# Improvement of MOSSE Object Tracking Using YOLO and Adaptive Multi-Filter

Chunhui Chung[1,*] and Yi-Lung Chen[2]

***Abstract*—This research aims to improve the object tracking algorithm, Minimum Output Sum of Squared Error (MOSSE). MOSSE is one of the correlation filter-based object tracking algorithms. Its processing speed is fast and easy to implement. However, it is easy to lose the target during tracking because of the deformation, rotation, and occlusion of the object. Two methods were proposed to improve MOSSE. They are re-tracking mode and adaptive multi-filter. Re-tracking mode utilizes YOLO to search the candidate objects on the image once the target is lost. The correlation would be manipulated between the filter and the possible object to determine the position of the lost target according to the peak to sidelobe ratio. In addition, adaptive multi-filter method cropped different images of the target appearance in the video to get the multiple templates. More than one filter was generated based on these templates to improve the tracking performance of MOSSE. These filters were created based on the change of the target appearance such as deformation, illumination variation, and rotation, which would cause the target lost because one filter could not satisfy all the target appearance in the video. The tracking performance was tested by 27 videos in the Object Tracking Benchmark 50 (OTB-50) database. The experimental results showed that the Spatial Robustness Evaluation (SRE) was improved from 0.215 to 0.291, and the Temporal Robustness Evaluation (TRE) was from 250 to 0.346 with the re-tracking mode and adaptive multi-filter.**

***Index Terms*—Visual object tracking, Correlation filter, Object detection.**

## I. Introduction

VISUAL object tracking is getting popular and essential in many applications. A drone with object tracking can help the operator to mark the location to land off [1] or choose a moving object to follow automatically [2-3]. The object tracking was integrated with the electro-optical targeting system for the military [4]. With the monitoring system on the road, it can record the traffic flow to help managers in traffic control. Object tracking is also applied to remote assistance system. The object can be marked and tracked for the communication between the engineers and customers to solve problems by sharing video and augmented reality [5]. With the development of computing technology, the real-time response can be realized for advanced applications. Once the target was labelled and analyzed by the tracking algorithm, the tracker would estimate the position of the target automatically in the next frames.

Many algorithms have been proposed for visual object tracking, and they are classified according to their characteristics [6-9]. Trackers can be categorized as generative [11-12] or discriminative [13-15], single-object [16] or multiple-object [17], online [18] or offline learning [19], correlation or noncorrelation filter. The challenge of object tracking is the complex scene in the frame, which would affect the tracking performance. There are several challenges such as illumination variation (IV), occlusion (OCC), scale variation (SV), deformation (DEF), motion blur (MB), fast motion (FM), in plane rotation (IPR), out-of-plane rotation (OPR), out of view (OV), background clutters (BC) and low resolution (LR) [7, 20].

Researchers are interested in the correlation-based tracking algorithms because of its simplicity and quick response. The concept of the correlation filter, Minimum Output Sum of Squared Error (MOSSE), was first proposed by Bolme et al. [21]. Correlation filters estimate the position of target with the peak correlation value of the response in the scene during tracking. The maximum correlation value indicates the maximum similarity with the template of the labeled target and where the object is. A lot of correlation-based algorithms are proposed after MOSSE, such as CSK [22], KCF [23], and DSST [24]. The complex scene affects the tracking performance and probably causes the tracker to lose the target. For the problem of losing target, Shin et al. [25] tracked the target with the KCF algorithm and defined the tracking state by the tracking failure detection. When the target is lost, it would be tracked again by using the multiple search windows surrounding the location where the target is lost. However, the re-tracking mechanism would not work if the target was not in the search field, and the searching region is limited by the target position. Yuan et al. [26] improved the KCF by relocating the position of the target using particle filter redetection when the result of tracking response is ambiguous or unreliable. The increase of the number of particles could improve the tracking performance but scarify the computational efficiency.

Minimum Output Sum of Squared Error (MOSSE) is fast and easy to be implemented. However, MOSSE is easy to lose the target if the scene was complex or the target was out of the screen. In this study, the re-tracking mode using YOLO was investigated to improve MOSSE in the performance of object re-tracking when the target is out of view or blocked. In addition, the adaptively multi-filter was integrated in MOSSE to create multiple filters instead of one to overcome the challenges of image deformation, out-of-plane rotation, illumination variation, etc. The multiple filters can help to identify the target under different object appearance. The experimental results showed that the Spatial Robustness Evaluation (SRE) improved from 0.215 to 0.291, and the Temporal Robustness Evaluation (TRE) was from 250 to 0.346.

[1]C. Chung is with the Department of Mechanical Engineering, National Cheng Kung University, Tainan 701, Taiwan (e-mail: chchung@mail.ncku.edu.tw).

[2]Y. L. Chen was with National Cheng Kung University, Tainan 701, Taiwan. He is now with Da Shang Automation Co., Ltd., Taoyuan 330, Taiwan (e-mail: leogash51813@gmail.com).

## II. Minimum Output Sum of Squared Error

MOSSE is one of the first correlation-based object tracking algorithms. The location of the target is tracked with high similarity to the template in the new frame. According to the convolution theorem, the convolution of two patches in the spatial domain is equivalent to the element-wise product in the frequency domain. Therefore, the correlation tracker has the advantages of low computation cost and high processing speed and is ideal for real-time applications.

The MOSSE tracker correlates the target by using a cropped target image, $f$, from the video frame as the input image. The filter, $h$, is obtained by calculation of the initial labeled object and ideal output response of the correlation. The symbol $g$ is the output response after correlation operation, and the value stands for the similarity between the filter $h$ and the input image $f$. The formula is:

$$g = f \circ h \tag{1}$$

where the symbol $\circ$ represents the correlation operator. To increase the processing speed, correlation is computed in the frequency domain by Fast Fourier Transform (FFT). According to the Convolution Theorem, the correlation becomes an element-wise multiplication in the frequency domain. The filter must be defined to correlate the image and estimate the position of the target. One of the methods to define the filter is minimum output sum of squared error between the actual output and the ideal output. The equation is as follows.

$$\min_{H^*} \sum_i |F_i \odot H^* - G_i|^2 \tag{2}$$

The uppercase variables $F_i$, $G_i$ and the template $H$ are the Fourier transforms of their lowercase counterparts., $H^*$ represents the Hermitian transpose of $H$, and the symbol $\odot$ represents the element-wise product. In the algorithm, it is $H^*$ being calculated and updated as the filter because the closed form expression of $H^*$ can be derived by the following equation:

$$H^* = \frac{\sum_i G_i \odot F_i^*}{\sum_i F_i \odot F_i^*} \tag{3}$$

With the (3), the target can be tracked by the following formula:

$$G = F \odot H^* \tag{4}$$

Here the $F$ represents the Fourier transform of the new patch cropped from the new frame, and $H^*$ is the template that contains the information about the appearance of the target. After multiplication by element-wise in the frequency domain, the output $G$ can turn back to the $g$ by the Inverse Fast Fourier Transform (IFFT). The peak of the response of $g$ indicates the position where the new patch is highly similar with the template of the target.

