SCIENTIFIC JOURNAL OF POLISH NAVAL ACADEMY

2020(LXI) 4 (223)

sciendo

DOI: 10.2478/sjpna-2020-0014

TRACKING OF UNMANNED AERIAL VEHICLES USING COMPUTER VISION METHODS: A COMPARATIVE ANALYSIS

Stanisław Hożyń[1] **Miłosz Wierszyło**[2]*

*[1] *Polish Naval Academy, Faculty of Mechanical and Electrical Engineering, Smidowicza 69, Str., 81-127 Gdynia, Poland; email: s.hozyn@amw.gdynia.pl; ORCID: 0000-0003-1422-0330;*

* [2] *Polish Naval Academy, Faculty of Navigation and Naval Weapons, Smidowicza 69, Str., 81-127 Gdynia, Poland; email: m.wierszylo1995@gmail.com; ORCID: 0000-0001-8277-2183.*

ABSTRACT

Tracking of small objects in any given airspace is an integral part of modern security systems. In these systems, there are embedded methods that employ the techniques based on either radio waves, or acoustic signals, or light radiation. The computer vision operation, springing from the light radiation-based technique, has prompted interest in its research. This operation has the advantage of being less expensive than radars and acoustic systems. In addition, it can solve complex security problems by detecting and tracking humans, vehicles, and flying objects. Therefore, this article evaluates the usefulness of the varying computer vision algorithms for tracking of small flying objects.

Keywords:
unmanned aerial vehicle (UAVs), computer vision, object tracking.

Research article

INTRODUCTION

Unmanned aerial vehicles' (UAVs) incursion of the critical infrastructure constitutes one of the most significant security-related challenges. Enhancing the detection, and tracking of UAVs increases the defence against such an incursion, and therefore provides the necessary protection for strategic facilities. There are three choices that aid the enhancement outcome, and they are either radio waves, acoustic signals, or visible light.

Radio waves, employed in radar systems, are used to detect objects in the airspace in both civil, and military aviation. Radar works on the principle of emitting electromagnetic waves, which, after reaching the object, are reflected, and returned to the receiver providing information about the object location. By utilising the phenomenon that a radio beam reflects from materials, such as metal and carbon fibre, the majority of flying objects can be detected [1]. In this process, the stable frequency generators and precise measurement of the time between sending a receiving signal are demanded [2]. Among the radio wave techniques, one of the state-of-the-art solutions uses the Phase-Interferometric Doppler Radar that facilitates creating 3D maps of the supervised area. The maps depict the detected objects, their range, velocity, and azimuth angles. The signal processing stage is divided into two parts: the Range-Doppler Processing, and Range Azimuth Processing. By using two receiving channels, the radar configured with phase-interferometry is able to estimate the angle of arrival (AOA) [3].

Acoustic systems analyse sound emitted by the object in order to detect the direction of signal emission. They do not require seeing the object during operation, which allows to use them in urban facilities, or behind hills [4]. In this respect, acoustic systems outperform other methods based on radio waves and visible light. What is more, the weather conditions, and the part of the day do not impede the detection process [5]. Acoustic systems usually comprise arrays of sensitive microphones. A large number of microphones, and advanced digital signal processing enable determination of azimuth, and elevation of the target in real-time. Another advantage involves modularity and scalability, which allow the use of this type of detection in combination with other systems [6].

The significant progress of computer vision technique has been demonstrated in many research areas, such as intelligent surveillance systems, autonomous vehicles, or industrial automation [7–10]. Cheaper cameras and faster computers, as well as more sophisticated algorithms, facilitate engaging computer vision in a wide range of real-time applications [11]. In case of protection against UAVs' attacks, these systems are responsible for the detection, and tracking of suspicious objects. They cost less than radars and acoustic systems. What is more, their highly complex solution systems have the ability to detect, recognise, and track various targets, such as humans, vehicles, or flying objects in any surrounding area [12].

The detection step often utilises advanced object detection algorithms, implemented in computer vision libraries. In [13], the comparative analysis of these algorithms, applied in the OpenCV libraries, was performed. The algorithms were divided into two main groups – supervised classifiers, and background subtractors. The first group included algorithms, such as the KNN, Codebook, and Codebook 2, while the second group comprised the MOG, MOG2, and GMG. According to the research, the MOG algorithm was the best choice for UAVs detection [13].

