International Journal of Research in Engineering, Science and Management
Volume-2, Issue-3, March-2019
www.ijresm.com | ISSN (Online): 2581-5792

Deep See Face Recognition System Dedicated to Visually Impaired People

P. Ansarshahin[1], P. Madhumitha[2], G. Sandra Karunya[3], A. Mathan Gopi[4], D. Rajini Girinath[5]
[1,2]*UG Student, Dept. of Computer Science and Engg., Sri Muthukumaran Institute of Technology, Chennai, India*
[3,4]*Assistant Professor, Dept. of Computer Science and Engg., Sri Muthukumaran Inst. of Tech., Chennai, India*
[5]*Professor, Dept. of Computer Science and Engg., Sri Muthukumaran Inst. of Tech., Chennai, India*

***Abstract*: Outwardly disabled individuals face parcel of challenges in their everyday life. Numerous multiple times they depend on others for help. A few innovations for help of outwardly hindered individuals have been created. Among the different innovations being used to help the visually impaired, Computer Vision based arrangements are rising as a standout amongst the most encouraging alternatives because of their moderateness and availability. This paper proposes a framework for outwardly debilitated individuals. The proposed framework intends to make a wearable visual guide for outwardly impeded individuals in which discourse directions are acknowledged from the client. Its usefulness tends to recognizable proof of articles and sign sheets. This will assist the outwardly weakened individual with managing everyday exercises and to explore through his/her environment. Raspberry Pi is utilized to execute fake vision utilizing python language on the Open CV stage.**

***Keywords*: Hear-like Features, Image Processing, Open CV, Python, Raspberry Pi, speech commands, Video capturing.**

1. Introduction

Visual deficiency is an issue that plagues a huge number of individuals all over the place. Daze individuals face numerous sorts of obstacles in playing out each day schedule works. Indeed, even in their very own homes they should show endeavors to explore starting with one spot then onto the next and to find person. As indicated by the World Health Organization (WHO), 253 million individuals live with visual impedance, 36 million of which are visually impaired and 217 million individuals have moderate to extreme vision debilitation. As of late, a few structures dependent on versatile stages what's more, devoted to social insurance administrations have developed. The novel advances created go for decreasing the expenses of the wellbeing part, by expanding the strengthening of individuals and, in a similar time, by improving the observing of patients with unending maladies. Through the persistent evaluation of side effects, such frameworks can assist the patients with managing their condition by their own, without requiring direct supervision of particular social insurance faculty. Right now, the patient checking frameworks dependent or digital physical frameworks (CPS) are pulling in impressive consideration from the scientific network. Such developing advancements have been utilized to different purposes: encourage smoking discontinuance, screen patients with constant heart disappointment, identify early indications of arrhythmia or on the other hand ischemia, give diabetes training or screen pertinent physiological markers. With a couple of outstanding exemptions (psychological wellness and mental imbalance), the general population with incapacities have not been the essential target of the developing versatile wellbeing applications. In any case, people with incapacities are probably going to participate in practices that can put their wellbeing in danger and there is a solid need of advances that can improve their day by day life conditions, empower social relations, and increment their level of independence what's more, wellbeing.

The outwardly disabled people adapt to normal life by using traditional assistive aids, for example, white sticks or strolling hounds. The white stick is preferred because it is easy to use, cheap and widely accepted by the visually impaired network. In any case, such an assistive component demonstrates rapidly its restrictions when faced with the high decent variety of circumstances that can happen in current urban scenes. Moreover, the white stick can't give extra data to clients, for example, the level of threat of the experienced obstructions or acknowledgment of people that are available in the scene.

Without such data, the VI dependably goes on known ways while attempting to figure the character of the people experienced. At the point when a VI client touches base in a social setting, the discussion must be hindered so as to declare which individuals are available. The proposed structure separate frame video-based highlights utilizing a profound face CNN show. The highlights are then totaled into a worldwide portrayal that can consider the varieties of the face appearance amid its life cycle. Furthermore, we present a hard negative mining stage intended to separate between known appearances and obscure characters. Such an issue is fundamental, so as to maintain a strategic distance from false alerts, when planning a customized learning system, where the client filter indicate their very own inclinations in terms of characters to be perceived. At last, the semantic

data about the nearness of a comfortable is conveyed with the assistance of acoustic cautioning messages, transmitted through bone conduction earphones.

