IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

SMART GADGET FOR VISUALLY IMPAIRED PERSON

Priti Nakhate, VJTI, Mumbai, India

Abstract: Across the world, come across many visually impaired persons and their limitations for experiencingsurrounding information about other people such as who are they, their gender, age, their facial experience. There are approximately 253 million people with visionimpairments, and assistive devices have constantly been in demand. Advanced research has led to the developmentof numerous assistive devices for blind people and visually impaired people (VIP) to improve their quality of life. As VIPs and blind people may have different behavior patterns, a criterion for classifying different types of visionimpairments which can be classify into the substitutive senses for visual perception into five categories: vision enhancement, audition, somatosensory, visual prosthesis, and olfactory and gustation. Two commonly used feedback forms, namely audition and vibration. In this project, focused on audition and working on personal assistance device with which blind person gets audio information about surrounding such as object, Gender and age of person, they can hear audio of text written. Computer vision techniques and image analysis can help improve visually-impaired people that allows for emotion detection, Age detection, Gender detection, facial recognition and detection of spoofing adapted to the needs of disabled people is proposed, implemented and validated.

Index Terms **- Face detection, CNN, D-CNN, R-CNN, Artificial intelligence, deep learning, image sensor, image processing, visually impaired, wearable devices, Storage device, raspberry pi, PyCharm, Python.**

I. INTRODUCTION

Statistics recently reported by the World Health Organization show that there are approximately 253 million people with vision impairments, among whom 36 million are blind and 217 million have moderate-to-severe vision impairments. The vision system is critical for humans to perceive the external world as more than 85% of external information can be obtained through the vision system. It largely influences our cognition and the progress of shaping spatial perception. Due to its vast significance for every individual, the absence of vision will reduce individuals' proficiency in various skills, which in turn may lead to severe livelihood problems. Owing to the ever increasing blind and visually impaired population, assistive devices have constantly been in a huge demand for recent years. For instance, as a traditional mobility aid, the white cane is the most popular among blind and visually impaired people (VIP) However, the performance is far from satisfactory as common white canes indicate limited information of obstacle position. For this reason, numerous state-of-the-art assistive devices have been developed, serving to gather more helpful clews of any obstacle, such as its category, volume and distance for VIPs and blind persons. There is need of some good assistive devices which help them in day-to-day life. Sensory modalities (e.g., audition) are used to provide environmental information normally by vision sense for VIPs and totally vision-deprived people. From the view of image processing, the purpose of image processing forvision replacement and vision substitution is to convert the visual information into the spatially resolved non-visual signals. Based on the feedback forms, we can conclude that the vision enhancement and vision replacement are, respectively, suitable for VIPs and blind people, while the vision substitution can be used for both VIPs and blind people.As vision replacement is option for assistance to VIPs and blind people, Assistance device having capability of giving environment information this project deals with this. In this following objective taken into consideration such as Facial expression object detection, gender detection, age detection, ability to read out text written, currency detection, all this information in the form of audio provide to both VIPs and blind people. Image processing play very important role in these objectives. CNN (convolutional neural network), deep learning modules, various model such as Haar cascade classifier mode used for finding result. Hardware part for initial testing used are raspberry pi with suitable camera and earphones to get 2 information processed by system. Raspberry pi a mini computer allows quick fixation of output irregularities by changing programming / codes.

II. RELATED WORKS

A Fernández [1] proposed facial detection, normalization of face extraction of characteristic through lbp (trough local binary pattern), shorting of facial recognition and spoofing detection through svm (Support vector machine) aggregated information generated into text this text converted into speech through text -to-speech converter.

Shekhar Singh [8] This study's six CNNs layer architecture performed competitively and achieved a FER2013test accuracy of 61.7%, without involving any pre-processing or feature extraction techniques. The state-of-the-art testaccuracy for the seven emotion categories using the ensemble of CNNs was 75.2%. We performed several experiments usingdifferent batch sizes and epochs but obtained the best testaccuracy using a batch size value of 512 and 10 epochs.

