JAST
Journal of Aerospace Science and Technology

Pages: 61-70/ Vol. 15/ No. 1 / Winter- Spring 2022

Journal of Aerospace Science and Technology

www:jast.ias.ir

Scientific- Research Article

Development of Vision-Based Human Tracking for Drone's Gimbal

Mohammad Hossein Bayat [1], **Mohammad Shahbazi** [2]*****, **Bahram Tarvirdizadeh**[3]

1- Advanced Service Robots (ASR) Lab., Department of Mechatronics Engineering, Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran
2- Mechanical Engineering, School of Mechanical Engineering, Iran University of Science and Technology
3- Advanced Service Robots (ASR) Lab., Department of Mechatronics Engineering, Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran

ABSTRACT

***Keywords**: Vision-Based Human Tracking, 2D Gimbal, Visual Servoing, Unmanned Aerial Vehicles, drone, Kalman Filter.*

The use of Unmanned Aerial Vehicles (UAVs) with different features and for various applications has grown significantly. Tracking generic targets and, in particular, humans using the UAV's camera is one of the most active and demanding fields in this area. This paper implements two vision-based tracking algorithms to track a human using a 2D gimbal which can be mounted on UAVs. To ensure smooth movements and reduce the effect of common jumps on the trackers' output, the gimbal motion control system is equipped with a Kalman filter followed by a Proportional-Derivative (PD) controller. Various experimental tests have been designed and implemented to track a human. The evaluation results show success in tracking the high-speed movements with one of the algorithms and high accuracy in tracking challenging movements in the other algorithm. In both methods, the tracking computation time is short enough and suitable for real-time implementation. The optimal performance of both algorithms indicates the ability of the designed system to be implemented on the UAVs for practical applications.

Introduction

In recent years, drones have been used for various applications. They have various applications in different fields, such as filming and preparing aerial images, mapping, aerial survey of agricultural land and construction operations, handling and delivery of postal cargo, aerial images for surveillance systems such as traffic control, or facilities such as oil and gas transmission lines. Reasonable price, easy access, diverse range of applications, and user-friendliness are some of notable features of drones [1, 2]

Among the various applications of drones, vision-based tracking is one of its most widely used applications [3]. Images are generally recorded by installing a camera on a 2D or 3D rotating gimbal system. This system reduces the effect of the drone's movement on the image, and a favorable image is obtained by reducing the drone's vibrations and rotations [4].

Identifying and tracking various targets in images with high accuracy and speed is possible with the ever-increasing development of algorithms in machine vision. Humans are one of the common

1 Msc
2 Assistant Professor (Corresponding Author) **Email: *** shahbazi@iust.ac.ir
3 Associate Professor

DOI: 10.22034/jast.2022.312829.1105
Submit: 31.10.2021/ Accepted: 23.01.2022
Print ISSN:1735-2134 Online ISSN: 2345-3648

targets in this field [5]. Different vision-based tracking systems are available with varied features which can track various targets. In recent years, with significant progress in artificial intelligence, powerful trackers have been developed in deep learning along the classical methods in machine learning [7], [6].

Today, active tracking has been implemented on some of the flagship products of UAV manufacturing companies. However, its technical detail is not publicly available. For the first time in 2016, DJI Company developed an active tracking system on one of its products. This system was implemented to help the user with professional imaging and easy application. First, this system determines a desired target in the image. Then the drone uses the camera and the three-dimensional measurement of the environment to create the correct imaging of the target and a safe fly [8].

Many activities carried out in this field include creating a user interface to help the user control the gimbal and prepare the desired images [9], [10], recording human movement to produce a skeletal model, animation production using a drone [11], [12] or three-dimensional positioning of the target movements [13]. Reference [14] tracked different targets using classic tracking and a drone camera without a gimbal. However, its performance was limited to soft and non-sudden movements, as it does not use a robust algorithm in target detection and tracking. Reference [15] first simulated and then tested different target tracking with a drone without a gimbal in an indoor environment. Its primary basis is to increase the performance of the UAV controller. Therefore, a tracking algorithm was used with moderate ability.

This article deals with implementing a 2D gimbal for active and real-time vision-based human detection and tracking. This action is carried out completely automatically and user-independently. For better evaluation, two different tracking algorithms have been used. The first algorithm was based on classical methods of image processing. Despite its very high processing speed, this algorithm has little accuracy in tracking. The second algorithm is selected from the most modern tracking algorithms based on deep learning. This algorithm has a real-time processing speed and high accuracy. The Kalman filter has been used to have smooth movement for the gimbal. Also, a PD controller is used to send the control commands to the gimbal.

Introducing the tracking system

The used system includes two parts, namely hardware and software. The hardware part includes a 2D gimbal, an Arduino board for connecting to a computer, and a camera connected to the gimbal. Two separate approaches are used for human tracking in the software part for image processing.