2. Related work

In existing methodology it is a technique to structure a Text to Speech transformation module by the utilization of Mat lab by basic network tasks. Right off the bat, by the utilization of amplifier some comparative sounding words are recorded utilizing a record program in the Mat lab window and recorded sounds are spared in ".wave" group in the index. The recorded sounds are then examined and the tested qualities are taken and isolated into their constituent phonetics. The isolated syllables are then connected to recreate the ideal words. By the utilization of different Mat lab directions for example wave read, subplot and so forth the waves are examined and removed to get the ideal outcome. This strategy is easy to execute and includes a lot lesser utilization of memory spaces. The current route frameworks for the visually impaired individuals require an exact GPS maps. This make them unusable in area where there are no GPS maps, they are not adequately exact. Calculation for GPS route for the outwardly weakened along a GPS track, which depict the way as a grouping of waypoints is proposed. The normal voice route, versatile to the speed and exactness of the GPS information, beginning of the route from any waypoint, connection of the bearing of development on the off chance that it is essential, return the client to the course if deviation is veered off, work with and without electronic compass, location of the development of the client the other way. The face discovery is performed utilizing the conventional Viola-Jones calculation with Hear-like highlights, while for acknowledgment the Local Binary Patterns Histograms calculation is utilized. The framework has been extended, where authors propose a CNN-based approach to perform both people detection and recognition. Even though the method returns good results for the detection module the performance of the recognition system is inferior 70% and is influenced by lighting condition or by user/camera motion. In addition, the system has never tested with actual visually impaired people and nothing is said about the hardware architecture or about the acoustic warning messages.

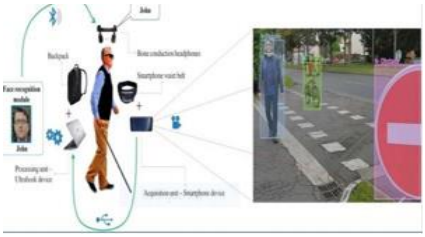

Fig. 1. Work

3. Literature review

A traditional strategy that has been utilized for quite a long time for help of the outwardly disabled is the utilization of guide hounds that are prepared to assist the visually impaired with navigating and strolling stick for maintaining a strategic distance from obstructions. Brilliant rehabilitative shoes and exhibitions proposed in encourage safe route and portability of visually impaired people. Each shoe is mounted with ultrasonic transducers to distinguish objects at various statures. The displays are instrumented with a couple of ultrasonic transducers mounted halfway over the extension, and with a ringer at one of the sanctuaries. The stick proposed in utilizes a ping ultrasonic sensor alongside Camshaft Position Sensor (CMP) compass sensor 511 so as to give data about potholes and hitches. The disservice of this framework is that the CMP compass sensor is powerless to any stray attractive fields and it is influenced by any iron items in the environment. It gives long separation route yet isn't useful in indoor route. Prior, sensor based methodology were simply to identify protests instead of remembering them. In this way, picture preparing ends up being a reasonable strategy in taking care of these circumstances. A Stereo Image Processing System for Visually Impaired is a framework that incorporates a wearable PC, stereo cameras as vision sensor and stereo headphones, all mounted on a protective cap. The picture of the scene before outwardly impaired is caught by the vision sensors. The caught pictures are handled to improve the critical highlights in the scene in front, for route help. So as to consolidate the separation data, stereo cameras are utilized . Be that as it may, the framework utilizes a stereo camera subsequently making the framework complex and cost ineffectual. Ongoing Visual Recognition with results changed over to 3D Audio is a framework which involves a few modules. Video is caught with a versatile camera gadget (Microsoft Kinect, or GoPro) on the customer side, and is gushed to the server for continuous picture acknowledgment with existing article discovery models. This framework expects time to process which does not make the framework to run/working progressively. A Hear Cascade based item distinguishing proof is the most generally utilized method for article recognizable proof. This orders the entire picture into huge and non-noteworthy classifications utilizing an administered procedure. Shading based item distinguishing proof is a strategy which utilizes shading highlights in recognizing and following the article. The current methodologies experience the ill effects of the downsides, for example, prerequisite of a few sensors, framework not being compact and neglect to do continuous preparing. In the proposed framework, genuine endeavors are made to address few of these issues.