The index to evaluate the confidence of the predicted position in MOSSE is peak to sidelobe ratio (PSR). The mathematical expression of PSR is as follows:

$$PSR = \frac{g_{max} - \mu}{\sigma} \tag{5}$$

where $g_{max}$ is the peak value of $g$, $\mu$ and $\sigma$ are respectively the mean and standard deviation of the $g$ values of the sidelobe. The sidelobe is the rest of the pixels excluding the 11 x 11 window around the peak. PSR is used to evaluate that the target position estimated by MOSSE is reliable or not.

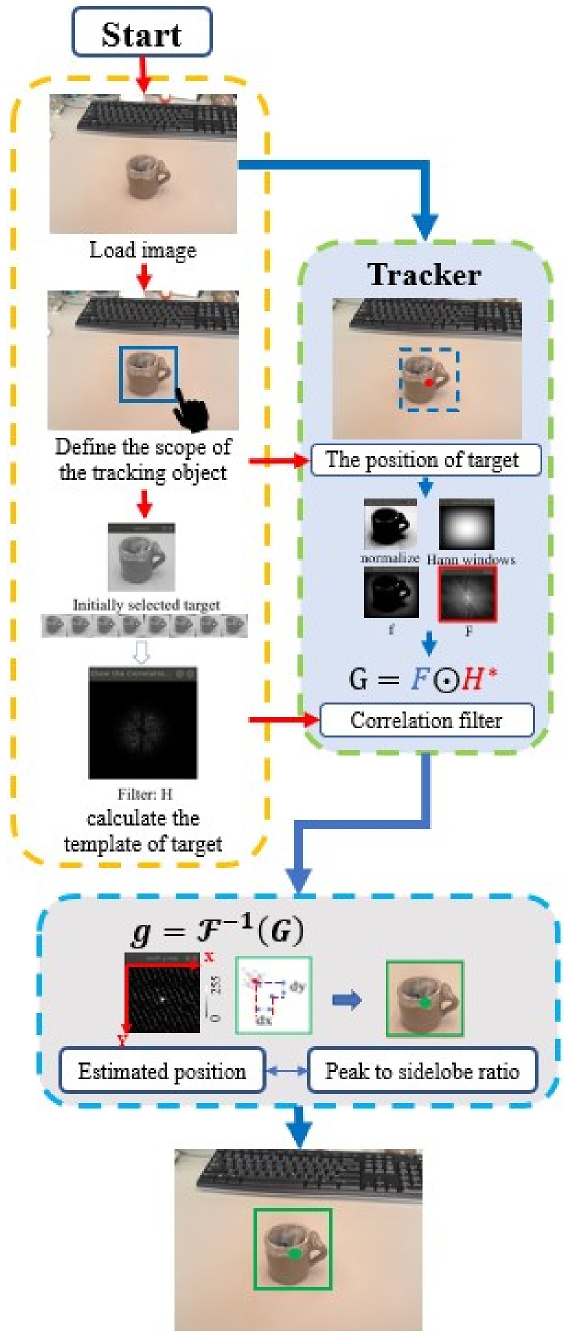
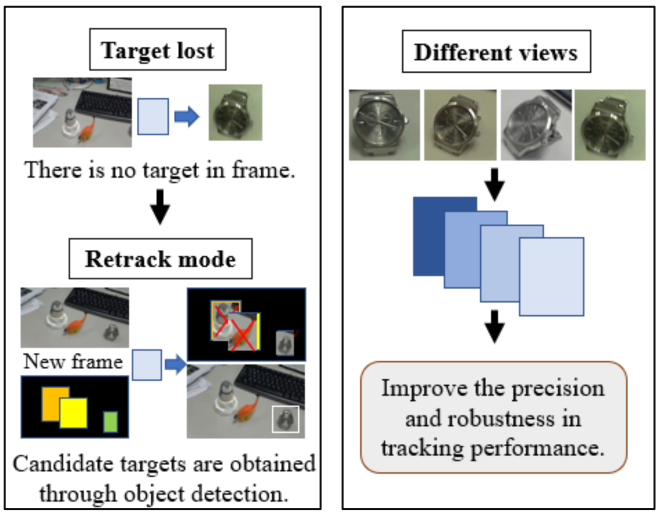



Fig. 1. The flow chart of MOSSE object tracking algorithm.

Figure 1 illustrated the flow chart of MOSSE. The procedure was divided into three parts: (1) initializing (orange), (2) calculating the correlation filter (green) and (3) estimating the position of the target (blue). The target is defined by the user at the beginning and the range of cropped image is used to generate the filter $H^*$. Next, the tracker starts tracking the target. It acquires the cropped image from the new frame based on the target position which is defined at initialization or the previous estimated position. The output $G$ was obtained by the element-wise product of the FFT of the cropped image, $F$, and the filter $H^*$. Finally, we can estimate the position of the target in the time domain by the peak value in the output $g$.

The MOSSE correlation filter is easily affected by the changing appearance of the target and to lose it. In this study, the tracking performance is improved by two approaches: (1) when the tracker loses the target, the re-tracking mode starts, and YOLO is employed to track the target again, as in Fig. 2(a); and (2) multiple filters were proposed to track the target in different appearance to improve the tracking performance, as in Fig. 2(b).




(a) Re-tracking mode with YOLO. (b) Adaptive multiple templates

Fig. 2. Illustration of (a) re-tracking mode with YOLO and (b) adaptive multi-filter method.

## III. Improvement of MOSSE

### *A. Re-tracking Mode with YOLOv4*

One of the disadvantages of MOSSE is that it cannot track the object back once the target is lost. The algorithm would determine whether the object is lost or not based on the PSR value. The PSR value could be low because the object was in different appearance or out of view. However, the lost object cannot be re-tracked by MOSSE even it shows up again in the video frame. To overcome this problem, the object detection algorithm, YOLO, was employed to determine the objects and their positions on the frame. Afterward, MOSSE is performed to determine correlation values for these objects. The detected object which has the highest correlation value with the filter was defined as the object.

