Balkan Journal of Electrical and Computer Engineering (BAJECE)
Volume 1 Number 1, September 2023

Vision Transformer Based Photo Capturing System

Abdulkadir Albayrak

1 Department of Computer Engineering, Dicle University, Diyarbakir, Turkey,(e-mail: abdulkadir.albayrak@dicle.edu.tr).

Manuscript received Feb 1, 2023; accepted August 18, 2023.

Abstract— **Portrait photo is one of the most crucial documents that many people need for official transactions in many public and private organizations. Despite the developing technologies and high resolution imaging devices, people need such photographer offices to fulfil their needs to take photos. In this study, a Photo Capturing System has been developed to provide infrastructure for web and mobile applications. After the system detects the person's face, facial orientation and facial expression, it automatically takes a photo and sends it to a graphical user interface developed for this purpose. Then, with the help of the user interface of the photo taken by the system, it is automatically printed out. The proposed study is a unique study that uses imaging technologies, deep learning and vision transformer algorithms, which are very popular image processing techniques in several years. Within the scope of the study, face detection and facial expression recognition are performed with a success rate of close to 100% and 95.52%, respectively. In the study, the performances of Vision Transformer algorithm is compared with the state of art algorithms in facial expression recognition.**

Index Terms— ***Deep learning, facial expression recognition, photo capturing system, single shot detection, vision transformer***

1. INTRODUCTION

At the present time, with the rapid development of mobile technologies, most of the persons have digital photo camera and can capture high-quality photos. However there is still a lack of applications that are useful for capturing portraits for passports and other legal documents. In this study, a sophisticated real time portrait capturing system that combines hand crafted image processing techniques with state of art deep learning approaches is proposed.

The proposed system involves automatic detection of frontal face, determining the face orientation through detected landmark points and facial expression analysis.

Face detection and determination of face orientation are two basic steps that should be performed for various computer vision applications. These tasks are also critical for the proposed portrait capturing system since they constitute a precondition for the subsequent facial expression analysis step. In order to develop a robust face detection system, Single-Shot-Multibox detector with ResNet-10 is used as a backbone architecture [1]. In the literature, face detection is generally focused on finding 68 points using the distinctive textural features of the face[2-3] Facial landmarks are found in order to localize eyes, nose, contour of the face and mouth. Landmark points are exploited for determining whether the eyes and mouth are open or closed. Histogram of Oriented G radients (HOG) and Support Vector Machine (SVM) are employed for selecting the images with opened eyes, while the proportion of the width and height of the mouth is calculated for determining weather the mouth is opened or closed. Frontal face images with open eyes and closed mouth are than processed for facial expression analysis.

Facial expression is one of the most effective channels of human communication, and therefore, automatic facial expression analysis systems can take place in various applications related to human-computer interaction. The success of the deep learning methods in modeling complex systems lead researchers to apply various deep learning approaches in this challenging task. Convolutional Neural Networks (CNN) based systems have proven their success in facial expression recognition problem [4]. However, facial expressions have high variability due to the nature of the human. This reality has demanded improvements in CNN-based systems, which require large amounts of training data to build a reliable model. Another drawback of the CNN based models is their relatively fragile structure against variant backgrounds and head poses [5]. Lozoya et.al aimed to improve generalization of their CNN based system by learning from mixed instances taken from different databases [6]. In [7], researchers employed CNN with Rectified Adam Optimizer in order to improve generalization.

Success of the Transformers in natural language processing pave the way for attempts to adapt transformers to computer vision problems. One of the key ideas of Transformer models is being pre-trained on a large corpus and fine-tuning on the target task with a smaller dataset [8]. Naseer et.al. compared CNN and Vision Transformer networks and stated that Vision Transformers are robust to occlusion and pose variant [9]. For all that reasons, in this study, a Vision Transformer based system is used for facial expression recognition. Proposed automatic portrait capturing system selects the neutral faces for further processing.

Selected proper portrait photographies (Frontal and neutral face images with opened eyes and closed mouth) are than post-processed for cleaning small speckles on the face and then the photographies are sent to the printer according to the preferences of the users.

The main contributions of the proposed system are given below:

- At the first time in the literature, a real time portrait capturing system that combines handcrafted image processing techniques with deep learning approaches is proposed.
- Performance of Vision Transformers in facial expression recognition task is evaluated.
- A comparison between the performance of Vision Transformer and state of art methods is done in facial expression recognition.

