ICIC Express Letters
Volume **17**, Number **5**, May **2023**

 ISSN 1881-803X
pp. 513–519

SURVEY OF VISUAL OBJECT TRACKING ALGORITHMS BASED ON CORRELATION FILTER

Xiaorong Qiu[1,*], Md Gapar Md Johar[2], Jacquline Tham[3] and Lilysuriazna Raya[1]

[1]Faculty of Information Sciences and Engineering
[2]Information Technology and Innovation Centre
[3]Faculty of Business Management and Professional Studies
Management and Science University
Section 13, Shah Alam, Selangor 40100, Malaysia
{ mdgapar; jacquline; lilysuriazna }@msu.edu.my
*Corresponding author: 012018070789@sgs.msu.edu.my

Received August 2022; accepted October 2022

Abstract. *With the development of correlation filter, visual object tracking algorithms have achieved a series of progresses and arise more and more attention in object tracking researches. In order to enable more scholars to further study, this paper summarizes the research status in this field. Firstly, the general correlation filter framework is introduced. Secondly, the popular algorithms are analyzed from three categories. Finally, the possible development trend is pointed out. Through comparative analyses, it can be concluded that under complex internal and external environments, there are broad application prospects on further effective fusions of correlation filter and deep neural network in the future.*
Keywords: Computer vision, Visual object tracking, Correlation filter, Feature representation, Scale updating

1. **Introduction.** As the basic technology of visual information analysis, visual object tracking algorithms have always been a hot research direction [1,2]. The researches related on visual object tracking algorithms are mainly to simulate human sensory cognition of visual objects in computers, and give computers the ability to track specific object stably, so as to provide an important technical basis for subsequent applications, such as pedestrian monitoring [3], dynamic gesture recognition [4], and human-computer interaction [5].

In computer vision's applications and related fields, the research of visual object tracking generally refers to the researches of single object tracking algorithm [6]. It usually refers to estimating the proper area of any object through the mouse or real data tag in the video initialization stage, and then the tracking algorithm analyzes the object online to determine the area of the object in each subsequent frame, so as to realize the online tracking of any single target. The determination of the target area is generally a rectangular box surrounding the object, which is used to estimate the position, rotation angle, scale and other information of the object in each frame of the video, and output the corresponding information to display the actual state of the object in each frame. This arbitrarily selected object tracking method makes the algorithm unable to obtain sufficient prior knowledge required for tracking, puts forward higher requirements for the comprehensive learning ability of object tracking algorithm, and also increases the difficulty of object tracking research.

However, due to the arbitrariness of the objects to be tracked and the complexity and diversity of the environment in which the object is located, it is still very challenging

DOI: 10.24507/icicel.17.05.513

to achieve stable real-time tracking of the object [7]. Nowadays, the researches of object tracking algorithms have been greatly developed, and various advanced object tracking algorithms continue to emerge. Scholars in various countries are committed to the improvement and development of tracking algorithms. Many state-of-the-art algorithms have emerged. Among them, the correlation filter algorithms, with its excellent real-time performance and excellent tracking effect, can complete object positioning with high tracking efficiency.

The subsequent contents of this paper are as follows. Section 2 introduces the general correlation filter framework. Section 3 analyzes the popular algorithms in three categories. Section 4 revisits the major findings of this paper and points out the possible development trend in the future.

2. Classical Correlation Filter Tracking Algorithms.

2.1. **Origin and development.** Bolme et al. successfully integrated correlation filters into tracking applications for the first time, and proposed Minimum Output Sum of Squared Error filter (MOSSE) [8], which is not only simple in structure, but also saves a lot of time by converting convolution in time domain into point multiplication in frequency domain. However, because the object in the first frame needs to be manually framed and can only use little information of the first frame, there are few training samples, and the subsequent learning uses the estimated current frame information to update the filter, so it is easy to produce drift phenomenon, which makes the tracker unrecoverable.

Circulant Structure with Kernels (CSK) [9] is an improvement of MOSSE algorithm. After minimizing the error function, a regular term is added, and a cyclic matrix is introduced to reduce the impact on the learning filter. At the same time, it avoids solving complex nonlinear transformations, and adopts the kernel ridge regression training template to improve the tracking performance.

Kernel Correlation Filter (KCF) [10] also makes an innovation on the benchmark of CSK, using kernel function to expand the features to multi-channel to adapt to more features, replacing the original gray-scale features, increasing the representation information carried in the template, so as to enhance the robustness of the algorithm. However, the target frame of the algorithm remains unchanged, so learning in a fixed range may lead to only local information or too much background information being integrated.