Detection of suspicious objects constitutes a crucial step in surveillance systems. However, when the object is detected, its location should be monitored. To perform this task, tracking methods, enabling real-time execution with high accuracy, are implemented. Plenty of algorithms dedicated to object tracking have been presented in the literature. Among the most popular ones can be enumerated: the Dense Optical Flow, Sparse Optical Flow, Kalman Filtering, Mean Shift, Cam Shift, Single Object Trackers, and Multiple Object Trackers. Since they differ in accuracy and complexity, the preliminary research was devoted to distinguishing the group of the methods which are promising in view of UAVs tracking. For that reason, each method was tested on short samples of movies, and evaluated using human eye examination. Consequently, the following methods were taken into further consideration: the CSRT Tracker [14], MIL Tracker [15], MOSSE Tracker [16], and KCF Tracker [17].

This article presents a comparative analysis of selected algorithms applied in object tracking in order to assess their usefulness for tracking of UAVs. First, the algorithms with their mathematical processes are described. Then the results of the experiments are presented, and, finally, the conclusions are formulated.

METHODS

The aim of the work was to assess the usability of the selected object tracking methods for the USVs defence system. A short description of the chosen algorithms is provided below.

CSRT Tracker

The CSRT algorithm, proposed by Alan Lukezic, is based on the DCF procedure. It improves the accuracy of the DCF tracker by adding spatial reliability maps. The use of the spatial reliability maps allows tracking more complex targets because the channels are combined to establish the final response map [5]. The execution of the algorithm can be divided into three stages:

- Scale location.
- Scale estimation.
- Updating operation.

Scale location and estimation select features from the search region, which is centred on the target estimated position in the previous time step, and correlated with the filter. In the updating step, the training region is centred on the previously estimated target location. Then, the foreground and background histograms are extracted, and updated by the exponential mean with the learning rate. After extracting the foreground $\tilde{\mathbf{c}}^f$ and background $\tilde{\mathbf{c}}^b$ histograms, the algorithm updates these histograms, utilising the formula presented below [6]:

$$\mathbf{c}_t^f = (1-\eta)\mathbf{c}_{t-1}^f - 1 + \eta\tilde{\mathbf{c}}^f, \tag{1}$$
$$\mathbf{c}_t^b = (1-\eta)\mathbf{c}_{t-1}^b - 1 + \eta\tilde{\mathbf{c}}^b, \tag{2}$$

where:
η – learning rate,
$\mathbf{c}_{t-1}^f$ – foreground histogram of the previous frame,
$\mathbf{c}_{t-1}^b$ – background histogram of the previous frame.

The next step is to calculate the channel reliability **w:**

$$\mathbf{w} = \mathbf{w}^{(\mathrm{lrn})} \times \mathbf{w}^{(\mathrm{det})}, \tag{3}$$

where:
$\mathbf{w}^{(\mathrm{lrn})}$ – reliability of the learning channel,
$\mathbf{w}^{(\mathrm{det})}$ – reliability of the detection channel.

The last step updates the filter, and channel reliability using the following equations:

$$\mathbf{h}_t = (1-\eta)\mathbf{h}_{t-1} + \eta\mathbf{h}, \tag{4}$$
$$\mathbf{w}_t = (1-\eta)\mathbf{w}_{t-1} + \eta\mathbf{w}, \tag{5}$$

where:
$\mathbf{h}_{t-1}$ – filter calculated for the previous frame,
$\mathbf{w}_{t-1}$ – reliability of the channel of the previous frame

MIL Tracker

The MIL Tracker algorithm represents supervised learning in which training instances are arranged in sets called bags. The training set consists of many bags containing instances. The bag is positively labelled if it has at least one positive case. Otherwise, the bag is marked negatively. The task of the method is to learn the concept from the training kit for the correct labelling of bags [7].

The MIL algorithm uses a gradient-increasing structure to train the classifier to maximise the occurrence of positive cases. This operation is executed according to the formula presented below [8]:

$$\log L = \sum_i (\log p(y_i|X_i)), \tag{6}$$

where:
L – plausibility function,
$p(y_i|X_i)$ – likelihood of a positive bag.

