GAZİOSMANPAŞA BİLİMSEL ARAŞTIRMA DERGİSİ (GBAD)
Gaziosmanpasa Journal of Scientific Research
ISSN: 2146-8168
http://dergipark.gov.tr/gbad
Araştırma Makalesi (Research Article)

Cilt/Volume : 12
Sayı/Number: 2
Yıl/Year: 2023
Sayfa/Pages: 93-104

Alınış tarihi (Received): 25.08.2023
Kabul tarihi (Accepted): 28.08.2023

Motion-Based Object Classification on UAV Images

Ahmet ER[1,*], Abdullah Hakan YAVUZ[2]

[1] Department of Mechatronics Engineering, Faculty of Engineering and Architecture, Tokat Gaziosmanpaşa University, Tokat, Türkiye
[2] Department of Electrical and Electronics Engineering, Faculty of Engineering and Architecture, Tokat Gaziosmanpaşa University, Tokat, Türkiye
**Corresponding author: er.ahmet@outlook.com*

ABSTRACT: Motion detection from video images is used for different purposes in various fields, especially in security. However, some difficulties may be encountered in detecting moving objects. Especially when the camera is moving, motion detection becomes difficult. This study, it is aimed to detect the moving object from the video images and to classify this detected object. It is recommended to use two different methods together for detecting the moving object. In the first method, reference points are determined on the image to separate the moving parts from the camera movement. These points were followed by optical flow along successive frames in the video image. Motion vectors are created from these traced points. The vectors that differed from the others in terms of slope and length were those belonging to the moving object. Then, in the second method, by calculating the transformation matrix between the frames, motion detection from the video image was performed with background stabilization. Finally, the parts that these two methods detected as moving were tried to be classified by adding a classifier layer to the pre-trained VGG16 model. With the study, moving objects on the VIVID Dataset could be detected and classified correctly.

Keywords – *Moving Object Detection, Object Tracking, Deep Learning, Anomaly Detection*

İHA Görüntülerinde Hareket Tabanlı Nesne Sınıflandırması

ÖZET: Video görüntülerinden hareket algılama başta güvenlik olmak üzere çeşitli alanlarda farklı amaçlarla kullanılmaktadır. Ancak hareketli nesnelerin tespitinde bazı zorluklarla karşılaşılabilinmektedir. Özellikle kamera hareket halindeyken hareket algılama zorlaşmaktadır. Bu çalışmada, video görüntülerinden hareketli nesnenin tespit edilmesi ve tespit edilen bu nesnenin sınıflandırılması amaçlanmaktadır. Hareketli nesnenin tespitinde iki farklı yöntemin bir arada kullanılması önerilmektedir. Birinci yöntemde hareketli kısımları kamera hareketinden ayırmak için görüntü üzerinde referans noktaları belirlenir. Bu noktalar, video görüntüsündeki ardışık kareler boyunca optik akış ile takip edilmektedir. İzlenen bu noktalardan hareket vektörleri oluşturulmaktadır. Eğim ve uzunluk açısından diğerlerinden farklılaşan vektörler hareketli nesneye ait vektörler olduğu tespit edilmiştir. Daha sonra ikinci yöntemde kareler arası dönüşüm matrisi hesaplanarak arka plan stabilizasyonu ile video görüntüsünden hareket tespiti yapılmıştır. Son olarak bu iki yöntemin kesiştirilerek hareketli olduğu tespit edilen bölgeler, önceden eğitilmiş VGG16 modeli ile bir sınıflandırıcı katman eklenerek sınıflandırılmaya çalışılmıştır. Çalışma ile VIVID Veri Seti üzerinde hareket eden nesnelerin doğru bir şekilde tespit edilip sınıflandırılması mümkün olmuştur.

Anahtar Kelimeler – *Hareketli Nesne Tespiti, Nesne Takibi, Derin Öğrenme, Anomali Tespiti*

1. Introduction

Moving object detection has an essential place in the field of computer vision. It also plays an important role in medical detection, virtual reality, industry, agriculture, and security fields. In particular, with the increase in the use and development of Unmanned Aerial Vehicles (UAV), an increase in the interest in detecting moving objects and the number of studies conducted in this context has been observed. Moving object detection and tracking; is critical when considering security applications such as creating safe zones, border security, and surveillance systems. Although there are many studies in this field in the literature, there are still issues waiting to be resolved (Bal et al. 2017). Studies for motion detection are divided into two groups in the literature. These are motion detection from fixed camera and motion detection from motion camera. The methods used can be listed as background subtraction (Carmona et al. 2008), optical flow (Jodoin and Mignotte, 2009), and the difference between frames. In addition to these, fuzzy logic and artificial neural network-based approaches are also used (Bal et al. 2017) (Zhu et al. 2020).