Gimbal mechanism

A 2D gimbal with the ability to connect to a UAV has been implemented for automatic human tracking. This gimbal has angular movement using two servo motors in both vertical and horizontal directions and is launched with a webcam with 3-megapixel image quality. It should be noted that this gimbal does not move in the environment. This set is connected to the computer using an Arduino Uno board, and control commands are sent to determine the gimbal's position in two directions. The implemented system can be seen in Figure 1.

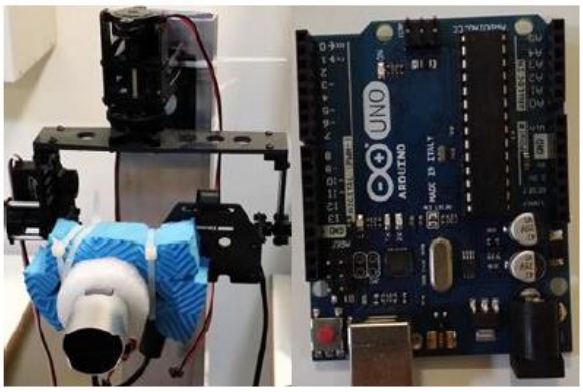

Figure 1. Right: Arduino board. Left: 2D gimbal and webcam.

Detection and Tracking Algorithms

A detection algorithm tries to find a specific target in the image, such as a human. This algorithm identifies and recognizes all targets in the image by processing each frame separately. However, regardless of the target type, the general tracking algorithm follows a target in the rest of the frames by receiving the bounding box around the target in the first frame [16]. The bounding box is a rectangle around the target in the image, and what is inside it is considered the tracking target. Figure 2 shows an example of the box enclosed around the target.

To use this system in real time, detection and tracking algorithms have been selected with high

accuracy and speed. For better evaluation and optimal performance, two methods have been used for tracking. In the first method, a combination of human detection and tracking algorithms with medium accuracy is used, and in the second method, a powerful tracker is used without a detection algorithm. These methods are explained below.

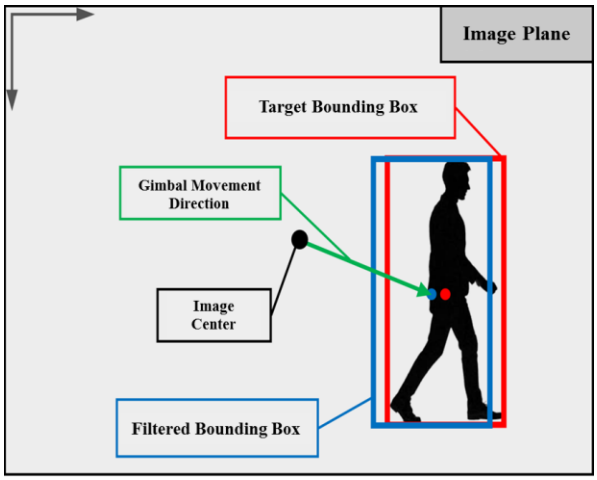

Figure 2. The image plane with the target bounding box, filtered box, centers of the boxes, the center of the image and the direction of the gimbal movement.

The first method is created by combining two detection and tracking algorithms, in which SSD [7] is used to identify the human in the image, and KCF [17] is used to track the detected target.

In general, detection algorithms are divided into two categories, single-stage, and two-stage. In the single-stage method, all processes are done in one step, including extracting image features, locating targets, predicting the bounding box, and classifying the identified classes. However, in the two-stage method, all the detections are made in one stage, and the classification of the results is applied in the other stage. The accuracy of the second method is more than the first method. However, its execution speed is lower due to the massive calculations. Therefore, the single-stage detection method is used in this study. At first, this algorithm extracts deep features of the input image using a convolutional neural network with the structure of MobileNet [18]. Several convolutional layers are used to reduce the dimension of the feature vector, then all probable bounding boxes and their scores are predicted. Finally, the target is determined after removing the boxes with the most overlap (Figure 3). This algorithm is trained to identify 20 types of targets, including human beings, and works in real-time with high accuracy.

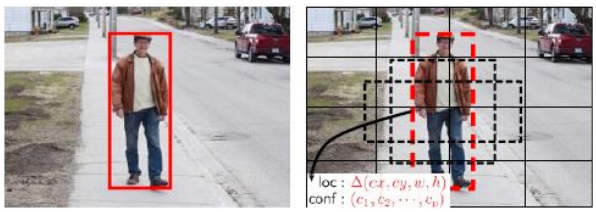

Figure 3. Detection with SSD algorithm. Right: Detection by dividing the image into 4x4 dimensions, the identified boxes for the target, removing the similar boxes and choosing the best one. Left: target detected in the original image.