The probability is determined for bags, not for instances, because the labels of instances are not known during training. The last step of the algorithm is to determine the probability of a positive bag. In this case, the Noisy-OR model, described by the formula below, is adopted [8]:

$$p(y_i|X_i) = 1 - \prod_j \left(1 - p\left(y_i|x_{ij}\right)\right). \tag{7}$$

MOSSE Tracker

The MOSSE tracker utilises correlation filters. By using adaptive correlation for object tracking, it generates stable correlation filters utilising only one frame. This leads to high performance by computing time-domain correlations. The tracking module is resistant to variable lighting conditions, as well as the object's movement and its deformation. The module calculates the minimum square error output to find the most precise location of the target [5].

The shape of the object is assumed to have a two-dimensional Gaussian distribution. Regarding the baseline image, the top of the distribution is located in the centre of the input picture. To find the filter template H, the algorithm uses the square of the sum of the output convolution, and determines the output error of the convolution using the following operation [9]:

$$H = \min_H \sum_i |F_i \odot H^* - G_i|^2, \tag{8}$$

where:
F_i – pre-processed template,
H^* - complex filter conjugate,
G_i – image of the object,
$\odot$ - elementary multiplication symbol.

In order to increase the learning speed, the algorithm raises the weight of the previous frame, and allows it to spread exponentially. The filter adapts to changes of the object's appearance by determining the coefficients using the following formulas [9]:

$$A_i = \eta G_i \times F_i^* + (1-\eta)A_{i-1}\,, \tag{9}$$
$$B_i = \eta F_i \times F_i^* + (1-\eta)B_{i-1}, \tag{10}$$

where:
F_i^* - the complex conjugate of the processed template.

The coordinates of the highest point in the result are recognised as the target at the frame's current position; thus, the new target is used to update the filter template. The above process is repeated for continuity in tracking [9].

KCF Tracker

The concept of this tracker assumes that many positive samples have large overlapping regions. The module uses these regions to speed up tracking, and to obtain the highest accuracy. The number of calculations is reduced by the properties of the circulating matrix and the kernel function. Additionally, the prediction of unidentified data is performed by linear ridge regression, based on the machine learning solution. The set of sample images is denoted as a_i, while the variable z_i takes the value +1 if the object is present in the image, and -1 if the object is imperceptible. Subsequently, the normalised ridge regression, based on square loss and square regulation, is used. It is a convex function with a unique solution, which can be represented as [11]:

$$w = (\mathbf{A}^T \boldsymbol{A} + \lambda I)^{-1} \boldsymbol{A}^T \mathbf{z}, \tag{11}$$

where:
$\boldsymbol{A}$ – matrix comprising of all the vectors used for training,
$\mathbf{z}$ – vector consisting of appropriate values z_i.

The circulating matrix $\mathbf{C}(\mathbf{a})$ obtains its values using the discrete Fourier transform [11]:

$$\mathbf{C}(\mathbf{a}) = U diag(\hat{a}) U^*, \tag{12}$$

where:
$diag(\hat{a})$ – the diagonal matrix containing the DFT coefficients of vector **a,**
U – the value of the discrete Fourier transform.

RESULTS AND DISCUSSION

This section presents a comparative analysis of the described algorithms. The algorithms were tested to determine their tracking accuracy and efficacy. In order to carry out research, the algorithms were implemented using the C++ language, as well as OpenCV and Qt libraries. The trackers were tested on video sequences, in which the movement of UAVs differed in respect of speed, dynamics, and distance from the camera. Additionally, various weather and lighting conditions were taken into consideration. One frame of the first sequence is shown in Fig. 1.