Motion detection from a fixed camera is relatively easier than when the camera is moving. The biggest problem in detecting moving objects from a moving camera is the difficulty in detecting a freely moving object due to background motion. The low resolution of the image or the large distance can be counted among the other reasons that make object detection and classification difficult (Delibaşoğlu, 2022). On the other hand, the weather conditions of the outdoor environment, the presence of any obstacle between the object and the camera, and the shadows of the objects are factors that affect the sensitivity of moving object detection (Zhu et al. 2020). The most preferred method for motion detection is optical flow. Optical flow is used for motion-based classification and feature extraction in object tracking, and Horn-Schunck and Lucas-Kanade methods are preferred (Bal et al. 2017).

In this study, reference points were determined by optical flow from the frames in the video images, and the movements of the said points were tried to be followed in successive frames. For the detection of the moving object, outlier detection was performed on the motion vectors obtained from the reference points determined before. Vectors that diverge abnormally from other vectors are considered moving parts. The transformation matrix was calculated using the associated reference points followed between the frames. Then, the images were warped and obtained a binary mask, and with the help of this mask, the image of the object was cropped from the image. Cropped images consisting of pixels of the moving object in the image were tried to be classified with the VGG16 network.

2. Related Studies

2.1. Moving Object Detection Methods

Many promising studies have been carried out for object detection and motion detection from video images. These studies aiming at detecting moving objects from a moving camera can be roughly evaluated under four headings as panoramic background subtraction, use of dual cameras, background stabilization, and artificial neural network-based approaches (Chapel and Bouwmans, 2020).

In the panoramic background subtraction method, it is used for both situations where the camera is both fixed and moving. With the images obtained from a moving camera, a larger picture called a panoramic or mosaic is created and the background or moving object is tried to be detected from this image. The method using dual cameras is similar to panoramic background subtraction. The only difference is that the image is taken from two different cameras simultaneously. Camera calibration plays an important role in this method (Chapel and Bouwmans, 2020).

In the motion or background compensation method, featured points are extracted from successive frames. Based on these points, the homography matrix is calculated for the frames, and the frames are warped with this matrix to keep the background constant. Then, it is tried to detect the moving object with morphological processes (Zhu et al. 2020). In addition, methods using the parallax method (Berker et al. 2017) or based on the temporal consistency of each target hypothesis by generating target hypotheses for different scales, both rejecting outliers and compensating for missing detections in a time-efficient manner were also used (Alkanat et al. 2015). Additionally, there are also studies such as RAFT (Recurrent All-Pairs Field Transforms) (Teed et al. 2020) and FlowNet (Fischer et al. 2015), in which models are trained directly over optical flow to detect the context between frames, and Convolutional neural networks (CNNs) for motion detection.

2.2. Object Detection and Object Classification

Deep learning methods are very useful in extracting the characteristics of objects and for object detection or tracking. So much so that with a well-trained CNN model, object detection can be done quickly and consistently. Models such as CNN-based Faster R-CNN (Girshick, 2015), SSD (Single Shot Multi-Box Detector) (Liu et al. 2016), and YOLO (You Only Look Once) (Redmon et al. 2016) are models that have fast and high accuracy in detecting objects.

3. Proposed Method

In this study, it has been tried to detect, classify and follow the optical flow-based moving object over the moving camera. The proposed method consists of three parts. In the first part, it is the detection of the image parts that have outliers from the movements followed along the frames with the optical flow method. In the second part, the homography matrix is determined from the features obtained with the SIFT algorithm and the image is warped. Moving areas are detected by subtracting the background from the warped image. In the third part of the study, the intersection of the moving areas detected in the first two parts is taken. In this way, it is ensured that the detection of the moving object is made more sensitively and successfully. The flow chart of the proposed approach is shown in Figure 1 and the proposed method is explained.

