



Anomaly Detection in Crowded Scenes With Machine Learning Algorithms

Makine Öğrenmesi Algoritmalarıyla Kalabalık Ortamlarda Anomali Tespiti

Hatice Kübra BOYRAZLI^{1*} Ahmet ÇINAR²

1 National Defence University, Turkish Military Academy Computer Engineering Department, 06550, Ankara, Turkey, hkboyrazli@kho.edu.tr
2 Fırat University, Engineering Faculty Computer Engineering Department, 23000, Elazığ, Turkey, acinar@firat.edu.tr

ARTICLE INFO

Article history:

Received: 30 December 2020
Revised: 7 March 2021
Accepted: 15 March 2021

Keywords:

Machine Learning, Anomaly
Detection, Artificial
Intelligence, Crowded Analyze

ABSTRACT

Camera systems have a very important place as they are widely used in providing security in crowded environments. The video images recorded by cameras are examined to check whether there is dangerous or unusual behavior. It is tried to develop appropriate measures according to the result of this control. Modeling human behaviors for the definition and detection of abnormal behavior has become a popular research area in recent years. This study was carried out by applying supervised learning algorithms, one of the machine learning methods, on five different scenes that in two open data sets. Normal and abnormal motion scenes were detected on the videos in the data sets. In these two data sets, abnormal motion was detected in a total of five different locations. Random Forest, Support Vector Machines and k Nearest Neighbor algorithms, which are among the supervised learning algorithms, were used in this process. The algorithms used were compared with performance criteria such as accuracy, sensitivity, precision and F1 score.

Doi: 10.24012/dumf.849923

MAKALE BİLGİLERİ

Makale geçmişi:

Geliş: 30 Aralık 2020
Düzeltilme: 7 Mart 2021
Kabul: 15 Mart 2021

Anahtar kelimeler:

Makine Öğrenmesi, Anomali
Tespiti, Yapay Zeka, Kalabalık
Analizi

ÖZET

Kamera sistemleri, kalabalık ortamlarda güvenliği sağlamada yaygın olarak kullanıldığı için oldukça önemli bir yere sahiptir. Kameralar aracılığıyla kaydedilen video görüntüleri incelenerek tehlikeli veya normalin dışında gerçekleşen davranış olup olmadığının kontrolü yapılmaktadır. Bu kontrolün sonucuna göre uygun tedbirler geliştirilmeye çalışılmaktadır. Anormal davranışların tanım ve tespiti için insan davranışlarının modellenmesi son yıllarda popüler bir araştırma alanı haline gelmiştir. Bu çalışma erişime açık iki adet veri setinde bulunan beş ayrı ortam videosu üzerinde makine öğrenmesi yöntemlerinden biri olan denetimli öğrenme algoritmaları uygulanara ile gerçekleştirilmiştir. Veri setlerinde yer alan videolar üzerinde normal ve anormal hareket sahnelerinin tespiti yapılmıştır. Bu iki veri setinde toplam beş farklı mekanda anormal hareket tespiti yapılmıştır. Bu işlemden denetimli öğrenme algoritmalarından olan Rastgele Orman, Destek Vektör Makineleri ve k En Yakın Komşu algoritmaları kullanılmıştır. Kullanılan algoritmalar, doğruluk, duyarlılık, kesinlik ve F1 skor gibi performans ölçütleri ile karşılaştırılmıştır.

* /Correspondence author / Sorumlu yazar
Hatice Kübra BOYRAZLI
✉ hkboyrazli@kho.edu.tr

INTRODUCTION

Machine learning, which is based on the understanding and processing of human actions by machines, is a complex, diverse and challenging field that has attracted considerable attention in recent years. Sensing action on human behavior, motion tracking, scene modelling and understanding of behavior (human activity recognition and determination of activity types) has become a field of great interest in computer vision and machine learning.