The rest of the paper is structured as follows: The proposed system is presented in Section 2. The experimental results are discussed in Section 3. The performance of the algorithms presented in this study is discussed in Section 4. Finally, the paper is concluded with Section 5.

2. MATERIALS AND METHODS

2.1 Performing Face Detection and Finding Face Orientation

First step of the this proposed system is to focus on face region in an image. Because one of the most basic conditions that must be met in passport photos or official documents is that the face should be detected and centered symmetrically. Most of the methods suggested in the literature try to find a total of 68 points belonging to the face including eyes, chin, nose and eyebrows. In this study, Single Shot Detection (SSD) method has been applied to detect face regions. The orientation of the face was tried to be calculated by using the positions of these points according to a line that passes through the points of nose vertically. Figure 1 shows a representation of the line dividing the face exactly in half from the vertical. The distance of this line to points 0 and 16 separately is calculated. The ratio of these distances to each other should be approximately 1. The acceptable range of face orientation is set to 0.9 and 1.1, but this range can be narrowed if this value is desired to be more precise. Equation 1 expresses the distance from point 0 to point 27 shown in Figure 1 at the nose level.

Equation 2 expresses the distance from point 16 to point 27 shown in Figure 1 at the nose level.

$$dist\_0\_27 = \sqrt{(x_0 - x_{27})^2 + (y_0 - y_{27})^2} \qquad (1)$$

Here x_0 and $y_0$represent the x and y coordinates of the point 0. x_{27} and y_{27} represent the x and y coordinate information of point 27. *dist_0_27* shows the distance between these two points.

Equation 2 expresses the distance between point 16 in Figure 1 and point 27 at nose level.

$$dist\_16\_27 = \sqrt{(x_{16} - x_{27})^2 + (y_{16} - y_{27})^2} \qquad (2)$$

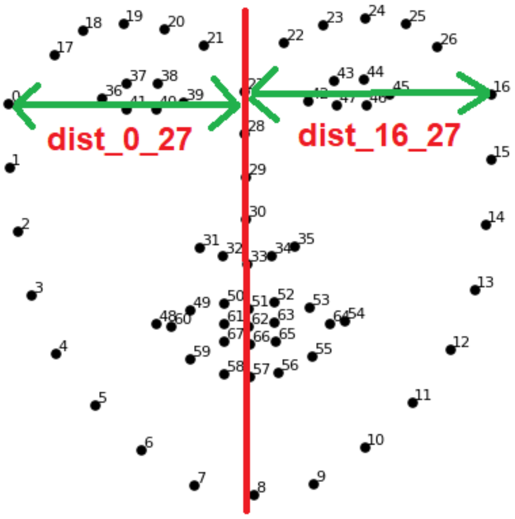


Figure 1. Sample image with 68 points used in face detection

Here, x_{16} and y_{16} represent the x and y coordinates of the point 16 on the front. x_{27} and y_{27} represent the x and y coordinates of the point 27 on the front. *dist_16_27* shows the distance between these two points.

$$dist = \frac{dist\_0\_27}{dist\_16\_27}, \qquad 0.9 \leq dist \leq 1.1 \qquad (3)$$

The ideal value of the dist value obtained in Equation 3 should be 1. In this study, the value range was chosen between 0.9 and 1.1. Thus, the system is able to take pictures in small value ranges without being bound by a very strict rule. Figure 2 shows the whole flowchart of the proposed system from turning on the camera to getting output from the system.

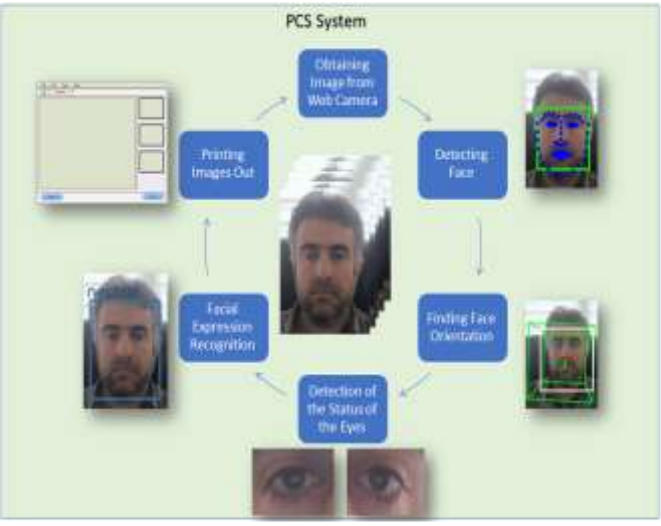

PCS System
Obtaining Image from Web Camera
Detecting Face
Finding Face Orientation
Detection of the Status of the Eyes
Facial Expression Recognition
Printing Images Out