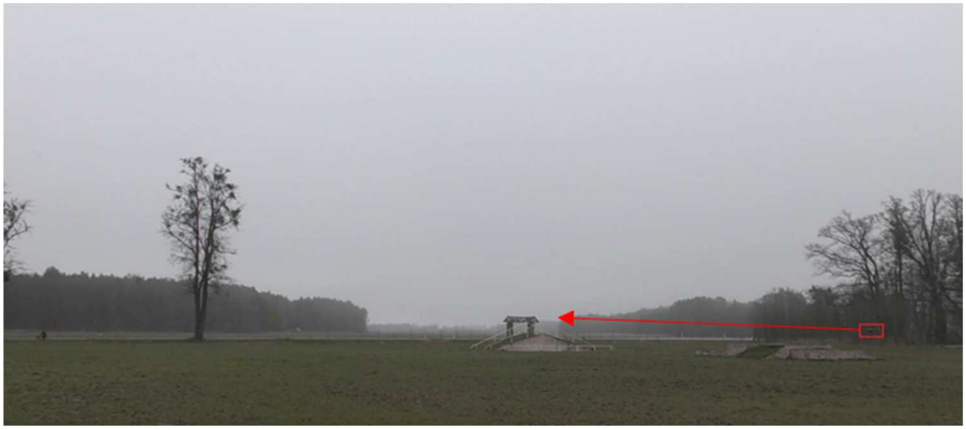

Fig 1. An exemplary frame of the first sequence. [own work]

Research results

Efficacy was the first analysed parameter. It was calculated as the ratio of the number of frames containing the correctly tracked object to the total number of frames in the sequence. Since in our work, we focused on tracking methods, the position of the object was manually marked in the first frame of the sequences.

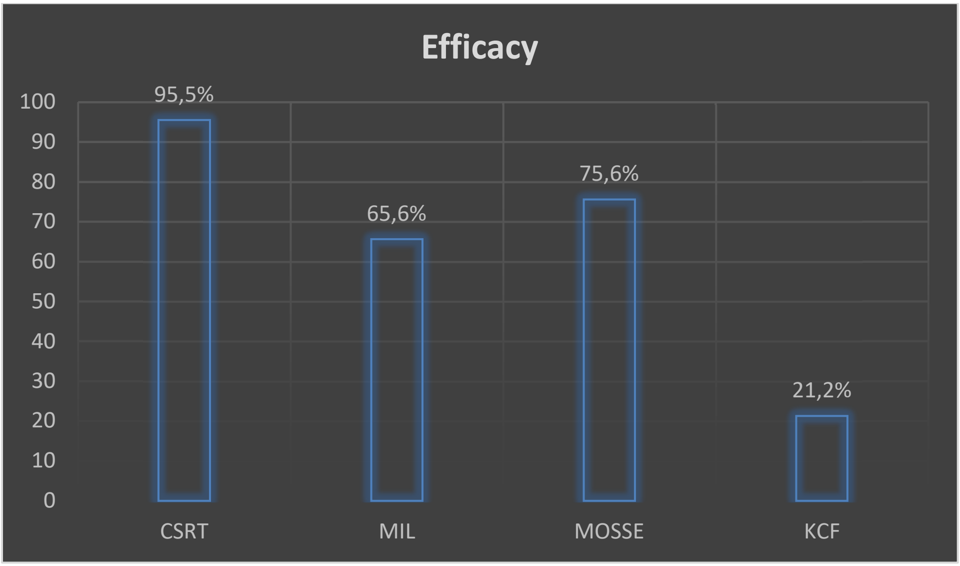

Fig 2. Efficacy of the tested algorithms. [own work]

The results of the conducted experiments are depicted in Fig. 2. They show that the CSRT algorithm tracked the object with very high accuracy, reaching ninety-six per cent. The MOSSE and MIL algorithms obtained about seventy per cent accuracy, which is a poor result for developing a reliable tracking application. The KCF method achieved only twenty-one per cent accuracy.

The next analysed parameter was the average processing time of a single frame in the analysed sequences.