KCF tracking algorithm is one of the algorithms with high execution speed and moderate accuracy. This algorithm uses classical machine vision methods to track any target, regardless of its type and class. At first, the initial target is delivered to this tracker as a template. Then the image is converted from the real space into frequency space and the similarity between the pattern and the set of neighboring pixels is checked using a fast but efficient correlation filter. Finally, the box with the highest score is considered as final target. The weakness of this algorithm is the lack of change in the dimensions of the received primary frame, which can be solved using the detection algorithm. As mentioned, the combination of these two detection and tracking algorithms is used in the first method. In this combined method, first, the detection algorithm processes the input image and sends the detection result in the form of a box enclosed around the target to the tracking algorithm. Next, the tracking algorithm follows the target in the subsequent frames.

To increase tracking accuracy, the detection algorithm is used every half second again. The re-detection procedure is performed in a smaller search region which is cropped from the current processing frame, using the target center position in the last frame. The detection result in this search area is again given to the tracking algorithm, and the target is followed by the tracker. By repeating this process in alternating time intervals, the possibility of losing the target is reduced and a more accurate bounding box surrounds the target.

In the second method, the Re3 tracker (6) is used. This tracking algorithm is one of the most modern and powerful trackers based on deep learning approach and has a high processing speed with

very suitable accuracy. Given two consecutive frames and a template determined in the first frame, a search zone is created in the second frame, using the template center position in the previous frame. In order to consider the displacement of the target between these two frames, the search field is formed by doubling the dimensions of the target. Next, using a convolutional neural network with the CaffeNet structure [19], the deep features of the template in the first frame and a search zone in the second frame are extracted and given to the fully connected layers. The task of these layers is to reduce the dimensions of the feature vector to determine the coordinates of the bounding box.

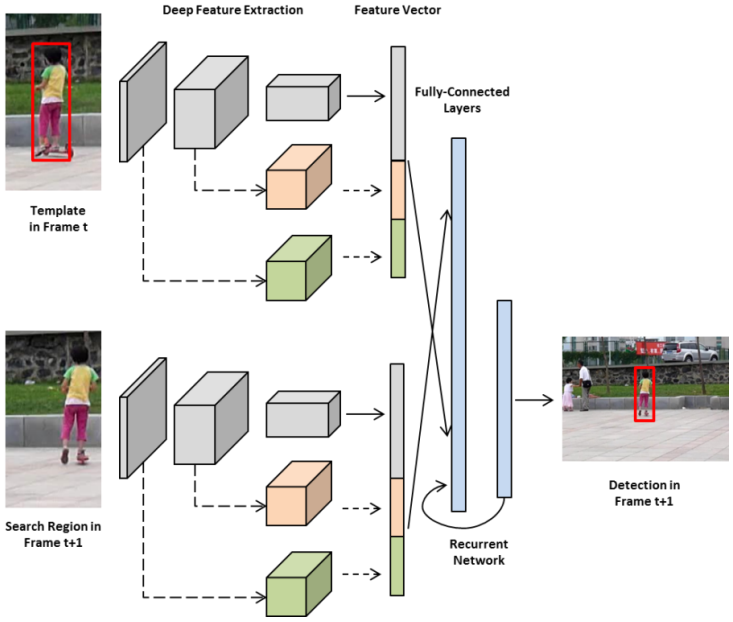

Figure 4. The structure of Re3 tracking algorithm. The use of all the convolutional neural network layers and also the recurrent neural network in order to increase the tracking ability and high processing speed.

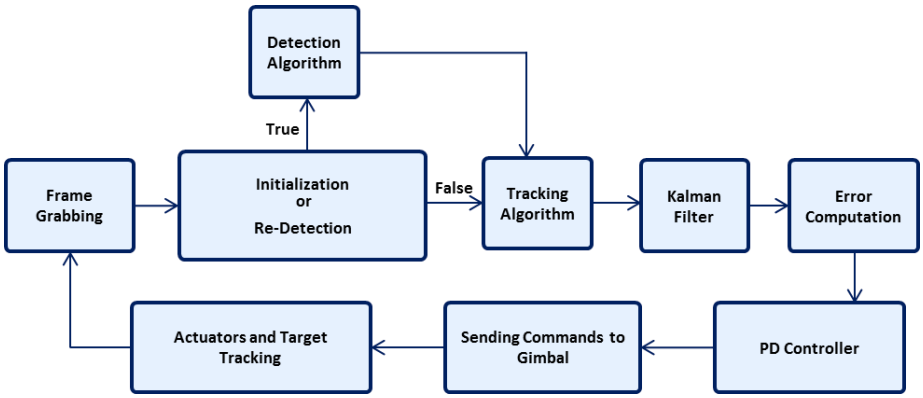

Figure 5. System performance schematic in software and hardware parts. If the detection algorithm is not used, the block will be ignored.

