JOURNAL OF AERONAUTICS AND SPACE TECHNOLOGIES
(ISSN : 1304-0448)
January 2021 Volume 14 Number 1
www.jast.hho.edu.tr

Research Article

Automated Real Time Detection of Suspicious Appearances using Deep Learning

Melek TURSUN[1], Ömer ÇETİN[2]

[1] *National Defense University, Hezarfen ASTIN Computer Engineering Department, 34000 Yeşilyurt , Istanbul, Turkey*, mtursun@hho.edu.tr, https://orcid.org/0000-0003-4789-4120
[2] *National Defense University, Hezarfen ASTIN Computer Engineering Department, 34000 Yeşilyurt , Istanbul, Turkey*, o.cetin@hho.edu.tr https://orcid.org/0000-0001-5176-6338

Article Info

Received: *December 01, 2020*
Accepted: *December 14, 2020*
Online: *January 25, 2021*

Keywords: Detection of Suspicious Behaviors, MTCNN, Deep Learning, Image Processing

Abstract

Security camera systems especially in public areas such as airports, courthouses or sports facilities etc. are used to find fugitive persons or detect suspicious behaviors manually under the monitoring of an operator. In hallway-like sections in public facilities, repeated appearances of an unknown ordinary person in a short span of time can be defined as suspicious behavior. However, the fact that multiple cameras are monitored by a single operator makes it harder to detect suspicious behaviors especially in crowded fields. Therefore, support decision systems are required to support operator. If individuals are detected on images automatically and their appearances on the camera are recorded on a database by giving them a temporary identity, suspicious behaviors can be reported to an operator as a support decision system. For this reason, two different methods are used together as a hybrid solution in the study; a MTCNN based facial detection is used on the real time security camera images that currently provide face images, and an identification method, created with facial landmarks produced with a deep learning algorithm that was trained with res-net, was used on the obtained person's face images. It has been presented that suspicious behaviors can be detected by interpreting the temporary identity information that was obtained. The success of the application was experimentally tested, and the causes of success and failures in the results were discussed.

To Cite This Article: M. Tursun, Ö. Çetin, "Automated Real Time Detection of Suspicious Appearances using Deep Learning", Journal of Aeronautics and Space Technologies, Vol. 14, No. 1, pp. 71-78, Jan. 2021.

Gerçek Zamanlı Derin Öğrenme Yaklaşımı ile Şüpheli Geçişlerin Otomatik Tespiti

Makale Bilgisi

Geliş: *01 Aralık 2020*
Kabul: *14 Aralık 2020*
Yayın: *25 Ocak 2021*

Anahtar Kelimeler: Şüpheli Davranış Tespiti, MTCNN, Derin Öğrenme, Görüntü İşleme

Öz

Havalimanı, adliye, spor salonu gibi kamusal alanlarda yer alan güvenlik kamera sistemleri ya aranan şahısların bulunması ya da bir operatör tarafından gözlemlenerek şüpheli davranışların tespiti için kullanılmaktadır. Bu tip alanlar içerisinde yer alan koridor gibi bölgelerde sıradan bir kişinin kısa bir süre içinde mükerrer geçişleri şüpheli bir davranış olarak tanımlanabilmektedir. Lakin çok sayıda kameranın tek bir operatör çalışanı tarafından takip edilmesi özellikle kalabalık alanlar içinde bu tip bir şüpheli davranışın tespitini zorlaştırmaktadır. Bu nedenle karar destek sistemlerine ihtiyaç duyulmaktadır. Görüntüler üzerinde kişilerin belirlenmesi, bu kişilerin bir geçici kimlik kazandırılarak bir veri tabanı üzerinde geçişlerinin tutulması gerçekleştirilir ise şüpheli davranışın bir karar destek sistemi tarafından operatöre rapor edilmesi sağlanabilir. Bu nedenle bu çalışma kapsamında halihazırda görüntü sağlayan gerçek zamanlı güvenlik kamera görüntüleri üzerinde MTCNN tabanlı yüz tespiti ve elde edilen şahıs imgeleri üzerinde ResNet ile eğitilmiş bir derin öğrenme algoritması ile gerçekleştirilmiş yüz noktalarından oluşturulmuş kimlik verme yöntemi bir arada kullanılmıştır. Elde edilen geçici kimlik bilgileri yorumlanarak şüpheli davranışların tespit edilebileceği ortaya konulmuştur. Uygulamanın başarısı deneysel olarak sınanmış ve elde edilen sonuçlardaki başarım ve hata nedenleri tartışılmıştır.