Depending on the development of technology, it is important to examine and analyse the behavior of people operating in crowded areas. Especially the security concern, which increases in direct proportion to the increase in the population, the camera systems; Its use has become widespread in health facilities such as stadiums, rally areas, airports, metro stations, shopping malls, military facilities, elderly nursing homes and homes of single living elderly people, such as monitoring daily activities and fall detection. Since the examination of video images recorded by camera systems requires an intense effort, it was necessary to make video surveillance systems under computer control and in an automatic way. The aim here is to recognize, identify, or learn interesting events that could be contextually defined as "suspicious event" [1], "irregular behavior [2] like in Figure 1," unusual behavior "[3]," unusual activity / event ". Different problems were encountered in the mentioned automatic video surveillance systems. The most important of these problems is that the definition of anomaly in videos is variable and uncertain according to the characteristics of the space and human community [4]. As a result of the ambiguity of the border between normal and abnormal behaviors in the images obtained from camera systems and the problems encountered in obtaining sample training data for abnormal behaviors, modelling possible behaviors in the scene in the image and accepting behaviors that do not conform to this model as abnormal is the most common anomaly detection approach. Abnormal events can be examined locally and globally in the videos obtained [1]. The behaviors exhibited individually are defined as different behaviors according to the crowd, as

local anomaly, and sudden changes displayed in the video for any reason (fire, etc.) as a global anomaly.

Anomaly detection; detection and monitoring is carried out with 3 different methods: behavior analysis and activity analysis of people. These methods process the data collected by cameras using closed circuit television (CCTV) cameras to capture and examine scenes involving human movements, through algorithms. [5]

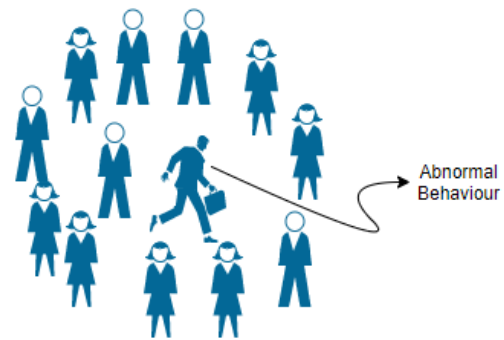


Figure 1. Exhibiting abnormal movement in a crowd

Different approaches have been proposed for the detection of anomalies in crowded and non-crowded scenes. These approaches are examined in three separate sections as individual, integral and hybrid approaches. Since individual approaches [2] [6] are based on object detection and tracking, they are suitable for uncomplicated scenes that do not involve much moving behavior. The holistic approaches consider the components in the scene where the movement takes place as a whole rather than examining them separately. Due to this feature, holistic approaches achieve greater success than individual approaches [7].

In this study, two data sets consisting of video images containing crowded scenes in five different environments, which are accessible on the internet, were studied. In these data sets, abnormal movements of people in crowded scenes were detected in video images. The block diagram of the study is shown in Figure 2.

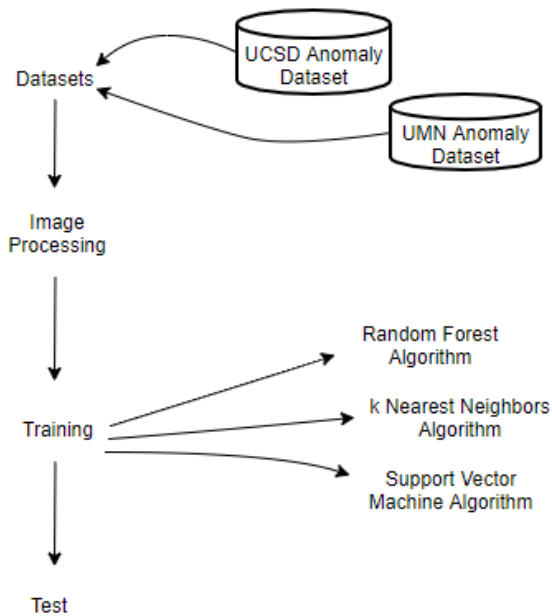


Figure 2 . Block diagram of the study

Firstly, the video images are divided into photo frames. Without a fixed rate, most of these divided squares were used as training and the remaining images as test data. In this study, Nearest Neighbour, Support Vector Machines and Random Forest algorithms were used to detect anomalies on these data sets. When the performance criteria of the algorithms are examined, it is seen that the Support Vector Machines and the k Nearest Neighbors algorithms show higher performance.