In order to increase tracking accuracy and deal with challenges such as occlusion, this tracker has also used a recurrent neural network. This neural network is responsible for remembering the both critical appearance and motion features of the target. Also, the parameters of this network are periodically updated during the tracking to adapt the tracker to the changes in the target's characteristics. Figure 4 shows the structure of this algorithm.

In the second method, first, the user draws the box around the target. Then the tracking algorithm initializes with this box and tracks the desired target without re-detection among the frames.
Figure 5 shows the system performance. For the first method, detection block is used; if it is not, the second method is obtained.

Kalman filter and gimbal control

Due to the difference between predicted boxes in consecutive frames, direct use of these results to calculate the error rate and generate the control command is unsuitable. The presence of permanent errors causes vibration and sudden jumps for the gimbal. Hence, to have a more uniform movement, the Kalman filter is used. Using this filter, the bounding box with less noise and smooth displacements is obtained, which is used to calculate the error between two consecutive frames and generate the control command. Figure 2 shows the filtered box and the centers of the box and image.

The Kalman filter considers the best choice between the observations over time and the output of the predictive model. Considering the center of the bounding box as the spatial coordinate x and also calculating the difference of this value between consecutive frames 1-k and k, the speed of movement is obtained based on the time interval Δt for both horizontal and vertical axes. By choosing the constant acceleration a_k from the normal distribution, with zero mean and standard deviation of σ_k, the kinematic equation of motion is written and converted into equation 1.

$$X_k = F_k X_{k-1} + G_k a_k \tag{1}$$

Where,

$$X_k = \begin{bmatrix} x_k \\ y_k \\ \dot{x_k} \\ \dot{y_k} \end{bmatrix} \tag{2}$$

is the speed and location of the center of target bounding box.

$$F_k = \begin{bmatrix} 1 & 0 & \Delta t_k & 0 \\ 0 & 1 & 0 & \Delta t_k \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3}$$

and

$$G_k = \begin{bmatrix} \frac{1}{2}\Delta t_k^2 & 0 \\ 0 & \frac{1}{2}\Delta t_k^2 \\ \Delta t_k & 0 \\ 0 & \Delta t_k \end{bmatrix} \tag{4}$$

On the other hand, covariance matrix is calculated by equation 5.

$$Q_k = G_k G_k^T \sigma_k^2 \tag{5}$$

By the mentioned equations for the target system and Kalman filter formulation, the state variable and covariance are calculated and updated for each frame.

To calculate the error and control the gimbal movement, the center of the image is considered as the optimal system input. After the target moved and left the center of the image, the difference between the center of the filtered box and the center of the image is calculated in both vertical and horizontal directions. Then, with the help of a PD controller and equation 6, a suitable control command is produced for the gimbal movement. In this regard, Kp is a proportional coefficient, and Kd is a derivative coefficient. Next, using the Arduino board, control commands are sent to the servo motors, and the movement continues until the center of the bounding box coincides with the center of the image.

$$u(t) = K_p e(t) + K_d \frac{de(t)}{dt} \tag{6}$$

System performance evaluation

In order to comprehensively evaluate the system performance, various criteria have been considered. Also, the obtained results were statistically analyzed, and the implemented algorithms were compared.

Evaluation criteria

Due to the lack of valid and common evaluation benchmark for the performance of this type of system, two general approaches have been considered for evaluation. In the first approach, "first failure," the user is asked to perform challenging and sudden movements in a limited time. These movements are called free movements, which cause the algorithm to fail. The algorithm's success rate is checked by recording the time of the first failure in tracking, which means the complete loss of the target. In the second part, tracking accuracy, a specific path is set for the user to travel at slow, fast, and faster speeds. In this case, the average distance between the target's location in the image and the center of the image, as well as the standard deviation, have been investigated. Also, the tracking accuracy benchmark [20] has been used to check the ability of the tracking algorithms. The benchmark will be explained in the next section. Figure 6 shows examples of processed images.

Evaluation results

In the evaluation using the first failure criterion, the first method (combination of SSD detector and KCF tracker) and the second method (Re3 tracker) followed the target successfully and did not miss it in the first 57% and 76% of the movements, respectively. However, re-detection in the first method led to the target being found later, nevertheless, the criterion of this evaluation is the first failure in tracking. Figure 7 shows an example of system performance using the first method. Despite the sudden and fast movements, the algorithm has performed well. Nevertheless, the target is lost at the 26th second.