1. INTRODUCTION

The number of studies conducted in the field of cyber security has been increasing in recent years. Deep learning is among the commonly used methods in applications such as image classification [1], medical image analysis [2], natural language processing [3], financial time series analysis [4]. Face recognition has been one of the efficient approaches commonly used in applications that aim to verify identity and in fields such as authorization approach. Image processing [5] and deep learning applications [6] are among the approaches that are also commonly used in the field of cyber security [7, 8]. Thanks to the interpretation of real time images with deep learning and image processing applications, measures and security structures are in the works of being developed against various security risks [9,10].

Within the scope of this study, in order to detect real time suspicious behaviors in areas such as airports and school hallways where people heavily pass through, a cyber security approach has been presented that detects and records different human faces and to detect the total number of appearances of a unknown person or the frequency of appearances by the same person in specified areas. What is intended to be detected with suspicious behaviors can be summarized as how many times individuals appear in areas where human behaviors are monitored using live optical cameras. For example, if an individual who is not an airport employee appears in the same airport hallway 3-4 times in a 15-20-minute period, it can be regarded as suspicious behavior. Within the scope of the application, a convolutional neural network (CNN) architecture, which is a deep learning approach, was employed for face recognition on real time images. In order to identify human faces and thus distinguish between different individuals, an image-processing based approach was used to calculate facial landmarks. In order to present a real time hybrid model, the study aims to demonstrate that the convolutional neural network model can be used at the stage of classifying human faces as an object, and identities of individuals can be detected with image processing methods at the same time. It is impossible to develop the application in real time using only convolutional neural networks as it requires being trained with a pre-prepared dataset to solve the problem. Therefore, detection of facial landmarks was presented using image processing methods to determine different identities and distinguish among different human faces.

In Section 2 of the paper, similar studies in the literature were examined, and contributions made to the literature by the approach developed here were presented. In Section 3, the presented method was discussed; in Section 4, the sample application was analyzed results were examined, and in last section conclusions are discussed.

2. LITERATURE REVIEW

Many studies have been carried out with regard to the problem of face recognition in the field of computer vision [11-15], and with the increased success of the deep learning approach in recent years, there has been increased tendency towards deep learning methods, which is one of the traditional approaches for facial recognition [12-15].

Setiowati et al. [11] comparatively presented the advantages and disadvantages of these methods by reviewing the studies in the literature that were conducted in the field of face recognition and analyzing them as approaches that were developed using deep learning techniques and traditional methods (non-deep learning). With the use of a rate of recognition, they demonstrated that, out of these methods, deep learning methods yielded better results with %94,67. It was stated that while the advantages of non-deep learning methods are that they are simple, efficient and produce quicker results due to their low computation complexity, their disadvantages are their longer computation duration in more complex situations, their limited recognition levels and the fact that support vector machines (SVM) method [16] yields the best results in small data sets. On the other hand, it was demonstrated that, among deep learning methods, CNN's advantages are that it can be implemented for problems that require high computation, yields the most accurate results and can make classifications among the faces that were found while its disadvantages are that process duration is long, computations are complex and determination of the parameters is not easy.

In the DeepFace method that they developed, Taigman et al. [12] used a deep learning approached trained with a data set of 4 million containing 4000 unique identities. At the same time, they used a Siamese network architecture where the same CNN is applied to identical face pairs in order to obtain identifiers to be later compared using the Euclidian distance. The aim of the training is to minimize this distance for identical faces and maximize the distance between unidentical faces, which is called metric learning. Authors demonstrated that, after the pretreatment of aligning the face images into a canonic pose using a 3D model in addition to a large volume data set, an architecture developed by using multiple CNNs yielded the best performance with an accuracy of %97,35 when it was trained on an LFW (Labeled Faces in the Wild) data set.

Sun et al. [13] brought some innovations to Taigman's method with their approach, DeepId. They form a very deep network with multiple CNN use, a metric extraction with Bayesian learning, multi-task learning for classification and verification processes and various CNN architectures that branch out to full connected layers. Contrary to the DeepFace method, DeepId performs a simpler 2D affine alignment as a pretreatment. Compared to DeepFace, its performance on LFW reached up to %99,53.

In their study, Google researchers, Shroff et al [14] trained a CNN by using 200 million face identities and 800 million pictures. What makes this study different is that they used triple base loss. The approach can be summarized as comparison of two identical faces (a, b) and a different face (c). The aim is to present that a resembles b more than it resembles c. Unlike other metric learning methods, comparison is always performed in accordance with a relative pivot. The performance of this approach on LFW achieved an accuracy of %99,63.