Figure 2. Process steps followed in the Photo Capturing System(PCS) developed within the scope of the study

2.1.1 Single Shot Detection (SSD) Deep Learning Algorithm

SSD is a feed-forward convolutional network-based deep learning method used to detect objects in images. The SSD approach makes a score estimation of the proportions of the boxes surrounding the objects that are intended to be detected. This approach does not make an estimation about the whole image like the methods in the early studies of deep learning, but it is used to determine in which part of the image the object to be classified is located. SSD architecture consists of 3 parts (parts):

Base Convolutions: Networks such as VGG, ResNet, which are suggested for image classification, are the name given to the part as the base.

Auxiliary convolutions: It is the part that is placed as the backbone to obtain higher level features.

Prediction Convolutions: The attributes of the object to be detected are classified in this section. It is the part that makes predictions about the location and score of the object in the image.

The deep learning model applied for face detection within the scope of this study is the SSD model, which is trained with 300x300 image sizes and 140000 iterations. This SSD model in OpenCV's DNN library uses ResNET-10 architecture as backbone.

2.2 Detection of the Status of Eyes

One of the conditions that must be met in passport photos (or passport photos) is to have eyes open. No matter how accurately the eye lines are determined with facial landmark detection, there is no control such as whether the eyes are open or not. In order to carry out this process, the Histogram of Oriented Gradients (HOG) algorithm, which is one of the traditional image processing methods, was used. A total of 200 eye picture systems, 100 closed and 100 open, were trained with HOG.

2.2.1 Histogram of Oriented Gradients (HOG)

Histogram of Oriented Gradients was first proposed by Dalal and Triggs for pedestrian detection [10]. HOG is used to obtain shape-based features of the regions whose attributes are desired to be extracted. It has been applied in many different areas of computer vision, particularly pedestrian detection [11]. In the HOG algorithm, the orientation of all the pixels in the image is calculated and the focus is on determining the silhouette of the object (or region) that is desired to be distinguished. The basic processing steps applied in the HOG algorithm are as follows:

First, edge information is obtained by applying a sobel filter on the horizontal and vertical axis of the image. For this, the following formula referred in Equation 4 is used:

$$I_x = I * S_x, \qquad I_y = I * S_y \tag{4}$$

Here, I denotes the input image, S_x the vertically applied sobel filter and Sy the horizontally applied sobel filter. I_x and I_y show the output images obtained after applying sobel vertically and horizontally, respectively. The I_x and $I_y$images are then used in the formula below to calculate the magnitude values.

$$|G| = \sqrt{{I_x}^2 + {I_y}^2} \tag{5}$$

Here $|G|$ value is expressed as a gradient and is calculated using the square root value of the sum of the squares of the I_x and $I_y$values specified in the previous formula. Finally, the magnitude value is |G| are obtained by calculating the arctan of the values. As a result of these operations, the orientation of the object or region is calculated and the process of distinguishing it from other objects or regions is performed.

2.3 Facial Expression Recognition with Vision Transformer

Neutral facial expression is one of the conditions that must be handled in portrait or passport photos. The vision transformer algorithm, which is inspired by the transformer algorithm, which has been very popular in the field of natural language processing in recent years, has been applied in automatic facial expression detection.

2.3.1 Vision transformer (ViT)

Transformer is a deep learning model that adopts self-attention mechanism. It can be expressed as the calculation of the relationship of each word of the data given as input in the attention mechanism with all the other words. It is primarily used in the fields of natural language processing (NLP) tasks such as machine translation and text summarization which has sequential input data. However, Transformers do not necessarily process data sequentially like Long Short Term Memory(LSTM) and Recurrent Neural Network(RNN). Instead, the attention mechanism provides context for any position in the input data. Inspired by the Transformer scaling achievements in NLP, we attempted to apply a standard Transformer directly to images with the least possible modification. To do this, we split an image into patches and provide the linear embedding order of those patches as an input to a Transformer. Image patches are treated in the same way as tokens (words) in an NLP application. We train the model on image classification in a supervised manner. Figure 3 shows the processing steps of the vision transformer method used in the proposed system for facial expression recognition.

