EyeHope

(A Real Time Emotion Detection Application)

Zulfiqar A. Memon(✉), Hammad Mubarak, Aamir Khimani, Mahzain Malik, and Saman Karim

Department of Computer Science, National University of Computer and Emerging Sciences (NUCES-FAST), Karachi, Pakistan
{zulfiqar.memon,k132113,k132072,k132004,k132071}@nu.edu.pk

Abstract. This paper describes about an android application named "EyeHope" whose sole purpose is to reduce the dependency level of the blind and visually impaired people. This application would help them decrease the communication gap by allowing them to perceive and identify the expressions of the person they are communicating with, whether the person is speaking or just listening to them. This would be an android based application. This paper also states about the real-time system communication with blind and visually impaired people using frames captured by mobile camera which would then be processed. Processing includes face detection and emotion detection using OpenCV. The output would be the emotion of the person whose image had been processed. This output would be converted from text to speech so that the blind person could listen to it through earphones connected to his/her phones. Its implementation and future enhancements will definitely be going to improve the life style of blind or visually impaired people by allowing them to effectively communicate with people around them.

Keywords: Emotion detection · Computer vision · Image processing · Assisting blind people

1 Introduction

Human senses are an essential part of human body through which a person can manage to survive in the current society. But in the case when a person is blind or visually impaired, the inability to identify the presence of people normally at home or during meetings becomes inconvenient for them [1, 29]. Real-time face recognition [4, 30], text recognition [2, 31] and object detection [5, 32] are some of the dominant developed applications. Machine learning employs algorithms that can find patterns from exemplars and make data driven predictions. Thus, computers have the ability to learn and act without being given explicit directions by mimicking the human cognitive framework of collecting and applying knowledge to make decisions [7, 33]. Furthermore, algorithms used for face detection are proposed by Voila and Jones [4, 34] and Eigenfaces (Principle Component Analysis) are used in order to identify gestures, facial expressions and emotions [3, 35]. The present work is focused on developing face

K. Arai et al. (Eds.): SAI 2018, AISC 858, pp. 615–623, 2019.
https://doi.org/10.1007/978-3-030-01174-1_47

detection and face recognition algorithms to be used by visually impaired people [1, 36]. The previous research also says that body movement can also become a part for detecting emotions of the humans. But the emotion communication through bodily expressions has been a neglected area for much of the emotion research history [6, 37]. Over the last two decades, researchers have significantly advanced human facial emotion recognition with computer vision techniques. Historically, there have been many approaches to this problem, including using pyramid histograms of gradients (PHOG) [8], AU aware facial features [9], boosted LBP descriptors [10], and RNNs [11, 38]. However, recent top submissions [12, 13, 39] to the 2015 Emotions in the Wild (EmotiW 2015) contest for static images all used deep convolution neural networks (CNNs). The classification of emotions is also a huge milestone to cover. As, in classification problems, good accuracy in classification is the primary concern; however, the identification of the attributes (or features) having the largest separation power is also of interest [14, 40, 51].

Below, Sect. 2 describes background work done by various researchers in the literature. Section 3 describes various attempts from the past to solve the emotion detection through software applications and how our approach is different from others. Section 4 describes our approach towards building the application and the workflow we used for developing the application. Section 5 describes the Methodology and tools and packages used to build the application and also the comparison between different tools. The section also explains the working of each feature of the application in detail. Section 6 includes experimental analysis conducted with different algorithms for calculating the accuracy and selection of algorithm. The section also provides an insight of our experiments and results derived from those experiments for building the application.