Because the traditional kernel correlation filter algorithms cannot perfectly deal with the scale transformation of the target, Li and Zhu and Danelljan et al. proposed Scale Adaptive with Multiple Feature (SAMF) [11] and Discriminative Scale Space Tracker (DSST) [12] respectively in the follow-up research process. These two algorithms extract a series of multi-scale foreground and background regions based on the central point of the tracking object, and then calculate and use the maximum response value samples as the object detection results.

In addition, Wang et al. proposed Large Margin Object Tracking with Circulant Feature Maps (LMCF) [13] for the short board of the online update strategy using a fixed ratio in the traditional kernel correlation filter algorithm. The algorithm uses multi peak detection and high confidence update mechanism. Only when Average Peak to Correlation Energy (APCE) is high, the online update of the target presentation model can be carried out, so as to better avoid the apparent model being polluted by the surrounding background.

2.2. **Algorithm framework.** At present, visual object tracking algorithms based on computer vision are generally composed of two modules: algorithm initialization and online tracking. In the initialization phase of the algorithm, it is usually necessary to define an initialization area in the first frame by dragging the mouse or algorithm, which is generally the rectangular area where the subsequent tracking target is located. Then one or more image feature extraction methods can be used to quantitatively describe

the presentation model of the tracking object, including the visual description of the size of the tracking object, motion posture, feature operator and other aspects, so that the presentation model can better adapt to various tracking environments. After building the presentation model, we enter the online tracking stage. In this stage, the position of the moving object in the current frame is detected in real time through the target search and recognition mechanism, and the presentation model of the tracking object is updated online. At this time, it may also involve the fusion of multiple features and algorithms, and finally get the motion trajectory of the object.

The framework of the classical correlation filter tracking algorithm is shown as the followings. Firstly, the rectangle of the target to be tracked is selected. Then the filter is initialized and trained. In order to effectively represent the presentation model of the target, all kinds of features, such as fusion features, depth features, manual features, can be extracted from the selected target image block. Secondly, the cosine window is used to smooth the image block boundary and the correlation filter operation is performed by the discrete Fourier transform. Finally, inverse Fourier transform calculates and generates the response map.

2.3. **Algorithm principle.**

1) **Kernel ridge regression.**

In the object tracking algorithms, using the principle of kernel function and ridge regression, the original linear model can be transformed into a nonlinear model for processing nonlinear image features. In the classical KCF algorithm, w to be solved in the original ridge regression correlation function $f(x) = w^T x$ is defined as the linear combination of training samples in high-dimensional space, as shown in Equation (1).

$$f(x) = w^T x \tag{1}$$

Thus, in the high-dimensional feature space, function $f(x) = w^T x$ can be transformed into Equation (2), where $\kappa(z, x_i)$ is a kernel function used to calculate the inner product of training samples $\phi(x_i)$ and $\phi(z)$. The parameter to be optimized changes from w in main space to dual space α.

$$f(z) = w^T z = \sum_{i=1}^{N} \alpha_i \phi^T(x_i)\phi(z) = \sum_{i=1}^{N} \alpha_i \kappa(z, x_i) \tag{2}$$

Thus, the closed solution of kernel ridge regression is obtained, as shown in Equation (3). K is an $N \times N$ matrix which is used to store the inner product of two feature samples in high-dimensional space. λ is a regularization parameter and y is a regression target.

$$\alpha = (K + \lambda I)^{-1} y \tag{3}$$

2) **Training sample construction.**

In the sample training stage, regardless of the type of kernel function, its corresponding $N \times N$ kernel matrix K is a cyclic matrix, which can be expressed as Equation (4), and k^{xx} is the first row vector of cyclic matrix K.

$$K = C(k^{xx}) \tag{4}$$

Combining Equation (3) and Equation (4), Equation (5) can be obtained:

$$\alpha = (C(k^{xx}) + \lambda I)^{-1} y = \left(F diag\left(\hat{k}^{xx}\right) F^H + \lambda I \right)^{-1} y \tag{5}$$

$diag(\,)$ is used to achieve the solution of diagonal matrix. Equation (5) can be simplified to Equation (6):

$$\hat{\alpha} = \frac{\hat{y}}{\hat{k}^{xx^*} + \lambda} \tag{6}$$

In Equation (6), the solution of α is converted from time domain operation to frequency domain operation, in which $\hat{\alpha}$, $\hat{k}$ and $\hat{y}$ respectively correspond to the discrete Fourier transform of the corresponding vector. Because there is no need to solve the inverse operation of the matrix, the algorithm can greatly reduce the related computational complexity when constructing training samples.

