clutters. On the whole in this problem the proposed tracking scheme outperforms KCF by introducing motion modeling via PKF within the KCF tracking framework. The proposed tracking scheme is efficient enough to track the target object in the presence of occlusion and can be recovered on a per frame basis to ensure the tracker stability even over long video sequences. In *Jogging2* sequence the proposed scheme outperforms all the tracking schemes, as shown in results all trackers are able to maintain the target object until frame#45, when the occlusion starts to occur. At this stage CT, KCF, MIL, STRUCK and TLD starts to lose the target object and in frame#80 the target completely appear again, the proposed tracking scheme consistently track the target object. Trackers like TLD were close to target object and recover the target in few frames later after the target object appears again but TLD suffers in presence of severe illumination variations in *CarDark* sequence leading the tracker to lose the target object completely. Whereas the proposed tracking scheme maintains stable tracking results, CT attains unstable performance and is followed by MIL in all sequences except in *couple* sequence where CT and MIL slightly outperforms STRUCK. However computational complexity as proposed in the KCF tracking framework of the proposed scheme does not hamper its efficiency for real-time applications. The tracking results are shown in Fig. 2 and the tracker precisions are summarized in Table II.

TABLE II: QUANTITATIVE COMPARISON OF PROPOSED METHOD. THE RESULTS ARE REPORTED IN PRECISION AT A THRESHOLD OF 20 USED IN [41]

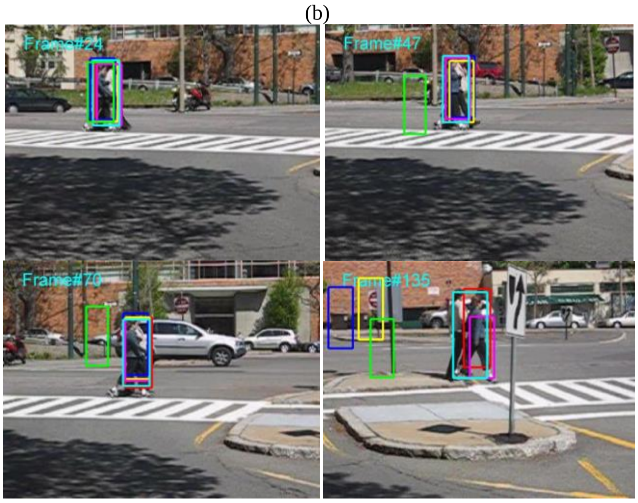
Sequence/Trackers	CT	KCF	MIL	STRUCK	TLD	Proposed
BlurCar1	0.06	0.90	0.08	0.89	0.98	0.98
Car2	0.39	1.00	0.13	0.97	0.98	1.00
CarDark	0.03	1.00	0.36	1.00	0.63	1.00
Couple	0.69	0.28	0.69	0.56	1.00	1.00
Jogging1	0.23	0.23	0.24	0.26	0.97	0.98
Jogging 2	0.17	0.16	0.19	0.23	0.86	0.96
Suv	0.26	0.97	0.21	0.58	0.93	0.97



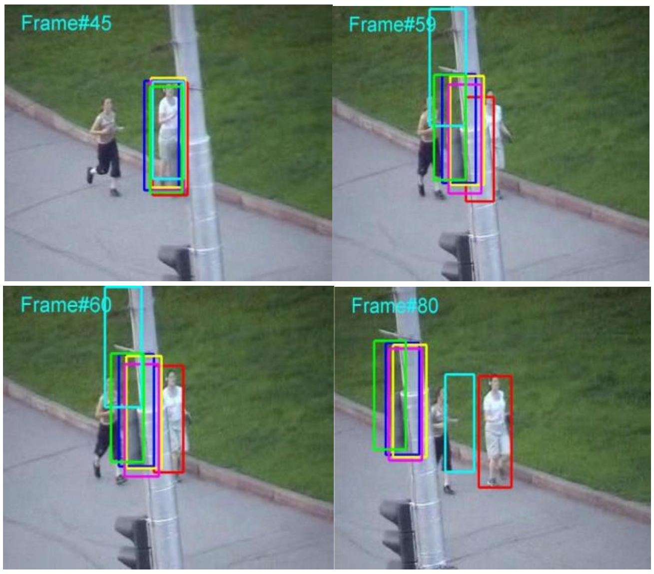
(a)