In another study conducted in this field, Liao and Gu [15] suggested a version of dictionary and subspace learning methods that have been inspired by low rank representation and aperture methods in order to compensate their deficiencies. Low rank presentation is resistant to noise and cloudiness on the face as it can save the clouded area on the face from subspace. New common subspace training and test sets were obtained by putting the training and test set into the subspace learning algorithm with this method. By training the obtained common subspace training set with high performance dictionary learning, they classified the ones in the common subspace test. The method they suggested yielded the best performance on LFW with %95,82 according to the latest technology face recognition techniques such as RRNN (recurrent regression neural network) and MDFR (multimodal deep face recognition).

Although these approaches were successful as they were trained and tested with a specific data set in all methods used above, their use in real time applications is not suitable for reasons explained above. In this study, an approach was developed for real-time recognition of faces that are encountered in a video for the first time, and the success of the approach was presented by being tested on an application.

3. METHODOLOGY

In this section of the study, process steps of the employed method were described in detail, and the processes that were performed were stated through inputs and outputs of each process step.

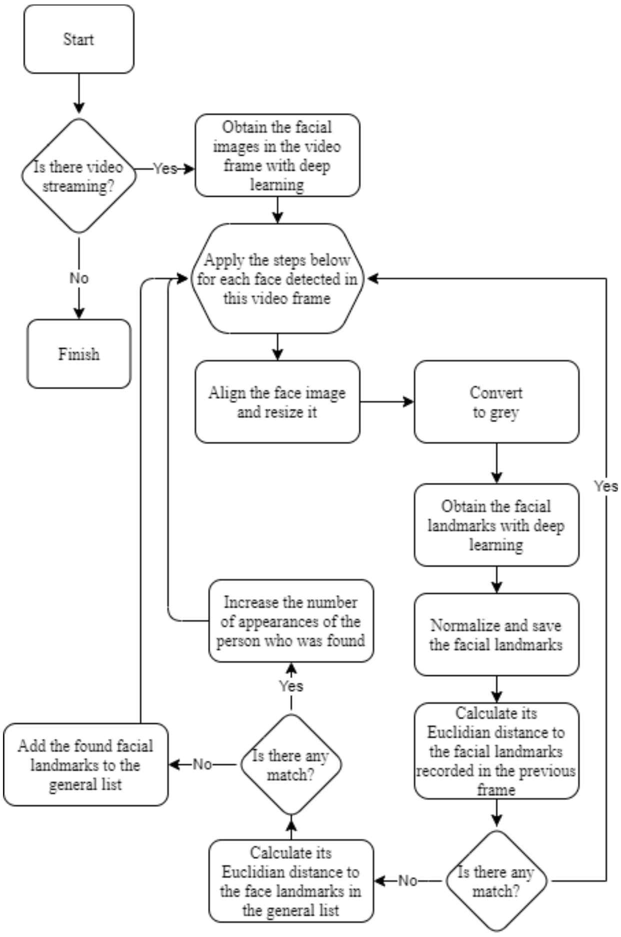

Start
Is there video streaming?
Yes
Obtain the facial images in the video frame with deep learning
No
Finish
Apply the steps below for each face detected in this video frame
Align the face image and resize it
Convert to grey
Yes
Obtain the facial landmarks with deep learning
Increase the number of appearances of the person who was found
Normalize and save the facial landmarks
Yes
Add the found facial landmarks to the general list
No
Is there any match?
Calculate its Euclidian distance to the facial landmarks recorded in the previous frame
Calculate its Euclidian distance to the face landmarks in the general list
No
Is there any match?

Figure 1. Flow chart of the presented method

Figure 1 shows the flow chart of the hybrid approach that is presented for identification using real time optic images that are based on deep learning supported with image processing. During the video stream, human faces were identified in real time using the deep learning approach on each video frame; identification was made with the help of image processing methods; and suspicious behaviors were detected based on the identity information.

