Visual Object Tracking with Hybrid Filter

Khin Ohnmar Maung
Department of Computer Engineering and Information Technology
Mandalay Technological University
Mandalay, Myanmar
khinohnmarmg@gmail.com

Theingi Myint
Department of Computer Engineering and Information Technology
Mandalay Technological University
Mandalay, Myanmar
drtgim@gmail.com

***Abstract*— Visual object tracking is to locate the target in the frame, which has a multitude of real-life application such as traffic monitoring, human computer interaction, autonomous vehicles tracking, robotics and more. However, there is still a challenging tasks such as illumination change, occlusion and re-entering into the camera's view, the tracking object's switched ID. In the tracking process, false positive tracks are changed the tracking object's ID and different detection class and tracking class are also problem for ID switching cases. To overcome those problems, the proposed system is used a Deep SORT object tracking algorithm with hybrid filter, which is combined a low confidence track and a class and active-time (LC-CA). According to the experimental results with two conditions, indoor and outdoor, the changes in the tracking object's ID of the proposed hybrid filter are 72% less than that of the conventional Deep SORT approach.**

Keywords*— *Visual tracking, YOLO, Deep SORT, Hybrid

I. INTRODUCTION (*HEADING 1*)

Visual object tracking has numerous applications, including video surveillance, autonomous driving, human-computer interaction, augmented reality, and robotics. The categorizations of the tracking are tracking-by-detection, tracking without detection, online tracking and offline tracking. To follow the moving target object correctly, the system must be able to reliably track the target. However, there are many challenges in the actual tracking scenarios that will lead to tracking performance decay, including the interaction between objects, occlusions, the high similarity between different objects, interference of the background, illumination, re-entering into the camera's view, etc. Under these challenges, undesirable errors such as tracking the wrong object class and ID switches are prone to occur, resulting in tracking performance decrease. Tracking systems have been broadly studied in numerous published works and different techniques have been proposed to solve some of these challenges.

A. Bewley [1] presented a simple online and real time tracking (SORT) using Kalman filter and Hungarian algorithm [2] as tracking components. Although SORT achieves best in class performance with both speed and accuracy, cannot handle ID switching problems. In 2017, N. Wojke et al. [3] presented a Deep SORT object tracking algorithm in which appearance information is integrated to improve the performance of SORT. Although Deep SORT can track multi objects in real time and solve occlusion problems, ID changing cases are still increased. In 2021, H. Wu et al. [4] presented a YOLOv4-tiny and motion prediction algorithm for multi object tracking to reduce ID switching cases in which YOLOv4 [5] is used as object detector and motion prediction is to predict the location of the lost objects. By adding this algorithm, although it is not in the real time condition, the tracking performance is significantly improved than the baseline Deep SORT.

X. Hou et al. [6] proposed a Deep SORT tracking algorithm with the extension of low confidence track filter (LCF) which can significantly reduce false positive tracks generated by Deep SORT. LCF is deleted the tracks with low average detection confidence in their initial several frames. By combining LCF, the detection confidence can be set to lower value and even zero to avoid missing detections. To handle the ID switching problem in the cases of detection class is not equal to the detection class, occlusion and re-entering into the camera view, in this proposed system, a hybrid filter is added Deep SORT tracker.

This paper is organized as follows: Section 2 explains the proposed methods in detail. Then, Section 3 shows the experimental results, and Section 4 concludes the manuscript..

II. PROPOSED SYSTEM WITH HYBRID FILTER

The proposed Deep SORT with hybrid filter in visual object tracking system is to overcome the ID switching problem of conventional Deep SORT. The main contribution is to enhance the tracking performance specially to reduce ID switching problems and to better deal with unreliable detection results such as low confidence true positive and high confidence false positive. In order to acquire the sequences of image, the video frames are captured by Pi-camera. This proposed system uses YOLOv4 to detect the generic objects such as Sports Ball and Person using 80 object classes of COCO dataset [7]. From the YOLO object detector, bounding boxes, scores and classes for objects are obtained. Then the Deep SORT algorithm with hybrid filter is applied to track the object. Finally, the objects are tracked with its IDs and bounding boxes.