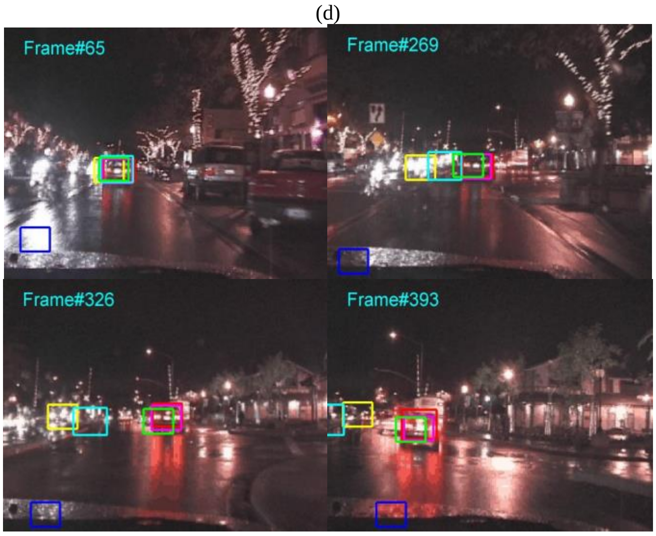
(b)



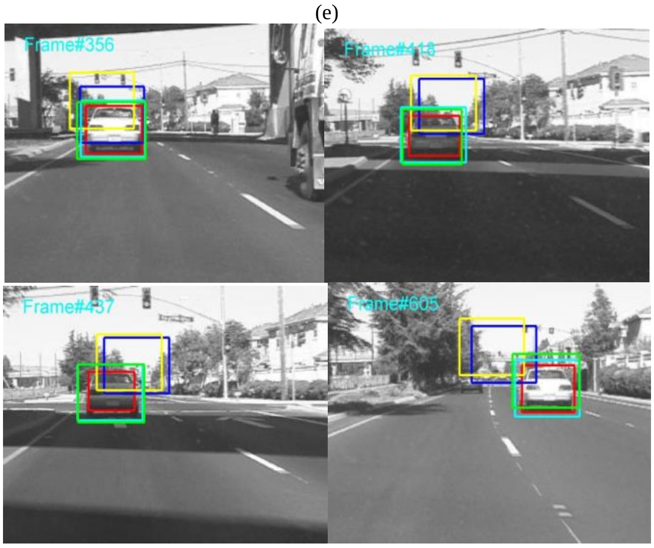
(c)



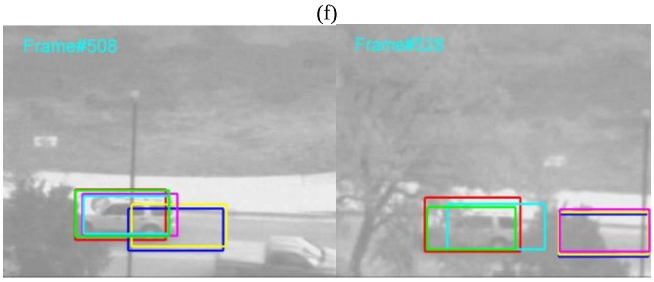
(d)

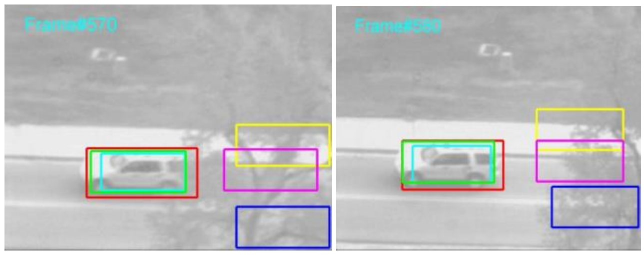


(e)



(f)





(g)

CT KCF MIL STRUCK TLD Proposed

Fig. 2. Tracking results on (a) BlurCar1 (b) Jogging1 (c) Couple (d) Jogging1 (e) CarDark and (f) Car2 (g) Suv.