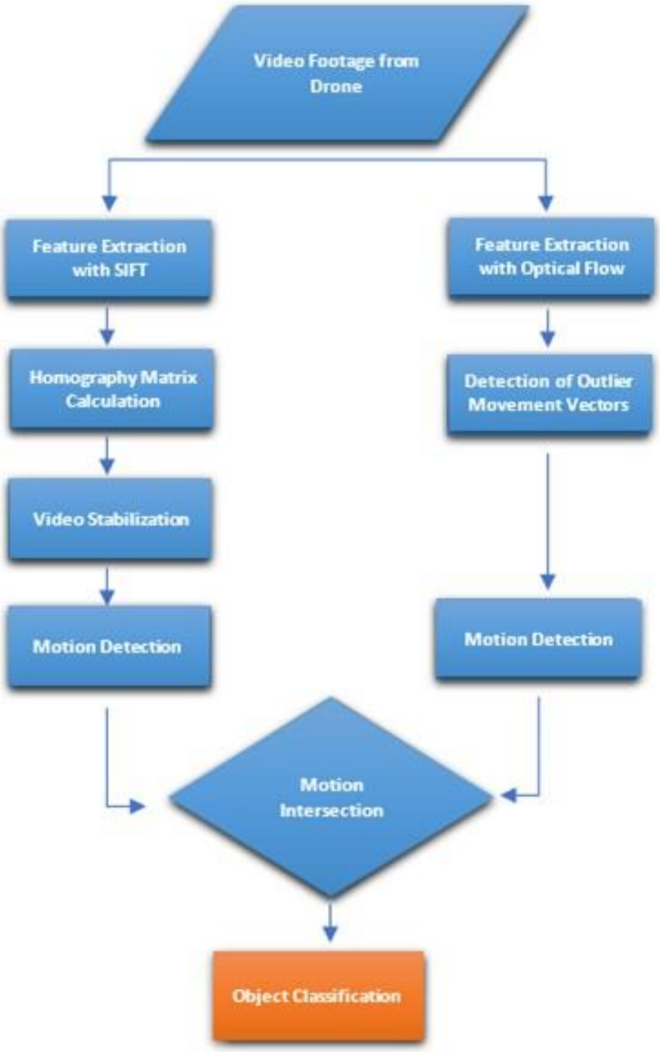


Figure 1. Flow chart of the proposed method.
Şekil 1. Önerilen yöntemin akış diyagramı

3.1. Motion Detection with Optical Flow

It is assumed that the intensity/brightness of the pixels of the relevant object does not change in the optical flow. As a result of this assumption, the equation $I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t)$ is obtained. When the first-degree Taylor Expansion is made to the said equation;

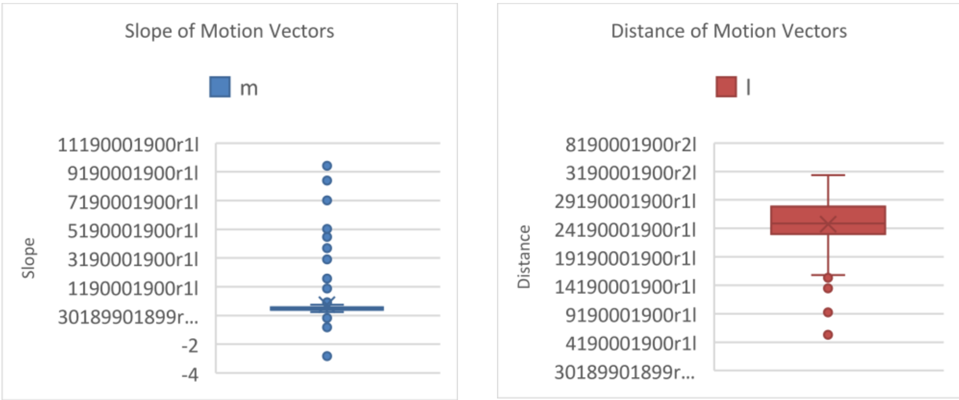
$$I(x + dx, y + dy, t + dt) = I(x, y, t) + \frac{\partial I}{\partial x} u + \frac{\partial I}{\partial y} v + \frac{\partial I}{\partial t} t + \cdots$$
$$= \frac{\partial I}{\partial x} dx + \frac{\partial I}{\partial y} dy + \frac{\partial I}{\partial t} dt = 0 \quad (1)$$

Equation (1) is reached. When this equation is divided by dt, the optical flow formula in equations (2), (3) is obtained (Lin, 2019).