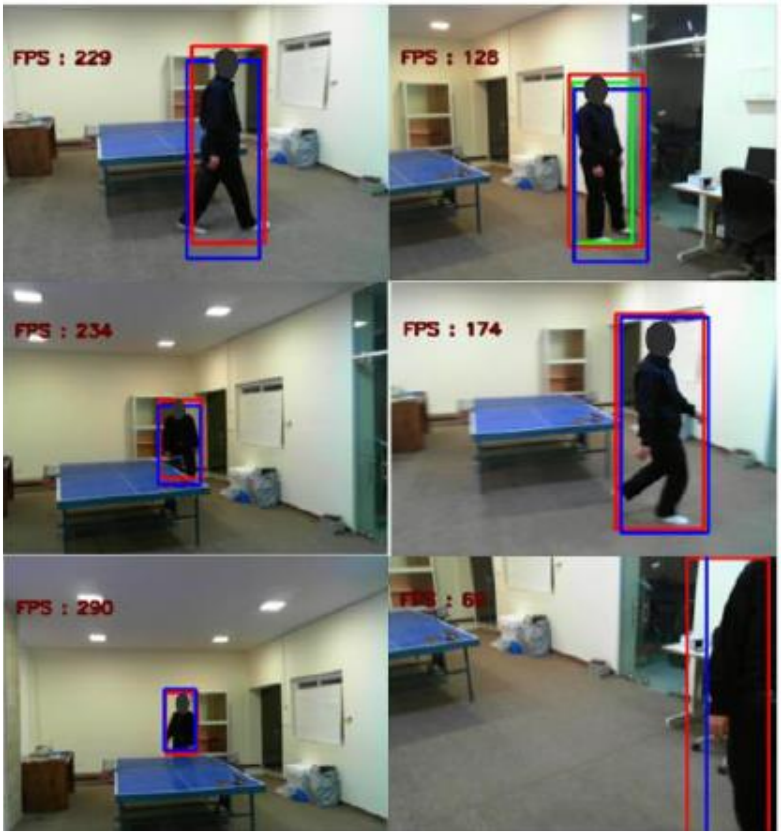

Figure 6. Sample images of system performance. Green box: detection algorithm, red box: tracking algorithm, blue box: Kalman filter.

The tracking accuracy method considers specific routes and different speeds in the evaluation. In Figure 8, the pixel location of the center of the filtered box on the image plane is drawn for all slow, fast, faster, and free movements. They are drawn using the first method in detection and tracking. The image plane refers to the frame received from the webcam with 240 x 320 pixels, and the center of this image is the 160th pixel on the horizontal axis and the 120th pixel on the vertical axis, which are shown with two perpendicular lines. Figure 9 displays the similar evaluation results using the second method in tracking. For a more accurate assessment, the Euclidean distance of the data to the center of the image and the standard deviation for all values were calculated. The results are reported in Table 1 and 2 for the first and second method, respectively.

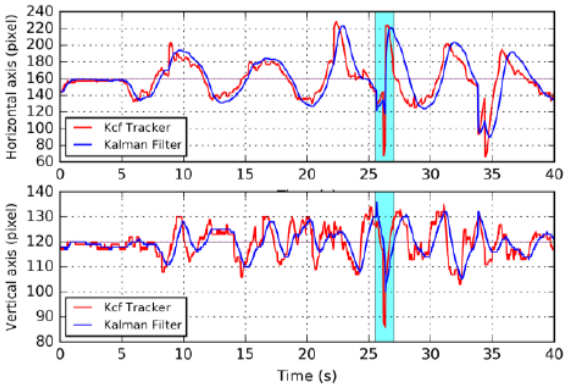

Figure 7. Diagram of the center location of the bounding box and filtered box on the horizontal and vertical axis with respect to time with the "KCF"

tracker in free movement (image center: horizontal axis at 160th pixel, vertical axis at 120th pixel).

Table 1. Average distance from the center of the image and standard deviation for the first method.

Type of movement	Slow	fast	Faster	Free
Average distance from the center	1.63	3.71	2.81	1.30
standard deviation	8.47	8.28	7.43	11.34

Table 2. Average distance from the center of the image and standard deviation for the second method.

Type of movement	Slow	Fast	faster	Free
Average distance from the center	1.64	2.95	5.57	2.07
standard deviation	6.76	8.73	9.61	9.27

In order to compare the accuracy of the two implemented methods, the tracking accuracy benchmark has been used. For this purpose, first, the difference between the center of the filtered box and the center of the image (at 120th and 160th pixel) are calculated for all slow, fast, faster, and free movements, as well as for all repetitions. The number of differences was counted for each threshold in the range of "0" to "50" pixels, then average values corresponding to each threshold were calculated. The average accuracy value for the threshold of "20" was reported according to the standard instructions.

