Current Journal of Applied Science and Technology

40(20): 7-16, 2021; Article no.CJAST.71827
ISSN: 2457-1024
(Past name: British Journal of Applied Science & Technology, Past ISSN: 2231-0843, NLM ID: 101664541)

Facial Expression Recognition Using Python Using CNN Model

Akash Kumar[1], Athira B. Nair[1], Swarnaprabha Jena[1], Debaraj Rana[1*] and Subrat Kumar Pradhan[1]

[1]Deptartment of Electronics and Communication, Centurion University of Technology and Management Bhubaneswar, India.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/CJAST/2021/v40i2031459
Editor(s):
(1) Ming-Chih Shih, Chinese Culture University, China.
Reviewers:
(1) Satyasis Mishra, Adama Science and Technology University, Ethiopia.
(2) Tanveer Aslam, Muhammad Nawaz Sharif University of Agriculture, Pakistan.
Complete Peer review History: https://www.sdiarticle4.com/review-history/71827

Original Research Article

Received 15 June 2021
Accepted 20 August 2021
Published 24 August 2021

ABSTRACT

Facial expressions are a vital part of human life. Each day has a number of instances and all instances include numerous amounts of communication. Every communication expressed with emotion tells us about the state of the person. The interpersonal as well as security purposes are solved through facial expressions. The mischievous intention of a person can be caught by his expressions.

The human mind can capture visual information faster. So, a machine recognizing it will be a challenge. As the saying goes- "A picture is worth a thousand words"- only when it is represented well. A machine being able to detect the atmosphere by the means of expression is less of a manual work.

This paper detects the faces, extract the features as well classify them into different categories which ultimately lead to expression recognition. We evaluate our proposed method with the dataset which we used and the recall of angry, fear, happy, neutral, sad, and surprise is 60%, 31%, 84%, 22%, 57% and 58% respectively and the f1-score is 51% 35%, 82%, 25%, 51% and 64% respectively. Experimental results demonstrate the competitive classification of our proposed system.

**Corresponding author: E-mail: debaraj.rana@cutm.ac.in;*

Keywords: Facial expression recognition; computer vision; neural networks.

1. INTRODUCTION

Judging mental state of a person is one of the difficult tasks. The Best way to understand an emotional state of a person is through facial expressions like happy, sad, fear, disgust, surprise and anger [1-6]. The automated analysis of facial expression (FER) is a challenging task in the field of computer vision. Its implementation is not restricted to mental state identification only, it is also applicable in the security domain, automatic counselling systems, face expression synthesis, lie detection, music for mood, automated tutoring systems, operator fatigue detection etc [7-13].

Facial expression is a major non-verbal means of expecting intentions in human communication. With the rapid development of artificial intelligence, automatic recognition of facial expressions has been intensively studied in recent years [14-20]. The study of Facial Expression Recognition (FER) has received extensive attention in the fields of psychology, computer vision, and pattern recognition. FER has broad applications in multiple domains, including human–computer interaction, virtual reality, augmented reality, advanced driver assistance systems, education, and entertainment [20-26].

This paper presents a quick survey of facial expression recognition. With the transition of facial expression recognition (FER) from laboratory-controlled to in-the-wild conditions and the recent success of deep learning in various fields, deep neural networks have increasingly been leveraged to learn discriminative representations for automatic FER

2. TECHNIQUE USED

2.1 Haar-Cascade

An algorithm which is used to identify the objects in an image or a video. It is a machine learning based approach where a lot of positive and negative images are used to train the classifier. It is then used to detect objects in other images.As haar cascade classifier allows real time detection so here we used haar cascade classifier to detect the face of human being to recognize their facial expressions.

Initially, the algorithm needs a lot of positive images and negative images to train the classifier. The positive images consist of images with the human faces. The negative images are the one which do not contain any faces Then we have to extract features from it. They are just like the convolutional kernel. The sum of pixels under the white and black rectangle are calculated respectively. Then each feature is obtained as a single value by subtracting the former from the latter.

2.2 Facial Pre-Processing

The image techniques use spatial information. The goal of this section is to understand some of the existing techniques out there that can be used to improve the current state for Facial Expression Recognition and applications that deal with facial expressions.