$$= \frac{\partial I}{\partial x} u + \frac{\partial I}{\partial y} v + \frac{\partial I}{\partial t} t = 0 \quad (2)$$

$$Ixu + Iyv = -It \tag{3}$$

Lucas-Kanade is the most preferred method in optical flow and its use was preferred in this study. The reference points determined by this method are followed along the frames. The optical flow is calculated between the current frame and the backward fifth frame held in the queue. Assuming that the moving object to be detected will be smaller than the background, it is predicted that the change in the vectors of the camera movement will be separated from the vectors of the moving object on the basis of slope and length. The outlier in the motion vectors observed in the image stream is observed from the Box Plot graphs of the slope and length of the motion vectors in Figure 2a and Figure 2b. Outliers are values that differ significantly from other observations or the rest of the dataset (Kriegel et al. 2011).



(a) (b)

Figure 2. (a) Box plot of slope. (b) Box plot chart of distance.
Şekil 2. (a) Eğime ait box plot grafiği. (b) Uzunluğa ait box plot grafiği.

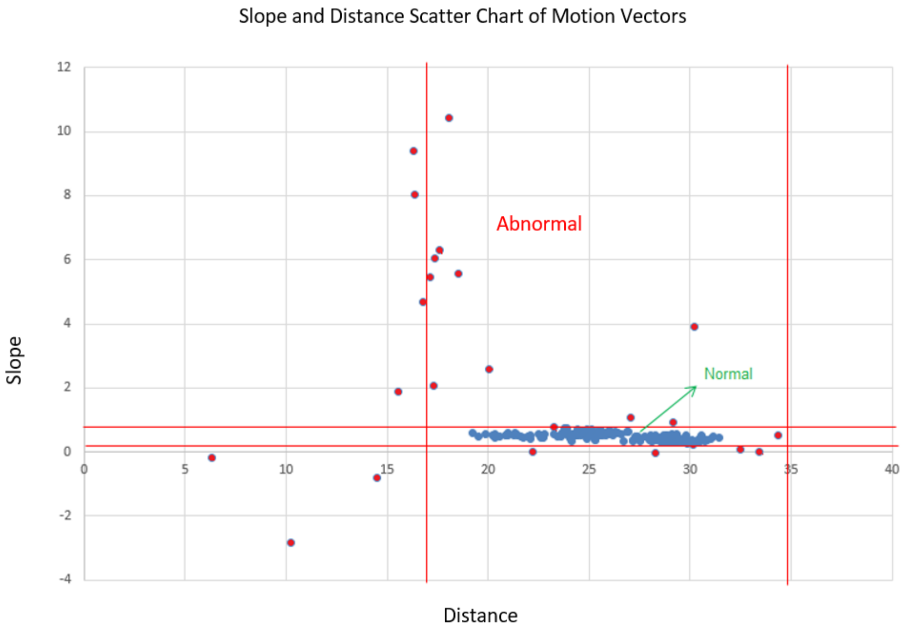


Figure 3. Slope and distance scatter plot chart.
Şekil 3. Eğim ve uzunluk scatter plot grafiği.

With the *Interquartile Range (IQR)*, the values that differed from the group in terms of both slope and magnitude of the vectors were determined. In equations (4)-(6), the quarterly gap formula is included.

$$Q1 = \frac{1}{4}\,[(n\ +\ 1)]\ th\ term) \tag{4}$$

$$Q3 = \frac{3}{4}\,[(n\ +\ 1)]\ th\ term) \tag{5}$$

$$IQR = Q3 - Q3 \tag{6}$$

In this part of the study, the slope of the motion vectors with normal values in the subject image is collected between 0.23 and 0.75 values. Distance values are between 16 and 34. Values outside of this range are considered outliers. The outlier calculation is calculated separately for both frames being compared. Figure 3 shows the scatter plot of the magnitude and slope values of the vectors of the reference points formed due to the camera movement. In Figure 4, the display of outliers on the image is given and it is observed that the motion vectors exhibiting abnormal changes are on the moving object.

Figure 4. Display of abnormal motion vectors on image.
Şekil 4. Anormal hareket vektörlerinin görüntü üzerinde gösterimi.