In this re-tracking mode, the algorithm is divided into three parts as shown in Fig. 3. (1) MOSSE is used as the basic object tracking algorithm to track the target in the video. (2) When it is under tracking, the two-stage classifier was designed to initiate the re-tracking mode if the target is lost. (3) In the re-tracking mode, the estimated position of the target could be not correct. YOLO object detection provides possible positions of candidate targets to the tracker to trace back the target. According to the positions detected by YOLO, we crop the patches and calculate the correlation values with the filter. The position with the maximum PSR value among all the positions is determined as the re-tracked target position if the PSR value is greater than the threshold. The algorithm would return into the tracking mode and resume to track the target with the new position. Or the algorithm would stay in the re-tracking mode until the target shows up or terminates the code. Although PSR has been used to determine whether the target is under tracking or not in the original MOSSE algorithm, it misses the target sometimes even when the target is still in the frame. The two-stage classifier with multilayer perceptron (MLP) is proposed to improve the inference of target existing or not. This can reduce the usage of the re-tracking mode and the unnecessary computational costs.

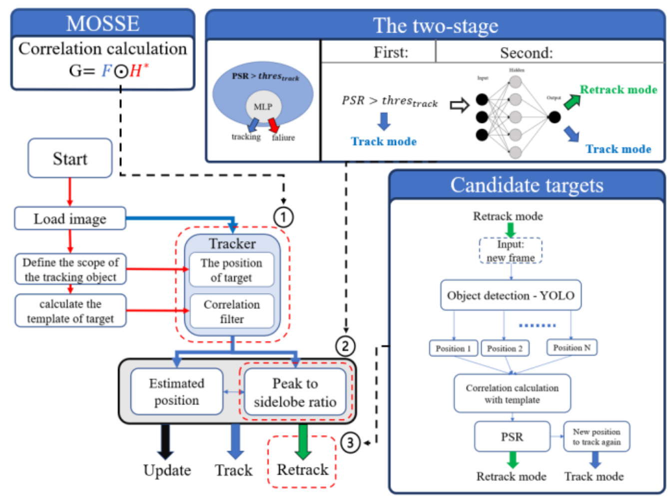



Fig. 3. The flow chart of re-tracking mode with YOLO divides into three parts: (1) MOSSE: The main object tracking algorithm. (2) The two-stage classification of tracking mode or re-tracking mode. (3) Possible positions of the candidate targets detected by YOLO.

According to Bolme et al. [21], when the PSR drops to about 7.0, it indicates that the object is occluded or the tracking is failed. The tracker would keep tracking if the PSR value is greater than the threshold. If it is not, the multilayer perceptron (MLP) model for the second stage classification was proposed here. The MLP model in this application is one hidden layer with five neurons. The input features are the PSR, the peak, average, and standard deviation of the output response. The output of the model is binary classification, where 0 means that the target is lost, and 1 is not. Compared the two-stage classification to the fixed threshold method, the two-stage classification can improve the accuracy to determine whether the target is lost or not. Once the target is lost, the re-tracking mode using YOLO would be initiate to tracks back the target. Using two-stage classification can reduce the misjudgment of the lost target and the initiation of YOLO algorithm, which cost extra computing resource.

### *B. Adaptive Multi-Filter*

Another disadvantage of MOSSE is that the object filter was established with 2-dimensional image. The tracking would be failed because of the deformation, rotation, and the appearance change of the object. Adaptive multiple-filter is proposed in this study to overcome these problems. The concept of this method is inspired by tracking the target from different perspectives or the appearance, as shown in Fig. 4.

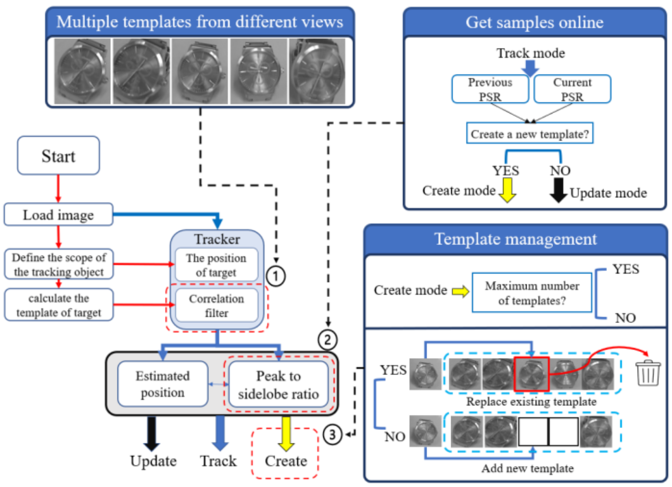



Fig. 4. The flow chart of adaptive multi-filter in MOSSE. (1) Use multiple filters to track the target. (2) Define how to get a new sample during tracking. (3) Define how to manage the multiple filters.

The filters are created during the tracking process except the initial one. The PSR value is evaluated during tracking. If the appearance of the target changes a lot, the PSR would drop sharply and usually below the tracking threshold in the next few frames. In this situation, the tracker would stop to estimate the position of the target and often miss the target in the end. To solve this problem, the idea is to adaptively create a new filter according to the change of PSR values. When the PSR in the previous frame, $PSR_{t-1}$, minus the PSR in the current frame, $PSR_t$, is larger than the preset threshold, as in (6), then the new content is cropped in the current position to create a new filter.

$$(PSR_{t-1} - PSR_t) \geq threshold_{create} \tag{6}$$

These filters are correlated to the object images took from different appearances of view. When the target changes a lot in the video, the PSR would drop sharply. A new filter is created immediately in current frames before losing the target. The sample we cropped is according to the estimated position in the previous frame, and it is highly prossible that the cropped region still has the target. The cropped image would include the latest target appearance information to help the tracker continue tracking adaptively before the target is lost. The way of managing the multiple filters during the adaptive tracking is the following. The maximum number of filters is given first, which was five in this study. If the filter number is below the set value, A new template is created directly if the criteria (6) is satisfied. If not, we need to choose an existing filter and replace it with the new one. The goal of the filter management is to keep the filter different from each other. The existing filter with the second largest PSR value is replaced. It is found that the performance is better than replacing the existing filter according to the largest PSR one. With the adaptive multi-filter, the performance of MOSSE is improved, especially for the video with the scene factors of deformation and illumination variance.