When multiple objects are tracked with more than single class, false positive tracks caused by unreliable tracking tends to be seriously ID switches. To handle the ID switching problem in the case of occlusion and re-entering into the camera view, the hybrid filter is added to the baseline Deep SORT tracker. The LCF filter is to delete low confidence tracks while the CAF filter for filtering out the track when the detection class is not equal to the tracking class and to check whether the occlusion time and active time is exceeded the threshold values.

In the case of LCF, the average detection confidence (ADC) of the new detections associated to the Tentative track in (f+1), (f+2), …, (f+ $t_{tentative}$) frames is calculated. In this proposed system, the average confidence threshold is set to 0.7. If the

ADC is larger than the predefined average detection confidence threshold (*ADC_threshold)* saved, the Tentative track (T_t) is updated to Confirmed track (T_c). Otherwise, the T_t is deleted. In the case of CAF, a T_t will be deleted in class filtering. And in active-time filtering, a T_t will delete if occluded time is greater than the predefined occlusion threshold (t_{OT}) and active time is less than the predefined active threshold (t_{AT}). Otherwise, the T_t can be updated to T_c. The proposed hybrid filter algorithm of low-confidence-track and class-and-active-time filtering is illustrated as follows:

Algorithm: Proposed Hybrid Filtering

Input: Tentative tracks T_t ; Tentative threshold $t_{tentative}$; Average detection confidence threshold t_{ave_d} ; Associated detection confidence p_t ; Tentative track class t_{cls} ; Associated detection class $T_{\det\_cls}$; Occlusion threshold t_{OT} ; Active threshold t_{AT} ; Active tracks T_a

Output: Confirmed tracks T_c ; Deleted tracks T_d

for sequential frames do
  for $t \in T_t$ do
    if t is new in T_t then
      $OT = 0$
      $AT = 0$
      $t_{cls} = T_{\det\_cls}$
      $hits = 0$
      $total\_prob = 0$
    $hits = hits + 1$
    $total\_prob = total\_prob + p_t$
    if $hits \geq t_{tentative}$ then
      if $\frac{total\_prob}{hits} < t_{ave_d}$ then
        $T_d = T_d \bigcup t$ and $T_t = T_t \setminus t$
      else
        $T_c = T_c \bigcup t$ and $T_t = T_t \setminus t$
    if t is missed in T_a then
      $OT = OT + 1$
    else if t is active in T_a then
      $OT = 0$
      $AT = AT + 1$
    if $t_{cls} \neq T_{\det\_cls}$ or ($OT > t_{OT}$ and $AT < t_{AT}$) then
      $T_d = T_d \bigcup t$ and $T_t = T_t \setminus t$
      $AT = 0$
    else
      $T_c = T_c \bigcup t$ and $T_t = T_t \setminus t$

Process flow of baseline Deep SORT tracker with hybrid filter for a novel proposed system is shown in Fig. 1, in which assigned parameter values are also illustrated. Firstly, detections of objects, which have bounding boxes, scores, classes and features, are the input for the tracking procedure. After detecting specified objects (person and/or sports ball), they are matching in matching cascade, and will be tentative if new one is observed at first time, or unmatched detection/ track if they are as unmatched as occurred before, or matched track as same as occurred before. The tentative tracks are observed for seven consecutive hits to be confirmed. According to confirmed tracks, matching cascade is used to match these features of track and detection. For matched tracks, tracking IDs keep going on. When tracks are missed, these are saved, and waited for further re-occurring. When observed missed tracks, IoU match is used to match these B-boxes of tracks and detections. In this novel object tracking system, consecutive hits are set to seven for confirming the tracks to be avoid of serious ID switches. In order to delete the track that is missed in the active track between ten frames and is not active in the active track between ten frames, occlusion time and active time are set to ten.