RELATED WORK

Chong et al. Studied the PED1 and PED2, Metro Station Entry and Exit image data of the UCSD data set. They detected abnormal movements on the images. Bicycle, skateboard and small vehicle users who enter the pedestrian path are defined as abnormal as can be seen in Figure 4. Cong et al. Learned the dictionary with the multi-scale optical flow histogram (MHOF) attribute obtained from normal behaviours. Anomaly detection was carried out according to the cost of the test sample calculated on dictionaries [8]. While this cost is lower for normal events, it results in higher values for abnormal events. In their study, Mehran et al. Tried to detect abnormal behaviours in crowded scenes using the social power model. The method deal with the ability of individuals to capture crowd

behavior dynamics based on interaction forces without the need to trace individual objects or segmentation. The results of the method they applied show that the method is effective in detecting and localizing abnormal behaviors of the crowd [9]. After Xu et al. Performed the learning processes with dynamic texture attributes obtained from normal behavioural patterns, they determined the anomaly by looking at the low learning coefficient and the reconstruction error [10]. Xu and his friends, after doing normal behavior of the learning process with the dynamic texture features they have obtained from the figures, less learning coefficient and have accomplished the detected anomalies to look again made that configuration errors [10].

MATERIAL AND METHOD

In this study, supervised learning algorithms were used for anomaly detection. Supervised learning trains a model to take a specific set of input data and known responses to the data, and then create plausible predictions that can respond to new data. In this study, data sets called UCSD (University of California San Diego) Anomaly Dataset [11] and UMN (University of Minnesota) Anomaly Dataset [12], which are accessible on the internet, were studied. Machine learning methods, a sub-discipline of artificial intelligence, have been used to detect abnormal movements in these data sets. Of these data sets, two separate scenes were examined in the UCSD data set and 3 separate scenes in the UMN data set. Video images in the scenes analyzed are in ".avi" format, so they were first converted into ".jpeg" format. Since HOG is a feature extraction method that gives successful results in object and pattern recognition, HOG (Histogram of Oriented Gradient) has been applied to the divided pixels. Perform feature extraction with gradient values and orientation angles of pixels in HOG method. The main purpose of this method is to represent the image in the form of local histograms [13]. HOG images of some scenes are given in Figure 3.

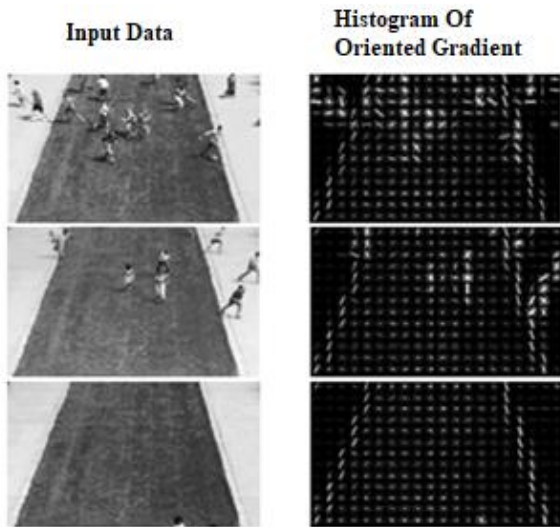


Figure 3. HOG images of scene 1 in UMN anomaly detection data set

The UMN anomaly dataset is contains 22 video (11 videos for training and 11 videos for testing) and 7739 frames with a 320×240 resolution. Videos are recorded in 1 indoor and 2 outdoor scenes. Each video starts with normal behaviour and ends with abnormal behaviour like escape. Example scenes are shown in Figure 4.



Figure 4. Normal / abnormal motion scenes belonging to UMN anomaly data set scene 1



Figure 5. Normal / abnormal motion scenes belonging to UCSD anomaly data set pedst1

7739 pieces of data were processed in the UMN anomaly data set. 7347 of these data were used for training purposes and the remaining 392 were

used to test the trained algorithm. Similarly, 5120 data out of 5780 data were used for educational purposes in the UCSD anomaly data set. 660 pieces of data were used to test the trained data. Scenes with skateboard or bicycle on the pedestrian path in the UCSD anomaly dataset are considered abnormal. An example is given in Figure 5.