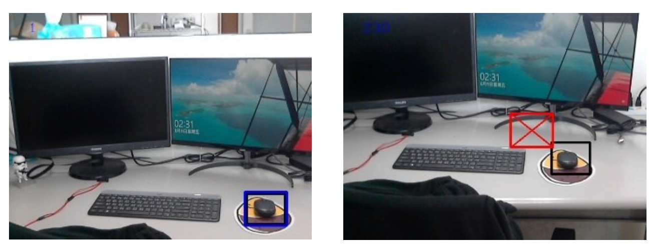
The re-tracking mode and the adaptive multi-filter method were studied with the MOSSE algorithm. The adaptive multi-filter creates new filters adaptively during tracking, and tracking with the multiple filters that are different greatly from each other improves the performance of the object tracking. With the addition of re-tracking mode, the algorithm can actively determine whether the target is lost or not by the two-stage classification. When the target is lost, it can be tracked again with the YOLO detection which provides the candidate positions. The experimental results are presented in the next section, which show that the re-tracking mode and multi-filter improve the capability of the MOSSE object tracking.

## IV. EXPERIMENTAL RESULTS

The re-tracking mode and multi-filter method were evaluated by self-made videos and 27 videos from OTB50 dataset. The temporal robustness evaluation (TRE) and spatial robustness evaluation (SRE) are two indices to compare the performance of the object tracking. The self-made videos were used to evaluate the performance of different target appearances.

### *A. Evaluation of Re-Tracking Mode and Milti-Filter*

The re-tracking mode using YOLO was evaluated first to test the improvement of the robustness and accuracy. The position of the target was labeled by the bounding box in every frame. In our video, there are a few frames in which the target does not exist and we cannot label it. In this situation, the bounding box (0,0,0,0) was assigned to the frame, which means that there is no target in the scene. As shown in Fig. 6, the picture is one of the frames in the testing video. A blue bounding box around the target represents good estimation of the tracker, as shown in Fig. 5(a). The tracking with low PSR value was presented by a red bounding box with a cross inside to indicate that the target is lost, as shown in Fig. 5(b). The black bounding box represents the ground truth. The criteria of successful tracking are defined as the center distance less than 20 pixels and the intersection over union (IOU) larger than 0.5 between the ground truth and the estimated bounding box.



(a) Successful tracking with a blue box.
(b) Target lost is indicated by a cross in the red box.

Fig. 5. Colors of bounding boxes.

There are total 322 frames in the self-made video. In the frames 67 to 71 and 101 to 103, the occlusion happens. In the frames 172 to 201, the target is out of view. The tracker would lose the target in these situations. The tracker performances with and without re-tracking mode are plotted in Fig. 7 and Fig. 6, respectively. The horizontal axis is the frame number, and the vertical axis is the PSR value. The green curve represents the PSR value in the corresponding frame. The red dots mark the frames in which the tracking was failed. The failure was defined by the distance between estimated the position and the ground truth based on the center distance and IOU. The blue dots indicate that the target is out of view in these frames. The dashed lines in black color and blue color represent the thresholds of updating and tracking, respectively. If the PSR was below the updating threshold, the algorithm will not update the template information. If the value was below the tracking threshold, the algorithm would not accept the estimated position in this frame, and the tracking is determined as failed. When the target is lost, it is difficult for the MOSSE tracker to track back the target. This is because of the complex scene factors such as occlusion and disappearance of the target. As shown in Fig. 7, the MOSSE tracker cannot re-track the target in the second half of the video. By contrast, the re-tracking mode can track the target again after losing the target, as shown in Fig. 6.

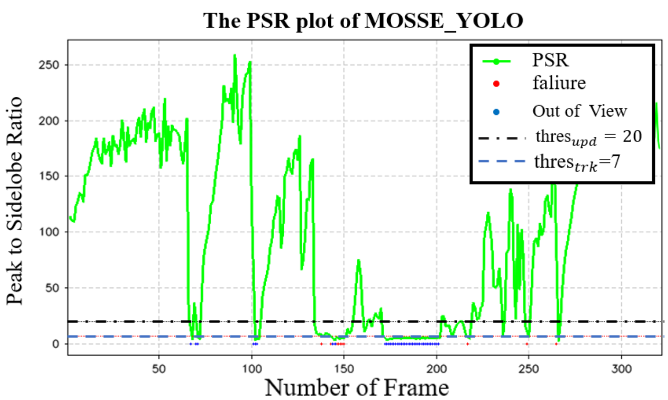



Fig. 6. The testing results of MOSSE with the re-tracking mode.

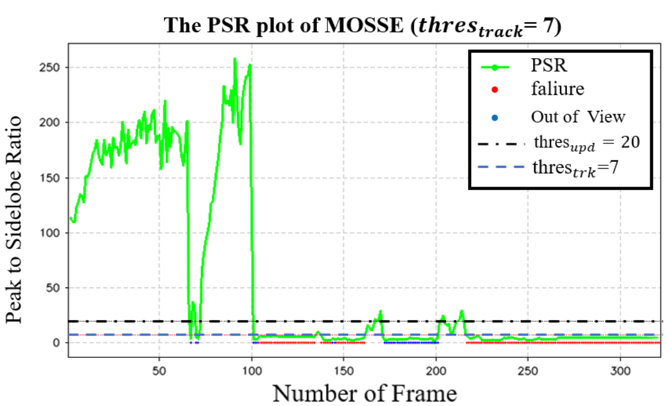



Fig. 7. The testing results of MOSSE without the re-tracking mode.

Fig. 8 illustrated the objects detected by YOLO detection after the target was lost because of the occlusion. The pro of using YOLO is that it can provide the positions of the candidate objects in the whole frame. The result of the YOLO detection helps to find the target by the correlations of the filter and the possible objects. The cons of using the YOLO detection are that it can only detect the object which has been trained in advance, and the object must be complete instead of part of it to be detected. In addition, YOLO consumes extra computational cost during the re-tracking mode.

(a) (b)

Fig. 8. (a)The objects detected by YOLO in the image and (b) the target tracked by the filter from the objects.