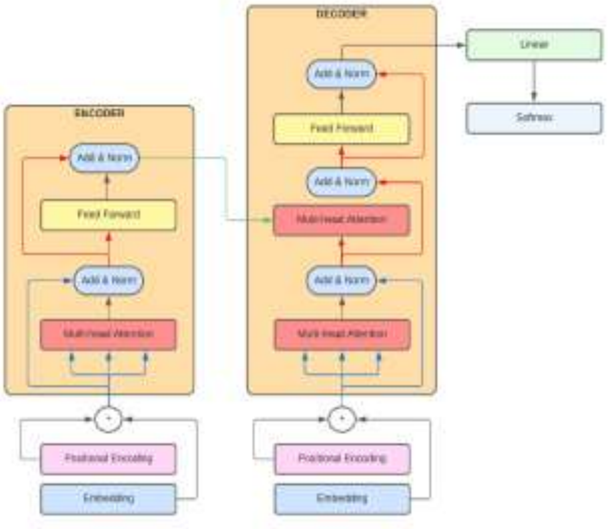


Figure 3. Vision Transformer

The classification header is implemented by an MLP with a hidden layer and a single linear layer of fine tuning at pre-training time. The MLP consists of two fully connected layers with an occasional GeLU nonlinear activation. The most effective part to highlight in the Transformer model is the attention mechanism. The attention mechanism looks at the input sequence and selects parts of the sequence that remain important at each step, preserving knowledge of which parts of the sequence are important. The attention mechanism instantly takes into account several other input data and assigns different weights to these inputs, providing prioritization. All encoder layers use an attention mechanism for each input that measures the fitness of all other inputs and retrieves the appropriate information to produce the output. Then, as a result of the attention mechanism, it takes the weights sent as output and the encoded string as input. These networks consist of repeated multi-headed attention blocks and feedforward layers. Multi-headed attention runs the processes in the attention mechanism in parallel and combines the results. Thus, it is ensured that different relationships are learned.

2.4 Graphical User Interface (GUI)

After the image processing steps were completed, a design that could be output to the printer was realized with the help of the user interface developed for the users. Face detection, face orientation, eye and mouth opening, facial expression recognition operations are performed in "CaptureFace" option which located in the top menu in GUI. When all the determined rules are fulfilled, 10 pictures are saved to a folder named for a specific user/customer. One of the captured image is given as input to the system using the GUI. Then, possible noises are cleaned by applying *3x3* median filter. Finally, after the user is asked whether he wants a digital copy, the interface page where the information is entered comes. On this page, the images is sent to the printer after the user fills all the contact information. Photos are printed from the printer by clicking the Confirm button. Figure 4 shows the graphical user interface (GUI) developed to proceed further processes and to print out the final images after billing.

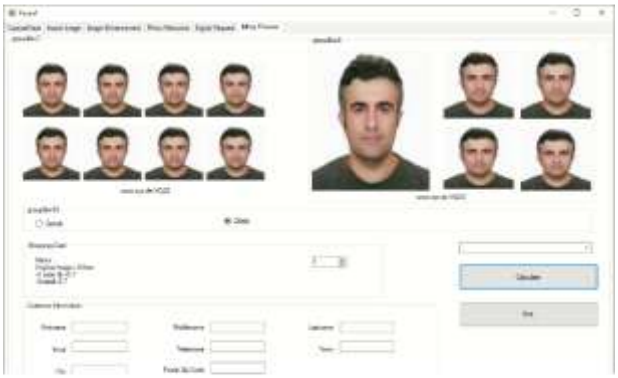
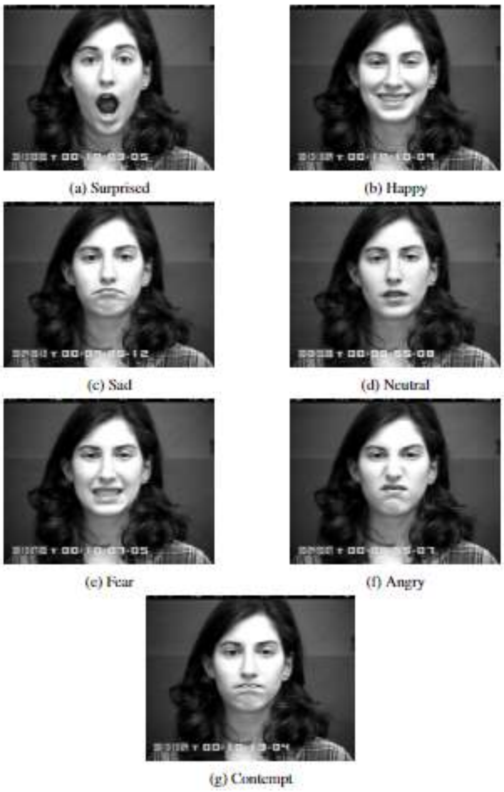