In this study, k Nearest Neighbors, Support Vector Machine and Random Forest algorithms are used. Random Forest (RF) is an algorithm that generates many decision trees while classifying, increasing the classification rate and combining them to obtain more accurate predictions. Randomly selected decision trees come together to form the decision forest. The Random Forest algorithm is used because it gives good results in data sets with a large number of variables, labeled classes and an uneven distribution. The biggest advantage of RF is that it can be used in classification and regression problems [14].

Support Vector Machine (SVM) theory assumes that data belonging to two classes have an infinite number of lines that can be separated from each other in a linear / non-linear manner optimally. This algorithm is one of the most used algorithms due to its ability to obtain very fast results in large data sets, to make the separation of data linearly or non-linearly, and to have the ability to choose the best among these distinctions [14].

This algorithm (K-Nearest Neighbors-KNN) is an algorithm that looks at the proximity of the new data to be classified to k of the previous data. During classification, test samples and training samples is compared. In these comparisons, the Euclidean relation is generally used to calculate the neighborhood distance. Estimates are based on the majority vote of neighboring samples. Care should be taken as it tends to over-conform to higher k values [14]. While applying the k-NN algorithm, the optimal k value was accepted as 2.

As a result of the applied algorithms, the desired abnormal behaviour was detected. Program outputs including these determinations are given in the Figure 6.



Figure 6. Outputs from data sets

The images obtained as a result of the algorithms applied to the images taken from a total of five separate scenes, including two data sets, are given in Figure 6. Confusion matrix of the data is obtained. Using the TP, TN, FP, and FN values in the confusion matrix, the precision, accuracy, sensitivity and performance criteria of the algorithms applied to the anomaly data sets were calculated. These calculated values are given in Table 1.

Table 1. Accuracy values of algorithms

Algorithm	Accuracy Value (%)				
	UCSD Anomaly Detection Dataset		UMN Anomaly Detection Data Set		
	PEDS1	PEDS2	Scene1	Scene2	Scene3
K Nearest Neighbour	70.48	99.69	95	92.80	75
Support Vector Machines	86.44	100	91.25	89.38	85
Random Forest	80.12	99.39	56.25	68.49	50

When Table 1 is examined, it is seen that the lowest accuracy values belong to the data in Scene 3 of the UMN anomaly data set. It has been determined that this is related to data quality and number of data. As the data quality decreases and the number of data decreases, the correct prediction ability of the algorithm also decreases. In Table 2, the values of the criteria such as accuracy, sensitivity precision and F1 Score belonging to the Support Vector Machine algorithm applied to the data sets are given.

Table 2. Performance criteria of support vector machines algorithm

Performance Metrics (%)	Data Sets				
	UCSD Anomaly Detection Dataset		UMN Anomaly Detection Dataset		
	PEDS1	PEDS2	Scene1	Scene2	Scene3
Accuracy	86.44	100	91.25	89.38	75
Recall	98	100	100	82.48	66.66
Precision	74.09	100	85	100	100
F1 Score	84.38	100	91.89	90.39	79.99

Table 3. Performance criteria of k nearest neighbors algorithm

Performance Metrics (%)	Datasets				
	UCSD Anomaly Detection Dataset		UMN Anomaly Detection Dataset		
	PEDS1	PEDS2	Scene1	Scene2	Scene3
Accuracy	70.48	99.69	95	92.80	85
Recall	66.66	99.39	90.90	88.81	76.92
Precision	81.92	100.0	100	97.94	100
F1 Score	73.50	99.69	95.23	93.15	86.95

Table 4. Performance criteria values of random forest algorithm

Performance Metrics (%)	Data Sets				
	UCSD Anomaly Detection Dataset		UMN Anomaly Detection Dataset		
	PEDS1	PEDS2	Scene1	Scene2	Scene3
Accuracy	80.12	99.39	56.25	68.49	50
Sensitivity	89.00	98.79	53.33	61.34	50
Precision	68.07	100.0	100	100	100
F1 Score	77.13	99.39	69.56	76.03	66.66

The values in Table 3 show that the K-Nearest Neighbours algorithm achieves the highest accuracy in the PEDS2 data set. The PEDS2 data set was found to have the highest accuracy and precision in all three algorithms. This is due to the fact that the data of the data set has sufficient number of views to train.