4. Proposed system

The block diagram of the proposed system is shown in

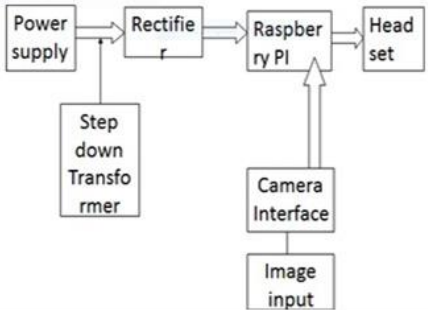

Fig. 2. Block diagram

The system uses speech commands as the user interface. A microphone is used to capture speech input. The obtained input is recognized using Google API. Also, it uses image processing as its primary technique to identify objects and sign boards. Video is captured by the Pi camera, from which the frames are extracted. The frames are preprocessed for better results. Image processing algorithms for object detection are applied on these frames and the object is detected. After the object is located, audio messages from headsets are given to the user to notify him/her about the location of the object. Obstacle detection is done using ultrasonic sensors. Any obstacle encountered in the path is notified to the blind person by producing a beep sound from the buzzer.

A. Raspberry Pi

The raspberry pi is a single-board computer based on Broadcom BCM2837 system on chip. It has a 1.2GHz CPU on board. It uses a 64-bit quad-core ARMv8 architecture based CPU. The raspberry pi version features 1GB of RAM. It uses an SD card to store the OS. It has USB port through which USB microphone is connected and speech input is given. It has a CSI port through which Pi camera is connected. The Raspberry PI features a 3.5 mm universal headphone jack for audio out. The Raspberry pi performs the task of taking video input, converting it to frames, does suitable image processing in Open CV platform using Python language.

B. Pi camera

Raspberry Pi camera module is used to take high resolution video, as well as still images. It has a resolution of 8 megapixel and 30 frames per second (fps). The output from the camera is fed to Raspberry Pi for further processing.

C. Ultrasonic Sensor

An Ultrasonic sensor measures the distance to an object by using sound waves. It does so by sending out a sound wave at a specific frequency and listening for that wave to bounce back. The elapsed time between the sound wave being generated and the wave bouncing back is recorded and the distance between the sonar sensor and the object is calculated. In this system, it is used to notify the user about any obstacle that is ahead of him/her.

D. Open CV and python

It is a library of programming functions mainly aimed at real time computer vision. It is used for various applications such as augmented reality, gesture recognition, feature matching etc. It is imported by using the command “import cv2” in python. Python is a widely used high level programming language which has a dynamic type system and automatic memory management and supports multiple programming paradigms including object oriented, imperative, functional programming and procedural styles. Python is a light-weight programming tool that has many built-in functions and does not consume many resources while operating on the Raspberry pi.

5. Modules

The DEEP-SEE FACE architecture that involves four independent modules: face detection, multiple people tracking, people identity recognition and acoustic feedback.

A. Face detection

The face detection module is based on the Faster R-CNN[28] with Region Proposal Networks (RPN) . Following the default settings, we have used 3 scales (128 × 128, 256×256 and 512×512 pixel blocks) and 3 aspect ratios (1:1, 1:2 and 2:1) that translate to n = 9 anchors at each possible location of a face. For a feature map of size W ×H (where W and H represent the width and height, respectively), we obtain a maximum number of W×H×n proposals. As indicated in, the RPN training is performed using the stochastic gradient descent (SGD) for both the classification and the regression branches. We train the face detection model using the pre-trained Image Net model of VGG. The training images are resized in order to fit the GPU memory constraints based on the following scheme: 1024/max(W, H),where W and H are the width and height of the image, respectively. The system is run for 100k iterations with a learning rate of 0.001 and for another 50k iterations at a learning rate of 0.00001.

B. Face tracking

The tracking system takes as input, at a given frame, the face bounding box indicated by the detection module (cf. Section III.A). Then, the goal is to determine the face position between consecutive frames. The tracking methodology is based on our previous ATLAS algorithm introduced in [13] that is adapted to work on face tracking scenarios and on multiple moving instances. We decided to use ATLAS due to its high performance and reduced computational costs. The ATLAS tracker is based on an offline- trained convolutional neural regression network that learns generic relations between various face appearances models and their associated motion patterns. The system receives as input the target and its associated search region and returns the target novel location (i.e., the

coordinates of the face bounding box). The process is based on a set of comparisons between high-level features representation extracted from both faces and search regions We need to emphasize that the CNN weights are modified uniquely during training (in the offline stage).In the online phase ,the network weights are frozen and no fine-tuning is required. The technique is robust to important deformation ,light change face motion and can function at more than 50fps when running on an Nvidia1050 GPU.