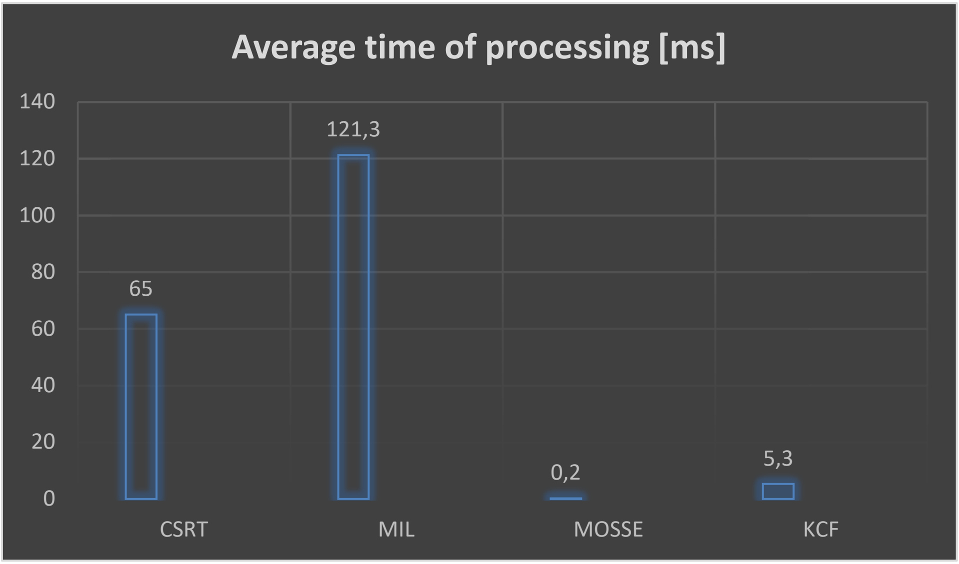

Fig 3. Average processing time of the tested algorithms. [own work]

In this case, the MOSSE module outperformed the other algorithms, achieving a time of 0.2 ms. The KCF algorithm also performed very quickly, while the CSRT and MIL methods were much slower (see Fig. 3). Even though the execution time of the MIL method amounted 121 ms, we assumed it as sufficient for real-time tracking applications.

The third analysed parameter was the shortest single frame processing time. The obtained results were in line with the average time, and proved that the MOSSE

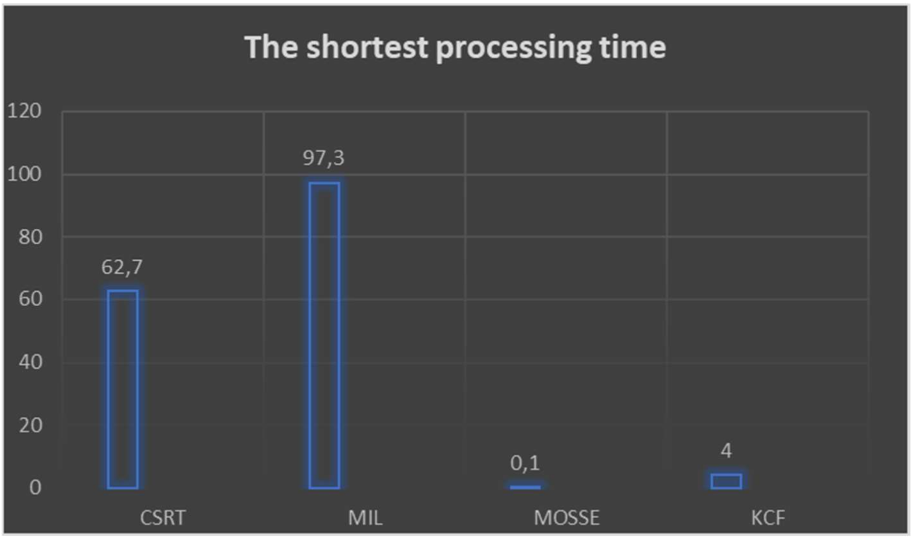

Fig 4. The shortest processing time of the tested algorithms. [own work]

and KCF methods are the fastest ones, while the MIL and CSRT methods execute much slower (Fig. 4).

The last investigated parameter was the longest processing time of a single frame in the tested sequences.

