Performance Analysis of Fuzzy-Weighted Multiple Instance Learning on Thermal Video-Based Visual Tracking

Nur Ibrahim, Arsyad R. Darlis, and Benyamin Kusumoputro
Department of Electrical Engineering Universitas Indonesia, Depok 16424, Indonesia
Email: {nur.ibrahim, arsyad.ramadhan}@ui.ac.id; kusumo@ee.ui.ac.id

***Abstract*—In this paper, performance analysis of fuzzy-Weighted Multiple Instance Learning (WMIL) with the fuzzy logic tracker on thermal video-based visual tracking is presented. Thermal cameras have been used recently in some pedestrian areas, cars, and surveillance areas that need to be monitored all day. A thermal camera with advantages over the other visual-based sensors in low-light conditions is utilized in this research. The paper presents an analysis of visual tracking with an experimental method in the low-light outdoor environment. The thermal camera is used to record object movement used as video sequences to analyze the performance of our proposed system that integrate Type-2 Fuzzy Logic System (T2FLS) and WMIL tracker. The WMIL-T2FLS tracker performance is shown in the failure rate and center location error. The results show that the object in the thermal video sequences can be tracked using WMIL-T2FLS tracker in the low-light outdoor environment with a low level of failure rate and center location error. Then, the WMIL-T2FLS tracker can track the object when it occluded with the other similar object quite accurately. This result was compared with the original WMIL and some state-of-the art of tracking algorithm: DSST, ECO, KCF, SRDCF, and BACF. The research results showed that the WMIL-T2FLS system significantly improved compared with the WMIL method only, with a success rate improvement of at least 35 % and precision of at least 0.2 in 15 m dan 10 m. WMIL-T2FLS tracker also outperform some state-of-the art method and showed good performance in visual tracking at low-light environments.**

***Index Terms*—thermal camera, type-2 fuzzy logic system, visual tracking, weighted multiple instance learning, low-light outdoor environment**

I. Introduction

Nowadays, thermal imaging technology has stimulated much research related to the visual-based, especially in the low light environment such at night time [1], [2] and severe weather conditions [3]. Object detection and object tracking are computer vision task-based developed by researchers using a thermal camera in many applications such as surveillance systems [4], ADAS (Advanced Driver Assistance System), and firefighter rescue [5]. Pedestrian tracking is one of the research areas that many researchers have used thermal cameras. Although many development methods have achieved significant results in thermal camera based-pedestrian tracking, there are still many challenges in tracking, e.g., low resolution, occlusion, and thermal crossover [6].

In general, the pedestrian tracking method is categorized into two kinds of approach, i.e., generative approach and discriminative approach. The generative approach-based tracking method is modeling the pedestrian object at the first frame and looking for the most similar in the second frame. The discriminative approach looks at the tracking as a binary classification, which separates the object from the background in the image. Most of tracking method that has been developed recently are using discriminative approach, such as multiple instance learning [7], [8], correlation filter [9], [10], and deep learning [11], [12]. However, most of tracking method still not have failure detection feature, which is possibly occurred at tracking process due to occlusion or low resolution. In this paper, the author will use WMIL as a tracking method that will be integrated with failure detection feature based on the fuzzy logic system.

WMIL is one of the methods that use a discriminative approach [8] and less computational cost. WMIL is an improvement of MIL method that selects a positive sample with has important information only into the learning process [8]. The idea of MIL method is to use bags to collect samples near the object for classifying and updating the boosting classifier using the likelihood function. WMIL adds the weighted instance probability in the bag probability function. The positive sample near the object location will have a larger weight than the positive sample far from the object. WMIL also optimized the bag likelihood function in MIL so that the weak classifier selection is more efficient than directly maximize the log-likelihood function used in MIL.

However, WMIL still cannot detect the failure in the object tracking process [13], which can be occurred due to low resolution and occlusion. In addition, the characteristic of the thermal image, which has less interval color than RGB image, is another challenge for the tracking process. Following [14], [15], our purpose is to add a failure detection function in WMIL based on color distribution and Euclidean distance between the center coordinate of the object in the first frame and next frame using a fuzzy

Manuscript received December 24, 2021; revised April 24, 2022.

doi: 10.18178/joig.10.2.88-94

logic system for thermal video-based tracking. The thermal camera used in this research is FLIR ONE Gen 3 with 80×60 thermal resolution and 8-14µm spectral range that gives 640×480 thermal image resolution. Fuzzy logic systems have been used in many applications due to its capability to handle uncertainties.