VI. Conclusion

In this paper we proposed a robust tracking scheme for surveillance applications that integrates prediction system into the kernelized correlation filter (KCF) tracking framework to achieve an efficient and robust tracking scheme to mitigate the problem of occlusion causing tracker drift. The proposed method uses PKF to predict the target object location in the presence of occlusion that further enhances the tracker's efficiency. The computational complexity proposed in the KCF tracking framework of does not impedes its efficiency for real-time application. Furthermore experimental results on publicly available benchmark dataset demonstrate that the proposed scheme has achieved a competitive performance in comparison with other state-of-the-art trackers in the presence of challenging factors. Furthermore, the qualitative and quantitative result on challenging dataset validates the performance of the proposed tracking method. Furthermore, hybrid methods i.e. combination of proposed method with deep learning based methods is worth exploring and may provide more room for further improvement in feature learning and visual understanding framework for surveillance systems that is more robust and efficient.

Conflict of Interest

The authors declare that the paper is carried out without a conflict of interest.

Author Contributions

Madah-Ul-Mustafa conducted the research; Prof. Zhu Liang Yu did review and editing of manuscript; Madah-Ul-Mustafa wrote the paper; all authors had approved the final version.

References

[1] S. Queiros, P. Morais, D. Barbosa, J. Fonseca, J. Vilaca, and J. D'Hooge, "MITT: Medical image tracking toolbox," *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, pp. 2547-2557, 2018.

[2] P. Patoliya and S. Bombywala, "Object detection and tracking for surveillance," *International Journal of Engineering and Technical Research (IJETR)*, vol. 3, pp. 18-24, 2015.

[3] N. Soe, T. Thein, and T. Aung, "GPS tracking and traffic monitoring system in urban transportation," in *Proc. the IEEE 7th Global Conference on Consumer Electronics*, 2018, pp. 804-805.

[4] Y. Li and R. Nevatia, "Key object driven multi-category object recognition, localization and tracking using spatio-temporal context," in *Proc. the European Conference on Computer Vision*, 2008, pp. 409-422.

[5] A. Ess, K. Schindler, B. Leibe, and L. Gool, "Object detection and tracking for autonomous navigation in dynamic environments," *International Journal of Robotics Research*, vol. 29, pp. 1707–1725, 2019.

[6] Y. Yeboah, Z. Yu, and Wei Wu, "Robust and persistent visual tracking-by-detection for robotic vision systems," *International Journal of Machine Learning and Computing*, vol. 6, no. 3, pp. 196-204, 2016.

[7] C. Chuang, J. Hsieh, and K. Fan, "Suspicious object detection and robbery event analysis," in *Proc. the 16th International Conference on Computer Communications and Networks*, 2007, pp. 1189-1192.

[8] L. Wang, N. Yung, and L. Xu, "Multiple-human tracking by iterative data association and detection update," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 1886-1899, 2014.

[9] A. Milan, K. Schindler, and S. Roth, "Multi-target tracking by discrete-continuous energy minimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.38, pp. 2054-2068, 2015.

[10] J. Bouguet. (1999). Pyramidal implementation of the Lucas Kanade feature tracker description of the algorithm. Intel Corporation, Microprocessor Research Labs. [Online]. Available: http://seagull.isi.edu/marbles/assets/components/workflow_portal/users/lib/opencv/share/opencv/doc/papers/algo_tracking.pdf

[11] K. Yucai, L. Shirong, W. Jiangping, and Z. Botao, "Person following based on HAAR-like feature and HOG feature in indoor environment," in *Proc. the 35th Chinese Control Conference*, 2016, pp. 6345-6349.

[12] D. Kang, Z. Ma, and A. Chan, "Beyond counting: Comparisons of density maps for crowd analysis tasks-counting, detection, and tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 5, pp. 1408-1422, 2019.

[13] L. Wang and N. Yung, "Three-dimensional model-based human detection in crowded scenes," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, pp. 691-703, 2012.

[14] B. Wu and R. Nevatia, "Detection and tracking of multiple, partially occluded humans by Bayesian combination of edgelet based part detectors," *International Journal of Computer Vision*, vol. 75, no. 2, pp. 247-266, 2007.