3) **Fast detection.**

In the object detection stage, it is necessary to detect the candidate image quickly to obtain the candidate target with the maximum response. At this time, it is necessary to build the kernel matrix K^z corresponding to the training sample x and the candidate image z. The relevant definitions are given as Equation (7).

$$K^z = C(k^{xz}) \tag{7}$$

Combined with Equation (2), the prediction response of the candidate image can be calculated as

$$f(z) = (K^z)^T \alpha \tag{8}$$

Here, the diagonalization of cyclic matrix K similar to Equation (4) can also be used. Then the predicted response value can also be converted into frequency domain operation through discrete Fourier transform, and we can get Equation (9):

$$\hat{f}(z) = \hat{k}^{xz} \odot \hat{\alpha} \tag{9}$$

4) **Kernel correlation calculation.**

In the process of training sample construction and rapid detection, k^{xx} and k^{xz} are used, respectively. Because the correlation of all input vectors needs to be calculated when calculating the two kernel correlation calculations, there is a certain computational complexity bottleneck. Therefore, the algorithm also needs to simplify the kernel correlation calculation based on the cyclic shift matrix.

For radial basis kernel, the kernel correlation function has the form for the specific function $h(\,)$ and is shown in Equation (10):

$$\kappa\left(x, x'\right) = h\left(\left\|x - x'\right\|^2\right) \tag{10}$$

Element $k^{x_i x_j}$ in $\kappa(x_i, x_j)$ is

$$\hat{k}^{xx'} = \kappa\left(P^{i-1}x, x'\right) = h\left(\left\|P^{i-1}x - x'\right\|^2\right) = h\left(\left\|P^{i-1}x\right\|^2 + \left\|x'\right\|^2 - 2x'^T P^{i-1}x\right) \tag{11}$$

According to Paseval's Theorem of signal energy conservation, the permutation matrix P^{i-1} does not affect the norm of x. Therefore, Equation (11) can be simplified as

$$\hat{k}^{xx'} = \kappa\left(P^{i-1}x, x'\right) = h\left(\left\|x\right\|^2 + \left\|x'\right\|^2 - 2x'^T x\right) \tag{12}$$

Then combined with the characteristic of correlation filter in Fourier domain: $x'^T x = F^{-1}\left(\hat{x}^* \odot \hat{x}'\right)$, it can be got as Equation (13):

$$\hat{k}^{xx'} = \kappa\left(P^{i-1}x, x'\right) = h\left(\left\|x\right\|^2 + \left\|x'\right\|^2 - 2F^{-1}\left(\hat{x}^* \odot \hat{x}'\right)\right) \tag{13}$$

In the classical KCF algorithm, Gaussian kernel $\kappa\left(x, x'\right) = \exp\left(-\frac{1}{\delta^2}\left\|x - x'\right\|^2\right)$ is used in the sample training stage. δ is the width parameter of Gaussian kernel function and controls the range of radial direction. Equation (13) can be transformed into Equation (14). Similarly, $\hat{k}^{xz}$ in the target tracking stage is shown in Equation (15):

$$k^{xx'} = \exp\left(-\frac{1}{\delta^2}\left\|\left\|x\right\|^2 + \left\|x'\right\|^2 - 2F^{-1}\left(\hat{x}^* \odot \hat{x}'\right)\right\|^2\right) \tag{14}$$

$$k^{xz} = \exp\left(-\frac{1}{\delta^2}\left\|\left\|x\right\|^2 + \left\|z\right\|^2 - 2F^{-1}\left(\hat{x}^* \odot \hat{z}\right)\right\|^2\right) \tag{15}$$

3. **Analysis of Current Research Progress.** After years of developments, the theoretical researches and application practices of object tracking algorithm have been expanded to many fields. However, in terms of environmental universality and long-term tracking, the current object tracking algorithm cannot fully suit the needs of practical applications. The main challenges are as follows: problems of internal and external environment influences, counterbalance improvements in accuracy and speed, applications of algorithms in real scenes.

In view of the above challenges and problems, the existing object tracking algorithms are analyzed and discussed from three categories: basic class, component class, and regularization class.

3.1. **Basic class.** The basic class takes the traditional KCF as the basic framework, aiming to improve the defects of the traditional KCF and deal with different tracking problems. Specifically, these trackers are optimized from the aspects of feature representation, scale change processing, kernel function, long-term tracking, response distribution and algorithm integration.