We trained and tested our models on the data set from the Kaggle Facial Expression Recognition Challenge, which comprises 48-by-48-pixel grayscale images of human faces, each labelled with one of 6 emotion categories: anger, fear, happiness, sadness, surprise, and neutral. We used a training set of 28,273 examples, and a validation set of 3,534 examples.

2.3 Face Detection

Computer vision, when all is said in done, means to double (or in a few cases compensate) human vision, and customarily, have been utilized as a part of performing routine monotonous undertakings, for example, classification in monstrous mechanical production systems. Today, scrutinize on machine vision is spreading gigantically so it is very nearly difficult to organize every last bit of its subtopics. Notwithstanding of this, one can rundown important a few provisions, for example, face processing (i.e. gesture recognition and face expression), machine human cooperation, swarm reconnaissance, and substance-based picture recovery. All of the applications stated above require detection of face, which can be simply viewed as a pre-processing step for obtaining the “object”. The face is our primary centre of consideration in social life assuming an imperative part in passing on feelings and character. We can perceive various appearances adapted all around our lifespan and distinguish faces considerably after numerous years of division.

2.4 CNN

A Convolutional Neural Network is a multi-layer neural network which has a special architectural design of detecting complex features in data. CNN has a great hand in image recognition, self-driving vehicles and powering vision in robots. CNN is used to classify the contents of different images. We need to load the images. CNN works exactly like a human brain. It classifies and identifies just like the human brain classifies and defines objects.

The Architecture of the model given below:

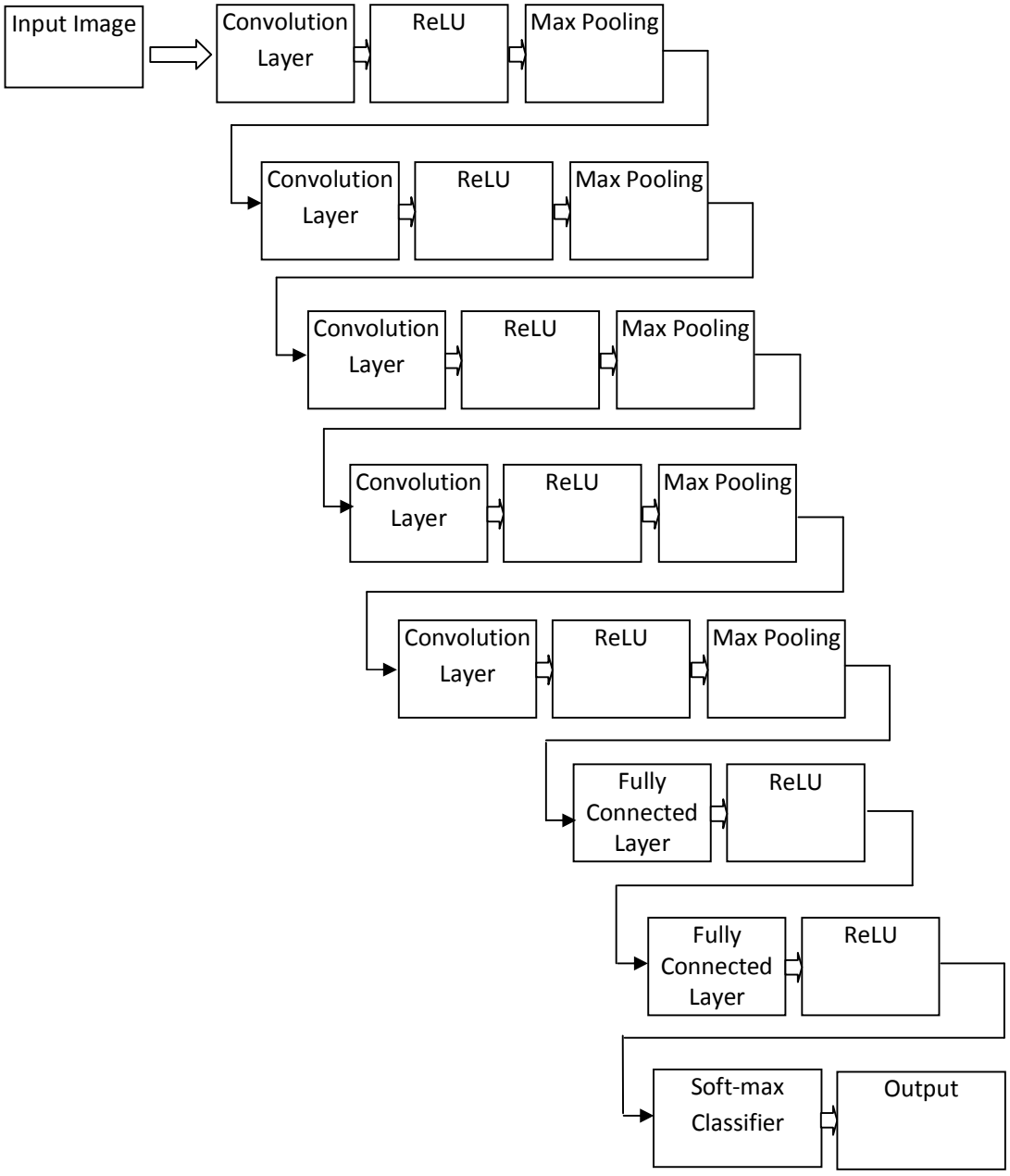


Fig. 1. Architecture of CNN model

The architecture of the proposed model consists of 8 layers where the first 5 layers are consecutive convolutional layers and max pooling layer through a ReLU function for feature extraction and feature map selection.. These 5 layers are followed by 2 numbers of fully connected layer with ReLU function for learning of the model. The final 8^{th} layer is the SoftMax classifier for the classification or the recognition purposes. The softmax classifier generally uses the cross-entropy loss, the name itself derived from the Softmax function

2.5 VGG Face

We implemented a first-pass CNN with a fixed depth of five convolutional layers. The model was trained using the architecture outlined and was trained using the following characteristics.

- Parametrized dropout rate, learning rate, and l2 regularization
- Batch normalization (optional) after each layer
- Adam update rule
- Weight initialization for using ReLU nonlinearities as presented by He et al.
- 3x3 convolutional filters with stride 0 and zero padding to preserve spatial size
- 2x2 max pools with a stride of 0

3. PACKAGES AND LIBRARIES USED

Tensor Flow is a free and open-source software library for machine learning. It can be used across a range of tasks but has a particular focus on training and inference of deep neural networks. TensorFlow is a symbolic math library based on dataflow and differentiable programming.

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together.

Keras is a minimalist Python library for deep learning that can run on top of Theano or TensorFlow. It was developed to make implementing deep learning models as fast and easy as possible for research and development. It runs on Python 2.7 or 3.5 and can seamlessly execute on GPUs and CPUs given the underlying frameworks. It is released under the permissive MIT license. We have built and trained a CNN in Keras from the scratch just to recognise the kind of expression portrayed. The data consists of 48x48 pixel grayscale image of faces. The aim is to classify among the most common kind of emotions which includes Neutral, Happy, Sad, Angry, Surprise, Fear.

Computer vision tasks include methods for acquiring, processing, analysing and understanding digital images, and extraction of high-dimensional data from the real world in order to produce numerical or symbolic information, e.g., in the forms of decisions. OpenCV, which is an image and video processing library with bindings in C++, C, Python, and Java. OpenCV is used for all sorts of image and video analysis, like facial recognition and detection, license plate reading, photo editing, advanced robotic vision, optical character recognition, and a whole lot more.

Matplotlib isa plotting library forthe Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK+. There is also a procedural "pylab" interface based on a state machine (like OpenGL), designed to closely resemble that of MATLAB. SciPy makes use of Matplotlib.

Scikit-learn (formerly scikits. learn and also known as sklearn) is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, *k*-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

NumPy, which stands for Numerical Python, is a library consisting of multidimensional array objects and a collection of routines for processing those arrays. Using NumPy, mathematical and logical operations on arrays can be performed. This tutorial explains the basics of NumPy such as its architecture and environment. It also discusses the various array functions, types of indexing, etc. An introduction to Matplotlib is also provided.

NumPy is often used along with packages like SciPy (Scientific Python) and Mat−plotlib (plotting library). This combination is widely used as a replacement for MATLAB, a popular platform for technical computing. However, Python alternative to MATLAB is now seen as a more modern and complete programming language.