Figure 9 compared the results of MOSSE with and without adaptive multi-filter. The self-made video was tested. The different box color in Fig. 9(b) presents the target position estimated by the different filters. The images show that the adaptive multi-filter method has a good performance than the original MOSSE algorithm. The PSR values were plotted in Figs. 10 and 11. The marks with a red circle and blue square represent the filter is updated or a new one is created, respectively. MOSSE can update the filter information during the tracking once the PSR value is large than the threshold, as shown in Fig. 10. In the frame 102, the target was lost because of the large change in the appearance of the target. The adaptive multi-filter not only updates the existing filter but also creates a new filter based on the newest appearance of the target. In Fig. 12, when the appearance of the target changes a lot, the PSR dropped sharply and a new filter was created before the PSR was below the tracking threshold. The created filter has the newest information of the target appearance and is helpful to continue the tracking. Fig. 12 shows the contents of the target which were obtained during tracking from different target appearances. These were adaptively created during the tracking by the adaptive multi-filter algorithm.

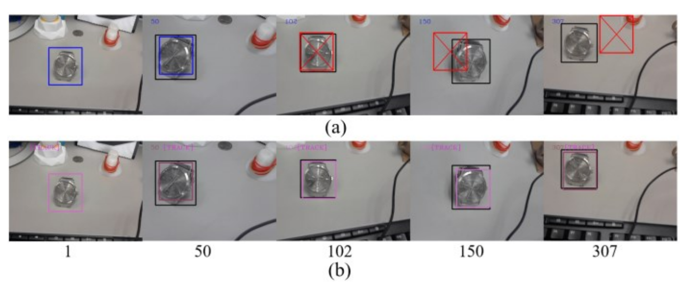



Fig. 9. The result of the tracking process with (a) the original MOSSE algorithm and (b) MOSSE with adaptive multi-filter.

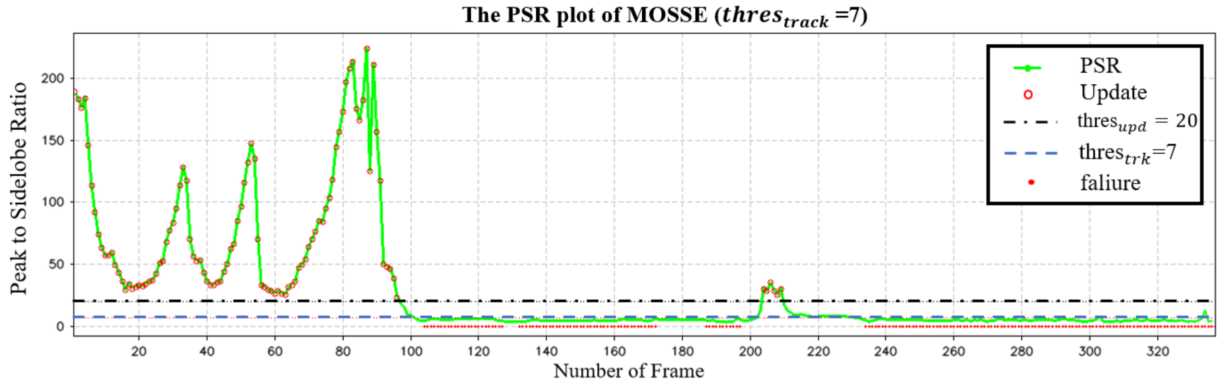



Fig. 10. The PSR plot of the testing results of the original MOSSE algorithm. Part of the tested video was presented in Fig. 9.

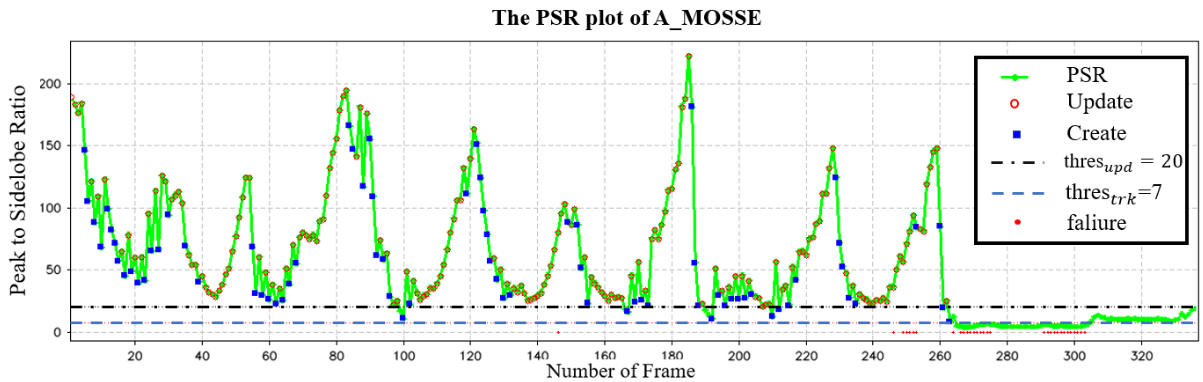



Fig. 11. The PSR plot of the testing results of the MOSSE algorithm with multi-filter method. Part of the tested video was presented in Fig. 9.

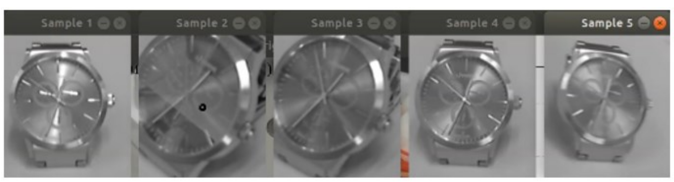



Fig. 12. Different appearances of the target selected to create the filters

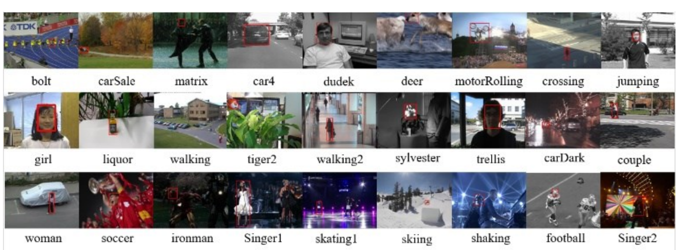



Fig. 13. The videos selected from the OTB50 dataset.