Figure 4. Graphical user interface (GUI) developed to proceed further processes and to print out the final image.

3. EXPERIMENTAL RESULTS

In this part of the study, face detection, recognition of facial expression, openness of the eyes and openness of the mouth were analyzed. While the publicly available facial expression recognition data set available in the literature was used within the scope of the study, the system performance was tried to be increased with a data set created within the scope of the study for eye opening/closed status.

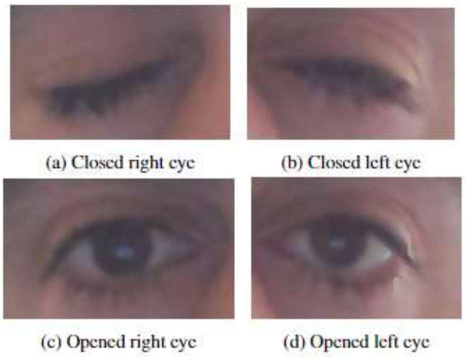
<u>CK+ data set for facial expression recognition</u>: The CK+(Cohn and Kanade) is an publicly available data set for facial expression recognition [12]. There are 7 classes in total in the data set: neutral, happy, disgusted, surprised, sad, angry and afraid. The data set includes 593 sequences from 123 individuals. These sequences begin with a neutral facial expression and end with the expression belonging to each class. Figure 5 shows sample facial expressions from the data set.



(a) Surprised
(b) Happy
(c) Sad
(d) Neutral
(e) Fear
(f) Angry
(g) Contempt

Figure 5. Sample images obtained from the CK+ dataset used for facial expression recognition within the scope of the study. The system is required to take pictures with neutral facial expressions.

<u>Eye Status (ES) Data set</u>: The points around the eyes obtained in the detection of facial regions are not sufficient to understand whether the eye is open or closed alone. In order to solve this problem, another data set was created within the scope of the study to evaluate the open/closed status of the eyes in the frames obtained from the camera (See Figure 6). A data set consisting of a total of 280 images, including 140 open-eye image sections and 140 closed-eye image sections, was created.



(a) Closed right eye
(b) Closed left eye
(c) Opened right eye
(d) Opened left eye

Figure 6. (a), (b) represent the close eyes and (c), (d) represents the opened eye image crops obtained from the data sets.

Evaluation: In order for the proposed system to be used successfully in real time, face detection must be performed with high success in the first stage. For this, the SSD network, which is frequently used in object detection in the literature, is used. At this stage of the study, the facial regions in all of the randomly shot sample videos were detected with SSD. Since the ambient lighting must be very good in passport photos, it is seen that the facial regions are easily detected. In all the trials conducted within the scope of the study, it was observed that face detection was detected in all images. However, face detection may not be performed when the face area is rotated up to a certain angle. This situation is not directly related to the ambient effect.

After the face detection was performed, it was tried to decide whether the image to be obtained was appropriate by taking into account the orientation of the head. An image taken with a certain angle or tilt is not valid. Therefore, in the proposed system, it is necessary to guide the user in images that do not comply with Equations 1,2 and 3. Provided that the face area is at the level of the camera, the user needs to adjust the angle and tilt of the person. Otherwise, the system will not automatically take a picture. If the user abides by the determined rules, the orientation phase will be successfully passed for the system to take a picture along with other conditions.