Our proposed key ideas are measuring the Bhattacharyya coefficient [14] and Euclidean distance between predicted object location in current frame and previous frame generated from WMIL. These two measurement parameters will be used as an input for the fuzzy logic system to determine whether the predicted object location in the current frame is successful or failed. First, WMIL will generate five candidates of predicted object location with the highest probability. A fuzzy logic system will detect the highest probability of predicted object location. If the fuzzy logic system detects it as failed, the system will continue to check another candidate of predicted object location until it detects a successful candidate. If the system still cannot find the successful candidate, the first candidate with the highest probability will be used for the current frame. The result will be compared with the result of WMIL using the correction method based on the fuzzy logic system, i.e., Type-2 Fuzzy Logic System (T2FLS).

II. Methods

Fig. 1 shows the implementation system stages. It shows that the proposed system is divided into two systems, i.e., detection system using Weighted Multiple Instance Learning (WMIL) method and correction system using Type-2 Fuzzy Logic System (T2FLS).

In Fig. 1, the thermal video obtained using the FLIR camera is used as input from the system. The process is continued by taking samples from each part of the video. And then, the samples will be processed using a classifier. The classifier used in this study is (WMIL), wherewith this method, it is possible to estimate the location of the desired object. In addition, in this study, a method for detecting failures based on a Type-2 Fuzzy Logic System is also applied. When a failure in location detection occurs, a correction process will be carried out, and then the model owned by the system will be updated based on the correction results. The correction process will continue to be carried out until no failure is detected from the system owned. This process will be carried out continuously until the last frame is processed so that a performance graph is produced from the visual tracking system using the built thermal video.

Fig. 2 shows the WMIL block diagram, which shows the process algorithm from WMIL. The first frame taken from the thermal video is used to obtain instance positive samples inside radius α from center object location, and instance negative samples between radius β and radius ζ, where $\alpha < \beta < \zeta$. Instance positive samples are put inside positive bag X^+ and instance negative samples is placed inside negative bag X^-.

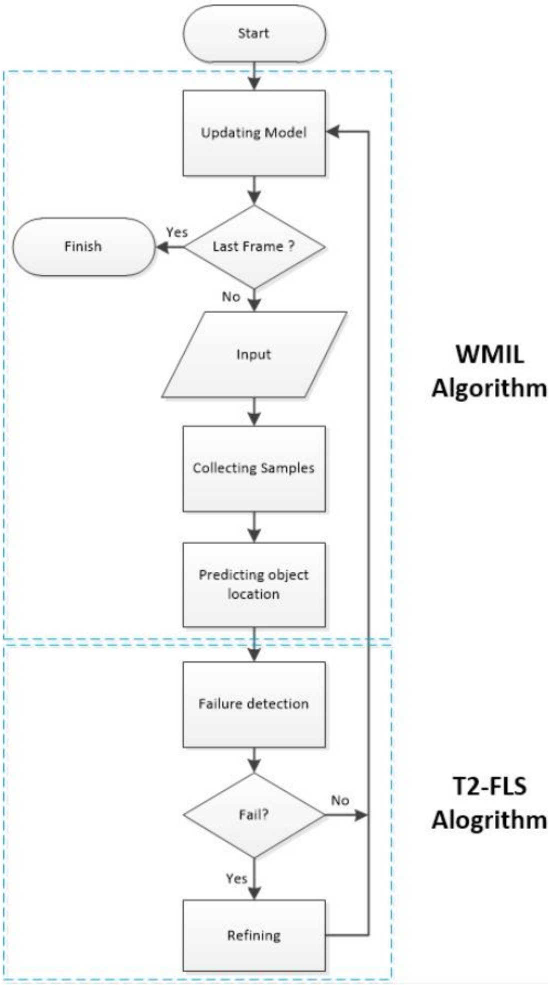

Figure 1. Flowchart of thermal video based visual tracking.

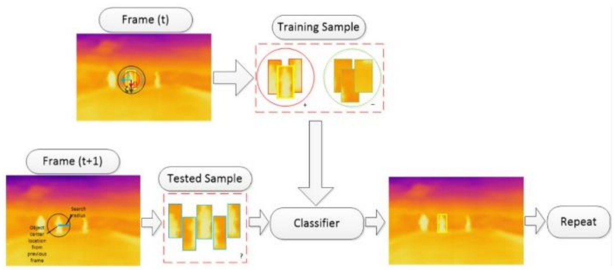

Figure 2. WMIL algorithm block.

