Facial Expression Recognition System for User Preference Extraction

Naoya Yamaguchi, Maria Navarro Caceres, Fernando De la Prieta and Kenji Matsui

Abstract This paper describes our preliminary study of facial expression recognition in order to extract user response information. We used Kinect to get real time facial expressions of the user to extract 6 facial expression categories (neutral, happiness, disgust, surprise, sadness, angry). As for the recognition process, we applied a multi-layer-perceptron to classify the face expressions. A total of 1,912 facial expression data sets were collected from 16 subjects. We performed holdout test using 80% of training data and 20% of test data. The recognition rate without "sadness" feature was around 90%, and the rate using every categories was around 80%. The positive results obtained shows this system as a proper one to measure user preferences in a visual test.

Keywords Face recognition · Kinect · Neural net · Categorical classification

1 Introduction

In our life, facial expression is an important role in communication. Therefore, observing listener's facial expression by computer may have a potential to solve the problems, and we might be able to obtain accurate user feedbacks automatically.

Face detection has been studied by many researchers, and latest powerful approaches are based on 3D images. However, face detection in video frames has not been extensively studied. Vinnetha et al. [5] described facial expression recognition using Kinect 3D features. A. Youssef et al. [6], described their attempt to recognize facial expressions using a 3D Kinect sensor. They constructed a

N. Yamaguchi · K. Matsui
Graduate School of Engineering, Osaka Institute of Technology, Osaka, Japan

M.N. Caceres · F. De la Prieta(✉)
Department of Computer Science, University of Salamanca, Salamanca, Spain
e-mail: Salamanca.fer@usal.es

S. Omatu et al. (eds.), *DCAI, 13th International Conference*,
Advances in Intelligent Systems and Computing 474,
DOI: 10.1007/978-3-319-40162-1_49

training data set containing time dimension. For individuals who did not participate in training the classifiers, the best accuracy levels were 38.8% (SVM) and 34.0% (k-NN). However, in case of closed test, the best accuracy levels that they obtained raised to 78.6% (SVM) and 81.8% (k-NN).

Puica et al. [4] showed emotion recognition from facial expressions using Kinect sensor with the Face Tracking SDK. The accuracy of emotion recognition with data outside the training set was off 80%.

However, the results of the systems described showed that the expression of sadness and disgust were more difficult than the others to recognize [5,6]. In order to solve this limitation, a facial expression detection is here performed by applying a trained neural net. The Kinect has been the tool to recognize the faces in video frames, using a multilayer perceptron to classify the different expressions detected in the users' faces. The positive results obtained showed the accuracy of this system to detect sad or disgust expressions.

Likewise, to evaluate some systems based on art, like music or visual art, an empirical test is usually performed. This test consists of a list of questions about several audio files or digital images where they give a punctuation of the quality of the art shown. However, in such situations the results obtained can differ from what listeners really feel when the sounds are played. Also, detailed subjective evaluation requires a lot of effort and time. Thus, a facial recognition application could be helpful to validate this kind of results more efficiently.

Therefore, the contribution of this work is twofold. On the one hand, detect several sentiments expressed by the users is aimed using a neural net. On the other hand, an application is proposed to carry out some test for other subjects, such as musical listenings or work of art evaluations.

This article is structured as follows. Section 2 contains the overall description of the system. Section 3 describes the experiments performed for the facial recognition. Section 4 explains the preliminary results obtained and Section 5 presents the conclusions and future work.

2 System Description

An automatic face expression recognition system falls into the following steps: face detection and location in a practical scene, facial feature analysis, and facial expression pattern matching. The application requires the user to be seated in front of the Kinect Sensor. Then, the device detects the human body and estimates the position of his head, drawing a face on this position. Once the face is isolated, 17 features are extracted to be the input of the multilayer perceptron. This classifier properly trained gives the facial expression of the user. This overall process is shown in Fig. 1.

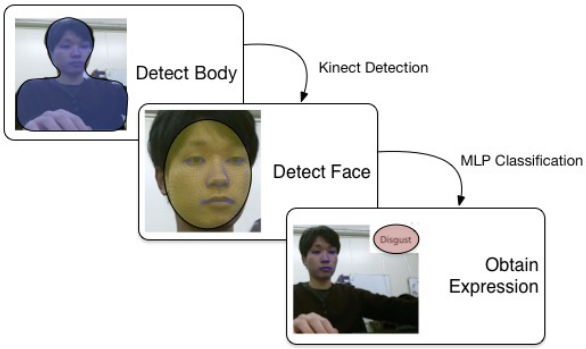

Fig. 1 Overall system process to detect facial expressions.

Section 2.1 describes some kinect details and the features extracted from the face. Section 2.2 explains the multilayer perceptron design and the training set.

2.1 The Architecture of Face Expression Recognition

The main tool used to retrieve the frames is the kinect [7]. This tool has been successfully applied in different research contexts. Khoshelham et al. [3] uses the kinect in map indoor data. Chang et al. [1] applies the device to improve skills in young adults with motor disabilities.