After the orientation phase is passed without any problems, the stage of determining the appropriate image according to the opening/closing status of the eyes and mouth is started. 140 images belonging to each class in the dataset were separated as training and test sets with 5-fold cross-validation method. A total of 28 test images belonging to each class were classified with HOG method and SVM (rbf kernel) with a success of 93%. While the open eye success rate was 96%, the detection success of the closed eye was 85%. When the picture was taken, the condition of having an eye opening ratio above 95% was accepted as successful. Since the success rate was high enough at this stage, it was not necessary to try alternative methods in order to determine the openness / closure of the eye. It has been seen that the HOG algorithm performs well in detecting the eye opening/closure state. A model was not trained for the aperture and closure of the mouth, but the points obtained by facial landmark detection were used. Since these points for the eyes did not show a significant difference, the data set was created, but the landmark points were sufficient within the scope of the study, since the points changed more significantly according to the shape of the mouth.

Finally, the facial expression recognition phase was carried out. Since the image accepted by the official authorities must be neutral, if the facial expression is not neutral, the picture is not taken. The method used for facial expression detection is the ViT method. The achievements after applying the ViT model to the CK+ dataset are shown in TableXX. As seen in the table, the expression recognition success of the ViT network is 95.52%. In the literature, the successes obtained in the CK+ dataset using traditional image processing techniques are relatively low. It is seen that deep learning networks, which have been performing successfully in many fields recently, have carried this success over 90%. It has been observed that the ViT model applied within the scope of the study gives as successful results as other deep learning methods. When all these conditions are fulfilled, the system automatically takes 10 images and these images are sent to the printer with the help of the user interface prepared within the scope of the study.

Table 1. The classification success obtained with the ViT model applied to the CK dataset and its comparison with the results obtained in the literature

Method	**Accuracy(%)**
3D SIFT [13]	81.35
LBP-TOP [14]	88.99
ITBN [15]	86.3
CERT [16]	87.21
MCF [17]	89.4
MSR [18]	91.4
TMS [19]	91.89
STM [20]	91.13
AUDN [21]	93.70
BDBN [22]	**96.7**
CNN [23]	92.73
Proposed ViT	95.52

4. DISCUSSION

For the necessary processes in public or private institutions and organizations, taking a passport photo has an important place in real life. People often go to places specialized for this process and passport photos are taken with high resolution cameras. Today, many technological devices have cameras capable of taking high-resolution pictures. By processing the images obtained with these devices, it is possible to obtain passport images with the same quality and similar standards. In this study, an autonomous system that takes passport photos with people's own devices or systems that can be installed in public environments is designed. The system focuses on major aspects such as face detection, detection of facial regions, eye and mouth condition, and facial expression. Since deep learning methods perform quite successfully in face detection processes, they can be used in such systems. The process of finding the orientation of the face is similarly possible.It may not be possible to determine whether the eye is open or not by using the points located around the eye. By using the coordinates of the points around the eyes, the image sections with the eyes were cropped and the status of the eyes was determined more easily with the help of the model trained with shape based HOG algorithm. A similar situation could be considered for the mouth, but since it is not as solid as the eye, the opening and closure of the mouth could be determined. Facial expression is very important in the process of taking passport photos. The CK+ dataset, which is one of the frequently used facial expression recognition datasets in the literature, was used for model training in this study. Since the expression recognition performance with traditional image processing methods is somewhat limited compared to deep learning methods, deep learning methods can be preferred. In this study, as an alternative to deep learning methods, the ViT model, which is inspired by the transformer method, which has been very successful in natural language processing, has been used. The use of the convolution layer in training the data in deep learning methods considerably extends the training time.

The absence of a training phase in ViT models enables data to be modeled in a relatively faster training period. The biggest handicap of ViT models is that they need larger resources to achieve greater success. Since the GPU used in this study has 8 GB of ram, the success is limited to 95.22%. It is thought that the success can be increased by increasing the RAM resource. Because after the image is divided into sections, the calculation of the relationship of each section with each other and with all other sections directly contributes to the success of classification.

5. CONCLUSION

In this study, a system that takes passport photos is proposed for use in official documents or in public and private institutions. The system is designed to combine traditional image processing methods and deep learning methods and have the capacity to perform real-time processing. In the study, the transformer method, which has recently shown a very high performance in natural language processing, has been applied for facial expression recognition. When the results obtained were evaluated, the success rates of facial detection, eye opening rate and facial expression recognition were obtained as 100%, 96% and 95.22%, respectively. The system works in real time and can take dozens of pictures depending on certain rules. Users can then select one of the saved images and take a printout with the help of a developed user interface. The system has a very important place in that it combines traditional image processing methods, deep learning methods and ViT models and works in real time. In future studies, it can be developed as a system where people can perform similar operations using their own mobile devices and the resulting image can be sent to users via mail. For this, designing it with a logic that will work on users' mobile devices will improve this system a little more.