RESULTS

Similar studies take place in the literature. However, differently, five different crowded scenes were examined in this study. The data sets obtained from these five different environments are divided into training and test data. Support Vector Machines, k nearest neighbour and random forest algorithms, which are among the supervised learning algorithms, have been applied. As a result, abnormal movements in crowded scenes were detected. In order to compare the algorithms used in the study, values such as accuracy, recall, precision and F1 score were calculated. The values of these performance criteria showed that the quality of the data is effective in the success of the algorithm. In future studies, it is aimed to detect anomalies on live broadcast images rather than recorded images. It is believed that the targeted study will enable measures to be taken against adverse events that may occur in events such as stadiums and demonstrating areas where many people are present at the same time.

REFERENCES

- [1] Cong, Y.; Yuan, J.; Liu, J. (2011). Sparse reconstruction cost for abnormal event detection, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, No. June 2014, 3449–3456
- [2] Wang, X.; Ma, X.; Grimson, W. E. L. (2009). Unsupervised Activity Perception in Crowded and Complicated Scenes Using Hierarchical Bayesian Models, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 3, 539–555.
- [3] Wiliem, A.; Madasu, V.; Boles, W.; Yarlagadda, P. (2008). Detecting uncommon trajectories, *Proceedings - Digital Image Computing: Techniques and Applications, DICTA 2008*, No. January, 398–404
- [4] Sezer, E. S.; Can, A. B. (2018). Anomaly detection in crowded scenes using log-Euclidean covariance matrix, *VISIGRAPP 2018 - Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, Vol. 4, No. Visigrapp, 279–286.
- [5] Popoola, O. P.; Wang, K. (2012). Video-based abnormal human behavior recognition review, *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, Vol. 42, No. 6, 865–878.
- [6] Zhou, B.; Wang, X.; Tang, X. (2012). Understanding collective crowd behaviors: Learning a Mixture model of Dynamic pedestrian-Agents, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2871–2878.
- [7] Hu, M.; Ali, S.; Shah, M. (2008). Learning motion patterns in crowded scenes using motion flow field, *Proceedings - International Conference on Pattern Recognition*, 2–6.
- [8] Chong, Y. S.; Tay, Y. H. (2017). Abnormal event detection in videos using spatiotemporal autoencoder, *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 10262 LNCS, 189–196.
- [9] Mehran, R.; Oyama, A.; Shah, M. (2009). Abnormal crowd behavior detection using social force model, *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*, Vol. 2009 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, No. 2, 935–942.
- [10] Xu, J.; Denman, S.; Sridharan, S.; Fookes, C.; Rana, R. (2011). Dynamic Texture reconstruction from

- sparse codes for unusual event detection in crowded scenes, *MM'11 - Proceedings of the 2011 ACM Multimedia Conference and Co-Located Workshops - JMRE 2011 Workshop*, J-MRE'11, 25–30.
- [11] UCSD Anomaly Detection Dataset, URL: svcl.ucsd.edu/projects/anomaly/dataset.htm (Accessing Time ; March, 18, 2019).
- [12] Monitoring Human Activity, URL: <http://mha.cs.umn.edu/> (Accessing Time; February, 10, 2019)
- [13] Scikit-Image, URL: <https://scikit-image.org/docs/dev/api/skimimage.feature.html#skimimage.feature.hog> (Accessing Time ; May, 5, 2019)
- [14] Aydın, C. “Makine Öğrenmesi Algoritmaları Kullanılarak İtfaiye İstasyonu İhtiyacının Sınıflandırılması”, *European Journal of Science and Technology* No. 14, pp. 169-175, December 2018