There is possibility that some instance positive sample inside positive bag X^+ is negative samples, and all the instance negative samples inside negative bag X^- should be negative samples. Then, the positive bag probability to be positive is defined as in:

$$p(y=1|X^+) = \sum_{j=0}^{N-1} w_{j0} p(y_1 = 1|x_{1j}) \quad (1)$$

where N is quantity of instance positive samples inside positive bag X^+, $y \in \{0,1\}$ is a binary label of sample x where 1 is labeled for positive sample and 0 is labeled for negative sample, $p(y_1 = 1|x_{1j})$ is the posterior probability, as described in MIL tracker [16], for j-th positive sample x_{1j} (see Fig. 3), and w_{j0} is weight described as in:

$$w_{j0} = e^{-|l(x_{1j})-l(x_{10})|} \tag{2}$$

It can be seen that the value of w_{j0} are the exponential function of distance between center coordinat of j-th sample and center coordinat of tracking result (see Fig. 3).

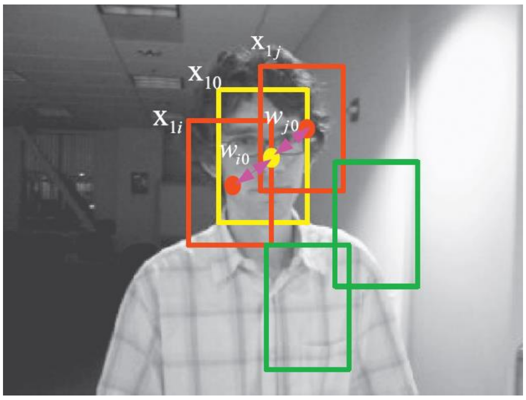

Figure 3. Illustration of the WMIL tracker. The tracking result is the location of yellow rectangle. Yellow and red rectangles are the positive samples. The solid circles are the central location of each sample.

Fig. 3 shows that all the instance negative samples inside the negative bag X^- is very far from the tracking result, so it can be assumed that all of the instance negative samples have same contribution to the negative bag X^-. All of these instance samples will be used to update classifier for sample x, as in

$$H_K(x) = \sum_{k=1}^{K} h_k(x) \tag{3}$$

where K is quantity of classifier $H_K(x)$, and $h_k(x)$ is a weak classifier of sample x weight that used in MIL Tracker [7]. Sample x is represented by a feature vector function f(x) that is generated using Haar-like features [17]. classifier selection is made by using Online weighted MIL Boost [8] to select classifiers with highest value of Online weighted MIL Boost, which will be implemented in the next frame to predict object location. The instance positive sample with the highest probability will be determined as the predicted object location in the current frame. Then, it will be used to update the classifier for the next frame. This process is repeated for the next frame until the last frame.

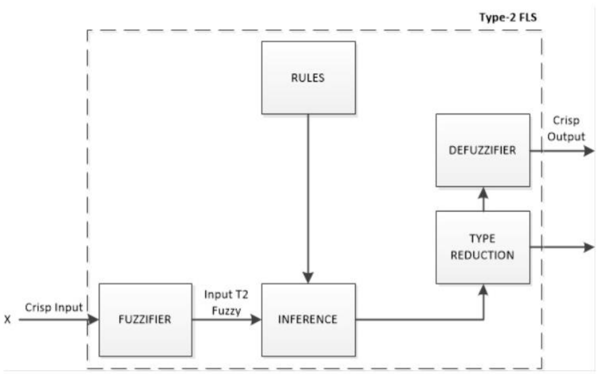

Figure 4. T2FLS algorithm block.

In our system, before the predicted object location of the current frame is to update the classifier for the next frame, it will go into the T2FLS block to check whether the predicted object location is a success or failure. Fig. 4 shows the T2FLS algorithm block system.

First, we compute the Bhattacharyya coefficient and Euclidean distance between the predicted object location in the current and previous frames.

The formula to calculate the Bhattacharyya coefficient is as in [14]:

$$\rho[p,q] = \sum_{u=1}^{m} \sqrt{p^{(u)} q^{(u)}} \tag{4}$$

where $p^{(u)}$ and $q^{(u)}$ is a discrete density of color histogram image p and image q respectively, $p = \{p^{(u)}\}$, $u=1...m$ and $q = \{q^{(u)}\}$, $u=1...m$. m is number of colour bin, which is 8 for RGB colour. if the two images is identical, the value of ρ is 1 which indicates perfectly identic. Euclidean distance is computed as in

$$Jc = \sqrt{(x_t - x_g)^2 + (y_t - y_g)^2} \tag{5}$$

where (x_t, y_t) is the center coordinate predicted object located at the current frame and (x_g, y_g) is the center coordinate of the predicted object located at the previous frame. These parameters are then fuzzified into membership function in the Fuzzifier block. The membership function is designed related to rules. Table I shows the parameter used in our T2FLS.

