Emotion recognition and drowsiness detection using Python

Anmol Uppal
Department of Computer Science and Engineering
ASET, Amity University
Noida, India
uppala1998@gmail.com

Shweta Tyagi
Department of Computer Science and Engineering
ASET, Amity University
Noida, India
shwetatyagi3978@gmail.com

Rishi Kumar
Department of Computer Science and Engineering
ASET, Amity University
Noida, India
rkumar25@amity.edu

Seema Sharma
Department of Computer Science and Engineering
ASET, Amity University
Noida, India
ssharma26@amity.edu

***Abstract*— Human emotions are natural expressions that people tend to make naturally, instead of any conscious effort that is accompanied by the reflexing of facial muscles. Some of the common emotions are Happy, sad, surprised, anger and stable (normal) which a human face can make according to the different situations one may find itself in. We present the software which detects and recognizes faces as well as tells a lot more about that person which could be used to get feedback from customers or to know if a person needs motivation. The objective of the project is to be an affordable and efficient product. Artificial Intelligence & Digital image processing technology used to make the system in python. Detection of eye blinking is important in certain scenarios where to avoid any accident or mishappening like in vehicles or in security vigilance. As the system also recognizes the identity card, this is a simple feature wherein the camera installed is trained in such way that it firstly focuses on the card and recognizes its shape and color.**

Keywords— Emotion recognition; Drowsiness detection; Face recognition; Digital Image Processing; Artificial Intelligence

I. INTRODUCTION

The field of Artificial Intelligence and Digital Image Processing is growing in our country slowly and steadily. Many areas of industry have started using the various techniques and applications of AI and DIP. The project can be implemented for marketing purpose also, as it let us know the feedback of any product. It provides accurate results as well as are easy to be implemented and understood in the most common systems. Also, these features can be installed in a cost effective and efficient manner in schools or colleges or any other area where surveillance is required but lack of finances is a major factor. So, using our proposed project, surveillance could be provided which results help in maintaining a regular health check and to understand he emotion of a person at work place. It can also be used as feedback of workers after making some changes at work place.

Artificial Intelligence & Digital image Processing technology used to make the system which contain face recognition, emotion recognition; drowsiness detection and id card detection. In face recognition conventional kNN algorithm is used. The given proposed work has shown us that the performance of face recognition technique can be improved much better by mixing Gabor wavelet and LBP for features extraction and the K Nearest Neighbour and Sparse Representation Classifier (KNN- SRC) for classification. We can understand a person's emotion if we are able to analyze it at different stages. For this purpose, we have an aim to develop a Convolutional -Neural Network (CNN) which is based on Facial Expression Recognition System (FER). The algorithm used for drowsiness detection detects the blinking of the eye through the camera installed using live video streams. The identity card detection is basically a feature which includes the use of 2 applications – First, the Shape Detection (Rectangle in this case) and Second, the color of the card(using RGB2HSV).

II. LITERATURE REVIEW

A. Face Recognition

Facial recognition is basically a technique that is capable of verifying or identifying the faces from an image or a video. The camera is trained in such a way that when a face comes in front of it, the first task it does is to capture at least 10 frames of the face. During this time, the face can be seen covered with a BLUE colored rectangle which depicts that the frame capture is taking place right now. When the rectangular frame disappears, it means that the face has been captured by the camera from every possible angle that was visible to it and now those frames have been saved in the database.

The k-Nearest-Neighbor (KNN) is an algorithm which is a classification technique that is considered to be non-parametric which has been used to show that it is effective in certain applications.[2] This technique achieves high accuracy in those problems which have unknown and non-normal distributions.

So this is based on projection vectors to make the classification process faster by eliminating the need to find large numbers of distances. It also uses linked list in order to retain the immediate k-nearest neighbors. Simulation of the result gives the effectiveness of the proposed algorithm.

As already discussed above, kNN algorithm is that classification technique which has been proven to be effective in certain applications. The given algorithm can give a high accuracy result to problems that consists of different distributions.[3]

Another rapid kNN classification technique is proposed in this term for recognition of various patterns and identification of textures. This technique identifies all the k nearest vectors in the training set of a kNN classifier for each input vector by performing the partial distance search in the wavelet domain. Simulated result shows us that, if we don't increase the rate of classification error, this technique requires only 12.97% of the computational time of the original kNN technique.