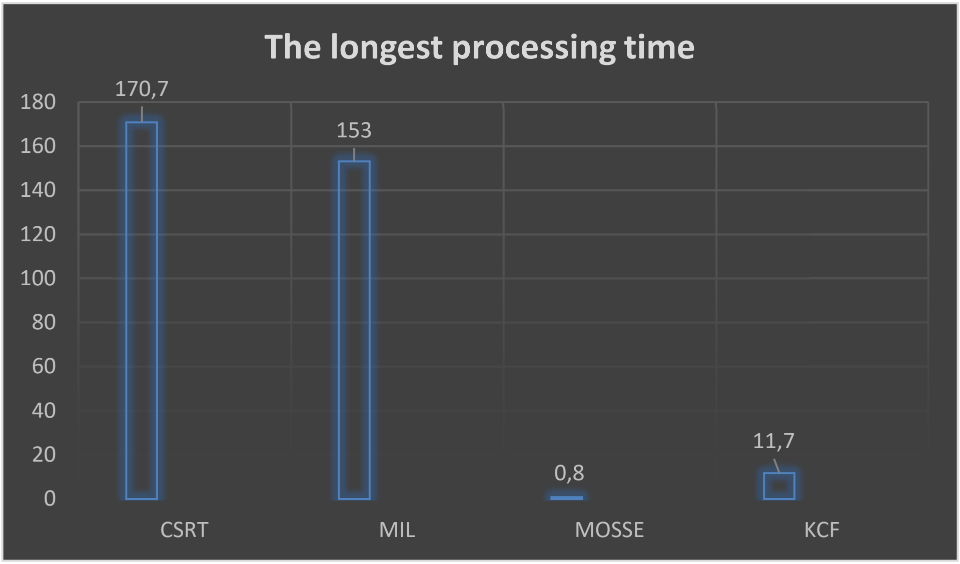

Fig 5. The longest processing time of the tested algorithms. [own work]

In this case, the CSRT and MIL algorithms executed significantly slower than the MOSSE and KCF methods. The longest processing time, amounted to 171 ms, was achieved by the CSRT algorithm, while the shortest time, equals to 0.8 ms, was obtained by the MOSSE method (see Fig. 5).

In our analysis, efficacy posed the most crucial factor because it determined the ability of the method to perform tracking of UAVs. Taking into consideration the characteristics of UAVs' movement, as well as weather, and lighting conditions, we acknowledged the CSRT method as the most appropriate for UAVs tracking applications. Even though its execution time varied from 63 to 171 ms, it may be applied for real-time application since longer time did not impede the UAVs tracking performance.

The MOOSE method executed very quickly. Consequently, improvement of its accuracy would make it suitable for fast real-time applications. The improvement could be achieved by adding additional image pre-processing steps, allowing a better distinction of UAVs from the background. We assume, that efficacy that is higher than ninety per cent will facilitate tracking with a high level of confidence. Additionally, in the case when higher efficacy was difficult to achieve, the cooperation of object detection and tracking algorithms would be considered. It would be convenient due to the short execution time of the MOOSE method. What is more, the efficacy of this algorithm could be improved with position estimators such as the Kalman, or Particle filters.

The performance of the MIL and KCF methods was lower than the CSRT and MOSSE methods. The MIL algorithm was the slowest one, whereas the KCF method achieved the lowest efficacy. Consequently, they are less suitable for real-time UAV applications.

CONCLUSIONS

Analysis of the above algorithms has shown that the CSRT algorithm is the best choice for UAVs tracking. Although the algorithm was not the fastest one, it outperformed other algorithms in the quality of efficacy. The MOSSE method, which was very fast, achieved seventy-six per cent efficacy. This efficacy is too low for tracking applications, but it can be improved using image pre-processing steps, cooperation with detection algorithm, or the Kalman and Particle filters. Therefore, additional experiments are needed to evaluate its effectiveness in real-time tracking applications. The MIL and KCF methods appeared to be unsuitable for UAVs tracking.

By combining the MOG algorithm for detection, and the CSRT algorithm for tracking, the backbone of an autonomous security system can be developed. In combination with modern vision systems, the effective, and efficient monitoring solution for essential infrastructure can be achieved.

The presented research has been carried out under daylight conditions. Since night conditions determine utilising an infrared camera, future work will focus on selecting the most suitable methods for UAVs detection, and tracking on thermal images.

REFERENCES

[1] Wąż, M.; Naus, K. *Wizualizacja wielowymiarowego obrazu radarowego*. Zesz. Nauk. Akad. Mar. Wojennej 2013, *195*, 99–116, doi:10.5604/0860889X.1097972.

[2] Blok, M.; Kaczmarek, S.; Młynarczuk, M. *Radar Data Fusion in the Stradar System*. Sci. J. Polish Nav. Acad. ***2019***, *218*, 43–56, doi:10.2478/sjpna-2019-0017.

[3] Yoon, J. H.; Xu, H.; Garcia Carrillo, L. R. *Advanced Doppler radar physiological sensing technique for drone detection*. In; Ranney, K. I., Doerry, A., Eds.; **2017**; pp. 10–21.