TABLE I. PARAMETER OF FUZZY LOGIC

Parameter	Members
Bhattacharyya coefficient	Similar
	Not Similar
Euclidean distance	Near
	Far

The range of the Bhattacharyya coefficient is 0 to 1, and the range of Euclidean distance is 0 to 100. Four rules that used in our T2FLS are:

1) If the Euclidean distance is near and the Bhattacharyya coefficient is similar, the output is a success.
2) If the Euclidean distance is near and the Bhattacharyya coefficient is not similar, the output is failed.
3) If the Euclidean distance is far and the Bhattacharyya coefficient is similar, the output is a success.
4) If the Euclidean distance is far and the Bhattacharyya coefficient is not similar, the output is failed.

Fig. 5 and Fig. 6 shows the membership function of the Bhattacharyya coefficient and Euclidean distance, respectively.

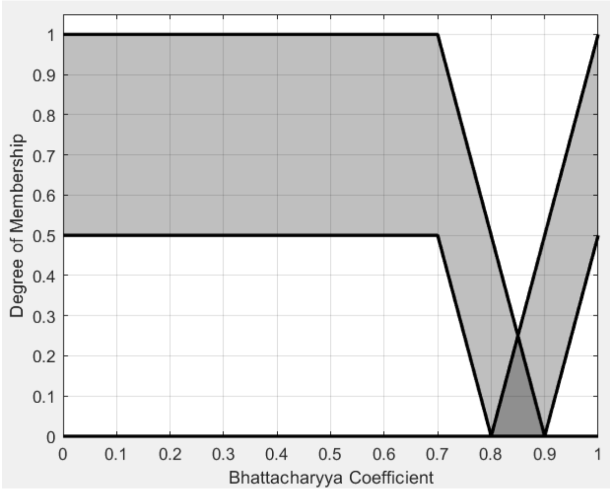

Figure 5. Membership function of Bhattacharyya coefficient.

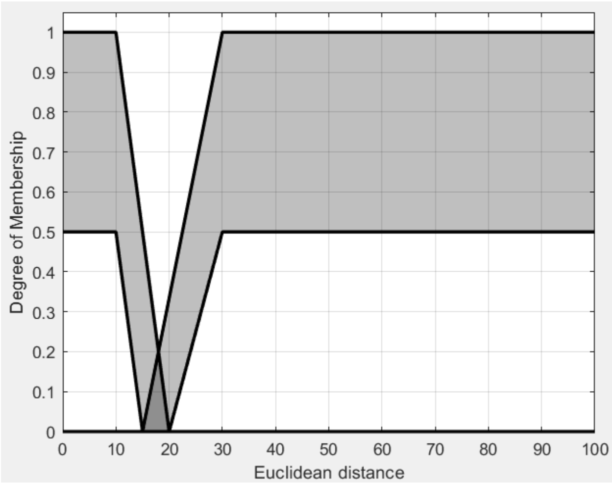

Figure 6. Membership function of Euclidean distance.

III. Result and Discussion

Our dataset was taken at night in clear weather and light rain weather, with 449 images for 10m distance and 445 for 15m distance from camera to object. The images consist of 3 people walking in front of thermal camera, which one of them is being tracked, with several conditions:

1) Each people walking together separately in the same direction, then one people change the direction to the opposite side.
2) Two people walking side-by-side in the same direction while one people waking in the opposite direction.
3) Three people walking side-by-side in the same direction.

Fig. 7 shows the Thermal Success Plot at a distance of 15m. The figure shows that the success rate will be inversely proportional to the increase in the IoU Threshold. By using WMIL, changing the IoU Threshold to 0.5, the success rate will be around 10%. The improvement can be seen in the graph by integrating the Weighted Multiple Instance Learning (WMIL) method with the Type-2 Fuzzy Logic System (T2FLS). At the same IoU Threshold of 0.3, there will be an increase in the success rate of 35% to 45%. However, the increase is not so significant on small and large IoU Thresholds. The figure shows that the average success rate is about 14%. These results show that the integration of the WMIL and T2FLS methods significantly increases the success rate in visual tracking.

In addition, Fig. 8 shows that the level of precision will increase linearly with the Location Error Threshold. It can be seen that when using WMIL on a system with a Location Error Threshold of 20, the precision shows a value of 0.05. The increase occurs when integrating WMIL and T2FLS, where there is an increase in precision of about 0.6 to 0.65. This result shows that the integration of T2FLS with WMIL shows very significant results, especially in the Location Error Threshold value range of 10-30, with an average increase of about 0.33.