Figure 10 shows an example of this benchmark's performance. In this figure, the location of the target's movement on the image plane is specified as in Figure 8, and the movement is of a free type obtained by the second method. In this figure, the outer box defines the boundaries with dimensions of 50 pixels. By counting the number of points located in this range and calculating its percentage compared to all the points that make up the movement, the accuracy value corresponding to the threshold of 50 will be obtained. The same process is done for the inner box, representing a range with a threshold of 20 pixels. By repeating this process and determining the accuracy percentage for all thresholds from "0" to "50" and calculating the total average, the algorithm's accuracy will be obtained.

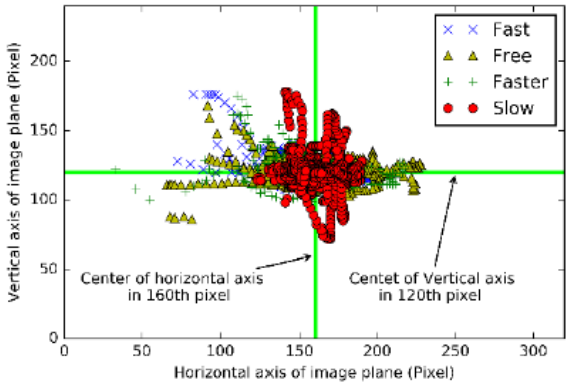

Figure 8. The horizontal center location of the filtered box compared to its vertical location on the image plane with the first method and for all movements.

Table 3. Tracking accuracy with a threshold of 20 and the average of all thresholds for the first method.

Type of movement	Slow	Fast	faster	Free
Tracking accuracy with a threshold of 20	85.1	81.3	88.8	74.1
average of all thresholds	81.6	82.5	85.9	75.8

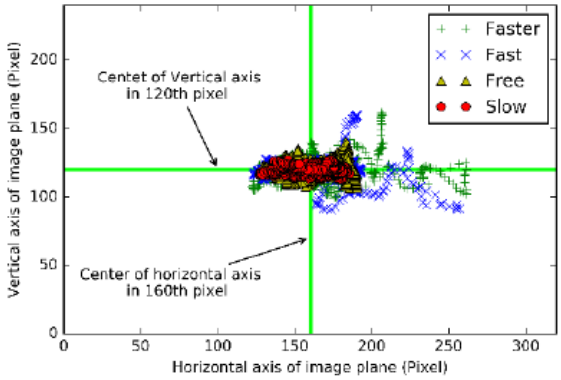

Figure 9. The horizontal center location of the filtered box compared to its vertical location on the image plane with the second method and for all movements.

Table 4. Tracking accuracy with a threshold of 20 and the average of all thresholds for the second method.

Type of movement	Slow	Fast	faster	Free
Tracking accuracy with a threshold of 20	88.7	80	85.8	83.4
average of all thresholds	85.1	82.5	84	81.9

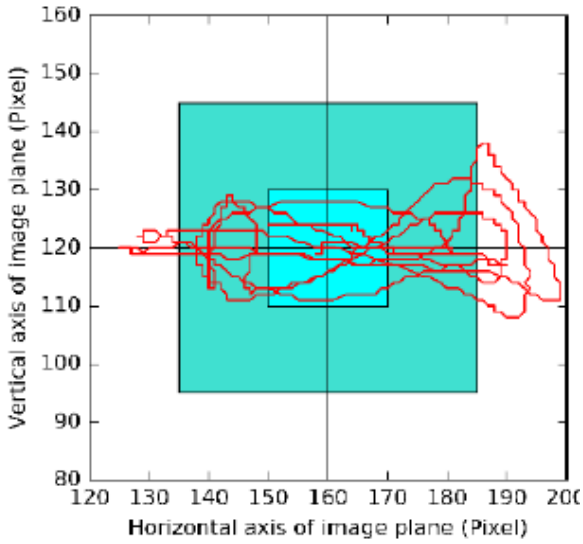

Figure 10. Thresholds of 20 and 50 in the image plane in order to evaluate the accuracy.

The tracking accuracy with a threshold of 20 for the four types of free, slow, fast, and faster movement, along with the average values of all thresholds, are listed for the first and second methods in Tables 3 and 4.

Analysis of the results

As mentioned in section 3, the output of the tracking algorithms is not directly used to calculate the error due to sudden jumps. For this purpose, the Kalman filter has been used to reduce jumps and smooth the gimbal movement. As an example, in Figure 11, the center location of the target bounding box and its filtered can be seen using the second method for faster movement in both horizontal and vertical directions. In this figure, the effect of the Kalman filter is evident. In the specified range in the figure corresponding to the horizontal axis, the time delay between the tracking algorithm and its filtered output is known due to the assumption of constant acceleration for motion. However, in the vertical axis, the optimal effect of using the filtered value is specified as an error calculation criterion and control command. In this figure, the output of the tracking algorithm is very disturbed, and this causes continuous vibrations in the gimbal. Therefore, the image will be blurred, and the target will be lost. Nevertheless, the filter box has solved the problem.