1) **Optimization of feature representation**

Chao et al. introduced rich hierarchical convolution features into the correlation filter framework, and then proposed Hierarchical Convolutional Features (HCF) [15]. Under the traditional framework, the tracker uses the specific three-layer features trained by VGG-19 in ImageNet to replace the original HOG features, and configures independent correlation filters for each layer of features to learn the template. After obtaining the confidence map, it performs weighted fusion to obtain the target location.

2) **Optimization of scale change processing**

Montero et al. proposed a rapid and scalable scheme based on the KCF [16], which deals with the scaling problem by introducing adjustable Gaussian kernel function and inter frame key point matching technology.

Huang et al. introduced edge boxes into KCF and proposed Kernelized Correlation Filter with Detection Proposal (KCFDP) to deal with the changes of object scale and aspect ratio by generating target candidate regions [17].

3) **Optimization of kernel function**

Tang and Feng proposed a tracking algorithm based on multi-kernel correlation filter to solve the problem that the basic correlation filter only uses one single kernel [18]. The algorithm makes full use of power spectrums of different features and their discriminant invariance to improve performance, and uses optimal binary search and fast feature estimation for scale estimation. At the same time, it uses the minimum number of layers of the feature pyramid to effectively reduce the amount of computation.

4) **Optimization of long-term tracking**

Chao et al. respectively adopted three kinds of filters based on DSST: translation, scale and confidence [19]. In the tracking process, the translation estimation is realized by modeling the time context related information, and scale pyramids are constructed by the presentation model to realize the scale estimation, and the on-line random ferns detector is used to realize the redetection in the case of target loss, which better improves the robustness of tracking in the case of removal of field of view and severe occlusion of the target.

5) **Optimization of response distribution**

Because the single centered Gaussian is commonly used as the target response, this traditional method will imped the performance of the tracking algorithm, and may cause unrecoverable drift. Bibi et al. set up a general algorithm framework and reduced the defect that the tracker cyclic shift cannot reliably approximate the conversion [20].

6) **Optimization of algorithm integration**

Because each tracking algorithm makes assumptions according to the different tracked

objects and environments, different algorithms can reflect their own advantages and disadvantages. Therefore, integrating different algorithms to achieve complementary advantages is an effective way to improve tracking performance.

3.2. **Component class.** Corresponding to the above global appearance model, many trackers set up target appearance model by local block strategy.

Xin et al. proposed a non-rigid object tracking method based on dynamic deformable component set [21]. Among them, the shape preserving kernel correlation filter is introduced into the level set framework to dynamically track a single target block, which has the ability to assume a complex topology. When deformable components capture a single target sub region, photometric discrimination and shape change are used to display the tracking performance of a single target sub region. Then the sub regions with good traceability are dynamically selected for likelihood estimation, and finally the target contour is determined.

The block model combines local image features with geometric features, which is a powerful example of visual object tracking, and has the ability to deal with the deformation of partially occluded objects and the change of viewing angle. The difficulties lie in the following aspects. How to effectively use the spatiotemporal confidence map of each component to estimate the global target location? How to deal with the spatial position relationship between components, the relationship between global targets and local blocks, and the calculation of component reliability in the case of occlusion deformation? The number of blocks and the relationship between spatial structure are closely related to the performance of tracking. How to achieve low computing cost in dense search is a very challenging problem.

3.3. **Regularization class.** The tracking performance of correlation filter is mainly limited by the following two aspects. Firstly, the condition that the filter size needs to be equal to the block size limits the detection range. Secondly, the fixed search area will cause the loss of negative samples in the training sample set, and it is difficult to restart the tracking when the object is severely occluded. Although choosing a larger search area can solve this problem, the use of too much background information will reduce the discrimination ability of the tracker. Therefore, researchers introduce regularization strategy to solve these problems.

Danelljan et al. proposed Spatially Regularized Discriminative Correlation Filters (SR-DCF) [22]. It reduces boundary effects by using the spatial weight function. Meanwhile, because the weight is fixed in the whole sequence, it cannot enhance the object with the change of the object's shape.

Lukezic et al. proposed Discriminative Correlation Filter Tracker based on Channel and Spatial Reliability-Discriminative Correlation Filter (CSR-DCF) [23]. It combines the color probability and optimizes it by using the direct summation of multi-channel features. Meanwhile, it also uses the spatial confidence map to support the filter to adaptively select the object region suitable for tracking, which reduces the boundary effect and the limitation of the rectangular hypothesis.

4. **Conclusion.** This paper analyzes and summarizes the visual correlation filter tracking algorithms. The research of these algorithms in basic class, component class, regularization class has become a medium and long-term research hotspot. Recently, the trackers which use deep neural networks have launched a strong challenge to correlation filter tracking. Because the correlation filter framework need not rely on the depth model and large-scale training data sets, it still maintains strong tracking accuracy and robustness to the visual object tracking with good real-time performance and wide adaptability. With the maturity and development of deep learning, how to effectively integrate correlation filter and deep neural network to achieve tracking performances with high accuracy, strong

robustness, good real-time performance and wide application is a long-term goal of visual target tracking algorithm.