2 Literature Review

Understanding emotional facial expressions accurately is one of the determinants in the quality of interpersonal relationships. The more one reads another's emotions correctly, the more one is included to such interactions. The problems in social interactions are shown in some psychopathological disorders may be partly related to difficulties in the recognition of facial expressions. Such deficits have been demonstrated in various clinical populations. Nonetheless, with respect to facial expressions, there have been discrepant findings of the studies so far [15, 23, 41, 52]. The process of emotion recognition involves the processing images and detecting the face then extracting the facial feature. Facial Expression Recognition consists of three main steps. In first step face image is acquired and detect the face region from the images and pre-processes the input image to obtain image that have a normalized size or intensity. Next is expression features are extracted from the observed facial image or image sequence. Then extracted features are given to the classifier and classifier provides the recognized expression as output [16, 24, 42]. In the past researches, it is also stated that Principal Component Analysis [17, 43], Local Binary Pattern (LBP) [18, 44], Fisher's Linear Discriminator [19, 45] based approaches are the main categories of the approaches available for feature extraction and emotion recognition [16, 28, 46]. There are various descriptors and techniques used in facial expression recognition like the Gradient faces, local features, local binary pattern (LBP), local ternary pattern (LTP), local directional pattern (LDiP) and Local derivative pattern (LDeP) [20, 47].

Many studies on facial expression recognition and analysis have been carried out for a long time because facial expressions play an important role in natural human-computer interaction as one of many different types of nonverbal communication cue [21, 48]. Paul Ekman et al. postulated six universal emotions (anger, disgust, fear, happiness, sadness, and surprise), and developed Facial Action Coding System (FACS) for taxonomy of facial expressions [22, 27, 49].

3 Approach

Advance technology is an integral part of the twenty first century. Initially it was thought that the fascinating ideas of upcoming scientists are reviving the lifestyle of those who are physically well and want to make their life even more convenient. There should be something for those who are already bereaved from their physical senses due to some reasons. There should be some technological initiatives which will pull them out from the sphere of blindness and will guide them towards self-reliant and independency. The purpose behind developing this application is to use the current advanced technological resources so that blind and visually impaired people can live their life conveniently and independently. Our Analyst team specifically targeted and started taking interviews from people related to different societal classes (such as literate or illiterate, employees or managers, etc.) who are blind or visually impaired. After concluding their answers and findings, it was observed that the most common and foremost problem they face in their daily routine is that they face difficulty while inferring about the number of people in front of them on a particular moment as well as they face hurdles while interpreting the expressions/emotions of people which are communicating with them on that moment. Most of the interviewee's said that when the audience in front of them is verbally communicating (talking), the blind one can barely judge what expressions currently are on the faces of audience. But when that verbal communication between them stops, it is quite hard to judge the expressions of that audience. These minor but daily occurring difficulties motivated us to develop something that could help blind or visually impaired people to overcome these hurdles.

Now days, each and every individual whether he/she is a youngster or older, owns a smart phone. A global research says that there are 2.32 billion people who owns and know how to use a smart phone [25, 26, 50]. Concluding all these findings and researches, it was decided that there should be an application developed for blind and visually impaired people, which will help them to interpret number of people in front of them as well as acknowledge them with the emotion on the faces of each detected person.

After analyzing the targeted user problems and setting project goal, technical tools, platform on which application should be built, algorithms which should be used for getting good results were discussed and implemented in detailed manner which is briefly described in next section.

4 Methodology

In the global market, there is 80% of the population which chooses android as their Smartphone platform. From this fact, it was decided to keep our application android based. Not only this, android also provides best prices to fit customer needs. Higher level of customization, multi-tasking and Google integration makes android the most preferred platform among consumers.