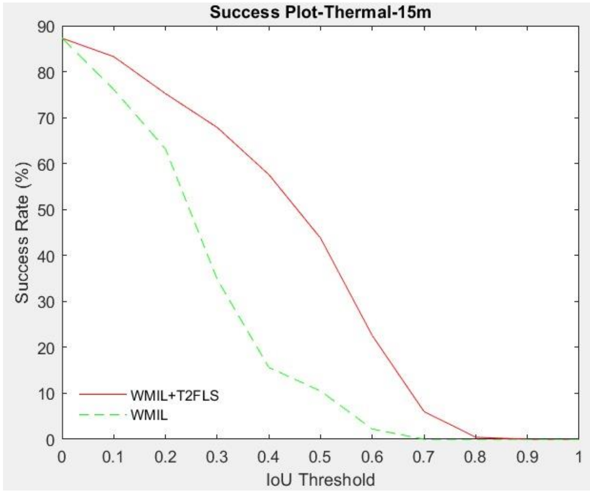

Figure 7. Success rate improvement of WMIL tracker with failure detection feature using T2FLS on 15m thermal video-based.

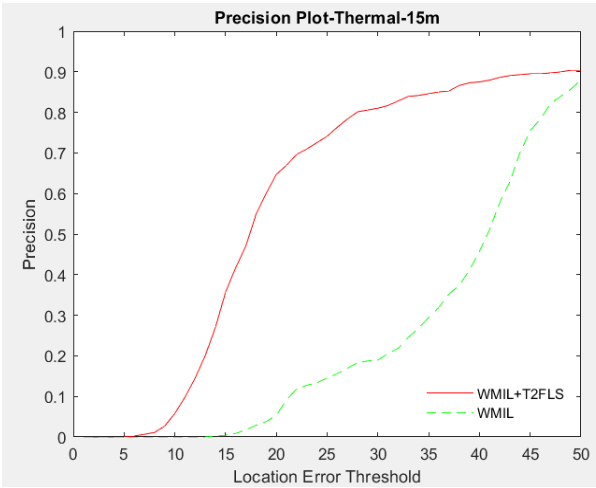

Figure 8. Precision rate improvement of WMIL tracker with failure detection feature using T2FLS on 15m thermal video-based.

When the distance is reduced to 10 m, there is a significant decrease in the success rate. It can be seen in Fig. 9 that with the IoU Threshold 0.3, the WMIL method will produce a success rate of only 20%. By decreasing the distance from 15 m to 10 m, there will be a decrease in the success rate of about 20% using the WMIL method and 50% with the use of the WMIL-T2FSL method. However, in contrast to Fig. 7, Fig. 9 shows that low and high IoU have significant differences in success rates. However, it is still seen that the effect of integrating WMIL and T2FSL at a distance of 10 meters shows an increase in the success rate in performing visual tracking using thermal video with an average increase of about 10%.

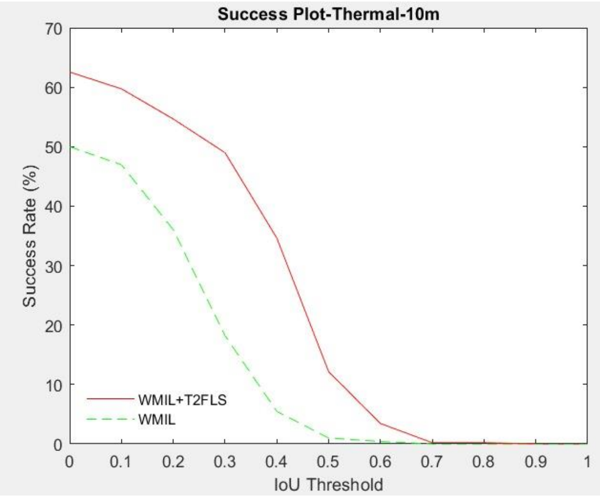

Figure 9. Success rate improvement of WMIL tracker with failure detection feature using T2FLS on 10m thermal video-based.