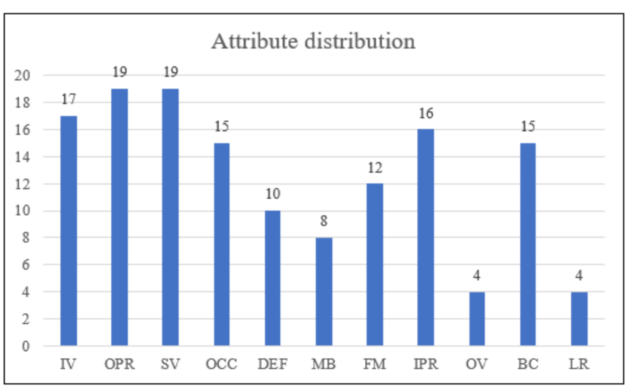
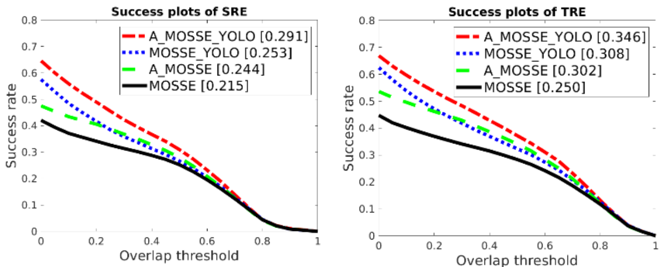



Fig. 14. The distribution of the scene factor in the selected videos.

The pro of object tracking with the adaptive multi-filter is that we can immediately create a new template with the newest target appearance before losing the target. It can track the target with the filters which come from the different perspectives of the target appearance. The con is that it will slightly slow down the processing speed because of the multiple filters and sometimes it might not improve the tracking capability.

### B. Evaluation by OTB50 Benchmark

Twenty seven videos from the OTB50 database were used to test the re-tracking mode, as shown in Fig. 13. The videos were specially chosen according to the tracking target to test the MOSSE with re-tracking mode and multi-filter. The YOLO detection provided candidate positions during the re-tracking, and the target must be detectable by the YOLO detection. The distribution of the scene factors in the selected videos is shown in Fig. 14., e.g., 19 videos have scale variation scene factor during the tracking. The object tracking algorithms in this evaluation are the original MOSSE algorithm, MOSSE with re-tracking mode, MOSSE with adaptive multi-filter, and MOSSE with both re-tracking mode and adaptive multi-filter, which are represented by MOSSE, MOSSE_YOLO, A_MOSSE, and A_MOSSE_YOLO, respectively. For the OTB50 database, the performances of the tracking algorithms were measured by the TRE and SRE indices. The evaluation matric we used here are the bounding box overlap and the success plot. The success plots show the ratios of success at the IOU threshold varied from 0 to 1 with the interval of 0.05. The area under curve (AUC) of each success plot was used to rank the tracking algorithms [20].

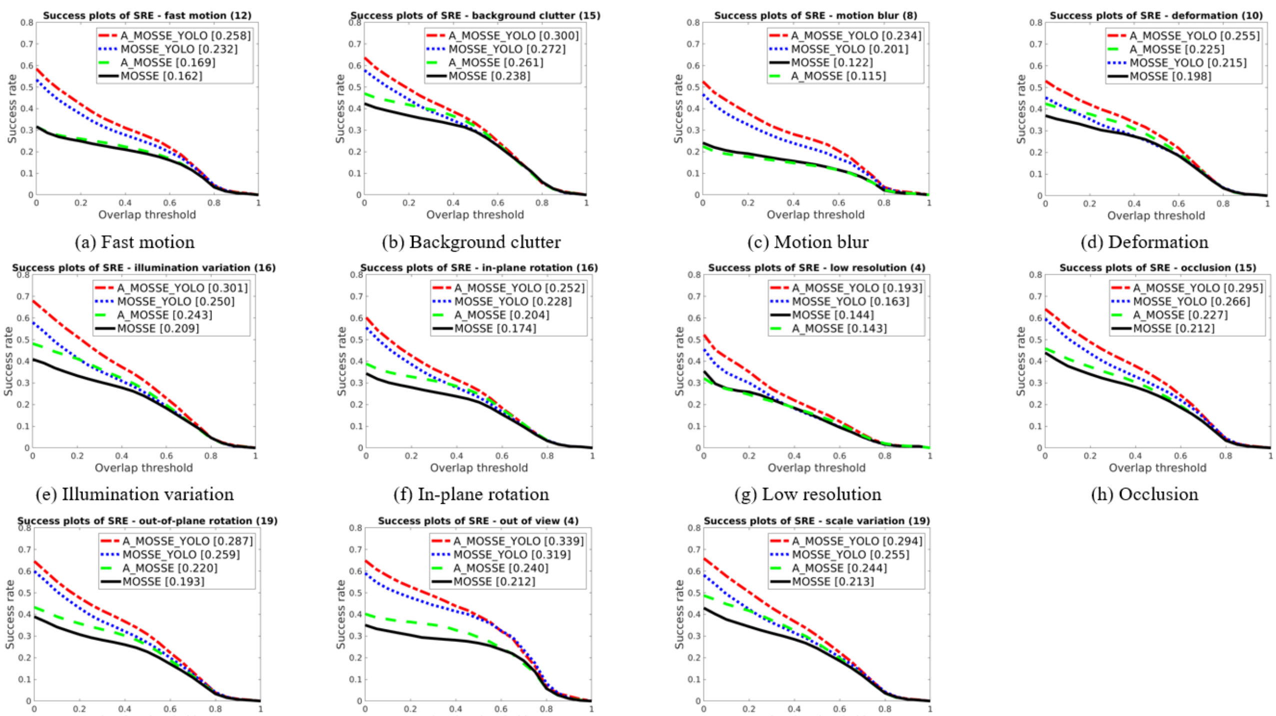


(a) The success plots of SRE (b) The success plots of TRE

Fig. 15. The success plots of SRE and TRE evaluations on 27 videos from OTB50 database.

The robustness evaluation is divided into the temporal robustness evaluation (TRE) and spatial robustness evaluation (SRE), which analyze the tracker capability by perturbed in the initialization temporally (i.e., start by different frames) and spatially (i.e., start by different initial bounding box). The SRE and TRE performances of the trackers are shown in Fig. 15. Including the re-tracking mode improves the SRE from 0.215 to 0.253 and TRE from 0.250 to 0.308. Adding the re-tracking mode improved MOSSE over 17.6% and 23.2%. The adaptive multi-filter improved MOSSE to 0.244 in SRE and 0.302 in TRE. Compared with the MOSSE, the adaptive multi-filter improved over 13.5% and 20.8%. Including both re-tracking mode and adaptively multi-filter achieves a score of 0.291 in SRE and 0.346 in TRE. Compared with the MOSSE, the method developed in this study improved over 35.3% and 38.4%.