As EyeHope is a computer vision related software system, OpenCV was suggested by many of the experienced professionals. OpenCV (Open source Computer Vision Library) is an efficient open source computer vision and ML (Machine Learning) library that provides a common infrastructure for computer vision based applications. OpenCV contains approximately 2500 optimized and efficient algorithms. These algorithms can be used to perform multiple computer vision related tasks. OpenCV has its user community which is based on more than 47 thousand users as well as its number of download is exceeding a count of 14 million. As the matter of its usage and popularity, the companies such as Google, Microsoft, IBM, Intel, Yahoo, Sony and many more have used OpenCV for their startup products. To acquire a commercial single user license for Matlab it would cost USD 2150 whereas the OpenCV's BSD license is free and easily available for users. These are only physical or social measures of OpenCV but what about its technical benefits? Most of the computer vision related projects or applications are confused between OpenCV and Matlab. The reasons behind choosing OpenCV as first priority of these applications are that if we compare these developmental tools, on the basis of speed, MATLAB is built on Java, and Java is built upon C. So, when you run a MATLAB program, your computer is busy trying to interpret all that MATLAB code. Then it turns it into Java, and then finally executes the code. OpenCV, on the other hand, is basically a library of functions written in C/C++. You are closer to directly providing machine language code to the computer to get executed. So ultimately you get more image processing done for your computers processing cycles, and not much interpreting. In the case of computer vision related programs, From Matlab, we would get 3 to 4 frames analyzed per second whereas from OpenCV, we would get at least 20 to 30 frames analyzed per second which result in real-time detection.

If these tools are compared on the basis of the resources it need, then due to the high-level nature of MATLAB, it uses a lot of your systems resources. MATLAB code requires over a gig of RAM to run. In comparison, typical OpenCV programs only require $\sim$70 mb of RAM to run in real-time.

There are several algorithms which are used for computer vision and image processing related systems. But the algorithm which best suit's in the case of EyeHope application is Fisher Face algorithm. When the goal of application is classification rather than representation, the eigenfaces/least-squares may not provide the best desirable results. For this, it is necessary that there should be a subspace that maps the sample vector of same class in a single point of feature representation. To find such subspace, the technique mostly used is known as Linear Discriminant Analysis. When linear discriminant analysis technique is used to find the subspace representation of a set of face images, the consequential basis vector defining the space is called fisher faces.

For computing Fisherfaces, it is to be assumed that each of the class is normally distributed. The multivariate normal distribution is denoted by Ni(μi, Σi), mean is denoted by μi, covariance matrix by Σi and its probability density function by fi(x|μi, Σi). In this case, we have more than two classes for which its procedure will minimize within class differences and maximize between class distances. For the class differences computed using the within class scatter matrix is given by:

$$Sw = \sum Cj = 1 \sum nji = 1(xij - \mu j)(xij - \mu j)T$$

Whereas between class differences are estimated using the between class scatter matrix is given by:

$$Sb = \sum Cj = 1(\mu j - \mu)(\mu j - \mu)T$$

Where, μ shows the mean of all classes. The above computations results in generalized eigenvalue decomposition which is represented by:

$$SbV = SwV\Lambda$$

Where V represents matrix of eigenvectors and Λ represents diagonal matrix of consequent eigenvalues. The eigenvector of V related to non-zero eigenvalues are the Fisherfaces.

One of the basic advantages of fisherface algorithm is that it is used for classification purpose and is quite faster than some of the existing algorithms. It uses discriminable linear projection model.

For the development of the application, first step was to configure and integrate OpenCV with android which required a bit research and tutorials for its successful completion. The next step was to create user friendly interface. The colors, font size, design and sound quality of application was properly checked so that the users (blinds and visually impaired people) can operate the application in a very convenient way. After this, the first feature of the application (detecting number of people adjusting in the frame of mobile screen) was focused in a very detailed manner. Specifically, Haar cascade classifiers were collected experimented and were deployed on the application. The face detection feature was successfully completed and then the focused was shifted towards emotion detection feature. For this, initially a large amount of dataset was required. This application was trained on both internal (self-made) and external (image repositories) datasets. After the dataset collection, the image data was divided for training and testing purposes. Application was trained and tested several times for getting accurate results. Several experiments were performed on the application (specifically on emotion detection feature) to increase its accuracy and to get desired results.

Finally, the application was tested by users in their respective environment.

5 Experiments

After developing the application using methodology described above, our focus was to increase the accuracy to the maximum level. The dataset we used to train and test the machine was Cohn Kanade Dataset which consists of the images of 123 subjects with 8 emotions. Using this dataset, we achieved an accuracy of 67%. Handling the noises in dataset and over-fitting of different classifiers our accuracy was close to 73.7%.