Regular expression is a special sequence of characters that helps you match or find other strings or sets of strings, using a specialized syntax held in a pattern. ... The Python module re provides full support for Perl-like regular expressions in Python. The re module raises the exception re.

OS module in python provides functions for interacting with the operating system. OS, comes under Python's standard utility modules. This module provides a portable way of using operating system dependent functionality. The *os* and *os. path* modules include many functions to interact with the file system.

Time module provides various time-related functions. For related functionality, see also the datetime and calendar modules. Although this module is always available, not all functions are available on all platforms. Most of the functions defined in this module call platform C library functions with the same name. It may sometimes be helpful to consult the platform documentation, because the semantics of these functions varies among platforms.

4. METHODOLOGY

In this part, an architecture-based approach and the methods we used for the completion of our facial expression model explained meticulously. For building the model which can detect a face, we used CNN- Convolution Neural Network. CNN is used to identify various objects of the input images to differentiate it from other. In CNN we used little vgg which uses 16 layers.

The dataset comprises of two parts: 1. Train 2. Validation.

The entire dataset comprises of 30,000 images with train and validation having 75% and 25% respectively. The dataset consists of folders.

Fig. 2. Sample data set

Table 1. Number of input images with male female ratio and age difference

Input images	Male	Female	Age difference
30,000	17,000	13,000	60 yrs

(i) MODEL BUILD:

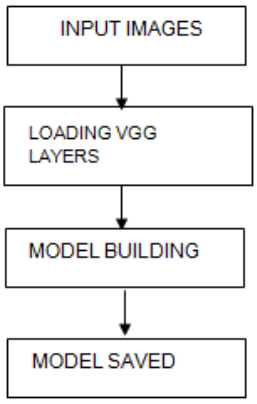


Fig. 3. Flow diagram for model design

(ii) MODEL IMPLEMENTATION:

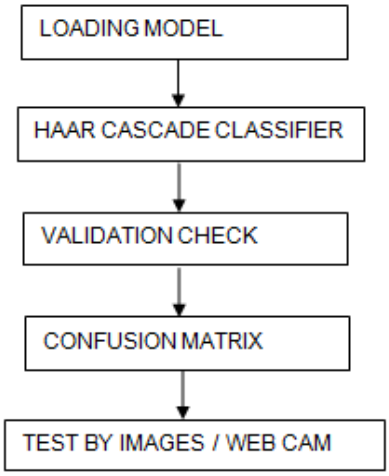


Fig. 4. Flow diagram for model implementation

4.1 Confusion Matrix

A confusion matrix is a term used in Machine Learning; also known as the error table. It is a tabular layout to elaborate the performance of the algorithm. We have used it to know the performance on prediction of different expressions defined by the classifier for a set of data provided.

		Predicted Values	
		Positive (1)	Negative (0)
Actual Values	Positive(1)	TP	FN
	Negative(0)	FP	TN

Fig. 5. Block diagram of confusion matrix

True Positive: The classifier predicted positive and it is actually true. The picture of a happy person is predicted happy and it is actually happy.

True Negative: The classifier predicted negative and it is true. The picture of a sad person is predicted sad and it is true.

False Positive: The classifier predicted positive and it is false. The picture of an angry person is predicted surprise and it is false.

False Negative: The classifier predicted negative and it is false. The picture of an angry person is predicted sad and it is false.

5. RESULT ANALYSIS AND DISCUSSION

The below image shows the confusion matrix obtained for the set of images we trained.

The matrix denotes the accuracy of each expression in the diagonal pattern. The diagonal from the top-left to bottom-right defines the accuracy values i.e., of correctly identified images by the classifier. The diagonal from the top-right to the bottom-left defines the ignored images by the classifier. In between lies the images that are confused with probability of half.