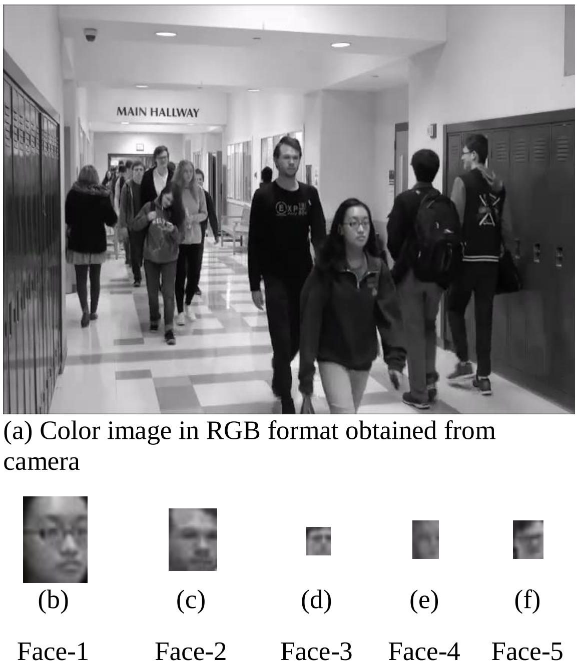
3.1. Detection of Face Images in the Video Frame Using Deep Learning

This study analyzes multi-task cascaded convolutional neural networks (MTCNN) [17] and Haar Cascade [21] methods which are used to obtain the face images that are in each frame of the video and taken as the input of the application. Haar Cascade is a model that is included in the Opencv library and used to find faces on images by using HOG (Histogram of Oriented Gradient) features and learning the brightness differences on the faces. In the MTCNN approach, an image taken as input is resized in different measurements and subjected to a three-step cascade process where a different CNN is used in each step. Refinement process was performed by obtaining all face candidates with the first CNN (P-Net) and ruling out the wrong candidates with the second CNN (R-

Net), and calibration was performed with bounding-box regression at the same time. In the last CNN (O-Net) stage, 5 important points on the face (right eye, left eye, nose, right side of the mouth, left side of the mouth) were extracted, and face coordinators were obtained with more control.

Table 1. Comparison of Face Detection Methods

Methods	Number of pictures in the data sate	Number of Obtained Faces	Number of Inaccurate Faces	Success
Haar Cascade	24111	19915	947	95.24%
MTCNN	24111	21666	428	98.02%



(a) Color image in RGB format obtained from camera

(b) Face-1 (c) Face-2 (d) Face-3 (e) Face-4 (f) Face-5

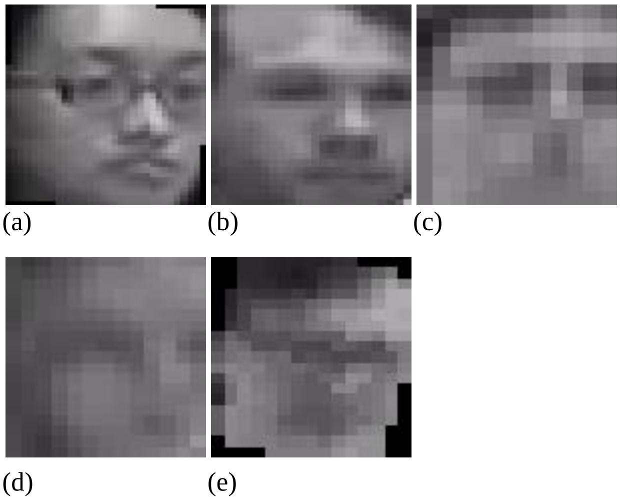
Figure 2. Demonstration of the face images detected in the sample video frame

Success of these two approaches were compared on the UTKFace data set that includes only one face image in each picture [18]. As shown in the results presented in Table 1, a success rate over 98% has been obtained using the MTCNN method. Therefore, the deep learning based MTCNN library was used for its performance to obtain the face images in each frame of the video taken as an input in the study. The MTCNNmodel used in the study was trained with VGGFace2 data set. This data set consists of approximately 3,3 million faces and 9000 classes.

Figure 2 shows face pictures obtained with the MTCNN method from a video frame that was chosen as a sample. Figure 2.a shows the sample video picture frame, and Figure 2.b-c-d-e-f show the detected face images that are in this frame.

3.2. Alignment and Resizing of the Detected Face Images

After face images are obtained from the real time video stream frame, images should be aligned and resized to a fixed size. For the alignment process, eye coordinates obtained in the previous step were used by taking the eyes as a basis. The angle divergence between the eyes were obtained using the arccosine process by finding the horizontal and vertical distance between eye landmarks and calculating the hypotenuse distance. The face image was rotated in accordance with the angle divergence between the eyes.