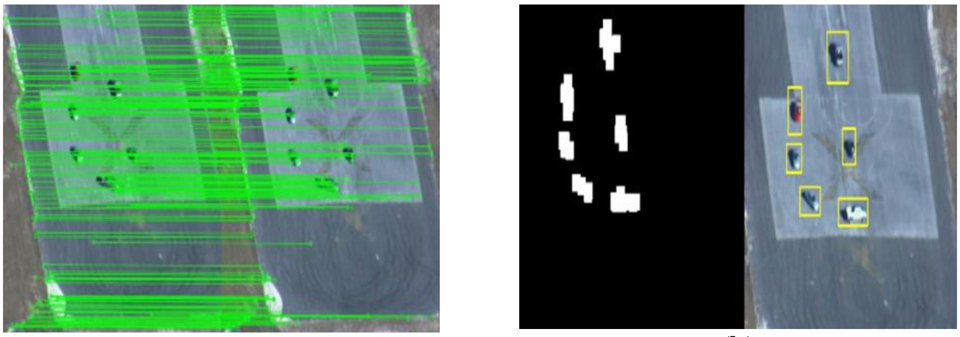
3.2. Motion Detection by Compensating the Background

Background stabilization is based on calculating the transform matrix by matching the features obtained in the first frame on successive frames and balancing the background of the image with this matrix. The compared frames are kept in a queue of five consecutive frames, and each frame is compared with the five frames before it. As a result of balancing the background in the image, moving parts can be brought to the fore.

The SIFT algorithm extracts features from the image compared to SURF, FAST, and Harris algorithms and is more advantageous than other algorithms in terms of speed (Zhu et al.

2020). For this reason, the SIFT algorithm is preferred. The interframe mapping of the features obtained with SIFT is shown in Figure 5a.

The images warped by the homography matrix are taken from the difference between the three frames before and the pixels of the moving object are obtained. Figure 5b shows moving object detection with background compensation.



(a) (b)

Figure 5. (a) Feature matching between frames. (b) Motion detection with background compensation.

Şekil 5. (a) Çerçeveler arası özellik eşleştirme. (b) Arka plan dengeleme ile hareket tespiti.

3.3. Taking Optical Flow and Background Compensation Intersection

With the intersection of the methods mentioned in the previous two sections, it is aimed to increase the accuracy in detecting the moving object. In this context, it was ensured that the binary image obtained by background compensation and the motion vectors detected using optical flow and having outliers compared to the others were overlapped. As seen in Figure 6, it has been observed that the intersection of the parts detected as moving in both methods consists of pixels belonging to the moving object.

Figure 6: Intersection of methods.

Şekil 6. Yönetmlerin kesişimi

3.4. Classification of Detected Moving Object

The moving parts extracted from the video image were tried to be classified with the developed deep learning model. The VGG16 neural network is applied to image classification and is frequently used in related studies (Da Rocha et al. 2022). ImageNet is a dataset of 15 million images with 22 thousand categories (Russakovsky et al. 2015). VGG16 is a model trained on this dataset (Pu et al. 2019). In cases where the data set is limited, it is a better choice to use pre-trained models (Chollet, 2018). In this context, the model is customized for the problem by adding a fully connected classifier layer on the pre-trained VGG16 model. As can be seen in Figure 7, it is predicted that using the pre-trained VGG16 network, it will be evaluated whether the moving object poses a danger or not.

Figure 7. Moving object classification.
Şekil 7. Hareketli nesne sınıflandırma.

As can be seen in Figure 8, a VGG16-based model has been developed for the classification of moving objects in the VIVID dataset.

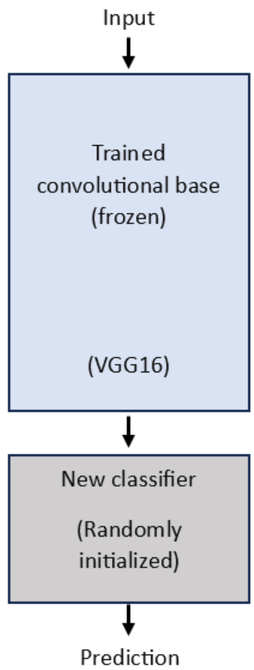


Figure 8. Adding a classifier to a pre-trained network (Chollet, 2018).
Şekil 8. Ön eğitimli ağ üzerine sınıflandırıcı eklemek (Chollet, 2018).