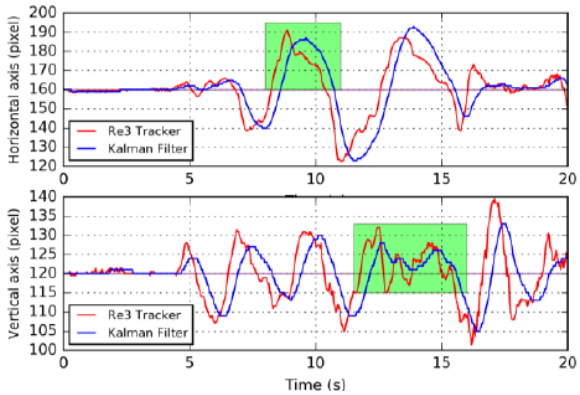

Figure 11 - The diagram of the center location of the bounding box and filtered box on the horizontal and vertical axis with respect to time with the "Re3" tracker in faster movement.

According to Figures 9 and 10, the dispersion of the data in the first method is more than the second method. According to the results in Tables 1 and 2, the slow movement with an average distance of 1.63 in the first and 1.64 in the second methods showed the best results among the three types of slow, fast, and faster movements. The average distance to the image center for the first method in the free movement was lower than in other movements. However, considering the amount of standard deviation, as the amount of dispersion and distance of the data to the average, it is clear that the performance of this algorithm was better in slow movement than in free movement.

On the other hand, a very small distance between the values related to the slow movement for both methods indicates the low impact of the tracking algorithm type in simple and unchallenged movements. On the other, the first method has shown better performance in fast and faster movements than the second method. It can be attributed to the combination of detection and tracking algorithms, while the second method does not use re-detection between frames.

According to the tracking accuracy criterion, the average of tracking accuracy for all thresholds and movements for the first and second methods was 81.45% and 83.34%, respectively. Also, the average value with a threshold of 20 for the first and second methods was 82.3 percent and 84.5, respectively. This shows that the second method has higher accuracy in tracking targets. Also, comparing the accuracy of both methods, it is clear that similar results are obtained concerning the accuracy of tracking in all four types of movement using the Re3 tracker. This shows the ability of this algorithm to provide more consistent performance

in four different types of movement compared to the KCF tracker. As a result, in general, for combined movements, a high-precision tracking algorithm, such as Re3, can produce better results. This can be due to the tracking continuity among consecutive frames.

Figure 12 plotted the horizontal position of the gimbal and the error of the horizontal center location of the filtered box image for the first method. At the beginning of the movement, there was an initial error in the target location, which was corrected by a slight gimbal movement. In the 4th second, the target movement has started, and when the error is generated, the gimbal movement starts with a constant speed. With each change of direction in the target's movement, the gimbal also changes direction with an acceptable reaction and aligns itself with the target. A noteworthy point is the stability and lack of gimbal oscillation, which is due to the filter and the smoothing of the bounding box movements.

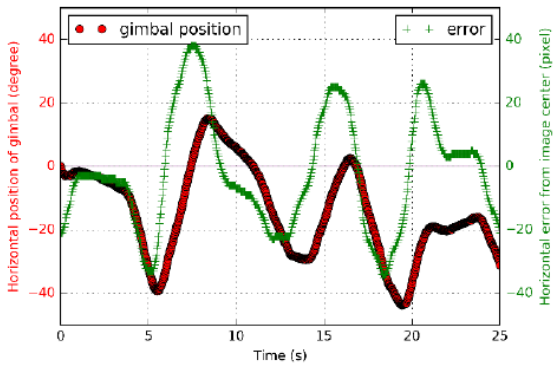

Figure 12. Left axis: graph of the gimbal's horizontal position with the first method and for fast movement, right axis: the difference between the horizontal position of the center of the filtered box and the center of the image in terms of time.

Summary and suggestions

This article investigated the implementation of a system for active and real-time human tracking on a 2D gimbal with the ability to connect to a UAV. In the image processing section, two different methods have been used to track the target in the image. In order to evaluate the performance of the system, two approaches of challenging movement and movement with different speeds in a fixed path have been considered. The second method has provided a better result in free movement due to its high ability to overcome tracking challenges. For slow and fast movements, both methods have similar results. However, in faster movements, the results have shown the optimal performance of the first method. This can be considered due to the combination of two detection and tracking algorithms in this method. The use of the Kalman filter has softened the movement of the gimbal. The obtained results show the acceptable ability of both methods to track humans in real conditions.