Lingling liu School of information Engineering [9] In thispaper, firstly, the idea of this project facial expression recognition is proposed. Then the author tries to find related literature about this topic as more as possible. Then the detailed network structures and experiments are introducedand last do the test work. The processing and results of the experiment are mentioned in the paper, also some potential method can be used to improve the efficiency and accuracy in the future in this topic are also proposed. These directions contain network structures, data processing, efficient loss functions.

III. METHODOLOGY

Figures presents the deep-see face architecture that involves six independent modules: object, emotion, age, gender, facial recognition, spoofing, text to speech to the needs of disabled people is proposed, implemented and validated.

1. Emotion detection

1.1 IMPLEMENTATION

The database used in the study consisted of facial expression images from the FER2013 database. Two types of parameters were extracted from the facial image: real valued and binary.A total of 15 parameters consisting of eight real-valued parameters and seven binary parameters were extracted from each facial image. The real valued parameters were normalized. Generalized neural networks were trained with allfifteen parameters as inputs. There were five output nodes corresponding to the five facial expressions (Natural, Happy, Sad, Angry, Calm). Based on initial testing, the best performing neural networks were recruited to form a generalized committee for expression classification. Due to a number of ambiguous and no- classification cases during the initial testing, specialized neural networks were trained forNatural, Happy, Sad, Angry, and Calm expression. The integrated committee neural network classification system wasevaluated with an independent expression dataset not used in training or in initial testing. A generalized block diagram ofthe entire system is shown in Figure1

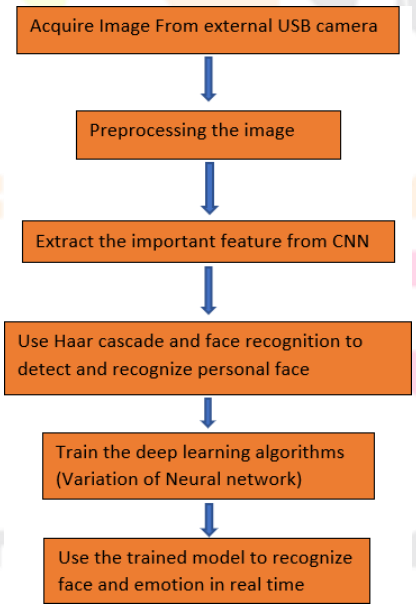

FIGURE 1: AN OVERALL BLOCK DIAGRAM

1.2 IMAGE PROCESSING AND FEATURE EXTRACTION

Two types of parameters were extracted from the facialimages of 97 subjects: (1) real valued parameters and (2)binary parameters. The real valued parameters have a definitevalue depending upon the distance measured. This definitevalue was measured in number of pixels. The binary measuresgave either a present (= 1) or an absent (= 0) value. In all,eight real valued measures and seven binary measures wereobtained. A number of parameters, both real-valued andbinary, were extracted and analysed to decide theireffectiveness in identifying a certain facial expression. Thefeatures which did not provide any effective information of thefacial expression portrayed in the image were eliminated and were not used in the final study. The real valued and binaryfeature selection was inspired by 8 the FACS. The followingreal valued and binary parameters were finally used in thestudy.

1.3 REAL VALUED PARAMETERS

1. **Eyebrow raise distance –** The distance between thejunction points of the upper and the lower eyelid and the lower central tip of the eyebrow.
2. **Upper eyelid to eyebrow distance –** The distance betweenthe upper eyelid and eyebrow surface.
3. **Inter-eyebrow distance –** The distance between the lowercentral tips of both the eyebrows.
4. **Upper eyelid – lower eyelid distance –** The distancebetween the upper eyelid and lower eyelid.
5. **Top lip thickness –** The measure of the thickness of the top lip.
6. **Lower lip thickness –** The measure of the thickness of the lower lip.
7. **Mouth width –** The distance between the tips of the lipcorner.
8. **Mouth opening –** The distance between the lower surface of top lip and upper surface of lower lip.

FIGURE 2: REAL-VALUED MEASURES FROM A SAMPLE NEUTRAL EXPRESSION IMAGE.