Major challenges that are faced in face recognition are related to the orientation being different, illumination factor like background lighting, various face expressions and aging factor.[4] Another method which is much more efficient is proposed for face recognition in an environment which can be considered to be uncontrollable where we mix the Gabor wavelets and Local Binary Patterns (LBP) in the feature extraction phase. Then, the next step is that we apply the various techniques for reduction to limit the vector patterns. At last, we mix both k Nearest Neighbor (KNN) and Sparse Representation Classifier for face recognition phase.

B. Emotion Recognition

In the current times, deep learning is that technique which is being used in many computer applications and studies for research work.[5] This technique is mainly being used in the image processing field, the whole technique can still be improved by using more accurate computer vision related applications. In the proposed study, we have an aim to develop a Convolutional -Neural Network (CNN) which is based on Facial Expression Recognition System (FER).

Figure 1: Different set of faces expressing various emotions[10]

A traditional image could have too much noise present in it which can cause a decline in the training dataset rate. To solve this problem, a new technique of recognition of expressions using CNN was introduced[6]. First thing is to decrease the range of face and for this purpose, image of the face can be identified from the original image by using the AdaBoost cascade classifier. After that, the various coordinates of eyes, mouth and other key parts can be traced by using the Haar features and the regression tree collection algorithm.

Expressing through faces is a way by which humans tend to communicate with each other. The different expressions on the face convey different information i.e. emotional, that are seen in humans. Before the face recognition part, detection of the face is important. This is a proposed technique in which detection of a face on an original image is detected using the AdaBoost cascade classifier. We obtain the human face from the original image by removing the interference data such as the background and the text. First, AdaBoost select these rectangular (weak classifier) which are the most representative of the facial features.

This is a proposed model in which there is a consideration for an improved face expression recognition (FER) technique which is based on the region of interesting (ROI) to let the convolutional neutral networks (CNN) focus only on those areas which are associated with that particular expression which the human face makes[7]. This method not only augments the given training data, it also identifies the relationship between the different ROI areas which are helpful in intensifying the accuracy, thereby making it reliable of the predicted targets. In test stage, we investigated two recognition methods: identify the test image directly; implemented decision fusion strategy on ROI areas.

Facial expressions recognition has been receiving many positive feedbacks since the last decade or so which is being used in many fields and under many circumstances. Despite all of its advantages and uses, the techniques that are traditionally present lack generalizability and flexibility when the images are captured in different backgrounds, therefore a misleading high-accuracy can be onserved. So, recognization of facial expression in real time with high accuracy is still a major problem due to differences in images that are due to light illumination, ageing etc.

Human emotions are natural expressions that people tend to make naturally, instead of any conscious effort that is accompanied by the reflexing of facial muscles.[8]

Some of the common emotions are Happy, sad, surprised, anger, stable(normal) etc which a human face can make according to the different situations one may find itself in. This is a proposed method to find the emotions of a person using deep Convolution Neural Network (CNN) and it also shows how the intensity of emotions changes on a given face from one level to another level of emotion.

C. Drowsiness Detection

Detection of eye blinking is important in certain scenarios where to avoid any accident or mishappening like in vehicles or in security vigilance. For instance in vehicles, this feature becomes important as whenever a driver may feel a bit drowsy and sleepy, then tracking on those emotions will immediately alarm the driver to be awake and stay focussed on driving.[9]

The algorithm used for this purpose detects the blinking of the eye through the camera installed using live video streams. The landmark detectors, that are trained on the various datasets shows terrific robustness against the orientation of the camera. The landmarks are detected precisely enough to accurately measure the eye opening.

As discussed earlier, facial recognition and detection through live videos using cameras can have lot of problems. The biggest could be the alignment of the face. The face should be in front of the camera in such an alignment that it is able to capture each and every feature of the face and then accordingly identify those and work accordingly.

III. SYSTEM DESIGN

A. Facial Recognition & Detection

The kNN algorithm trains the camera in such a way that after capturing the frames, the next task is to take the input, the name of that person whose face is captured and then save it to the database. This task will obviously be the first in its process as for the first time when a new face is captured, the details of it are needed to be entered. This process will continue for every new face the camera captures. When the name has been entered and saved in the database and again the same face appears in front of the camera then the screen will display the name of that person on top of the face successfully recognizing the face and capturing each and every frame successfully.

For face recognition we have used the algorithm KNN which expands to k Nearest Neighbour. We are going to show step by step how it captures, trains and predict along with its advantages over existing algorithms.

We start the program by importing the libraries. We then start the camera app in the laptop to start capturing the face and then we call the predefined the library to detect only face, and not eyes or any other feature because the quality of the camera plays an important role in this as we are only able to detect face and not the retinas too so we are going to settle with the face, using the haarcascade_frontalface.