It also exists a variety of versions whithin the kinect devices. In this study we utilized KinectV2 Sensor for the face tracking and detection as well as feature extraction. KinectV2 Sensor was released in 2014 to be connected to windows devices, thus it permits to work with a computer. It has a RGB camera, depth sensor, and a microphone array (See Fig. 2). Together with Kinect, Microsoft provides a SDK as a software development tool kit and a software development kit HD Face API for face tracking. That makes it possible to track the user's face in the real time

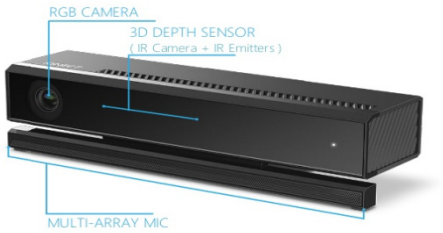

Fig. 2 Main external features of the kinect device

The first step consists of detecting the users' head from the frames obtained by the kinect camera. To do so, it has been used a method called "Time of Flight (TOF)", which relies upon an image sensor to measure the time of laser light pulses to travel to a target surface. Based on that depth information, the face position is calculated and tracked in each frame emitted by the kinect. Then, the system starts to extract the information from the face expressions. For this purpose, the Face API provides various useful properties for the face expression recognition. In this study, we used FaceShapeAnimation Enum. This Enum property has Shape Units (SU) and Animation Units (AU). SU represent shape features of the detected 3D face model, while AU represents motion features of the face. After examining the behavior of those features, AU values were used in this study. In the API property, there are 17 AUs as shown in Fig. 3: 1) JawOpen, 2) LipPucker, 3) JawSlide, 4) LipStretcherRight, 5) LipStretcherLeft, 6) LipCornerPullerRight, 7) LipCornerPullerLeft, 8) LipCornerDepressorLeft, 9) LipCornerDepressorRight, 10) LeftCheekPuff, 11) RightCheekPuff, 12) LeftEyeClosed, 13) RightEyeCloed, 14) RightEyebrowLowerer, 15) LeftEyebrowLowerer, 16) LowerLipDepressorLeft, 17) LowerLipDepressorRight.

Most of the AUs are expressed as a numeric weight varying between 0 and 1, three of them, JawSlide, RightEyebrowLowerer, and LeftEyebrowLowerer, vary between -1 and +1. Kinect updates those AU values every frame.

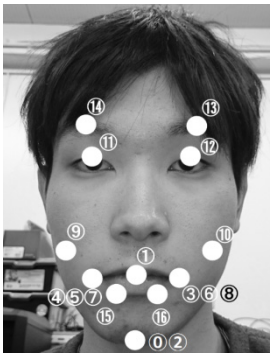

Fig. 3 The locations of the 17 Animation Units.

These data extracted from the kinect are used as input for the MLP module to classify the different facial expressions.

2.2 The Architecture of Face Expression Recognition

The architecture of our Face Expression Recognition system is based on multilayer-perceptron implemented by Weka and applied the default configurations.

Figure 4 shows the basic architecture of multilayer-perceptron classifier. As we can see, the inputs are the AU values obtained with the kinect. The outputs are the different type of faces we target in this article. Due to the difficulty of recognizing

sad faces, we develop one first system without sadness recognition, and then we added this feature in a more complex design. The number of hidden layers are also changed in order to study the classification accuracy. This system is able to recognize six different categories of facial expressions, namely, neutral, happy, disgust, surprise, sad and angry, equivalent to closed-eyes.

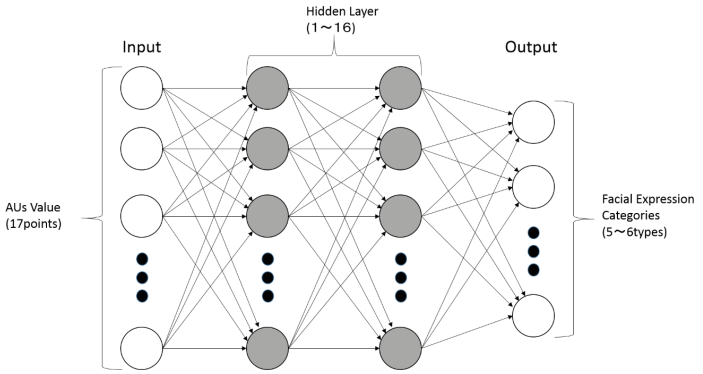

Fig. 4 Basic architecture of multilayer-perceptron based classifier

Figure 5 shows the flow chart of the MLP design. The training data are obtained from the kinect information collected. In this study, 80% of the data were used for the training, and 20% of the data were used for recognition validation. The MLP is trained with the corresponding data. The evaluation of the trained network is achieved by using the validation data. This data contains the AU values used as input for the MLP. Then, the results are compared with the expected ones and a measure of accuracy is obtained. This facial recognition accuracy depends on the number of hidden layers. Thus, we designed several iterations to change dynamically the number of hidden layers and select the best option according to the accuracy measure.