1-eyebrow raise distance, 2-upper eyelid to eyebrow distance, 3-inter eyebrow distance, 4- upper eyelid to lower eyelid distance, 5-top lip thickness, 6-lower lip thickness, 7-mouth width, 8-mouth opening. (Facial expression image from the Cohn-Kanade database. Used with permission).

1.4 BINARY PARAMETERS

1. **Upper teeth visible –** Presence or absence of visibility of upper teeth.
2. **Lower teeth visible –** Presence or absence of visibility of lower teeth.
3. Forehead lines – Presence or absence of wrinkles in the upper part of the forehead.
4. **Eyebrow lines –** Presence or absence of wrinkles in the region above the eyebrows.
5. **Nose lines –** Presence or absence of wrinkles in the region between the eyebrows extending over the nose.
6. **Chin lines –** Presence or absence of wrinkles or lines on the chin region just below the lower lip.
7. **Nasolabial lines –** Presence or absence of thick lines on both sides of the nose extending down to the upper lip.

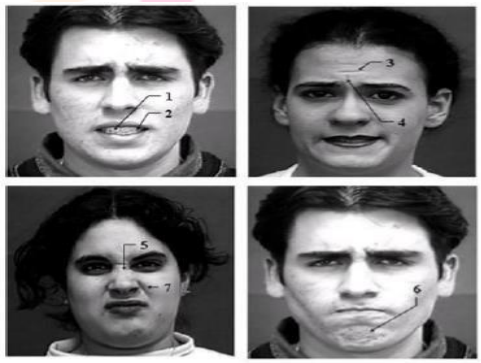

FIGURE 3: Binary measures from sample expression images. 1-upper teeth visible, 2-lower teeth visible, 3-forehead lines, 4-eyebrow lines, 5-nose lines, 6-chin lines, 7-nasolabial lines. (Facial expression image from the Cohn-Kanade database. Used with permission).

The real valued parameters were the distances (in number ofpixels) measured between specified facial features. In case of parameters involving features, which were symmetrically present on both sides of the face, an average of both themeasurements was obtained. Real valued measures wereobtained for expressions including the neutral image. The realvalued parameters were then normalized in the following manner:

$$\text{Normalized Value} = \frac{\text{(Measured Value-Neutral Value)}}{\text{Neutral Value}}$$

All the parameters were extracted by manual and/or semi- automatic techniques. The purpose of the present study was to evaluate the efficacy of committee neural networks. Therefore, no effort was made to develop automated techniques for feature extraction. The binary parameters were characterized by the presence or absence of the facial muscle contractions or the facial patterns formed dueto these contractions. An edge detection algorithm was appliedto the image to determine if the pattern was present or absent. A simple canny edge detector was used to determine whethera pattern of lines existed which further decided the binary feature was true (1) or false (0). The eight normalized real valued parameters together with the seven binary parameters were fed to neural networks. The entire dataset from 97 subjects (467 images) was divided into three groups: 25 subjects (139 images) for training, 10 subjects (46 images) for initial testing, and 62 subjects (282 images) for final evaluation.

1.5 TRAINING OF GENERALIZED NEURAL NETWORKS

Several multi layered, fully connected, feed forward neural networks were trained to classify different expressions. A total of 105 networks were trained using different number of hidden layers (2, 3, 4, 5), different initial weights, different number of

neurons in the hidden layers (7, 14, 15, 28, 45, 60), and different transfer functions. Each network had fifteen input nodes, each corresponding to the fifteen input parameters. Each of these networks had seven output nodes, each corresponding to one of the seven expressions (Natural, Happy, Sad, Angry, Calm). Since the normalized input data was in the range of -1 to 1, the "tansig" function was used for the hidden layer neurons. The output of the neural network has to be in the 0 to 1 range. Thus, the "logsig" function was used as the transfer function for the output layer neurons. The output of each node was converted to a binary number (either 0 or 1). An output of 0.6 or more was forced to 1 and an output of less than 0.6 was forced to 0. An output of 1 indicated that particular expression was present and output of 0 indicated that particular expression was absent. We have varied the threshold from 0.55 to 0.9 and found that a threshold of 0.6 gave better results.