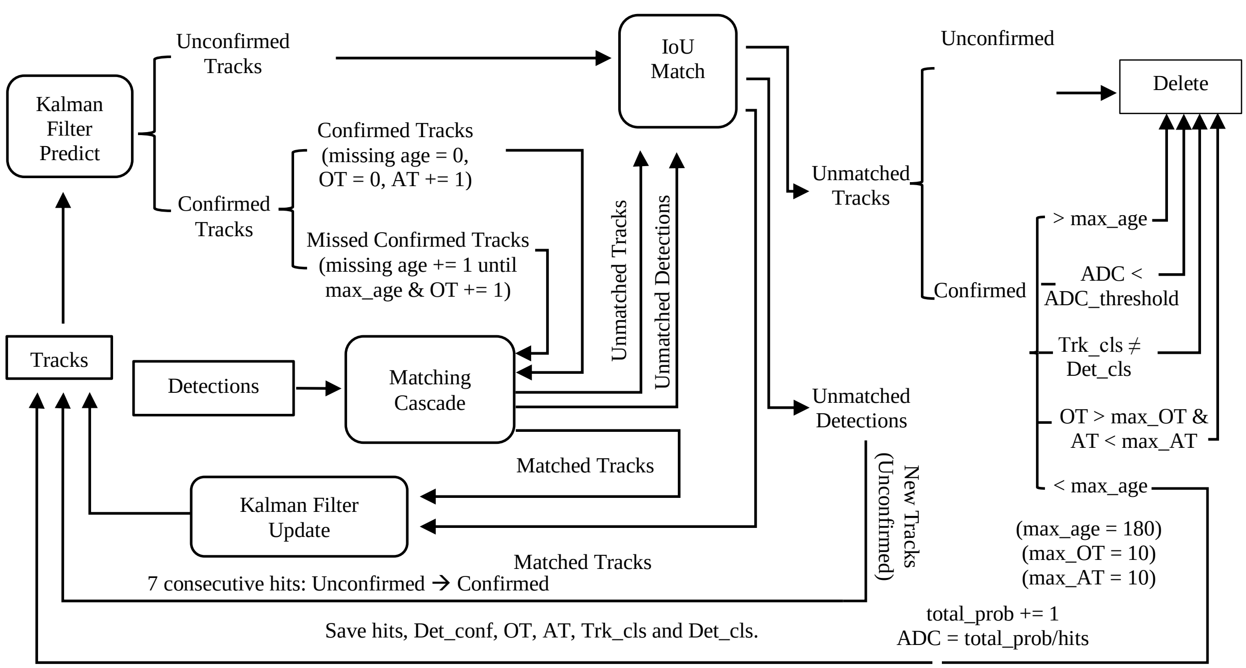


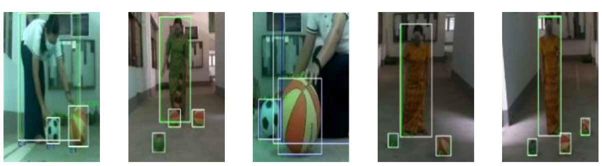
Fig. 1. Process flow of proposed Deep SORT with Hybrid Filter

For maintaining a track ID in an experimental video as long as possible, Max_age is set to 180, which will delete the track state after the maximum number of consecutive misses (180). And if the detection class is not equal to the tracking class, that frame will be discarded. Finally, the tracks will be deleted in four conditions: *If age > max_age*, *If ADC < adc_threshold*, *If Trk_cls ≠ Det_cls* and *If OT > max_OT & AT < max_AT*.

III. EXPERIMENTAL RESULTS

In this system, experiments are conducted to investigate the identity switches in indoor and outdoor conditions. Ten different experiments for indoor and outdoor environments are shown in the Fig. 2 (a) and (b). To investigate different situations, five different videos in indoor and another five different videos in outdoor are recorded. All video files are concerning with the illumination, occlusion and identity switches. Each video file is 30_second long and 720 frames. These video files including three objects (such as a person and two balls) or four objects (such as two persons and two balls) are experimented with both conventional and proposed Deep SORT.