After that we collect the data from video capture object and classifies them to a face or no face and then returns tuple (faces in image, frame read). During the extraction of the frames it changes the colour of the frame from rgb to gray so that the execution becomes much faster as the memory used to store gray images is comparatively very less than the memory used to contain rgb image. Hence, the execution becomes faster.

In collecting the data we make an array of images which is named as data. Here we make a file of .npy denoting the name of certain person for the prediction. We start with capturing every 10th frame and appending it to the name file as part of the database. We obviously resize the frame, as we are only interested in the face we capture the face and then just resize that much area and then make it in a file. We use a rectangle to denote that the frame is captured. Once the size is reached at the maximum limit which in this case is 10 frames then it automatically stops or if we press "Esc" then it closes the window. After collecting the data it just closes the window.

After starting and collecting the frames now it's time to train the data and associate the name with the face. Using the list of names it finds associated .npy files on system and trains KNN. By that I mean that the list of names is mentioned at the bottom which means the person at number one would be present for collection at first and then second and so on. That's how the system associates the frames with the name based on the serial number of the name. If we do something wrong in that then a wrong name would be associated with a certain face. It returns a KNN object.

After training the data we need to predict the name according to the face. For that we start capturing the frames and keep matching it with the existing .npy files. After that we show the name on the face with which the existing face matches. We could use the "Esc" key to stop the program.

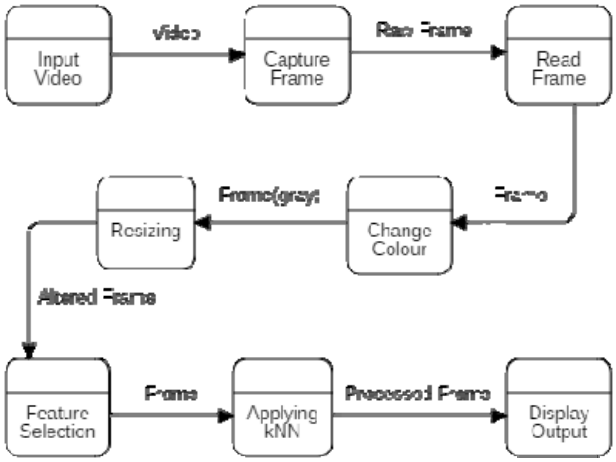

Figure 2: DFD for Face Recognition

B. Emotion Recognition

The process of Emotion recognition, as the name suggests itself, is basically trying to capture the various moods or emotions that a person makes through their faces to communicate with one another. The various moods can be termed as the facial expressions.

We simply define the facial functions as the way in which people tend to communicate with each other by using their face instead of directly talking or writing. In technical terms, it is the movement of muscles that exist under the skin of human face which tend to move when we express ourselves through our faces.

We instantiate different directories the most important one being Keras. It is essentially a high level neural network API written in python. It has the capability of running over Tensorflow which is used in this program. First we start the program by starting the webcam and calling the pre-saved file represented by "/models/emotion_model.hdf5". Then we call the different labels representing different emotions. Then we load the boundary box in which the system will detect face in which it will read the emotion. After that we load models and then we get input model shapes for inference and then finally we start lists for calculating modes. We named the window "Window Frame".

After starting the webcam we convert the bgr image into a gray image and also into rgb format. We use face_cascade.detectMultiscale to detect multiple faces and store them in "face" in which we later apply coordinates to check for emotions. We then resize the gray image.

After converting the frame into grayscale we then put it through the emotion classifier where it is supposed to match the emotion with the pre-saved emotions and then it checks what is the probability is there for the frame to be classified as a certain emotion based on the probability. Then we add emotion text which means the label corresponding to the emotion is added.

In case the emotion window is greater than window frame then it won't be able to detect the emotion. We check whether the given emotion exists in the databse and the match the probability of a certain emotion and the one with the highest probability is chosen to be displayed as the result.

After matching the probability and deciding which particular emotion it's been exhibited we just add the name in a square box along with a particular colour of the box, for example angry is shown by red box and happy as yellow and so on. In order to break the loop of videocapture the key "Q" is pressed.