4. Results and Discussion

In this research, it has tried to provide a solution to the challenging problem of automatic object detection and identification from surveillance videos recorded by motion cameras. Although an increase has been observed in the studies on moving object detection today, there are still deficiencies in terms of a correctly labeled dataset containing moving objects from low altitudes. The proposed method was applied and evaluated on VIVID Dataset. The algorithm was developed with a computer with Intel Core i5-8250U @ 1.60 GHz processor and 16 GB RAM. The developed algorithm was prepared in Python programming language and Keras, OpenCV, Numpy, and Pandas libraries were used.

4.1. Datasets

The dataset used in the study is the VIVID Dataset, which was created under the Video Verification of Identity (VIVID) program by The Defense Advanced Research Projects Agency (DARPA) for monitoring air-to-ground vehicles. This data set consists of images of ground vehicles, which were obtained from an unmanned aerial vehicle and in environments with different terrain structures. In this dataset, there are five different video images with 640x480 pix resolution and 10,019 frames.

4.2. Evaluation Metrics

In this section, the Confusion Matrix results of the algorithm are given to measure the performance of the study. The algorithm was evaluated separately for motion detection and separately for object classification.

Table 1. Confusion matrix.
Tablo 1. Karışıklık matrisi.

		Predicted Values	
		Class A	**Class B**
Actual Values	**Class A**	True Positive (TP)	False Positive (FP)
	Class B	False Negative (FN)	True Negative (TN)

A confusion matrix is a tabulation technique used to measure the performance of classification algorithms (Bulut, 2017). This technique details the correct and incorrect predictions made by the classification model. The metrics used in the Confusion Matrix are shown in equations (7)- (10).

$$\boldsymbol{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{7}$$

$$\boldsymbol{Precision} = \frac{TP}{TP + FP} \tag{8}$$

$$\boldsymbol{Recall} = \frac{TP}{TP + FN} \tag{9}$$

$$\boldsymbol{F_1} = 2 * \frac{precision * recall}{precision + recall} \tag{10}$$

4.3. Experimental Results

The success of the proposed method on the data set used is given in this section. When the developed algorithm was applied to the VIVID data set, an average of 5 FPS values were obtained with the computer used.

The success of the algorithm in detecting moving objects from moving video images was evaluated with the Confusion Matrix. The evaluation was made over the first 1,000 frames of the video images in the dataset. According to the evaluation results in Table 1, the proposed method showed a successful performance. However, it has been observed that the success of video images with low resolution decreases.

Table 2. Accuracy, Precision, Recall, F1-score Values for Motion Detection
Tablo 2. Hareket tespiti için Doğruluk, Kesinlik, Duyarlılık, F1-score değerleri

Sequence	**egtest01**	**egtest02**	**egtest03**	**egtest04**	**egtest05**
Accuracy	0.85	0.89	0.67	0.71	0.80
Precision	0.97	0.99	0.82	0.87	0.92
Recall	0.88	0.90	0.75	0.76	0.85
F1-score	0.92	0.94	0.78	0.81	0.89

Table 3 shows the classification success of correctly detected moving objects. However, classification success can be mentioned when the object is not closed. As seen in Table 3, the precision value for the classification step ranges from 89% to 99%. The average precision for five videos is 94%.

Table 3. Precision rates for classification.
Tablo 3. Sınıflandırma için Precision oranları.

Sequence	**egtest01**	**egtest02**	**egtest03**	**egtest04**	**egtest05**
Precision	0.99	0.97	0.94	0.89	0.91

The comparison of previous studies on motion detection on the same data set with the results of this study is shown in Table 4. As can be seen from Table 4, it has been understood that the proposed method is successful in detecting motion from a moving camera.

Table 4. Comparison of success rates of proposed method.
Tablo 4. Önerilen yöntemin başarı oranlarının karşılaştırılması.

Authors	**Year**	**egtest01**	**egtest02**	**egtest03**	**egtest04**	**egtest05**
Alkanat et al.	2015	0.99	0.89	0.81	0.95	0.91
Logoglu et al.	2017	0.93	0.85	-	0.71	0.71
Delibasoglu	2022	0.87	0.74	0.47	0.68	0.52
Proposed Method	2023	0.97	0.99	0.82	0.87	0.92

5. Conclusion

This study has tried to detect and classify moving objects from noisy images caused by camera movement in UAV images. In this context, it is thought that the algorithm developed will contribute to the creation of a safe zone for security purposes and the observation of this zone. In the study, the sparse optical flow algorithm is used to keep the processing performance at a reasonable level. It has been observed that the detection made by combining feature-based object detection and background compensation-based object detection is more successful. The VGG16 network was used as the classification model. It is evaluated that the success of the model formed with the classification layer built on the pre-training network will increase in direct proportion to the increase in the number of samples. In future studies, it is thought that more sensitive determinations can be made by extracting features for each pixel by using a system with better processing power and dense optical flow. It is considered to conduct a study in which higher performance values can be obtained in the future.