We then worked on ways to improve the accuracy of our application. Through in-depth analysis and research, we found out that FisherFace classifier usually performs better than the results we acquired. So, we switched our focus on the dataset we were working on. Our team paid a visit to community schools and other places to collect our own realistic dataset in order to help the algorithm correctly classify the emotions. Our dataset was diversified as it includes subjects belonging to almost all age groups. We then filtered the dataset to remove any discrepancies. We designed the final dataset by cleansing and then merging both of the datasets, i.e. the Cohn Kanade dataset and our own dataset. We tested our application on the combination of some emotions, so as to know which emotions are classified better and which of them need improvements. We acquired the following results (Table 1):

Table 1. Accuracy results of the combination of some different emotions

Emotions	Accuracy achieved
Neutral and happy	96%
Neutral and disgust	90%
Neutral and angry	80%
Neutral and fear	82%
Neutral and surprise	88%
Neutral and sad	80%

After many optimizations and observing our application by applying various approaches, our final accuracy was around 85.53%. This reflects that the emotions detected by EyeHope application are more likely to match the interpretation of a normal human observer.

6 Conclusion

Our aim was to create a solution to assist blind people and help them communicate effectively using the technological advancements. We made use of OpenCV library for Android and Android Studio to develop an Android application. This application works real time. It captures, on average, 30 frames per second, from the mobile camera and uses Haar Cascade Classifiers to detect the frontal face and crops it. The picture is then converted to gray scale and tested on the set of emotions. We used the FisherFace Classifier provided by the OpenCV for Machine Learning. Moreover, the dataset we used is a merger of Cohn Kanade dataset and a dataset prepared by us. Both of the datasets were filtered and cleaned prior to applying the machine learning techniques. The application was tested on a set of emotions and was optimized constantly to improve the accuracy. We are able to achieve around 85.53% accuracy in terms of correctly classifying the emotions in real time environment. We believe this project is an important step in creating a sense of self reliability and sustainability among the disabled community.

7 Future Enhancements

The system of EyeHope application can be enhanced by integrating more features in it such as, text detection by processing the frames having any kind of text in it, describing the scenario in the frame, face recognition, object detection and object recognition. The system can also be enhanced by modifying or polishing the current features like, the frontal face detection feature can be extended by detecting the face even if camera is facing only left side of the face or only right side of the face and settings module can also be modified to implement more filtering options.

References

1. Chillaron, M., et al.: Face detection and recognition application for Android (2015). https://www.researchgate.net/publication/284253914_Face_detection_and_recognition_application_for_Android
2. Bourlard, H., Chen, D., Odobez, J.-M.: Text detection and recognition in images and video frames. Pattern Recognit. **37**, 595–608 (2004)
3. Krishna, Sh., et al.: A wearable face recognition system for individuals with visual impairments, pp. 106–113 (2005)
4. Viola, X., Jones, M.: Robust real-time face detection. Int. J. Comput. Vis. **57**, 137–154 (2004)
5. Yuille, A.L., Chen, X.: A Time-Effect Cascade for Real-Time Object Detection: With Applications for the Visually Impaired (2005)
6. Konar, A., Chakraborty, A.: Emotion Recognition: A Pattern Analysis Approach (2014)
7. Cosgrove, C., Li, K., Lin, R., Nadkarni, S., Vijapur, S., Wong, P., Yang, Y., Yuan, K., Zheng, D.: Developing an Image Recognition Algorithm for Facial and Digit Identification (1999)
8. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS, vol. 1, p. 4 (2012)
9. Yao, A., Shao, N.M., Chen, Y.: Capturing AU-aware facial features and their latent relations for emotion recognition in the wild. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, ICMI 2015, New York, NY, USA, pp. 451–458. ACM (2015)
10. Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on local binary patterns: a comprehensive study. Image Vis. Comput. **27**(6), 803–816 (2009)
11. Kahou, S., Michalski, V., Konda, K., Memisevic, R., Pal, C.: Recurrent neural networks for emotion recognition in video. In: ICMI, pp. 467–474 (2015)