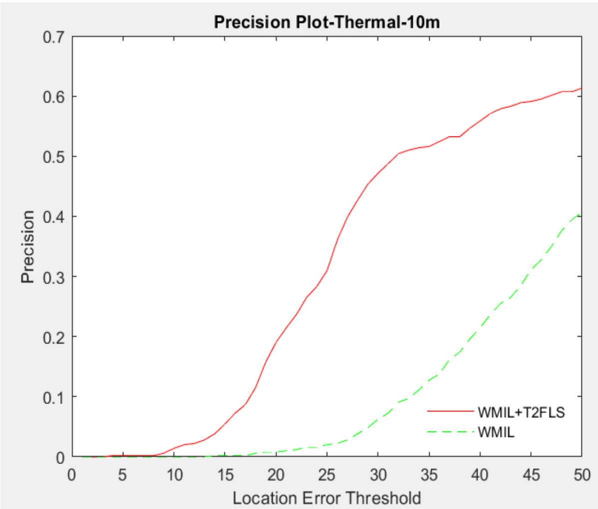

Figure 10. Precision rate improvement of WMIL tracker with failure detection feature using T2FLS on 10m thermal video-based.

The same thing happens in the graph that shows the level of precision of the system. It can be seen in Fig. 10 that with a location error threshold of 20, the precision with the WMIL method shows a value of around 0.01. It is decreased by about 0.04 compared to the 15 m distance. However, there is still an increase with an average of 0.21 by integrating WMIL with T2FSL. Then, our proposed method is compared with several state-of-the-art tracking method: SRDCF [18], ECO [11], BACF [10], KCF [9], DSST [19], DCF [20], MOSSE [21].

Fig. 11 and Fig. 12 shows that our proposed method outperform DSST, KCF, DCF, MOSSE, BACF, ECO-HC at a distance 15m for IoU threshold < 0.55 with 45% success rate and at a location error threshold > 20 pixel with 0.65 precision. Although WMIL+T2FLS performance is still below ECO, SRDCF, and BACF, our proposed method still faster with 79.3 fps compared with ECO, SRDCF, and BACF with 1 fps, 4.57 fps, and 28 fps, respectively.

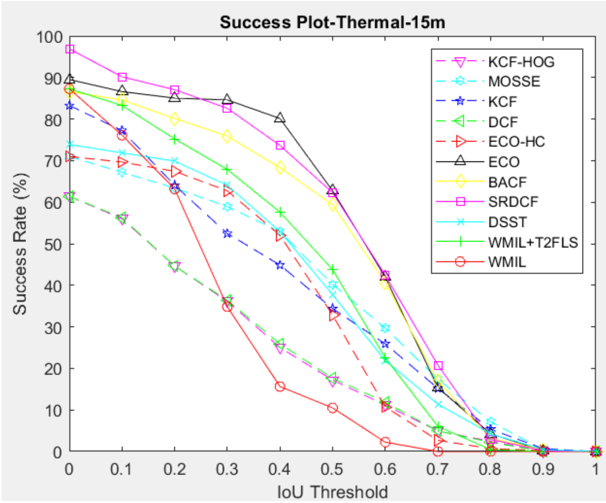

Figure 11. Comparation of success rate performance of WMIL+T2FLS with several state-of-the-art tracking method on 15m thermal video-based visual tracking.

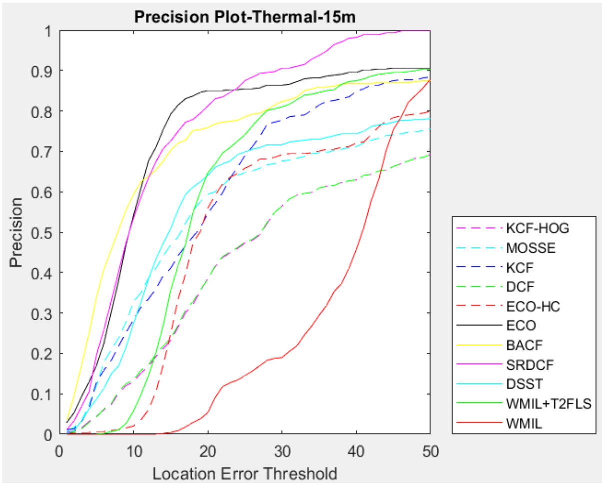

Figure 12. Comparation of precision rate performance of WMIL+T2FLS with several state-of-the-art tracking method on 15m thermal video-based visual tracking.