C. *Face recognition*

Each face identified by the detection module is represented as a set of features extracted from the last layer before the classification layer of a traditional CNN. In our implementation, we have adopted the VGG16 network architecture with the batch normalization strategy . Let us note that other CNNs topologies can be employed. In our work, we have preferred to use a relatively standard representation, without focusing on any optimization at this stage. Instead, we have put forward the adaptation/ personalization strategies. Notably, we show that such stages can be accomplished uniquely by considering the final layers of the network, with a light re-learning process. The VGG output is a 4096-dimensional feature vector representation (corresponding to the penultimate layer) of the face, which is further normalized to a unit vector.

6. Experimental setup

A. *The benchmark*

Due to the novelty of the application and the unavailable free data that can be used for testing the performance of the proposed architecture, we have created a video dataset of 30 video sequences, with an average duration of 10 minutes, recorded at a resolution of 1280×720 pixels and with 30 fps. Five video streams have been recorded with a regular smart phone by real visually impaired users, walking in indoor/outdoor scenes, while 25 image sequences have been provided by the France national television. We need to highlight that the videos recorded in real-world conditions by the VI users are highly challenging: they are trembled, noisy, include different lighting conditions, motion blur, rotation and scale changes.

B. *CNN training for face recognition*

In the training phase, we have considered a dataset with categories of known persons that contain faces representing user family members and friends and also some celebrities (politicians, movie stars or singers) appearing on TV. For each person, a maximum number of 800 face instances were stored in the dataset. The faces have been detected (cf. Section III.A) and aligned using the facial landmarks. The input image size plays an important role in the training process since it can bring additional information and samples for the convolutional filters. Even though the system accuracy depends linearly on the image size, the computational resources grow quadratically. In our case, we have considered input images of size 224 × 224 pixels. Then, we applied batch normalization (BN) that solves the gradient exploding or vanishing problem and guaranties near optimal learning regime for the convolutional layers following the BN. Regarding the image batch size, this is always a tradeoff between the computational resources and the system accuracy. Experiments that keeping a constant learning rate for different min-batch sizes has a negative impact on the system'sperformance.Batchsizessuperiorto512orbatche s with single examples can lead to a significant decrease in performances. The learning rate is one of the most important hyper-parameter that needs to be adjusted when training deep neural networks, since it controls the weight variation in the direction of the gradient foramina -batch . In our case, we used for training 50k iterations, at a learning rate of 0.0001 and a batch size of 64. Based on transfer learning, the initialization of the CNN weights is performed using the pre-trained VGG face model that achieves state of the art results in face recognition tasks. Based on the observation of that copying all but last layer of the CNN is generally the best practice for fine tuning on new small data sets ,in our work we have trained only the last layer of the CNN. So, in the training stage only the weights of the final layer of the model are updated. After training, the CNN weights remain fixed.

7. Implementation and algorithm

The video is captured using a Camera which is then divided into a sequence of frames. Face recognition is done using Hear cascade classifiers and color based face recognition technique. Hear Cascade

Algorithm: The Open CV library in Python has functions specifically to detect faces. It delivers software packages that are used to train classifiers for their face recognition system, called Hear Training. Hear-like features: Face recognition using Hear feature based cascade classifiers is a machine learning based approach where a cascade function is trained from a lot of positive and negative images. It is then used to detect face in other images. The algorithm extracts images using a lot of positive and negative images. A Hear-like feature can be considered as a template of several white and black rectangles interconnected. The features value for the given mask is calculated as the weighted sum of intensity of the pixel intensities covered by the whole mask. But all the features extracted will not be useful for the purpose at hand.

8. Results

The proposed system focuses on detection of person. The system is made wearable and is portable. The system is mounted on the chest of the person. The Pi camera connected to Raspberry Pi captures the video of the scene and this is converted into frames by the processor.