2. GENDER AND AGE DETECTION

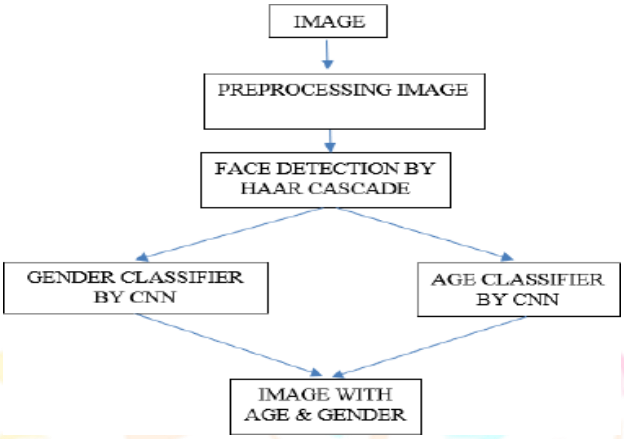

FIGURE 4. BLOCK DIAGRAM FOR THE AGE GENDER DETECTION

In this Python Project, we will use Deep Learning to accurately identify the gender and age of a person from a single image of a face. The predicted gender may be one of 'Male' and 'Female', and the predicted age may be one of the following ranges- (0 – 2), (4 – 6), (8 – 12), (15 – 20), (25 –32), (38 – 43), (48 – 53), (60 – 100) (8 nodes in the final soft max layer). It is very difficult to accurately guess an exact age from a single image because of factors like makeup, lighting, obstructions, and facial expressions. And so, we make this a classification problem instead of making it one of regression.

2.1 FACE DETECTION WITH HAAR CASCADES

This is a part most of us at least have heard of. Open CV provide direct methods to import Haar-cascades and use them to detect faces.

2.2 GENDER RECOGNITION WITH CNN

Gender recognition using OpenCV's fisher faces implementation is quite popular and some of you may have tried or read about it also. But, in this example, I will be using a different approach to recognize gender. This method was introduced by two Israel researchers, Gil Levi and Tal Hassner in 2015. I have used the CNN models trained by them in this example. We are going to use the OpenCV's DNN package which stands for "Deep Neural Networks". In the DNN package, OpenCV has provided a class called Net which can be used to populate a neural network. Furthermore, these packages support importing neural network models from well- known deep learning frameworks like caffe, tensor flow and torch. The researchers I had mentioned above have published their CNN models as caffe models. Therefore, we will be using the Caffe Importer import that model into our application.

2.3 AGE RECOGNITION WITH CNN

This is almost similar to the gender detection part except that the corresponding prototxt file and the caffe model file are "deploy_agenet. prototxt" and "age_net. Caffe model". Furthermore, the CNN's output layer (probability layer) in this CNN consists of 8 values for 8 age classes ("0–2", "4–6", "8–13", "15–20", "25–32", "38–43", "48–53" and "60-") A caffe model has 2 associated files,

1. **PROTOTXT-** The definition of CNN goes in here. This file defines the layers in the neural network, each layer's inputs, outputs and functionality.
2. **CAFFEMODEL -**This contains the information of the trained neural network (trained model).

2.4 THE CNN ARCHITECTURE

The network comprises of only three convolutional layers and two fully-connected layers with a small number of neurons. This, by comparison to the much larger architectures applied. Our choice of a smaller network design is motivated both from our desire to reduce the risk of overfitting as well as the nature of the problems we are attempting to solve: age classification on the audience set requires distinguishing between eight classes; gender only two. This, compared to, e.g., the ten thousand identity classes used to train the network used for face recognition in.