6. References

Alkanat, T., Emre T., Sinan Ö., 2015. A Real-time, Automatic Target Detection and Tracking Method for Variable Number of Targets in Airborne Imagery, In VISAPP (2) 61-69.

Bal M., Karakaya İ., Başeske E., 2017. Point based motion detection on UAV cameras, 2017 25th Signal Processing and Communications Applications Conference (SIU), 1–4. https://doi.org/10.1109/SIU.2017.7960468.

Bulut F., 2017. A new clinical decision support system with instance based ensemble classifiers, Journal of the Faculty of Engineering and Architecture of Gazi University, 32 (1) 65-76.

Carmona, E. J., Martínez-Cantos, J., & Mira, J., 2008. A new video segmentation method of moving objects based on blob-level knowledge, Pattern Recognition Letters, 29(3) 272-285.

Chapel M-N, Bouwmans T., 2020. Moving objects detection with a moving camera: A comprehensive review., Computer Science Review, 38 (100310). https://doi.org/10.1016/j.cosrev.2020.100310.

Chollet, F., 2018. Deep Learning with Python. Shelter Island, NY: Manning Publications.

Da Rocha, D. A., Ferreira, F. M. F., Peixoto, Z. M. A., 2022. Diabetic retinopathy classification using VGG16 neural network, Research on Biomedical Engineering, 38 (2) 761-772.

Delibaşoğlu İ., 2022. Vehicle Detection from Aerial Images with Object and Motion Detection, Turkish Journal of Mathematics and Computer Science, 14 174–83. https://doi.org/10.47000/tjmcs.1002767.

Fischer, P., Dosovitskiy, A., Ilg, E., Häusser, P., Hazırbaş, C., Golkov, V., Van der Smagt, P., Cremers, D., Brox, T., 2015. Flownet: Learning optical flow with convolutional networks.

Girshick, R., 2015. Fast r-cnn, In Proceedings of the IEEE international conference on computer vision, 1440-1448.

Jodoin, P. M., 2009. Mignotte, M., Optical-flow based on an edge-avoidance procedure, Computer Vision and Image Understanding, 113 (4) 511-531.

Kriegel, H. P., Kroger, P., Schubert, E., Zimek, A., 2011. Interpreting and unifying outlier scores, In Proceedings of the 2011 SIAM International Conference on Data Mining, Society for Industrial and Applied Mathematics 13-24.

Lin, C., 2021. Introduction to Motion Estimation with Optical Flow, (2019). Available at: https://nanonets.com/blog/optical-flow/. [Accessed 11 August 2021].

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., Berg, A. C., 2016. Ssd: Single shot multibox detector, In European conference on computer vision, Springer, Cham 21-37.

Logoglu, K. B., Lezki, H., Kerim Yucel, M., Ozturk, A., Kucukkomurler, A., Karagoz, B., Erdem, E., Erdem, A., 2017. Feature-based efficient moving object detection for low-altitude aerial platforms, In Proceedings of the IEEE International Conference on Computer Vision Workshops, 2119-2128.

Pu, Y., Yang, H., Ma, X., Sun, X., 2019. Recognition of Voltage Sag Sources Based on Phase Space Reconstruction and Improved VGG Transfer Learning. Entropy, 21 (10) 999.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection, In Proceedings of the IEEE conference on computer vision and pattern recognition 779-788.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., 2015. Imagenet large scale visual recognition challenge, International journal of computer vision, 115 (3) 211-252.

Teed, Z., Deng, J., 2020. Raft: Recurrent all-pairs field transforms for optical flow, In European conference on computer visioni, Springer, Cham, 402-419.

Zhu J., Wang Z., Wang S., Chen S., 2020. Moving Object Detection Based on Background Compensation and Deep Learning, Symmetry, 12 (1965). https://doi.org/10.3390/sym12121965.