12. Yu, Z., Zhang, C.: Image based static facial expression recognition with multiple deep network learning. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, ICMI 2015, New York, NY, USA, pp. 435–442. ACM (2015)
13. Kim, B., Roh, J., Dong, S., Lee, S.: Hierarchical committee of deep convolutional neural networks for robust facial expression recognition. J. Multimodal User Interfaces **10**, 1–17 (2016)
14. Visa, S., Ramsay, B., Ralescu, A., van der Knaap, E.: Confusion matrix-based feature selection (2011)
15. Dursun, P., Emül, M., Gençöz, F.: A review of the literature on emotional facial expression and its nature (2010)
16. Raval, D., Sakle, M.: A literature review on emotion recognition system using various facial expression (2017)
17. Sagarika, S.S., Maben, P.: Laser face recognition and facial expression identification using PCA. IEEE (2014)
18. Happy, S.L., George, A., Routray, A.: A real time facial expression classification system using local binary patterns. IEEE (2012)
19. Bhadu, A., Tokas, R., Kumar, V.: Facial expression recognition using DCT, gabor and wavelet feature extraction techniques. Int. J. Eng. Innov. Technol. **2**(1), 92 (2012)
20. Bele, P.S., Mohod, P.S.: A literature review on facial and expression recognition with advanced image descriptor template (2015)
21. Suk, M., Prabhakaran, B.: Real-time Mobile Facial Expression Recognition System – A Case Study. Department of Computer Engineering, The University of Texas at Dallas, Richardson (2014)
22. Ekman, P., Friesen, W.: Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto (1978)
23. https://karanjthakkar.wordpress.com/2012/11/21/what-is-opencv-opencv-vs-matlab/2017
24. https://www.researchgate.net/post/Which_is_the_best_opencv_or_matlab_for_image_processing/2017
25. https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/2017
26. http://topmobiletrends.com/apple-vs-android/2017
27. Lyons, M.J., Akemastu, S., Kamachi, M., Gyoba, J.: Coding facial expressions with gabor wavelets. In: 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200–205 (1998)
28. Kanade, T., Cohn, J.F., Tian, Y.: Comprehensive database for facial expression analysis. In: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG 2000), Grenoble, France, pp. 46–53 (2000)
29. Bosse, T., Duell, R., Memon, Z.A., Treur, J., van der Wal, C.N.: Computational model-based design of leadership support based on situational leadership theory. Simul. Trans. Soc. Model. Simul. Int. **93**(7), 605–617 (2017)
30. Bosse, T., Duell, R., Memon, Z.A., Treur, J., van der Wal, C.N.: Agent-based modelling of emotion contagion in groups. Cognit. Comput. J. **7**(1), 111–136 (2015)
31. Hoogendoorn, M., Klein, M.C.A., Memon, Z.A., Treur, J.: Formal specification and analysis of intelligent agents for model-based medicine usage management. Comput. Biol. Med. **43** (5), 444–457 (2013)
32. Bosse, T., Memon, Z.A., Treur, J.: A cognitive and neural model for adaptive emotion reading by mirroring preparation states and hebbian learning. Cognit. Syst. Res. J. **12**(1), 39–58 (2012)
33. Bosse, T., Hoogendoorn, M., Memon, Z.A., Treur, J., Umair, M.: A computational model for dynamics of desiring and feeling. Cognit. Syst. Res. J. **19**, 39–61 (2012)