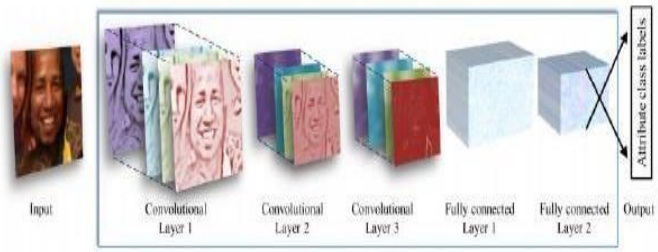

FIGURE 5. CNN ARCHITECTURE

All three-color channels are processed directly by thenetwork. Images are first rescaled to 256 × 256 and a crop of 227 × 227 is fed to the network. The three subsequentconvolutional layers are then defined as follows.

- Convolutional layer; 96 nodes, kernel size 7
- Convolutional layer; 256 nodes, kernel size 5
- Convolutional layer; 384 nodes, kernel size 3

It has 2 fully connected layers, each with 512 nodes, and afinal output layer of soft max type.

1. 96 filters of size 3×7×7 pixels are applied to the input in thefirst convolutional layer, followed by a rectified linear operator (ReLU), a max pooling layer taking the maximalvalue of 3 × 3 regions with two-pixel strides and a localresponse normalization layer.
2. The 96 × 28 × 28 output of the previous layer is then processed by the second convolutional layer, containing 256 filters of size 96 × 5 × 5 pixels. Again, this is followed by ReLU, a max pooling layer and a local response normalizationlayer with the same hyper parameters as before.
3. Finally, the third and last convolutional layer operates on the 256 × 14 × 14 blob by applying a set of 384 filters of size 256 × 3 × 3 pixels, followed by ReLU and a max pooling layer.

The following fully connected layers are then defined by:

4. A first fully connected layer that receives the output of the third convolutional layer and contains 512 neurons, followed by a ReLU and a dropout layer.
5. A second fully connected layer that receives the 512- dimensional output of the first fully connected layer and again contains 512 neurons, followed by a ReLU and a dropoutlayer.
6. A third, fully connected layer which maps to the finalclasses for age or gender. Finally, the output of the last fully connected layer is fed to a soft-max layer that assigns a probability for each class. The prediction itself is made by taking the class with the maximal probability for the given testimage.

2.5 TESTING AND TRAINING INITIALIZATION

The weights in all layers are initialized with random values from a zero mean Gaussian with standard deviation of 0.01. To stress this, we do not use pre-trained models for initializing the network; the network is trained, from scratch, without using any data outside of the images and the labels available by the benchmark. This, again, should be compared with CNN implementations used for face recognition, where hundreds of thousands of images are used for training. Target values for training are represented as sparse, binary vectors corresponding to the ground truth classes. For each training image, the target, label vector is in the length of the number of classes (two for gender, eight for the eight age classes of the age classification task), containing 1 in the index of the ground truth and elsewhere.

3. OBJECT DETECTION

3.1 DETECTING OBJECTS USING R-CNN

To overcome the problems faced in CNN, R-CNN was adopted. Girshick et al. proposed a method where Selective Search is used to extract just 2000 regions from the image which are known as region proposals. While in CNN we have to deal with the classification of a large number of regions but now in R-CNN we only need to work with 2000 regions.

These 2000 region proposals are generated using the selective search algorithm. The 4 regions which form an object can be regarded as - varying scales, colors, textures, and enclosure. The selective search tries to identify such patterns in the image and using this it proposes different regions. Selective search comprises of various steps-

Step1- Input an image and generate initial sub- segmentations to obtain multiple regions from the image.

Step2- Combine similar regions to form a larger region that is based on the color similarity, texture similarity, size similarity, and shape compatibility.

Step3- The regions now produce the final object locations (Region of Interest).