The SRE and TRE performances of the trackers on each scene factor are shown in Fig. 16 and Fig. 17, which demonstrate the tracking performance on the 11 attributes. For the attributes of fast motion, motion blur, occlusion, and out of view, adding re-tracking mode has a great improvement over the MOSSE performance. The main reason is that these scene factors make the tracker lose the target easily. With the proposed re-tracking mode, the target can be re-tracked with the candidate positions provided by YOLO from the whole frame to improve the tracking performance in these attributes. For the attributes of background clutter, deformation, and in-plane rotation, the appearance of target changes a lot. The adaptive multi-filter method has achieved pretty performance, which can track the target with multiple filters or create a filter from the newest appearance of the target before losing the target.

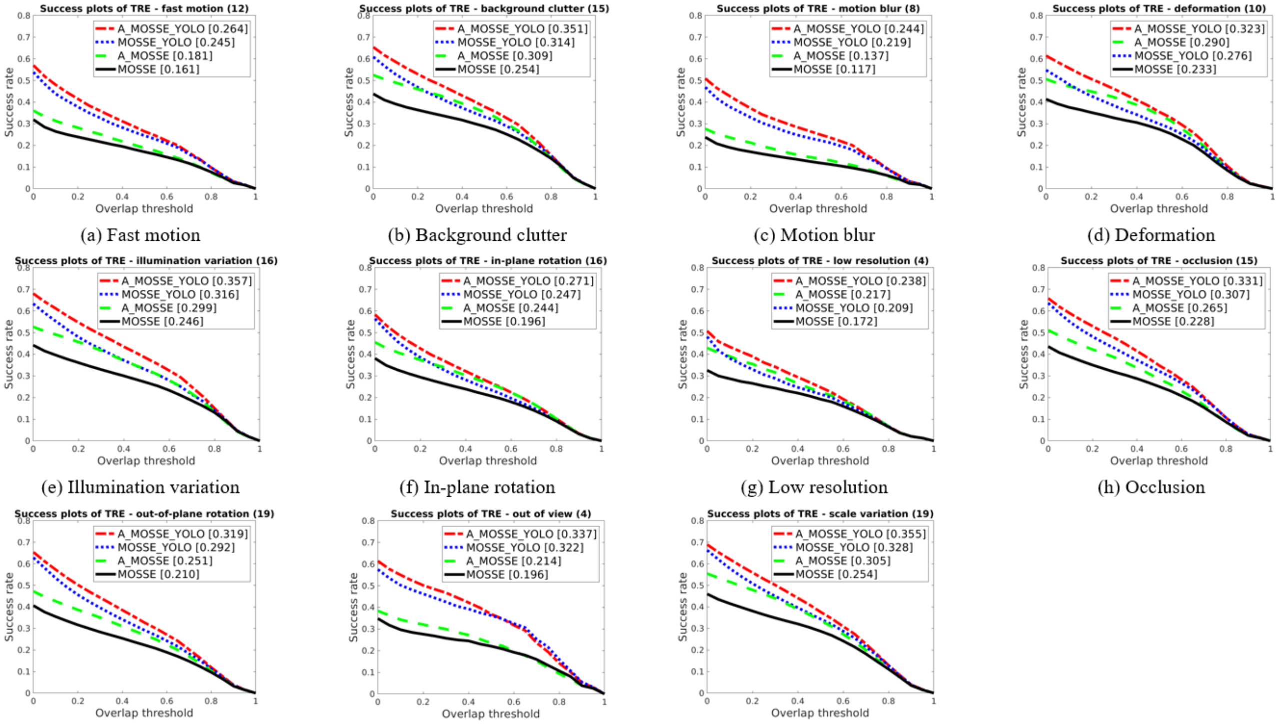


(a) Fast motion (b) Background clutter (c) Motion blur (d) Deformation

(e) Illumination variation (f) In-plane rotation (g) Low resolution (h) Occlusion

(i) Out-of-plane rotation (j) Out-of view (k) Scale variation

Fig. 16. The success plots of SRE evaluation of different attributes. The number at the end of the title of each figure shows how many videos are included in the corresponding scene factor.




(a) Fast motion (b) Background clutter (c) Motion blur (d) Deformation
(e) Illumination variation (f) In-plane rotation (g) Low resolution (h) Occlusion
(i) Out-of-plane rotation (j) Out-of view (k) Scale variation

Fig. 17. The success plots of TRE evaluation of different attributes. The number at the end of the title of each figure shows how many videos are included in the corresponding scene factor

## V. CONCLUSIONS

In this study, two methods, re-tracking mode with YOLO and adaptive multi-filter, were proposed to improve the MOSSE object tracking algorithm. The re-tracking mode help to improve the attributes of losing target, and the adaptive multi-filter improves to track the target in background clutter, deformation, and in-plane rotation. Integrating the methods with the MOSSE algorithm can significantly improve the tracking capability. The SRE was improved from 215 to 291, and the TRE was from 250 to 346. Most of the videos are based on the 2-dimensional images, as well as the image processing methods. However, the objects are 3-dimensional and the appearance would change because of the deformation, illumination, and rotation of the target. The proposed adaptively multi-filter stored the filters of different appearances of the object for the object tracking. The results show the significant improvement of the MOSSE. In addition, the re-tracking mode with YOLO quickly provides the possible targets and their positions. This method helps to re-track the target once it is lost. Although the re-tracking mode costs more computing resource, it significantly improves the performance of MOSSE in both SRE and TRE.

## REFERENCES

[1] S. He, H. Wang, S. Zhang, and B. Tian, "Vision Based Autonomous Tracking and Landing on an Arbitrary Object of Quadrotor," in *Proc. IEEE CYBER*, Suzhou, China, 2019, pp. 474-479, DOI: 10.1109/CYBER46603.2019.9066614.

[2] M. Jiang, R. Li, Q. Liu, Y. Shi, and E. Tlelo-Cuautle, "High Speed Long-Term Visual Object Tracking Algorithm for Real Robot Systems," *Neurocomputing*, vol. 434, pp. 268-284, Apr. 2021, DOI: 10.1016/j.neucom.2020.12.113.

[3] Z. Jie, L. Lei, G. Xiaqing, and L. Shanmei, "Tracking and Landing of QUAV under Complex Conditions based on Ground-Air Information Fusion," in *Proc. IEEE- ICCASIT*, Weihai, China, Oct. 2020, pp. 313-322, DOI: 10.1109/ICCASIT50869.2020.9368815.

[4] B. H. Kim, A. Lukezic, J. H. Lee, H. M. Jung and M. Y. Kim, "Global motion-aware robust visual object tracking for electro optical targeting systems," *Sensors*, vol. 20, no. 2, Art. no. 556, Jan. 2020, DOI: 10.3390/s20020566.