(a) (b) (c)

(d) (e)

Figure 3. Scaled and rotated face images

Figure 3 shows image outputs where faces obtained in the previous step and shown in Figure 2 are aligned and sized to 100x100 pixels. Horizontal and vertical distances ($X_{distance}$ ve $Y_{distance}$) between the right and left eyes were calculated in the face aligning process, and the angle required for alignment was calculated as follows using the equation 1-3:

$$X_{distance} = Right\ Eye_x - Left\ Eye_x \quad (1)$$

$$Y_{distance} = Right\ Eye_y - Left\ Eye_y \quad (2)$$

$$Angle = \cos^{-1} \frac{X_{distance}}{\sqrt{(X_{distance}{}^2 + Y_{distance}{}^2)}} \quad (3)$$

The face images that were aligned and scaled, shown in Figure 3, constitute the inputs of the next step.

3.3. Determination of Facial Landmarks

At this stage, images that are cropped, resized and aligned after being detected from the video frame with deep learning methods are converted into single channel grey images, and the inputs which are obtained this way are processed. Dlib library was used to obtain the landmarks of grey-colored face pictures [19]. Dlib library offers open source tools to detect the objects in the images, including face detection and object pose estimation.

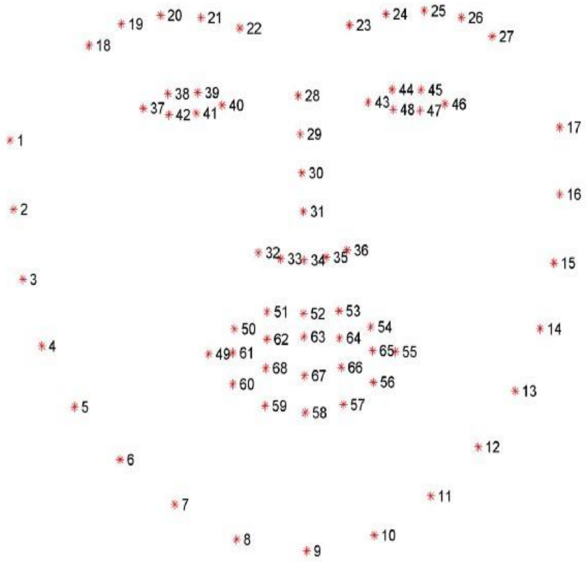


Figure 4. Facial Landmarks

The method uses the pre-trained model, dlib_face_recognition_resnet_model_v1, which can be obtained from the website of dlib free of charge. ResNet, which is the abbreviation of Residual Networks, is a classical neural network that is used as a backbone for various computer vision tasks. This model won the 2015 ImageNet contest. The main breakthrough of ResNet is that it allows us to successfully train very deep neural networks with more than 150 layers. In the standard LFW face recognition comparison, the facial landmarks created by using the dlib library are obtained in the way shown in Figure 4.

3.4. Normalization and Preservation of Facial Landmarks

At this stage, 68 facial landmarks that were obtained were placed in a frame whose resolution was 100x100 pixels, and normalized independently of the landmarks' horizontal and vertical positions on the video frame.

At this stage, it was regarded necessary to stabilize the facial landmarks due to difficulties such as two dimensionality of face images and changes in head movements and facial gestures in the streaming video.

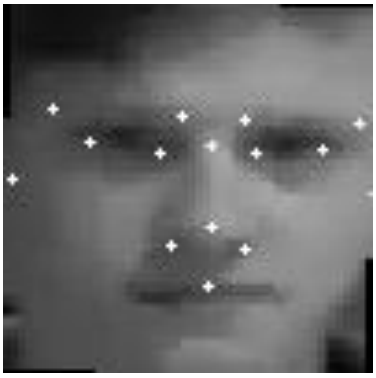
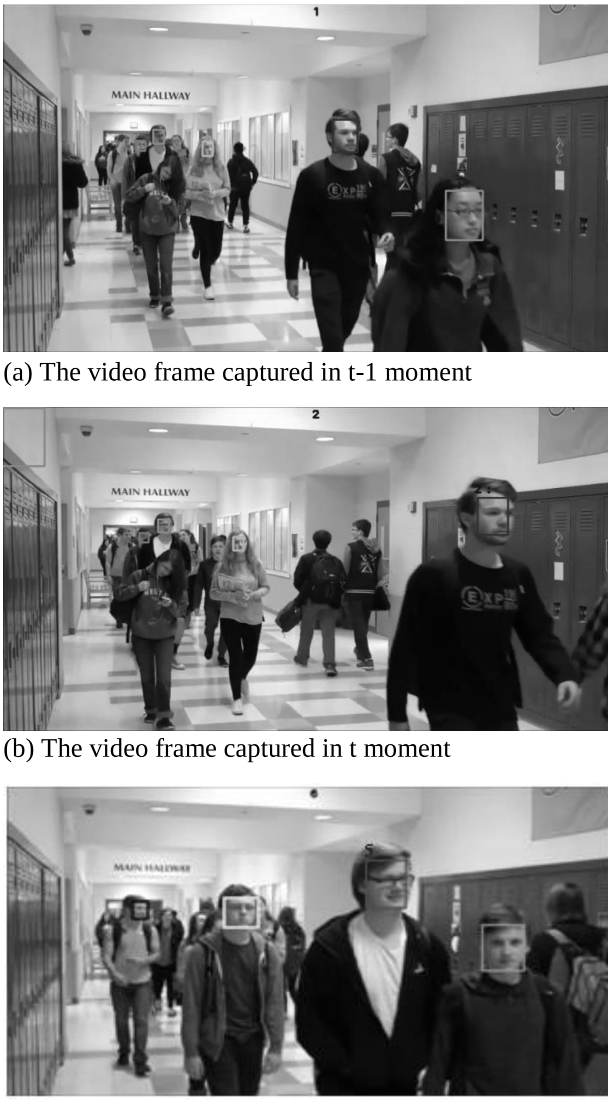