The main contribution of the proposed system is the reduction of ID switching problems. There are two types of ID switching problems in occlusion and re-entering cases: formerly same class but different IDs, and different classes but formerly defined ID. In the experimental videos, there are two or three or four real objects but two different classes (person and sports ball in this study) entering and departing the video clips. In tracking processes, an object is specified as matched with its unique ID number when it is detected in seven consecutive frames of the video. In case of missing the object in many video frames (less than 180 frames), it will have to be re-identify with its ID number when it is occurred again in a current frame. The entering object is labelled as their formerly defined ID number to know it as an existing object.



(a) Five indoor videos

(b) Five outdoor videos

Fig. 2. Experimental videos

The experimental results for ID switching cases are as shown in Table. 1; the first five videos are indoor experiments and the rest five videos are outdoor experiments. Less ID switching is better in object tracking. In video 2 (IV1) and video 4 (IV4) of indoor, person is partially appeared and this causes serious ID switches. In these video experiments, the original Deep SORT approach has total seven ID switches in 615 of 720 frames in video 2 (IV2) and three ID switches in 871 of 720 frames in video 4 (IV4). Meanwhile, in our approach, there are no ID switches of four objects and three objects in IV2 and IV4, respectively. It can be clearly seen that the indoor experiment results are better in the proposed system than the conventional Deep SORT. In outdoor experiments, where there are challenges in illumination changes, clothes color and sun-drops which are badly effect on ID switches, the proposed system can successfully reduce ID switches, less than the conventional approach. In OV3 video, there are no ID switches in the proposed system, while there are 54% in the conventional. According to the indoor and outdoor experimental results, the proposed system can overcome former challenges with better results than the conventional approach, about 72% less in ID switches.

TABLE I. EXPERIMENTAL RESULTS FOR ID SWITCHING

	Indoor Experiments (%)					**Outdoor Experiments (%)**				
	IV 1	***IV 2***	***IV 3***	***IV 4***	***IV 5***	***OV 1***	***OV 2***	***OV 3***	***OV 4***	***OV 5***
Conventional	36	38	83	40	58	56	41	54	40	45
Proposed	**25**	**0**	**16**	**0**	**9**	**5**	**30**	**0**	**17**	**37**

IV. CONCLUSION

This paper is implemented a real-time object tracking using YOLO detector and Deep SORT along with hybrid filter. Detection in YOLOv4 is able to solve for illumination cases. For long-term occlusions and ID switching in tracking process, the proposed novel tracking system can re-identify the objects in the experiments by setting maximum-ages of track to 180. Experimental results have demonstrated that the proposed system can significantly reduce the number of identity switches compared with the conventional approach. The proposed system will be tested on mobile robot to track the selected target moving object as future research.

REFERENCES

[1] A. Bewley, Z. Ge, L. Ott, F. Ramos and B. Upcroft, “Simple online and real-time tracking”, 2016 IEEE International Conference on Image Processing (ICIP), 2016, pp. 3464-3468.

[2] B. Sahbani and W. Adiprawita, “Kalman filter and Iterative-Hungarian Algorithm implementation for low complexity point tracking as part of fast multiple object tracking system,” 2016 6th International Conference on System Engineering and Technology (ICSET), 2016, pp. 109-115.

[3] N. Wojke, A. Bewley, and D. Paulus, “Simple online and real-time tracking with a deep association metric”, IEEE International Conference on Image Processing (ICIP), IEEE 2017, pp. 3645–3649.

[4] H. Wu, C. Du, Z. Ji, M. Gao and Z. He, “SORT-YM: An algorithm of multi-object tracking with YOLOv4-tiny and motion prediction”, 2021, Electronics 10, no. 18, pp. 2319.

[5] A. Bochkovskiy, C. Y. Wang and H. Y. M. Liao, “YOLOv4: Optimal speed and accuracy of object detection”, Computer Vision and Pattern Recognition (cs.CV); Image and Video Processing(eess.IV), arXiv:2004.10934 [cs.CV].

[6] X. Hou, Y. Wang and L. -P. Chau, "Vehicle tracking using Deep SORT with low confidence track filtering", in: Proc. of 16th IEEE International Conf. on Advanced Video and Signal Based Surveillance (AVSS), 2019, pp. 1-6.

[7] T. Petrosyan, https://opencv.org/introduction-to-the-coco-dataset, October 12, 2021.