[5] R. Petruse, A. Matei, M. Kayser, M. Maier, S. Ćuković, "Academia-industry collaboration for augmented reality application development," in *Proc. BRCEBE*, Sibiu, Romania, Oct. 2019, pp. 243-250, DOI: 10.2478/cplbu-2020-0028.

[6] R. Verma, "A Review of Object Detection and Tracking Methods," *Int. J. Adv. Eng. Res. Dev.*, vol. 4, pp. 569-578, Oct. 2017.

[7] Z. Soleimanitaleb, M. A. Keyvanrad and A. Jafari, "Object tracking methods: A review," in *Proc. ICCKE*, Oct. 2019, pp. 282-288, DOI: 10.1109/ICCKE48569.2019.8964761.

[8] S. M. Jainul Rinosha, and M. Gethsiyal Augasta, "Review of recent advances in visual tracking techniques," *Multimed. Tools Appl.*, vol. 80, pp. 24185-24203, 2021, DOI: 10.1007/s11042-021-10848-6.

[9] M. Fiaz, A. Mahmood, S. Javed and S. K. Jung, "Handcrafted and deep trackers: A review of recent object tracking approaches," *ACM Comput. Surv.*, vol. 52, no. 2, pp. 1-44, Mar. 2020, DOI: 10.1145/3309665

[10] A. Kumar, G. S. Walia, and K. Sharma, "Recent trends in multicue based visual tracking: A review," *Expert Syst. Appl.*, vol. 162, no. 30, Art. no. 113711, Dec. 2020, DOI: 10.1016/j.eswa.2020.113711.

[11] X. Jia, H. Lu and M.-H. Yang, "Visual tracking via adaptive structural local sparse appearance model," in *Proc. IEEE-CVPR*, Providence, RI, USA, 2012, pp. 1822-1829, DOI: 10.1109/CVPR.2012.6247880.

[12] S. He, Q. Yang, R. Lau, J. Wang and M.-H. Yang, "Visual tracking via locality sensitive histograms," in *Proc. IEEE-CVPR*, Portland, OR, USA, 2013, pp. 2427-2434, DOI: 10.1109/CVPR.2013.314.

[13] M. Danelljan, F. S. Khan, M. Felsberg and J. van de Weijer, "Adaptive color attributes for real-time visual tracking," in *Proc. IEEE-CVPR*, Columbus, OH, USA, 2014, pp. 1090-1097, DOI: 10.1109/CVPR.2014.143.

[14] Q. Lei, S. Hichem and F. Abdallah, "Object Tracking Using Adaptive Covariance Descriptor and Clustering-Based Model Updating for Visual Surveillance," *Sensors*, vol. 14, no. 6, pp. 9380-9407, May 2014, DOI: 10.3390/s140609380.

[15] S. Hare, A. Saffari and P. Torr, "Struck: Structured output tracking with kernels," in *Proc. Int. Conf. Comput. Vis.*, pp. 263-270, Barcelona, Spain, 2011, DOI: 10.1109/ICCV.2011.6126251.

[16] I. Leang, S. Herbin, B. Girard and J. Droulez, "On-line fusion of trackers for single-object tracking," *Pattern Recognit.*, vol. 74, pp. 459-473, Feb. 2018. DOI: 10.1016/j.patcog.2017.09.026.

[17] L. Leal-Taixé, A. Milan, K. Schindler, D. Cremers, I. Reid and S. Roth, "An analysis of the state of the art in multiple object tracking," 2017, [online] Available: http://arXiv:1704.02781.

[18] M. Y. Abbass, K.-C. Kwon, N. Kim, S. A. Abdelwahab, F. E. A. El-Samie and A. A. M. Khalaf, "A survey on online learning for visual tracking," *The Visual Computer*, vol. 37, pp. 993–1014, May 2021, DOI: 10.1007/s00371-020-01848-y.

[19] R. Jiang et al., "Object tracking on event cameras with offline–online learning," *CAAI Trans. Intell. Technol.*, vol. 5, no. 3, pp. 165-171, 2020, DOI: 10.1049/trit.2019.0107.

[20] Y. Wu, J. Lim and M. H. Yang, "Online Object Tracking: A Benchmark," in *Proc. IEEE-CVPR*, Portland, OR, USA, 2013, pp. 2411-2418, DOI: 10.1109/CVPR.2013.312.

[21] D. S. Bolme, J. R. Beveridge, B. A. Draper and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proc. IEEE-CVPR*, San Francisco, CA, USA, 2010, pp. 2544-2550, DOI: 10.1109/CVPR.2010.5539960.

[22] J. F. Henriques, R. Caseiro, P. Martins and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Proc. ECCV*, 2012, vol. 7575, pp. 702-715, DOI: 10.1007/978-3-642-33765-9_50.

[23] J. F. Henriques, R. Caseiro, P. Martins and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 583-596, Mar. 2015, DOI: 10.1109/TPAMI.2014.2345390.

[24] M. Danelljan, G. Häger, F. S. Khan and M. Felsberg, "Discriminative scale space tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 8, pp. 1561-1575, Aug. 2017.

[25] J. Shin, H. Kim, D. Jeong, and J. Palk, "Automatic failure detection and correction for real-time object tracking with kernelized correlation filter," in *Proc. IEEE-ICCE*, Las Vegas, NV, USA, Jan. 2019, pp. 1-3 DOI: 10.1109/ICCE.2019.8661938.

[26] D. Yuan, X. Lu, D. Li, Y. Liang and X. Zhang, "Particle filter re-detection for visual tracking via correlation filters," *Multimed. Tools Appl.*, vol. 78, no. 11, pp. 14277-14301, Jun. 2019, DOI: 10.1007/s11042-018-6800-0.

**Chunhui Chung** received his Ph.D. degree in Mechanical Engineering from State University of New York at Stony Brook in 2010. He is currently an Associate Professor at the Department of Mechanical Engineering, National Cheng Kung University, Taiwan. His research interests include Metal Cutting, Abrasive Machining, Computer Aided Manufacturing, Vibration, and Prognostics and Health Management.



**Yi-Lung Chen** received his M.S. degree in Mechanical Engineering from National Cheng Kung University, Tainan, Taiwan in 2021. He is currently an Engineer with Da Shang Automation Co., Ltd., Taoyuan, Taiwan. His current research interests and publications are in the areas of industrial automation.