[15] L. Sun, G. Liu, and Y. Liu, "3D hand tracking with head mounted gaze-directed camera," *IEEE Journal Sensors*, vol. 14, no. 5, pp. 1380-1390, 2014.

[16] F. Zheng, L. Shao, and J. Han, "Robust and long-term object tracking with an application to vehicles", *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 10, pp. 3387-3399, 2018.

[17] S. Avidan, "Support vector tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 8, pp. 1064-1072, 2004.

[18] S. Avidan, "Ensemble tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 261-271, 2007.

[19] J. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Proc. the European Conference on Computer Vision*, 2012, pp. 702-715.

[20] D. Bolme, J. Beveridge, B. Draper, and Y. Lui, "Visual object tracking using adaptive correlation filters," in *Proc. the IEEE International Conference on Computer Vision*, 2010, pp. 2544-2550.

[21] Y. Li and J. Zhu, "A scale adaptive kernel correlation filter tracker with feature integration," in *Proc. European Conference on Computer Vision*, 2014, pp. 254-265.

[22] T. Liu, G. Wang and Q. Yang, "Real-time part-based visual tracking via adaptive correlation filters", in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4902-4912.

[23] X. Li, Q. Liu, Z. He, H. Wang, C. Zhang, and W. Chen, "A multi-view model for visual tracking via correlation filters," *Knowledge-Based Systems*, vol. 113, pp. 88-99, 2016.

[24] J. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed Tracking with Kernelized Correlation Filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583-596, 2015.

[25] N. Funk, "A study of the Kalman filter applied to visual tracking," Dissertation, University of Alberta, Canada, 2003.

[26] P. Li, T. Zhang, and B. Ma, "Unscented Kalman filter for visual curve tracking," *Image and Vision Computing*, vol. 22, no. 22, pp. 157-164, 2004.

[27] K. Hossain and C. Lee, "Visual object tracking using particle filter," in *Proc. the International Conference on Ubiquitous Robots and Ambient Intelligence*, 2013, pp. 98-102.

[28] X. Li, S. Lan, Y. Jiang, and P. Xu, "Visual tracking based on adaptive background modeling and improved particle filter," in *Proc. the IEEE International Conference on Computer and Communications*, 2017, pp. 469-473.

[29] X. Li, W. Hu, C. Shen, Z. Zhang, A. Dick, and A. Hengel, "A survey of appearance models in visual object tracking," *ACM Transactions on Intelligent Systems and Technology*, vol. 4, no. 4, pp. 1-48, 2013.

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

[31] D. Ross, J. Lim, R. Lin, and M. Yang, "Incremental learning for robust visual tracking," *International Journal of Computer Vision*, vol. 77, no. 1, pp. 125–141, 2008.

[32] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 564-577, 2003.

[33] A. Adam, E. Rivlin, and I. Shimshoni, "Robust fragments-based tracking using the integral histogram," in *Proc. the IEEE Conference on Computer Vision Pattern Recognition*, 2006, pp. 798-805.

[34] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, "Robust visual tracking via multi-task sparse learning," in *Proc. the IEEE Conference on Computer Vision Pattern Recognition*, 2012, pp. 2042-2049.

[35] B. Han, D. Comaniciu, Y. Zhu, and L. Davis, "Sequential kernel density approximation and its application to real-time visual tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30 no. 7, pp. 1186-1197, 2008.

[36] Y. Qi, K. Suzuki, H. Wu, and Q. Chen, "EK-means tracker: A pixel-wise tracking algorithm using Kinect," in *Proc. the Third Chinese Conference on Intelligent Visual Surveillance*, 2011, pp. 77-80.

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

[38] Y. Bai and M. Tang, "Robust tracking via weakly supervised ranking SVM," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1854-1861.

[39] S. Hare, A. Saffari, and S. Torr, "Struck: Structured output tracking with kernels," in *Proc. the IEEE International Conference on Computer Vision*, 2011, pp. 263-270.

[40] K. Zhang, L. Zhang, and M. Yang, "Real-time compressive tracking," in *Proc. the European Conference on Computer Vision*, 2012, pp. 864-877.

[41] B. Babenko, M. Yang, and S. Belongie, "Robust object tracking with online multiple instance learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1619-1632, 2011.