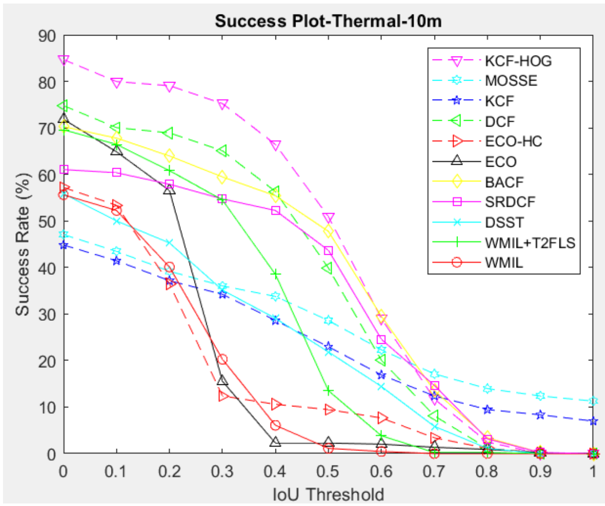

Figure 13. Comparation of success rate performance of WMIL+T2FLS with several state-of-the-art tracking method on 10m thermal video-based visual tracking.

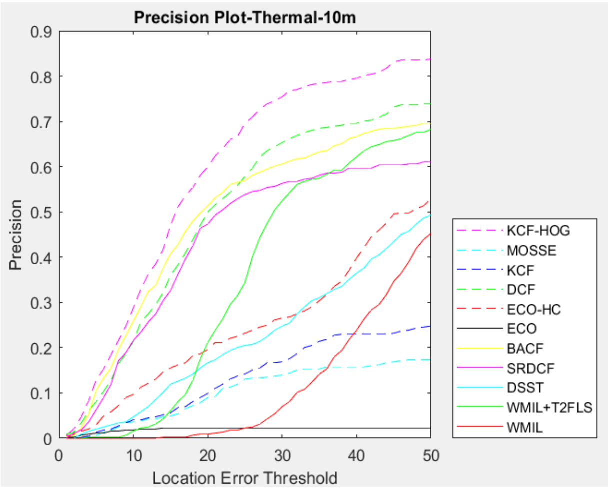

Figure 14. Comparation of precision rate performance of WMIL+T2FLS with several state-of-the-art tracking method on 10m thermal video-based visual tracking.

However, when the distance is reduced to 10 m, there is a significant decrease of performance for most tracking method used in this research. Fig. 13 shows that the success rate of WMIL+T2FLS is 58% for IoU threshold < 0.3, while it outperform ECO, MOSSE, KCF, DSST. Fig. 14 also indices a significance decrease of performance. When location error threshold > 20 pixel, our proposed method have 0.24 precision, slightly outperform DSST, KCF, and MOSSE.

It is shows that the performance of mostly tracking method in thermal image is decrease when the object size is large and occlusion happened.

IV. Conclusion

The paper presents an analysis of visual tracking with an experimental method in a low-light outdoor environment. The thermal camera is used to record object movement that will be used as video sequences to analyze performances of Weighted Multiple Instance Learning (WMIL) tracker with failure detection features using Type-2 Fuzzy Logic System (T2FLS). This result was compared with the original WMIL and several state-of-the-art tracking method. In this study, the performance of a system that uses WMIL+T2FLS in tracking thermal video has been obtained. From the results of the study, it was found that the system using WMIL-T2FLS showed a significant improvement compared to using only the WMIL method. At a distance of 15 m and 10 m, the success rate increased about 15% by integrating WMIL and T2FLS. The changing of the distance from 15 m to 10 m causes the success rate decrease around 10%, and the precision rate was around 0.04. Our proposed method outperform DSST, KCF, DCF, MOSSE, BACF, ECO-HC tracker. Although WMIL+T2FLS performance is not better than SRDCF, ECO, and BACF, it still faster with 74-79 fps. For further work, failure detection features with T2FLS could be optimized and implemented in other tracking methods to improve their performance on tracking task.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Benyamin Kusumoputro and Nur Ibrahim designed and implemented the experiments; Benyamin Kusumoputro, Arsyad R. Darlis, and Nur Ibrahim analyzed the data and wrote the paper.

References

[1] A. Nowosielski, K. Malecki, P. Forczmanski, A. Smolinski, and K. Krzywicki, "Embedded night-vision system for pedestrian detection," *IEEE Sensors Journal*, vol. 20, pp. 9293-9304, 2020.

[2] M. Kristo, M. Ivasic-Kos, and M. Pobar, "Thermal object detection in difficult weather conditions using YOLO," *IEEE Access*, vol. 8, pp. 125459-125476, 2020.

[3] P. Tumas, A. Nowosielski, and A. Serackis, "Pedestrian detection in severe weather conditions," *IEEE Access*, vol. 8, pp. 62775-62784, 2020.

[4] M. Ivasic-Kos, M. Kristo, and M. Pobar, "Human detection in thermal imaging using YOLO," in *Proc. 5th Int. Conf. Comput. Technol. Appl*., New York, 2019, pp. 20-24.