ACKNOWLEDGMENT

The authors state no conflicts of interest and have nothing to disclose.

REFERENCES

[1] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “Ssd: Single shot multibox detector,” in European conference on computer vision. Springer, 2016, pp. 21–37.

[2] S. Ren, X. Cao, Y. Wei, and J. Sun, “Face alignment at 3000 fps via regressing local binary features,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1685–1692.

[3] O. Déniz, G. Bueno, J. Salido, and F. De la Torre, “Face recognition using histograms of oriented gradients,” Pattern recognition letters, vol. 32, no. 12, pp. 1598–1603, 2011.

[4] I. M. Revina and W. S. Emmanuel, “A survey on human face expression recognition techniques,” Journal of King Saud University-Computer and Information Sciences, vol. 33, no. 6, pp. 619–628, 2021.

[5] H. Li, M. Sui, F. Zhao, Z. Zha, and F. Wu, “Mvt: Mask vision transformer for facial expression recognition in the wild,” arXiv preprint arXiv:2106.04520, 2021.

[6] S. M. González-Lozoya, J. de la Calleja, L. Pellegrin, H. J. Escalante, M. Medina, A. Benitez-Ruiz et al., “Recognition of facial expressions based on cnn features,” Multimedia Tools and Applications, vol. 79, no. 19, pp. 13 987–14 007, 2020.

[7] D. O. Melinte and L. Vladareanu, “Facial expressions recognition for human–robot interaction using deep convolutional neural networks with rectified adam optimizer,” Sensors, vol. 20, no. 8, p. 2393, 2020.

[8] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, “Transformers in vision: A survey,” ACM Computing Surveys (CSUR), 2021.

[9] M. M. Naseer, K. Ranasinghe, S. H. Khan, M. Hayat, F. Shahbaz Khan, and M.-H. Yang, “Intriguing properties of vision transformers,” Advances in Neural Information Processing Systems, vol. 34, 2021.

[10] P. Scovanner, S. Ali, and M. Shah, “A 3-dimensional sift descriptor and its application to action recognition,” in Proceedings of the 15th ACM international conference on Multimedia, 2007, pp. 357–360.

[11] G. Zhao and M. Pietikainen, “Dynamic texture recognition using local binary patterns with an application to facial expressions,” IEEE transactions on pattern analysis and machine intelligence, vol. 29, no. 6, pp. 915–928, 2007.

[12] Z. Wang, S. Wang, and Q. Ji, “Capturing complex spatio-temporal relations among facial muscles for facial expression recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2013, pp. 3422–3429.

[13] G. Littlewort, J. Whitehill, T. Wu, I. Fasel, M. Frank, J. Movellan, and M. Bartlett, “The computer expression recognition toolbox (cert),” in 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG). IEEE, 2011, pp. 298–305.

[14] S. W. Chew, S. Lucey, P. Lucey, S. Sridharan, and J. F. Conn, “Improved facial expression recognition via uni-hyperplane classification,” in 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012, pp. 2554–2561.

[15] R. Ptucha, G. Tsagkatakis, and A. Savakis, “Manifold based sparse representation for robust expression recognition without neutral subtraction,” in 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops). IEEE, 2011, pp. 2136–2143.

[16] S. Jain, C. Hu, and J. K. Aggarwal, “Facial expression recognition with temporal modeling of shapes,” in 2011 IEEE international conference on computer vision workshops (ICCV workshops). IEEE, 2011, pp. 1642–1649.

[17] M. Liu, S. Shan, R. Wang, and X. Chen, “Learning expressionlets on spatio-temporal manifold for dynamic facial expression recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 1749–1756.

[18] M. Liu, S. Li, S. Shan, and X. Chen, “Au-inspired deep networks for facial expression feature learning,” Neurocomputing, vol. 159, pp. 126–136, 2015.

[19] P. Liu, S. Han, Z. Meng, and Y. Tong, “Facial expression recognition via a boosted deep belief network,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 1805–1812.

[20] X. Sun, M. Lv, C. Quan, and F. Ren, “Improved facial expression recognition method based on roi deep convolutional neutral network,” in 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII). IEEE, 2017, pp. 256–261