[4] Piskur, P.; Szymak, P. *Algorithms for passive detection of moving vessels in marine environment.* J. Mar. Eng. Technol. 2018, *16*, 377–385, doi:10.1080/20464177.2017.1398483.

[5] Sutin, A.; Salloum, H.; Sedunov, A.; Sedunov, N. *Acoustic detection, tracking and classification of Low Flying Aircraft.* In *2013 IEEE International Conference on Technologies for Homeland Security (HST)*; IEEE, 2013; pp. 141–146.

[6] Dumitrescu, C.; Minea, M.; Costea, I. M.; Cosmin Chiva, I.; Semenescu, A. *Development of an Acoustic System for UAV Detection*. Sensors 2020, *20*, 4870, doi:10.3390/s20174870.

[7] Hożyń, S.; Zalewski, J. *Shoreline Detection and Land Segmentation for Autonomous Surface Vehicle Navigation with the Use of an Optical System*. Sensors 2020, *20*, 2799, doi:10.3390/s20102799.

[8] Praczyk, T.; Hożyń, S.; Bodnar, T.; Pietrukaniec, L.; Błaszczyk, M.; Zabłotny, M. *Concept and first results of optical navigational system*. Trans. Marit. Sci. 2019, *8*, 46–53, doi:10.7225/toms.v08.n01.005.

[9] Hożyń, S.; Żak, B. *Local image features matching for real-time seabed tracking applications*. J. Mar. Eng. Technol. 2017, doi:10.1080/20464177.2017.1386266.

[10] Maan, S.; Tyagi, A.; Kumar, S.; Kumar, A. *Security and Surveillance using Computer Vision*. Int. J. Comput. Appl. 2020, *176*, 30–35, doi:10.5120/ijca2020920280.

[11] Hożyń, S.; Żak, B. *Moving Object Detection, Localization and Tracking Using Stereo Vison System.* Solid State Phenom. 2015, *236*, 134–141, doi:10.4028/www.scientific.net/SSP.236.134.

[12] Gautam, K. S.; Parameswaran, L.; Thangavel, S. K. *Computer Vision Based Asset Surveillance for Smart Buildings.* J. Comput. Theor. Nanosci. **2020**, *17*, 456–463, doi:10.1166/jctn.2020.8691.

[13] Hożyń, S.; Przybysz, M. *Detection of Unmanned Aerial Vehicles Using Computer Vision Methods: A Comparative Analysis*. Sci. J. Polish Nav. Acad. 2020, *220–221*, 5–13, doi:10.2478/sjpna-2020-0001.

[14] Lukežič, A.; Vojíř, T.; Čehovin Zajc, L.; Matas, J.; Kristan, M. *Discriminative Correlation Filter Tracker with Channel and Spatial Reliability*. Int. J. Comput. Vis. **2018**, *126*, 671–688, doi:10.1007/s11263-017-1061-3.

[15] Babenko, B.; Ming-Hsuan Yang; Belongie, S. *Robust Object Tracking with Online Multiple Instance Learning*. IEEE Trans. Pattern Anal. Mach. Intell. 2011, *33*, 1619–1632, doi:10.1109/TPAMI.2010.226.

[16] Bolme, D.; Beveridge, J. R.; Draper, B. A.; Lui, Y. M. *Visual object tracking using adaptive correlation filters.* In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; IEEE, **2010**; pp. 2544–2550.

[17] Henriques, J. F.; Caseiro, R.; Martins, P.; Batista, J. *High-speed tracking with kernelised correlation filters*. IEEE Trans. Pattern Anal. Mach. Intell. 2015, *37*, 583–596, doi:10.1109/TPAMI.2014.2345390.

ANALIZA METOD ŚLEDZENIA BEZZAŁOGOWYCH STATKÓW POWIETRZNYCH WYKORZYSTUJĄCYCH TECHNIKI WIDZENIA KOMPUTEROWEGO

STRESZCZENIE

W artykule przedstawiono analizę metod śledzenia bezzałogowych statków powietrznych, wykorzystujących techniki widzenia komputerowego.

Słowa kluczowe: Bezzałogowy statek powietrzny, widzenie komputerowe, śledzenie obiektów.