[5] M. Bhattarai and M. MartíNez-Ramón, "A deep learning framework for detection of targets in thermal images to improve firefighting," *IEEE Access*, vol. 8, pp. 88308-88321, 2020.

[6] Q. Liu, Z. He, X. Li, and Y. Zheng, "PTB-TIR: A thermal infrared pedestrian tracking benchmark," *IEEE Transactions on Multimedia*, vol. 2, pp. 666-675, 2020.

[7] B. Babenko, M. Yang, and S. Belongie, "Robust object tracking with online multiple instance learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, pp. 1619-1632, 2011.

[8] K. Zhang and H. Song, "Real-Time visual tracking via online weighted multiple instance learning," *Pattern Recognition*, vol. 46, no. 1, 2013.

[9] J. F. Henriques, *et al.*, "High-Speed tracking with kernelized correlation filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, pp. 583-596, 2014.

[10] H. K. Galoogahi, A. Fagg, and S. Lucey, "Learning background-aware correlation filters for visual tracking," in *Proc. the IEEE International Conference on Computer Vision*, 2017, pp. 1135-1143.

[11] M. Danelljan, G. Bhat, F. S. Khan, and M. Felsberg, "Eco: Efficient convolution operators for tracking," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6638-6646.

[12] B. Li, J. Yan, W. Wu, Z. Zhu, and X. Hu, "High performance visual tracking with Siamese region proposal network," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8971-8980.

[13] S. A. Wibowo, H. Lee, E. K. Kim, T. Kwon, and S. Kim, "Tracking failures detection and correction for face tracking by detection approach based on fuzzy coding histogram and point representation," in *Proc. International Conference on Fuzzy Theory and Its Applications*, 2015, pp. 34-39.

[14] K. Nummiaro, E. Koller-Meier, and L. V. Gool, "An adaptive color-based particle filter," *Image and Vision Computing*, vol. 21, pp. 99-100, 2003.

[15] H. Liu, L. Wang, and F. Sun, "Mean-Shift tracking using fuzzy coding histogram," *International Journal of Fuzzy System*, vol. 16, pp. 457-467, 2014.

[16] A. Nowosielski, K. Malecki, P. Forczmanski, and A. Smolinski, "Pedestrian detection in severe lighting conditions: Comparative study of human performance vs thermal-imaging-based automatic system," in *Proc. International Conference on Computer Recognition Systems*, 2020, pp. 174-183.

[17] D. Geronimo, "Haar-Like features and integral image representation," Universitat Autonoma de Barcelona, 2009.

[18] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Proc. the IEEE International Conference on Computer Vision*, 2015, pp. 4310-4318.

[19] M. Danelljan, G. Häger, F. Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," in *Proc. British Machine Vision Conference*, Nottingham, 2014.
[20] A. Lukezic, T. Vojir, L. C. Zajc, J. Matas, and M. Kristan, "Discriminative correlation filter with channel and spatial reliability," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6309-6318.
[21] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2544-2550.

Nur Ibrahim was born in Bandung, Indonesia in 1987. He received a Bachelor's Degree in Telecommunication Engineering from Institut Teknologi Bandung, Indonesia, in 2009, and he received a master's degree from Institut Teknologi Bandung, Indonesia, in 2011. He is currently a Doctoral Candidate at the Department of Electrical Engineering, Universitas Indonesia, Indonesia. His research interests are related to Artificial Intelligence, Machine Learning, and Signal Processing.

Arsyad R. Darlis was born in Bandung, Indonesia in 1971. He received a Bachelor's Degree in Electrical Engineering from Institut Teknologi Nasional Bandung, Indonesia, in 2009, and he received a master's degree from Institut Teknologi Bandung, Indonesia, in 2011. He is currently a Doctoral Candidate at the Department of Electrical Engineering, Universitas Indonesia, Indonesia. His research interests are related to Wireless Communication, Optical Communication, Artificial Intelligence, and Signal Processing.

Benyamin Kusumoputro was born in Bandung, Indonesia in 1957. He received a Bachelor's Degree in Physics from Institut Teknologi Bandung, Indonesia, in 1981. While he received a master of engineering science from Universitas Indonesia, Indonesia, and Tokyo Institute of Technology, Japan, in 1984. He received a Doctor of Engineering from Tokyo Institute of Technology, Japan. In 1993. He is currently a Professor in The Department of Electrical Engineering, Universitas Indonesia, Indonesia. His research interests are related to Computational Intelligence and Control Engineering.