Figure 5. 15 facial landmarks that were detected

In the picture shown in Figure 5, changes of the facial landmarks of the same person were observed, and the most stable 15 landmarks were studied. These landmarks were sides of the eyes, edges of the eyebrows, corners of the nose, temples and upper points of the mouth as shown in Figure 5. These landmarks are recorded on a list by being selected from amongst the 68 landmarks using the number tags specified in Figure 4. These landmarks were chosen as they were the fixed landmarks that were the least affected by facial movements unlike the landmarks around the eyes and sides of the lips which constantly change when eyes are closed and opened or when the person is speaking. These landmarks are recorded on a list and compared with those obtained from the next video frame. A total of three lists are used in the algorithm, namely the list that contains the landmarks in the current video frame, the list that contains the faces in the previous frame and the list that contains all single faces that were separately detected from the beginning of the video.

3.5. Individual Analysis

As shown in Figure 6, the algorithm compares the facial landmarks it detected in the frame t with the facial landmarks in the previous frame t-1 by calculating the Euclidian distance one by one; and if the facial landmarks detected in the new video frame match with one of the faces in the previous frame, it is considered that the person continues his/her behaviors in the camera area and the action is ignored. However, if they do not match with any of the faces in the previous frame, then the same comparison is made with the facial landmarks that are contained in the general list. If there is no match in the general list either, this face is added to the general list, and the number of newly found faces is increased by 1.



(a) The video frame captured in t-1 moment

(b) The video frame captured in t moment

(c) Video frame captured in t+1 moment

Figure 6. Face detection in video frames captured in different moments

If there is a match with one of the landmarks in the general list, the number of appearances of the person on the camera is increased by one. If the number of appearances of a person within a certain time period is detected to be over a certain value, this behavior is considered suspicious. This process is shown on the sample video stream in Figure 6.

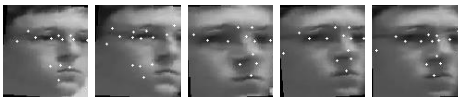
$$\sqrt{(p_1-q_1)^2+(p_2-q_2)^2+\cdots+(p_n-q_n)^2}=\sqrt{\sum_{i=1}^{n}(p_i-q_i)^2} \qquad (4)$$

Euclidian distance is used to compare the facial landmarks in different clusters with each other. Euclidian distance is the linear distance between two points. It is also named L2 norm or L2 distance. Euclidian distance between n sets of p and q point clusters is calculated as shown in equation 4. If the resulting Euclidian distance is lower than the tolerance value determined for similarity, it is regarded the same; if it is higher, it is regarded different.

4. IMPLEMENTATION

Success of the approach presented in the study was tested in an environment that was simulated to create inputs that represent a live video [20] image source that contains 30 image frames at 854x480 pixel resolution per second on a computer running with 2.5 GHz processor and 8 GB RAM. Detection and recognition of the faces that pass in front of the camera in the video start when faces are captured at the size of at least 70x70. The aim is to capture the people walking towards the camera at the clearest moment and avoid detecting faces and making calculations before that moment.