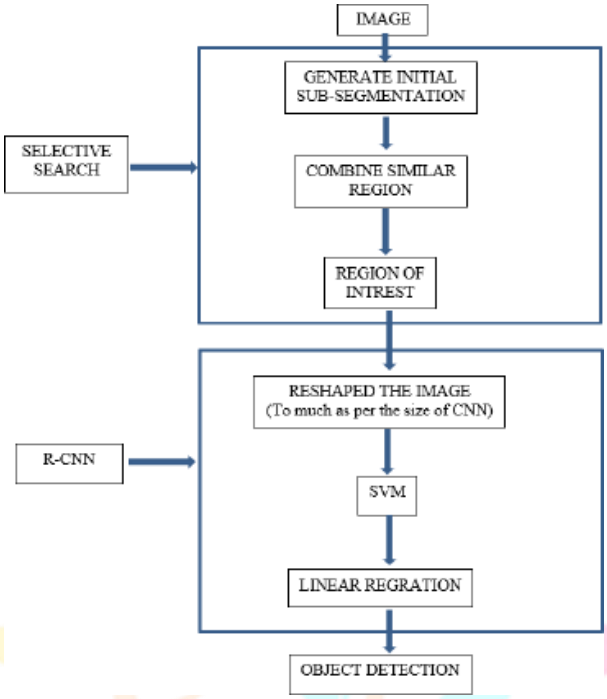

FIGURE 6. BLOCK DIAGRAM FOR THE OBJECT DETECTION

During each iteration, larger segments are formed & added to the list of region proposals. Thus, a bottom-up approach is incorporated to create region proposals from smaller segments to larger segments and this is referred to as computing "hierarchical" segmentations using Felzenszwalb and Huttenlocher's over segments.
Now, to detect objects using R-CNN a series of steps takes place-
A pre-trained CNN is taken, which is retrained on the last layer of the network based on the number of classes which are needed to be detected.
Now, the ROI (Region of Interest) for every image is taken & then these regions are reshaped to match as per the size of CNN.
Now regions are obtained, so a Linear Support Vector Machine (SVM) classifier is trained to classify the objects and background, i.e., for each class, one binary SVM is trained.
In the last step, a linear regression model is trained to output tighter coordinates for the box once the object has been classified in the image.

IV. CONCLUSION

The neural networks offer a potential tool for image-based mood detection. Eight real valued and seven binary parameters were successfully extracted from 97 subjects (467 facial images) for seven different expressions (Natural, Happy, Sad, Angry, Calm). CNN can be used to provide improved age and gender classification results, even considering the much smaller size of contemporary unconstrained image sets labeled for age and gender. The simplicity of our model implies that more elaborate systems using more training data may well be capable of substantially improving results beyond those reported here.
The R-CNN was adopted for object detection due to a large number of regions in CNN. However, it still takes much time in R-CNN to predict for a new test image. Thus, it leads to variations of R-CNN like Fast R-CNN, Faster R-CNN, Mask R-CNN, which are more efficient compared to previous versions.

V. REFERENCES

[1] Fernández, J. L. Carús, R. Usamentiaga and R. Casado "Face Recognition and Spoofing Detection System Adapted to Visually- Impaired People" IEEE LATIN AMERICA TRANSACTIONS, VOL. 14, NO. 2, FEB. 2016, DOI 10.1109/TLA.2016.7437240

[2] He Zhang and Cang Ye, Senior Member, IEEE "An Indoor Way finding System based on Geometric Features Aided Graph SLAM for the Visually Impaired" Citation information: IEEE Transactions on Neural Systems and Rehabilitation Engineering. DOI 10.1109/TNSRE.2017.2682265,

[3] WAN-JUNG CHANG, (Member, IEEE), LIANG-BI CHEN, (Senior Member, IEEE), CHIA-HAO HSU, JHEN-HAO CHEN, TZU-CHIN YANG, AND CHENG-PEI LIN "MedGlasses: A Wearable Smart-Glasses-Based Drug Pill Recognition System Using Deep Learning for Visually Impaired Chronic Patients" Citation information: IEEE Access, VOL-8, Jan 2020, DOI 10.1109/ACCESS.2020.2967400

[4] Bogdan Mocanu1,2, (Member, IEEE), Ruxandra Tapu1,2, (Member, IEEE), and Titus Zaharia1 "DEEP-SEE FACE: A Mobile Face Recognition System Dedicated to Visually Impaired People" Citation information: IEEE Access, VOL. 6, Sep 2018DOI 10.1109/ACCESS.2018.2870334