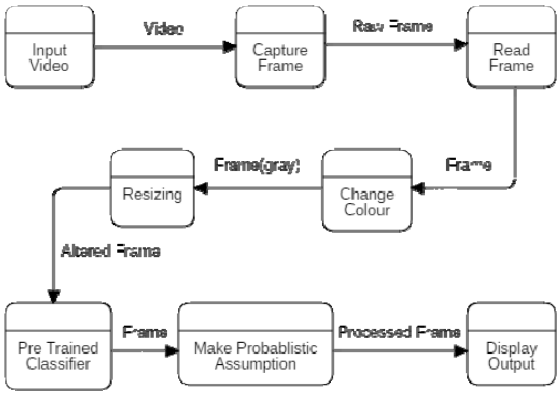

Figure 3: DFD for Emotion Recognition

C. Drowsiness Detection

Various computer vision techniques have been used in order to implement these features.Computer vision is basically a digital field that is concerned on how the computers can understand the given photo or a video in an effective way.

Computer vision involves various tasks that are important for the understanding of digital images such as processing, analyzing the given image so that the computers can interpret those images as the human does and is able to understand the features that are seen in the photo or the video. Understanding of the photo in this context means that the computer will be able to interpret the things that it is able to see in the photo as a human will do when it sees the same photo.

Algorithms like Video Processing- Video processing involves various processes and techniques that the modern day computers uses for wider range of accessibility in videos such as video tracking, image stabilization, detection of objects etc.

Detection & Tracking – Various algorithms can be used which can detect and track the objects or a person, identification of shapes and figures around that person and trying to identify and notify if something is missing from the frame. It can also be used to detect some kind of defects, if any, that are present in the given photo or video.

Morphology -The term Morphological transformation is used to on the shape of the images and performs some of the basic operations involved in it. These operations re done on the images(binary). This transformation operation involves 2 input steps-

a) the original image

b) Structuring element(kernel) which is used to decide the nature of the implemented operation.

Two basic morphological operators are Erosion and Dilation.

Image Enhancement – We can use various operations of image enhancement to basically increase the effects of an image or simply to make the image look better than the original one like reduction of noise, gray level transformation, color conversions etc.

The project uses dlib to detect the face using the get_frontal_face_detector() function, further on the shape predictor function is used to find the facial landmarks. The indexes for left and the right eye are found which are used are used t extract eye region from the frame.

All the frames are applied to preprocessing first, this includes resizing as well as converting the image to grayscale. The eye aspect ratio for the left and the right eye are found. The average of both of these values is the final eye aspect ratio.

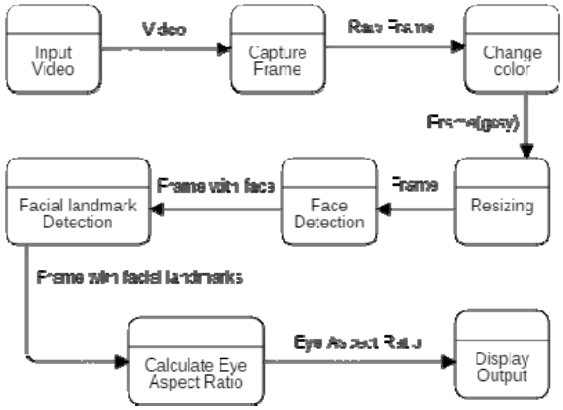

Figure 4: DFD for Drowsiness detection

IV. RESULT

A. Face Recognition

The conventional kNN algorithm finds out the k-nearest neighbors of each given sample that is present in the dataset. But in the proposed paper, the improved version of kNN algorithm not only finds the present nearest neighbors, but also the neighbors of the unknown test objects. Comparisons show that for larger values of k (around 20), the traditional algorithm shows larger, fluctuating trends but the proposed algorithm gives more adequate results.[1]

The given proposed work has shown us that the performance of face recognition technique can be improved much better by mixing Gabor wavelet and LBP for features extraction and the K Nearest Neighbour and Sparse Representation Classifier (KNN- SRC) for classification.

Table 1: For different value of k, distance calculated is shown.

Method	k	Number of distance calculations	Time in PC clocks
VQ-KNN	for all	512	22832
Proposed algorithm	5	71.19	1972
	10	81.38	2503
	15	89.21	2954
	20	95.95	3384
	25	102.34	3795

Table 2: Rate E for classification error and their time computing

	kNN (r=9216)	fast kNN (t=9216)	VQ-kNN (r=768)	fast VQ-kNN (r=768)	VQ-kNN (t=1536)	fast VQ-kNN (t=1536)
T[s]	471.02	60.93	40.68	8.47	79.01	12.37
E	5.44%	5.44 %	6.51%	6.51 %	5.63 %	5.63%

B. Emotion

The assessing process through the given technique gives increasingly good results and the accuracy obtained gives the researchers positive results for the models to be made in future of computer based emotion recognition system.