Table 2. Examination of Inaccurate Findings

Person	Repeated inaccurate detections
A	
B	
C	
D	

Each face contained in the application is resized to 100x100 pixels as described above, and facial landmarks are calculated on these images in a way that is described in the previous section. The threshold value is set as 21 when comparing the Euclidian distances calculated in person analysis. This value was obtained based solely on experience. It means that the maximum tolerated pixel movement is 21 for 15 landmarks obtained from a face picture at the size of 100x100, because facial landmarks may slightly change in each picture frame due to head movements and facial gestures. In the end, faces belonging to 18 different people were detected in a video that was known to include 25 different people. However, it was observed that there were 9 repeated person records among the facial findings that were recorded as 18 different faces. Table 2 shows examples where same persons are identified as different people.

When the results in Table 2 are examined, it can be seen that two different identifications were made for each person known as A, B and D, and 6 different identifications were made for the person known as C. The reason is that the changes in the landmarks on detected face images were over the specified threshold value due to their head and body movements when passing in front of the fixed optical camera. When the results in Table 2 are examined, it is observed that these faces were detected as different faces due to inaccurate extraction of the facial landmarks or because of changes in the distance (for example, the distance between eye landmarks and the root of the nose, or the distance between the tip of the nose and the sides of the nose) between facial landmarks resulting from head movements. These errors can be deemed normal due to the lack of depth information on 2D images of 3 dimensional objects. The shift value of the landmarks on the sides of the eyes inside the face that occurred due to head movements and the shift value of the nose landmarks on the outer part of the face differ from each other based on the anatomy of each face.

Table 3. Evaluation of the Success of the Application

Number of Facial Landmarks Used	Number of Known Persons	Number of Persons Detected After the Application	Number of Persons Detected Repeatedly	Success
15	25	18	9	5%
68	25	18	17	50%

In addition, the test results that are obtained by using all 15 facial landmarks are more successful than those that are obtained using 68 facial landmarks when examined in accordance with the amount of errors specified in Table 2. However, when the number of facial landmarks is decreased below 15, it becomes harder to detect the differences between individuals.

Table 3 shows the errors and performance results of the approach presented in this study, which were obtained from the outcomes of the experiment that was carried out with the same input video by using different facial landmarks.

5. CONCLUSION

In order to detect suspicious behaviors on live images within the scope of this study, an identification method, created with facial landmarks produced with a deep learning algorithm trained with res-net, was applied in real time on person images obtained with a MTCNN based face detection approach, and the success of this hybrid approach was experimentally tested on a video where the number of persons was known beforehand. When the results of the study were analyzed, it was observed that the approach could detect appearances of different persons with a success rate of 65%. It was seen that the error margins stemmed slightly from the inability to detect the faces, and largely from the success of identification. It was determined that the main cause of the errors is that the processed images were obtained from a single 2D camera in face detection and identification approaches, and that individuals made head and body movements with an angle that prevented the production of reliable face values in front of a single camera.

In order to decrease the error margin, it is necessary to obtain multiple camera images in the same area, which can produce 3 dimensional images and allow frontal face images to be obtained despite various behaviors of individuals, and subject them to a similar process. Moreover, tagging people who hide their faces or pass in front of the camera with a behavior that prevents their faces from being seen from the front as suspicious may be presented as a method that increases success.

It has been presented in this study that places such as airports, customs gates or entrances of public buildings can be monitored; people who display unusual behaviors such as multiple appearances in certain areas in a short span of time can be detected; and this process can even be performed on real video images with computer systems that run with low power.

It has been concluded that, in future studies, it would be suitable to use different methods that are thought to produce higher success in determining facial landmarks and try to perform a 3D analysis of the images using multiple cameras. It is deemed that it would increase the success of the application. It has also been concluded that it will be possible to obtain more successful results even with 2D images by developing a deep learning algorithm that can reproduce the facial landmarks in a linear and aligned way even when the head is moving right, left, upwards and downwards.

6. REFERENCES

[1] D. Ciregan, U. Meier and J. Schmidhuber, "Multi-column deep neural networks for image classification," in Proc. of 2012 IEEE Conf. on Computer Vision and Pattern Recognition, June 16-21, 2012, pp. 3642-3649.

[2] A. Memiş, S. Albayrak and F. Bilgili, "A brief overview of medical software tools used in MR image segmentation," in Proc. of 2018 Medical Technologies National Congress, TIPTEKNO 2018, Nov. 8-10, 2018, pp. 1-4.

[3] C. Hark et al., "Doğal dil işleme yaklaşimlari ile yapisal olmayan dökümanlarin benzerliği," in Proc. of 2017 International Artificial Intelligence and Data Processing Symposium, IDAP 2017, Sept. 2017, pp. 1-6.