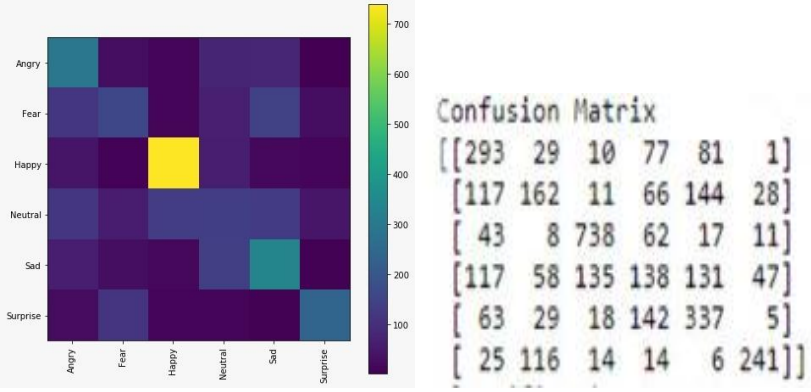
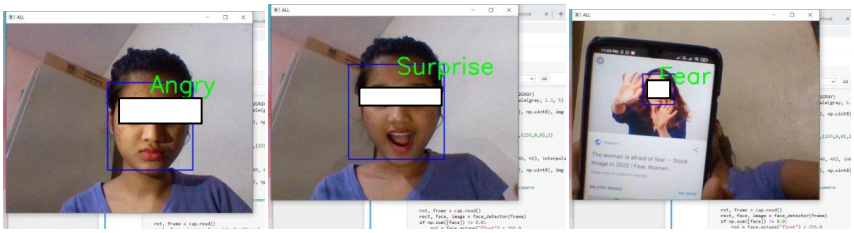


Fig. 6. Confusion matrix

Table 2. Classification report

Classification Report				
	Precision	**Recall**	**Fil-score**	**Support**
Angry	0.45	0.60	0.51	491
Fear	0.48	0.31	0.35	526
Happy	0.88	0.84	0.82	879
Neutral	0.28	0.22	0.25	626
Sad	0.47	0.57	0.51	594
Surprise	0.72	0.58	0.64	416
Accuracy			0.54	3534
Macro avg	0.52	0.52	0.51	3534
Weighted avg	0.53	0.54	0.53	3534

[a] [b] [c]



[d] [e] [f]

Fig. 7. (a)-(f) Some result for facial recognition

The colour of the graph is coordinated with the values with the help of the bar alongside the matrix in Fig. 6. The more the values in the left-to-right diagonal results in higher accuracy as well as better prediction of the expressions. The less the values in the right-to-left diagonal will results to more accuracy as the classifier have ignored less image.

Similarly, The less the values in the left-to-right diagonal results in lower accuracy as well as weak prediction of the expressions. The more the values in the right-to-left diagonal will results to pale accuracy as they have ignored more images.

As per the classification report precision was maximum 0.80 in case of happy, where as recall was maximum of 0.84 in the same happy expression. Similarly the happy emotion has high f1 score of 0.82, where as Natural expression has lowest precision, recall and f1 score as 0.28, 0.22, and 0.25 respectively.

This section deals with the outcome of all the trials and tests done. In this, we need to present the effectiveness of the system whether it actually works perfect with recognizing the expressions or not. The facial expressions are basically categorized into: **Neutral, Happy, Sad, Angry, Surprise, Fear**. The dataset consists of multiple photos of a person who oculd bring out the best of emtions of all the above mentioned category. The dataset consists of total **30,000** images. The dataset is divided into **train** and **validation;** where the former consists **27000** and later consists of **3000** images. We have taken 35 epochs and have achieved an accuracy of 53%. The results can be further improvised taking total of 280 epoch to receive an accuracy higher than 65%. The categories are explained below (Fig. 7).

5. CONCLUSION AND FUTURE SCOPE

Facial expressions play an important role in finding the roots of causes and issues in our day-to day life. In the earlier era, we had thick fat registers to store details. This way was totally manual where the technology hadnt rise high. Then we had this whole new technology of CCTV cameras and then the launch of biometric systems. In the near future, the most widely used technology will be th eface recognition nad feature classification.

- While the truck drivers drive night long for continuous days, they need absolute rest during the day time. In cases if they don't, this may cause mishaps to happen. At the poll gates we can have this facial expression system which will detect the fatigued weary expression of the driver and prevent the mishaps.
- The facial expression recognition can be used for ASD-Autism Specific Disorder; where the person suffering cannot express their emotions due to some abnormalities they face; hence this Machine learning sytem comes into use.

DISCLAIMER

The products used for this research are commonly and predominantly use products in our area of research and country. There is absolutely no conflict of interest between the authors and producers of the products because we do not intend to use these products as an avenue for any litigation but for the advancement of knowledge. Also, the research was not funded by the producing company rather it was funded by personal efforts of the authors.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Shima Alizadeh, 11.88. Stanford University. Convolutional Neural Networks for Facial Expression Recognition; 2017.