REFERENCES

[1] H. Fan, L. Lin, F. Yang et al., LaSOT: A high-quality benchmark for large-scale single object tracking, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.5369-5378, 2019.

[2] M. Kristan, J. Matas, A. Leonardis et al., The ninth visual object tracking VOT2021 challenge results, *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pp.2711-2738, 2021.

[3] J. Li and Z. Wu, The application of YOLOv4 and a new pedestrian clustering algorithm to implement social distance monitoring during the COVID-19 pandemic, *International Conference on Advances in Optics and Computational Sciences (ICAOCS)*, vol.1865, no.4, 2021.

[4] M. Oudah, A. Al-Naji and J. Chahl, Hand gesture recognition based on computer vision: A review of techniques, *Journal of Imaging*, vol.6, no.8, 2020.

[5] J. Huang, Z. Zhang, G. Xie et al., Real-time precise human-computer interaction system based on gaze estimation and tracking, *Wireless Communications and Mobile Computing*, 2021.

[6] Y. Wu, J. Lim and M. H. Yang, Object tracking benchmark, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.37, no.9, pp.1834-1848, 2015.

[7] A. Li, M. Lin, Y. Wu et al., NUS-PRO: A new visual tracking challenge, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.38, no.2, pp.335-349, 2016.

[8] D. S. Bolme, J. R. Beveridge, B. A. Draper et al., Visual object tracking using adaptive correlation filters, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.2544-2550, 2010.

[9] J. F. Henriques, R. Caseiro, P. Martins et al., Exploiting the circulant structure of tracking-by-detection with kernels, *Proc. of the Euopean Conference on Computer Vision (ECCV)*, pp.702-715, 2012.

[10] J. F. Henriques, R. Caseiro, P. Martins et al., High-speed tracking with kernelized correlation filters, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.37, no.3, pp.583-596, 2015.

[11] Y. Li and J. K. Zhu, A scale adaptive kernel correlation filter tracker with feature integration, *Proc. of the Euopean Conference on Computer Vision (ECCV)*, pp.254-265, 2014.

[12] M. Danelljan, G. Häger and F. S. Khan, Accurate scale estimation for robust visual tracking, *Proc. of the British Machine Vision Conference (BMVC)*, 2014.

[13] M. Wang, Y. Liu and Z. Huang, Large margin object tracking with circulant feature maps, *IEEE Computer Society*, pp.4021-4029, 2017.

[14] M. Kristan, J. Matas, A. Leonardis et al., The seventh visual object tracking VOT2019 challenge results, *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp.2206-2241, 2019.

[15] M. Chao, J. B. Huang, X. Yang et al., Hierarchical convolutional features for visual tracking, *2015 IEEE International Conference on Computer Vision (ICCV)*, pp.3074-3082, 2015.

[16] A. S. Montero, J. Lang and R. Laganière, Scalable kernel correlation filter with sparse feature integration, *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pp.587-594, 2015.

[17] D. Huang, L. Luo, M. Wen et al., Enable scale and aspect ratio adaptability in visual tracking with detection proposals, *British Machine Vision Conference (BMVC)*, pp.1-12, 2015.

[18] M. Tang and J. Feng, Multi-kernel correlation filter for visual tracking, *2015 IEEE International Conference on Computer Vision (ICCV)*, pp.3038-3046, 2015.

[19] M. Chao, X. Yang, C. Zhang et al., Long-term correlation tracking, *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.5388-5396, 2015.

[20] A. Bibi, M. Mueller and B. Ghanem, Target response adaptation for correlation filter tracking, *Proc. of the European Conference on Computer Vision (ECCV)*, pp.419-433, 2016.

[21] S. Xin, N. M. Cheung, H. Yao et al., Non-rigid object tracking via deformable patches using shape-preserved KCF and level sets, *2017 IEEE International Conference on Computer Vision (ICCV)*, pp.5495-5503, 2017.

[22] M. Danelljan, G. Häger, F. S. Khan et al., Learning spatially regularized correlation filters for visual tracking, *2015 IEEE International Conference on Computer Vision (ICCV)*, pp.4310-4318, 2015.

[23] A. Lukezic, T. Vojir, L. C. Zajc et al., Discriminative correlation filter with channel and spatial reliability, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.6309-6318, 2017.