[4] W. Bao, J. Yue and Y. Rao, “A deep learning framework for financial time series using stacked autoencoders and long-short term memory,” *journals.plos.org*, Jul. 14, 2017. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0180944. [Accessed: Aug. 10, 2020]

[5] J.C. Russ, "Image Processing," in Computer-Assisted Microscopy, 1nd ed. Reading, Eds. Boston, MA:Springer, 1990, pp. 33-69.

[6] Y. LeCun, Y. Bengio and G. Hinton, "Deep Learning," *Nature*, vol. 521, pp. 436-444, May 2015.

[7] A. Makandar and A. Patrot, "Malware class recognition using image processing techniques," in Proc. of 2017 International Conference on Data Management, Analytics and Innovation, ICDMAI 2017, February 24-26, 2017, pp. 76-80.

[8] J. Kour, M. Hanmandlu, A.Q. Ansari, "Biometrics in Cyber Security," *Defence Science Journal*, vol. 66, pp. 600-604, October 2016.

[9] E. Saykol, M. Bastan, U. Gudukbay, O. Ulusoy, "Keyframe Labeling Technique for Surveillance Event Classification" *Optical Engineering*, vol. 49, pp. 1-12, November 2010.

[10] A. Sukumar, V. Subramaniyaswamy, L. Ravi, V. Vijayakumar and V. Indragandhi, "Robust image steganography approach based on RIWT-Laplacian pyramid and histogram shifting using deep learning," *link.springer.com*, Jul. 01, 2020[Online]. Available: https://link.springer.com/article/10.1007/s00530-020-00665-6. [Accessed: Aug. 10, 2020]

[11] S. Setiowati, Zulfanahri, E. L. Franita and I. Ardiyanto, "A review of optimization method in face recognition: Comparison deep learning and non-deep learning methods," 9th International Conference on Information Technology and Electrical Engineering, ICITEE 2017, Phuket, Thailand, October 12-13, 2017, pp. 1-6.

[12] Y. Taigman, M. Yang, M. Ranzato and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," Conference on Computer Vision and Pattern Recognition, IEEE 2014, Columbus, OH, USA, June 23-28, 2014, pp. 1701-1708.

[13] Y. Sun, L. Ding, X. Wang, and X. Tang. "Deepid3: Face recognition with very deep neural networks," *arxiv.org,* Feb. 3, 2015. [Online]. Available: https://arxiv.org/abs/1502.00873. [Accessed: Jul. 2, 2019].

[14] F. Schroff, D. Kalenichenko and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015, pp. 815-823.

[15] M. Liao and X. Gu, "Face recognition based on dictionary learning and subspace learning," *Digital Signal Processing*, vol. 90, pp. 110-124, July 2019.

[16] V. N. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed. Reading, Eds. New York, MA: Springer, New York, 2000. [Online] Available: SpringerLink.

[17] K. Zhang, Z. Zhang, Z. Li and Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," *IEEE Signal Processing Letters*, vol. 23, pp. 1499-1503, October 2016.

[18] "Face Detection: Haar Cascade vs. MTCNN," datawow.io, May. 20, 2020. [Online]. Available: http://datawow.io/blogs/face-detection-haar-cascade-vs-mtcnn. [Accessed: Aug. 10, 2020].

[19] D. E. King, "Dlib-ml: A Machine Learning Toolkit," *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, December 2009.

[20] In the Mind Films. "High School Hallways", *YouTube*, Mar. 1, 2017 [Video file]. Available: http://www.youtube.com/watch?v=1ojGb6QU7qY&t=33s&ab_channel=IntheMindFilms. [Accessed: Jul. 21, 2019].

[21] P. Viola and M. Jones, "*Rapid object detection using a boosted cascade of simple features,*" Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, HI, USA, December 8-14, 2001, pp. 511-518.

VITAE

Melek TURSUN received her B.Sc. degree in Computer Engineering from Faculty of Engineering, Fırat University, Turkey in 2013. She received her M.Sc. degree in Business Informatics from Paris-Saclay University, France in 2017.

Ömer ÇETİN is currently an assistant professor at the Computer Engineering Department of National Defense University (NDU), Turkey. He received his B.Sc. degree in Computer Engineering from Turkish Air Force Academy, Istanbul, in 2003. He received his M.Sc. degree in Software Engineering from Aeronautics and Space Technologies Institute (ASTIN), Istanbul, Turkey, in 2008. Asst. Prof. Dr. Ömer ÇETİN received his Ph.D. degree in Computer Engineering Program in Department of Computer Engineering of ASTIN in 2015. He is currently researching related with cyber security, deep learning, and autonomous systems.