9. Conclusion

In this paper we have presented a face-acknowledgment assistive gadget supposed DEEP-SEE FACE, intended to improve comprehension of outwardly impeded individuals while collaborating with different people in social experiences. The proposed methodology does not require any from the earlier learning about the situation of different individuals existent in the scene and mutually misuses PC vision calculations and profound convolutional neural systems (CNNs) so as to improve perception of VI clients. By utilizing the VGG CNNs design joined with locale proposition structure the framework that gets as information the whole video outline can accurately distinguish, track and perceive, progressively different people arranged at self- assertive areas. The semantic interpretation of the recognized person identity is transmitted to the VI user as a set of acoustic warnings.

From the methodological perspective, the center of the methodology depends on a novel video-based face acknowledgment structure ready to build a viable worldwide, fixed-measure face portrayal strategy, which is free of the length of the picture grouping. A weight adjustment conspire is proposed, ready to adaptively dole out a weight to each face example relying upon the video content variety. Furthermore, a hard negative mining stage is recommended that encourages us separate among known and obscure face characters. The exploratory assessment performed on an extensive dataset of 30 recordings obtained with the assistance of VI individuals approve the proposed procedure, which can restore an acknowledgment rate better than 92% in any case on the lighting conditions, face present or different kinds of movement existent in the scene. For further work and developments, visage to further expand the DEEP-SEE assistive gadget with extra functionalities that includes: illuminate the client when a perceived individual exists the clients field-of- see, route direction, crossing location or shopping help inside huge general stores. In addition, when taking a gander at the rising patterns in the cell phone industry, we can see that different constructors start to propose equipment models devoted to CNN applications. Inside this unique situation, let us notice the artificial knowledge chips as of late propelled by CEVA (e.g., NP4000) or Samsung (e.g., Exynos 9 Series 9810) at the Consumer Electronics Symposium (CES'2018). We hope that such advancements will allow us, in the ongoing future, to self-rulingly run the DEEP SEE FACE system on a cell phone gadget.

References

[1] J. L. Obermayer, W. T. Riley, O. Asif, and J. Jean Mary,‘‘College smoking cessation using cell phone text messaging,’’ J. Amer. College Health, vol. 53, no. 2, pp. 71–78, 2004.

[2] S. Haul, C. Meyer, G. Schorr, S. Bauer, and U. John, ‘‘Continuous individual support of smoking cessation using text messaging: A pilot experimental study,’’ Nicotine Tobacco Res., vol. 11, no. 8, pp. 915–923, 2009.

[3] D. Scherr, R. Zweiker, A. Kollmann, P. Kastner, G. Schreier, and F. M. Fruhwald, ‘‘Mobile phone-based surveillance of cardiac patients at home,’’ J. Telemedicine Telecare, vol. 12, no. 5, pp. 255–261, 2006.

[4] P. Rubel et al., ‘‘Toward personal eHealth in cardiology. Results from the EPI-MEDICS telemedicine project,’’ J. Electrocardiol., vol. 38, no. 4, pp. 100–106, 2005.

[5] S. C. Wangberg, E. Årsand, and N. Andersson, ‘‘Diabetes education via mobile text messaging,’’ J. Telemed. Telecare, vol. 12, no. 1, pp. 55–56, 2006.

[6] Give Andriana Mutiara,Gita Indah Hapsari and Ramanta Rijalul,”Smart Guide Extension for Blind Cane,” IEEE Int. Conf. Information and Communication Technologies,2016.

[7] Rainer Lienhart and Jochen Mandat, “An Extended Set of Hear-like Features for Rapid Objection Detection,” IEEE Int .Con .Image Processing, 2014.

[8] Pritpal Singh ,B.B.V.L.Deepak ,Tanjot Sethi and Meta Dev Prasad Murthy, “Real Time Object Detection and Tracking Using Color Feature and Motion,” IEEE Int .Conf,Communication and Signal Processing 2015.

[9] A.Rodríguez,J.J.Yebes,P.F.Alcantarilla,L.M.Bergas a,J.Almazán,and A. Cela, ‘‘Assisting the visually impaired: Obstacle detection and warning system by acoustic feedback,’’ Sensors, vol. 12, no. 12, pp. 17476–17496, 2012.

[10] R. Tapu, B. Mocanu, and T. Zaharia, ‘‘DEEP- SEE: Joint object detection, tracking and recognition with application to visually impaired navigational assistance,’’ Sensor, vol. 17, no. 11, p. 2473, 2017.