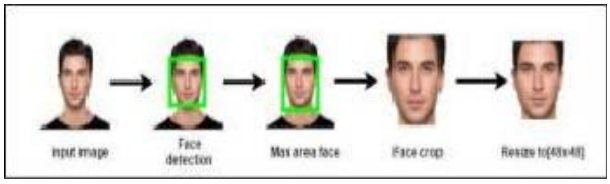

Figure 6: Image pre-processing for Emotion Detection.

The expressions on a face of determine the current state, moods and current feelings of a person through the state of nonverbal communication. We can understand a person's emotion if we are able to analyze it at different stages. In different stages the emotions of a person are significantly varying.

C. Drowsiness detection

To view the eye region it find the contour of the frame and we do this through convex hull which finds if a curve is convex. On each frame we check if the calculated eye aspect ratio is less than the given threshold value, if its more it saves the result. If it gets the eye aspect ratio as more than threshold for consecutive three frames, the person into consideration is declared to be drowsy and a warning signal is generated.

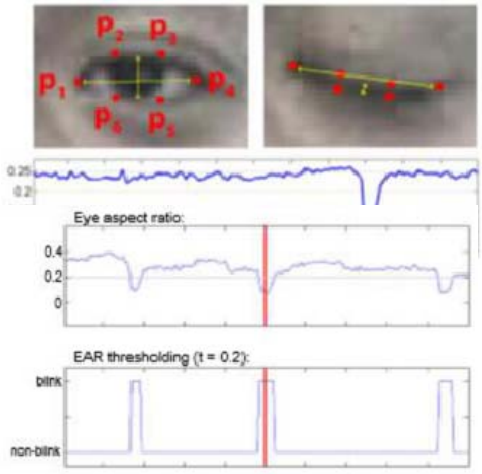

Figure 5: Landmark detection for open and closed eyes & Eye blinking examples with plots for EAR

V. FUTURE PROSPECTS

Researchers are working day in and day out to get better results from the previous times. The field of Artificial Intelligence and Digital Image Processing is growing exponentially, so the future prospective is really very bright. Many algorithms and techniques have been already proposed but the better versions of them are on their way. Letting the machine decide what actions are needed to be taken in certain circumstances is as simple as it sounds, but the work to be done behind it is extremely complex.

Training machines to think like humans can be an extremely difficult task but the way the researches and other related work that has been carried out till now is fantastic and in the future, it will only tend to get better with the advancements in the field of computer sciences.

REFERENCES

[1] Nian Zhang, Welezane Karimoune, Lara Thompson, and Hongmei Dang. “A Between-Class Overlapping Coherence-Based Algorithm in KNN Classification”, 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC) Banff Center, Banff, Canada, October 5-8, 2017

[2] SeongJoon Baek and Koeng-Mo Sung, “Fast K-nearest-neighbour search algorithm for nonparametric classification.”, ELECTRONICS LETTERS 12th October 2000 Vol. 36, No 21

[3] Wen-Jyi Hwang and Kuo-Wei Wen. “Fast kNN classification algorithm based on partial distance search.”, ELECTRONICS LETTER 15th October 1998 Vol. 34, No. 21

[4] Bilel Ameur, Sabeur Masmoudi, Amira Guidara Derbel, Ahmed Ben Hamida. “Fusing Gabor and LBP Feature Sets for KNN and SRC-based Face Recognition.” 2nd International Conference on Advanced Technologies for Signal and Image Processing

[5] Vedat TÜMEN, Ömer Faruk SÖYLEMEZ, Burhan ERGEN. “Facial Emotion Recognition on a Dataset Using Convolutional Neural Network.”

[6] Lin-Lin Xu, Shu-Mei Zhang, Fu-Xing Wang. “CNN Expression Recognition Based on Feature Graph.”

[7] Xiao Sun, Man Lv, Changquin quang. “Improved Facial Expression Recognition Method Based on ROI Deep Convolutional Neutral Network.”

[8] Rajesh Kumar G A, Ravi Kant Kumar, Goutam Sanyal. “Facial Emotion Analysis using Deep Convolution Neural Network.” International Conference on Signal Processing and Communication (ICSPC’17) – 28th & 29th July 2017.

[9] Tereza Soukupov´a and Jan ˇCech. “Real-Time Eye Blink Detection using Facial Landmarks.” 21st Computer Vision Winter Workshop Luka ˇCehovin, Rok Mandeljc, Vitomir ˇStruc (eds.) Rimske Toplice, Slovenia, February 3–5, 2016

[10] Hongwei Ng, “Deep Learning for Emotion Recognition on Small Datasets using Transfer Learning”, 2015