[42] Z. Kalal, J. Matas, and K. Mikolajczyk, "P-N learning: Bootstrapping binary classifiers by structural constraints," in *Proc. the IEEE Conference on Computer Vision Pattern Recognition*, 2010, pp.49-56.

[43] A. Saffari, C. Leistner, J. Santner, M. Godec, and H. Bischof, "On-line random forests," in *Proc. the IEEE 12th International Conference on Computer Vision Workshops*, 2009, pp. 1393-1400.

[44] N. Jiang, W. Liu, and Y. Wu, "Learning adaptive metric for robust visual tracking," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2288-2300, 2011.

[45] M. Danelljan, G. Hager, F. Khan, and M. Felsberg, "Discriminative scale space tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 8, pp. 1561-1575, 2017.

[46] R. Patnaik and D. Casasent, "Fast FFT-based distortion-invariant kernel filters for general object recognition," in *Proc. the SPIE*, 2009, vol. 7252, pp. 1-18.

[47] C. Xie, M. Savvides, and B. Kumar, "Kernel correlation filter based redundant class-dependence feature analysis (KCFA) on FRGC2.0 data," in *Proc. the 2nd International Conference on Analysis and Modeling of Faces and Gestures*, 2005, pp. 32-43.

[48] A. Rani, V. Malik, and B. Chithravathi, "Robust object tracking using kernelized correlation filters (KCF) and Kalman predictive estimates," in *Proc. the 2nd International Conference on Recent Trends in Electronics Information and Communication Technology*, 2017, pp. 587-591.

[49] W. Ou, S. Yu, G. Li, J. Lu, K. Zhang, and G. Xie, "Multi-view non-negative matrix factorization by patch alignment framework with view consistency," *Neurocomputing*, vol. 204, pp. 116-124, 2016.

[50] Y. Li, J. Zhu, and S. Hoi, "Reliable patch trackers: Robust visual tracking by exploiting reliable patches," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 353-361.

[51] T. Liu, G. Wang, and Q. Yang, "Real-time part-based visual tracking via adaptive correlation filters," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4902-4912.

[52] S. Liu, T. Zhang, X. Cao, and C. Xu, "Structural correlation filter for robust visual tracking," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4312-4320.

[53] Y. Wu, J. Lim, and M. Yang, "Object tracking benchmark," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37 no. 9, pp. 1834-1848, 2015.
[54] S. Yeqing, G. Huajun, and X. Yan, "Application of adaptive extended kalman filter for tracking a moving target," *Computer Simulation*, vol. 24, pp. 210-213, 2007.
[55] S. Yin, J. Na, J. Choi, and S. Oh, "Hierarchical Kalman-particle filter with adaptation to motion changes for object tracking," *Computer Vision and Image Understanding*, vol. 115, pp. 885-900, 2011.
[56] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1409-1422, 2012.

Madah-Ul-Mustafa received the B.Eng. degree in electronics engineering from NFC Institute of Engineering and Technology, Multan, Pakistan, in 2010 and the M.Eng. in electrical and computer engineering from the South China University of Technology, Guangzhou, China in 2013. He is currently pursuing his Ph.D. in pattern recognition and intelligent systems, School of Automation Science and Engineering, South China University of Technology. He is currently with the Center for Brain Computer Interfaces and Brain Information Processing, South China University of Technology. His current research interests include pattern recognition, intelligent systems, brain–computer interfaces and machine learning.

Zhu Liang Yu received the BSEE in 1995 and the MSEE in 1998, both in electronics engineering from the Nanjing University of Aeronautics and Astronautics, Nanjing, China and the Ph.D. in 2006 from Nanyang Technological University, Singapore. He joined the Center for Signal Processing, Nanyang Technological University in 2000 as a research engineer, then as a group leader. In 2008, he joined the College of Automation Science and Engineering, South China University of Technology, China. He was promoted to be a full professor in 2010. Currently he is the deputy dean of the School of Automation Science and Engineering, the head of the Department of Intelligent Science and Technology and the deputy director of Center for Brain Computer Interfaces and Brain Information Processing. His research interests include signal processing, brain–computer interfaces, machine learning, computer vision and applications in biomedical engineering and intelligent robotics.