[5] Laurindo Britto Neto, Felipe Grijalva, Student Member, IEEE, Vanessa Regina Margareth Lima Maike,Luiz C´esar Martini, DineiFlorencio, Fellow, IEEE, Maria Cec´ılia Calani Baranauskas,Anderson Rocha, Senior Member, IEEE,

and Siome Goldenstein, Senior Member, IEEE "A Kinect-Based Wearable Face Recognition System to Aid Visually Impaired Users" IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS(Volume: 47, Issue: 1, Feb. 2017) DOI 10.1109/THMS.2016.2604367

[6] J.M. S´aez, F. Escolano, M.A. Lozano "Aerial obstacle detection with 3D mobile devices" Citation information: IEEE Journal of Biomedical and Health Informatics (Volume: 19, Issue: 1, Jan. 2015) DOI 10.1109/JBHI.2014.2322392

[7] Muiz Ahmed Khan , Pias Paul , Mahmudur Rashid , Student Member, IEEE, Mainul Hossain , Member, IEEE, and Md Atiqur Rahman Ahad , Senior Member, IEEE "An AI-Based Visual Aid With Integrated Reading Assistant for the Completely Blind" IEEETRANSACTIONS ON HUMAN-MACHINESYSTEMS(Volume: 50, Issue: 6, Dec. 2020), DOI: 10.1109/THMS.2020.3027534

[8] Shekhar Singh Computer Science University of Nevada Las VegasLas Vegas, USA, Fatma Nisus Computer Science University of Nevada Las Vegas Las Vegas, USA "Facial Expression Recognition with Convolutional Neural Networks", 2020 10th Annual Computing and Communication Workshop and Conference (CCWC)

[9] Lingling liu School of information Engineering, Wuhan University of technology, China "Human Face Expression Recognition Based on Deep Learning- Deep Convolutional Neural Network", 2019 International Conference on Smart Grid and Electrical Automation (ICSGEA)

[10] S. Lawrence, C.L. Giles, A.D. Back et al., "Face Recognition: A Convolutional Neural Network Approach", IEEE Trans Neural Netw, vol. 8, no. 1, pp. 98-113, 1997.

[11] A. Shaukat, M. Aziz and U. Akram, "Facial Expression Recognition Using Multiple Feature Sets", 2015 5th International Conference on IT Convergence and Security ICITCS), pp. 1-5, 2015.

[12] J.Y.R. Cornejo, H. Pedrini and F. Florez-Revuelta, "Facial expression recognition with occlusions based on geometric representation in: Progress in Pattern Recognition Image Analysis Computer Vision and Applications", Proceedings of the 20th Iberoamerican Congress (CIARP2015), pp. 263-270, 2015.

[13] Robert Katzschmann, Member, IEEE, Brandon Araki, Member, IEEE, and Daniela Rus,NFellow, IEEE "Safe Local Navigation for Visually Impaired Users with a Time-of-Flight nd Haptic Feedback Device" TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING.

[14] He Zhang and Cang Ye, Senior Member, IEEE "An Indoor Way finding System based on Geometric Features Aided Graph SLAM for the Visually Impaired" Citation information: IEEE Transactions on Neural Systems and Rehabilitation Engineering, DOI 10.1109/TNSRE.2017.2682265

[15] Harshitha S, Sangeetha N, Shirly Asenath P, Abraham Chandy D Department of Electronics and Communication Engineering Karunya Institute of Technology and Sciences Coimbatore, India "Human facial expression recognition using deep learning technique", 2019 International Conference on Signal Processing and Communication (ICSPC-2019), March. 29– 30, 2019, Coimbatore, INDIA

[16] Jascha Sohl-Dickstein*, Santani Teng*, Benjamin M. Gaub, Chris C. Rodgers, Crystal Li, Michael R. DeWeese and Nicol S. Harper "A device for human ultrasonic echolocation" This article has beenaccepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: IEEE Transactions on Biomedical Engineering.