34. Bosse, T., Memon, Z.A., Treur, J.: A recursive BDI-agent model for theory of mind and its applications. Appl. Artif. Intell. J. **25**(1), 1–44 (2011)
35. Bosse, T., Memon, Z.A., Oorburg, R., Treur, J., Umair, M., de Vos, M.: A software environment for an adaptive human-aware software agent supporting attention-demanding tasks. Int. J. Artif. Intell. Tools **20**(5), 819–846 (2011)
36. Memon, Z.A.: Designing human-awareness for ambient agents: a human mindreading perspective. J. Ambient Intell. Smart Environ. **2**(4), 439–440 (2010)
37. Memon, Z.A., Treur, J.: On the reciprocal interaction between believing and feeling: an adaptive agent modelling perspective. Cognit. Neurodynam. J. **4**(4), 377–394 (2010)
38. Kashif¸ U.A., Memon, Z.A., et al.: Architectural design of trusted platform for IaaS cloud computing. Int. J. Cloud Appl. Comput. (IJCAC) **8**(2), 47 (2018)
39. Laeeq, K., Memon, Z.A., Memon, J.: The SNS-based e-learning model to provide smart solution for e-learning. Int. J. Educ. Res. Innov. (IJERI) **10**, 141 (2018)
40. Samad, F., Memon, Z.A., et al.: The future of internet: IPv6 fulfilling the routing needs in Internet of Things. Int. J. Future Gener. Commun. Netw. (IJFGCN) **11**(1) (2018, in Press)
41. Memon, Z.A., Samad, F., et al.: CPU-GPU processing. Int. J. Comput. Sci. Netw. Secur. **17** (9), 188–193 (2017)
42. Samad, F., Memon, Z.A.: A new design of in-memory file system based on file virtual address framework. Int. J. Adv. Comput. Sci. Appl. **8**(9), 233–237 (2017)
43. Memon, Z.A., Ahmed, J., Siddiqi, J.A.: CloneCloud in mobile cloud computing. Int. J. Comput. Sci. Netw. Secur. **17**(8), 28–34 (2017)
44. Abbasi, A., Memon, Z.A., Jamshed, M., Syed, T.Q., Rabah, A.: Addressing the future data management challenges in IOT: a proposed framework. Int. J. Adv. Comput. Sci. Appl. 8(5), 197–207 (2017)
45. Waheed-ur-Rehman, Laghari, A., Memon, Z.: Exploiting smart phone accelerometer as a personal identification mechanism. Mehran Univ. Res. J. Eng. Technol. **34**(S1), August 2015, pp. 21–26 (2015)
46. Memon, Z.A., Treur, J.: An agent model for cognitive and affective empathic understanding of other agents. Trans. Comput. Collect. Intell. (TCCI) **6**, 56–83 (2012)
47. Duell, R., Memon, Z.A., Treur, J., van der Wal, C.N.: Ambient support for group emotion: an agent-based model. In: Agents and Ambient Intelligence, Ambient Intelligence and Smart Environments, vol. 12, pp. 261–287. IOS Press (2012)
48. Hoogendoorn, M., Memon, Z.A., Treur, J., Umair, M.: A model-based ambient agent providing support in handling desire and temptation. Adv. Intell. Soft Comput. **71**, 461–475 (2010)
49. Batra, R., Memon, Z.A.: Effect of icon concreteness, semantic distance and familiarity on recognition level of mobile phone icons among e-literate and non e-literates. Int. J. Web Appl. **8**(2), 55–64 (2016)
50. Siddiqi, S.S., Memon, Z.A.: Internet addiction impacts on time management that results in poor academic performance. In: Proceedings of the 14th International Conference on Frontiers of Information Technology, pp. 63–68. IEEE Computer Society (2016)
51. Laghari, A., Waheed-ur-Rehman, Memon, Z.A.: Biometric authentication technique using smartphone sensor. In: Proceedings of the 13th International Conference on Applied Sciences & Technology, IBCAST 2016, pp. 381–384 (2016)
52. Kashif, U.A., Memon, Z.A., Balouch, A.R., Chandio, J.A.: Distributed trust protocol for IaaS cloud computing. In: Proceedings of the 12th International Conference on Applied Sciences & Technology, IBCAST 2015, pp. 275–279 (2015)