Implementation of the gimbal on a UAV and the evaluation of the algorithms in challenging situations can be explored in future works. Among these challenges, we can mention environmental effects, including weather conditions, crowded and clutter backgrounds, and long or short-term occlusions. Each of these challenges contributes to the system's performance in different conditions. Therefore, the developed system can be subjected to richer evaluations by producing large and diverse labeled datasets or using previous appropriate evaluation sets. Therefore, it is possible to evaluate its performance in different environments, different weather conditions, and different light intensities.

References

[1] C. H. Lin, F. Y. Hsiao, and F. Bin Hsiao, "Vision-based tracking and position estimation of moving targets for unmanned helicopter systems," *Asian J. Control*, vol. 15, no. 5, pp. 1270–1283, 2013, doi: 10.1002/asjc.654.

[2] V. N. Dobrokhodov, I. I. Kaminer, K. D. Jones, and R. Ghabcheloo, "Vision-based tracking and motion estimation for moving targets using unmanned air vehicles," *J. Guid. Control. Dyn.*, vol. 31, no. 4, pp. 907–917, 2008, doi: 10.2514/1.33206.

[3] Y. Wu, Y. Sui, and G. Wang, "Vision-Based Real-Time Aerial Object Localization and Tracking for UAV Sensing System," *IEEE Access*, vol. 5, pp. 23969–23978, 2017, doi: 10.1109/ACCESS.2017.2764419.

[4] R. Cunha *et al.*, "Gimbal Control for Vision-based Target Tracking," no. 3, pp. 1–5.

[5] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," *Proc. - Int. Conf. Image Process. ICIP*, vol. 2017-Septe, pp. 3645–3649, 2018, doi: 10.1109/ICIP.2017.8296962.

[6] D. Gordon, A. Farhadi, and D. Fox, "Re3 : Real-Time Recurrent Regression Networks for Object Tracking," *IEEE Robot. Autom. Lett.*, vol. 3, pp. 788–795, 2018.

[7] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 9905 LNCS, pp. 21–37, doi: 10.1007/978-3-319-46448-0_2.

[8] DJI, "DJI Active Track: Make the Drones Follow You," 2017. https://store.dji.com/guides/film-like-a-pro-with-activetrack (accessed Jan. 15, 2021).

[9] H. Kang, H. Li, J. Zhang, X. Lu, and B. Benes, "FlyCam: Multitouch Gesture Controlled Drone Gimbal Photography," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 3717–3724, 2018, doi: 10.1109/LRA.2018.2856271.

[10] C. Huang *et al.*, "ACT: An Autonomous Drone Cinematography System for Action Scenes," *Proc. - IEEE Int. Conf. Robot. Autom.*, pp. 7039–7046, 2018, doi: 10.1109/ICRA.2018.8460703.

[11] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "P finder: real-time tracking of the human body," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 780–785, 1997, doi: 10.1109/34.598236.

[12] X. Zhou, S. Liu, G. Pavlakos, V. Kumar, and K. Daniilidis, "Human Motion Capture Using a Drone," *Proc. - IEEE Int. Conf. Robot. Autom.*, pp. 2027–2033, 2018, doi: 10.1109/ICRA.2018.8462830.

[13] H. Zhang, Z. Lei, G. Wang, and J. N. Hwang, "Eye in the sky: Drone-based object tracking and 3D localization," in *MM 2019 - Proceedings of the 27th ACM International Conference on Multimedia*, 2019, no. 1, pp. 899–907, doi: 10.1145/3343031.3350933.

[14] R. Bartak and A. Vykovsky, "Any object tracking and following by a flying drone," in *Proceedings - 14th Mexican International Conference on Artificial Intelligence: Advances in Artificial Intelligence, MICAI 2015*, 2016, pp. 35–41, doi: 10.1109/MICAI.2015.12.

[15] A. Chakrabarty, R. Morris, X. Bouyssounouse, and R. Hunt, "Autonomous indoor object tracking with the Parrot AR.Drone," in *2016 International Conference on Unmanned Aircraft Systems, ICUAS 2016*, 2016, pp. 25–30, doi: 10.1109/ICUAS.2016.7502612.

[16] D. Held, S. Thrun, and S. Savarese, "Learning to track at 100 FPS with deep regression networks," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9905 LNCS, pp. 749–765, 2016, doi: 10.1007/978-3-319-46448-0_45.

[17] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012, vol. 7575 LNCS, no. PART 4, pp. 702–715, doi: 10.1007/978-3-642-33765-9_50.

[18] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2017, [Online]. Available: http://arxiv.org/abs/1704.04861.

[19] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *MM 2014 - Proc. 2014 ACM Conf. Multimed.*, 2014, pp. 675–678, doi: 10.1145/2647868.2654889.

[20] Y. Wu, J. Lim, and M. H. Yang, "Online object tracking: A benchmark," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 2411–2418, 2013, doi: 10.1109/CVPR.2013.312.