2. Shan C, Gong S, McOwan PW. Robust facial expression recognition using local binary patterns. In Image Processing, 2005. ICIP 2005. IEEE International Conference on. IEEE. 2005;2:II-370.
3. Matusugu Masakazu, Katsuhiko Mori, Yusuke Mitari, Yuji Kaneda. Subject independent facial expression recognition with robust face detection using a convolutional neural network" (PDF). Neural Networks. 2003; 16(5).
4. Raghuvanshi Arushi, Vivek Choksi. Facial expression recognition with convolutional neural networks. Stanford University; 2016.
5. Alizadeh Shima, Azar Fazel. Convolutional neural networks for facial expression recognition. Stanford University; 2016.
6. YI Tian, T Kanade, JF Cohn. Recognizing action units for facial expression analysis IEEE Transactions on pattern analysis and machine intelligence.
7. B. Graham. Fractional max-pooling; 2015.
8. Omkar AV, Parkhi M, Zisserman A. Deep face recognition; 2015.
9. Xiaoming Zhao, Xugan Shi, Shiqing Zhang. Facial expression recognition via deep learning, IETE Technical Review. 2015;32(5):347-355.
10. Kołakowska A, Landowska A, Szwoch M, Szwoch W, Wrobel MR. ´ Human-computer systems interaction: Backgrounds and applications 3, ch. Emotion Recognition and Its Applications, Cham: Springer International Publishing. 2014;51–62.
11. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks, NIPS. 2012;1(4).
12. Chibelushi CC, Bourel F. Facial expression recognition: A brief tutorial overview. CVonline: On-Line Compendium of Computer Vision, 9; 2003.
13. Schmedhuber J. Deep Learning in convolutional neural networks: An summary, Neural Networks. 2015;61:85–117.
14. Balekrishnan A, Reige A. Recognizing Emotions from Speech mistreatment Deep Neural Networks; 2017.
15. Rusehforth CK. Emotion Detection mistreatment Linear prophetical cryptography," IEEE Trans. Acoust. 1977;25(5):361–367.
16. Seehapich T, Woengthanavasu S. Emotion Recognition mistreatment Support Vector Machines 2013 fifth Int. Conf. Knowl. Good Technol. 2013; 86–91.
17. Polzein TS, Waeibel A. Emotion-sensitive human-computer interfaces, ISCA Tutor. Res. Work. Speech Emot; 2000.
18. Suiny S, Shaw DP, Jacab KP. Feature Extraction ways supported linear prophetical cryptography and riffle Packet Decomposition, 2012 Int. Conf. Adv. Comput. Commun. 2012;27–30.
19. Yin L, Chen X, Sun Y, Worm T, Reale M. A High-Resolution 3D Dynamic Facial Expression Database, The 8th International Conference on Automatic Face and Gesture Recognition (FGR08). 2008;17-19.
20. Lyons MJ, Budynek J, Akamatsu S. Automatic classification of single facial images, IEEE Transactions on Pattern Analysis and Machine Intelligence. 1999;12:1357-1362.
21. Wallhoff F. Facial expressions and emotion database, Technische Universität München, Tech. Rep; 2006.
22. Li S, Gu L. Real-time multi-view face detection, tracking, pose estimation, alignment, andrecognition. In: IEEE Conf. on Computer Vision and Pattern Recognition Demo Summary; 2001
23. Lie, JJ, Kanade T, Cohn J, Li C. Subtly different facial expression recognition andexpression intensity estimation. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. 1998;853–859.
24. Lien JJ, Kanade T, Cohn J, Li C. Detection, tracking, and classification of action unitsin facial expression. J. Robot. Auton. Syst. 2000;31:131–146.
25. Zhang Y, Ji Q, Zhu Z, Yi B. Dynamic facial expression analysis and synthesis withmpeg-4 facial animation parameters.

IEEE Trans. Circuits Syst. Video Technol. 2008;18(10):1383–1396.

26. Zhao L, Pingali G, Carlbom I. Real-time head orientation estimation using neural net-works. In: Proc of the 6th International Conference on Image Processing; 2002.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle4.com/review-history/71827