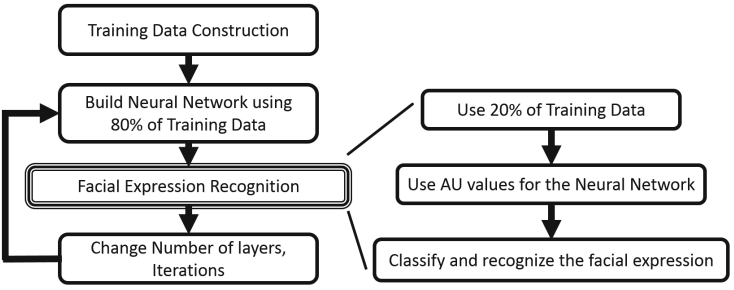

Fig. 5 Flow chart of facial expression recognition system.

3 Experimental System Overview

To train our network and evaluate our design, a preliminary experiment is performed. 16 subjects were participated in this study. They were asked to place their face 1.5 meter from the KinectV2 sensor. We captured for each individual 120 images, 20 for each facial expression (happy, sad, disgust, angry, neutral and surprise), taken in twenty seconds. Thus, a total of 1,920 images are used to train and validate the system, with a set of 17 AU values for each image. We measure the accuracy by using the number of recognized faces against the number of analysed faces, according to the number of hidden layers and the number of training iterations of the neural network.

The number of nodes N on each layer has been stablished following the next equation:

$$N = \frac{a+c}{2}$$

Where a means the number of attributes measure (in this case, 17) and c are the number of classes to classify (in the present study, 5 or 6 classes). We obtain then a total of 11 nodes for each hidden layer.

In order to separate the dataset in training and validation data, the images were randomized and 1536 (80%) were selected to train the system, while 384 (20%) were used to validate it.

To measure the accuracy, the following equation is applied [2]:

$$Acc = \frac{TP+TN}{TP+TN+FP+FN}$$

Where TP means True Positive values, TN True Negative rates, FP False Positive values and FN False Negative Values according to [2].

4 Results and Discussion

The results are plotted in Fig. 6 and 7. Horizontal axes represent the number of hidden layers applied in each execution. Vertical axes represent the recognition rate (accuracy) in percentages. Color lines represent the number of iterations used. Blue line indicates 500 iterations for each number of hidden layers. Likewise, orange lines corresponds to 1000 iterations, whereas grey lines mean 1500 iterations are applied to the system.

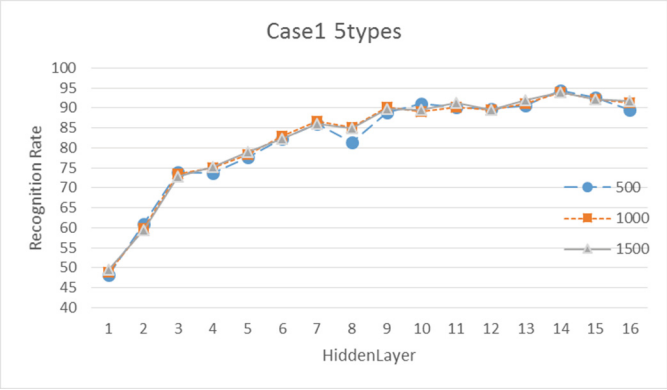

Fig. 6 Recognition results in the case of 5 facial expression categories. Vertical axis represents the recognition rate in percentages. The horizontal axis represents the number of hidden layers. The color lines corresponds to different number of iterations for each number of hidden layers.

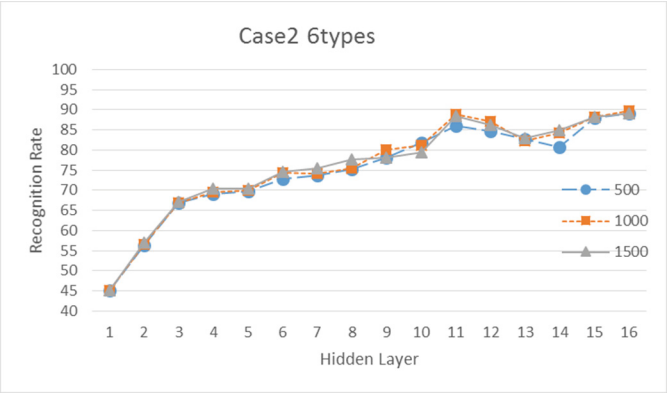

Fig. 7 Recognition results in the case of 6 facial expression categories. Vertical axis represents the recognition rate in percentages. The horizontal axis represents the number of hidden layers. The color lines corresponds to different number of iterations for each number of hidden layers.

During the system test operation, we noticed that the sadness facial expression seems confusable compare with other categories. Therefore, we decided to test both cases, i.e. 5 categories and 6 categories. Figure 6 shows the recognition results in the case of 5 facial expression categories, excluding "Sadness". Figure 7 shows the ones in the case of 6 facial expression categories. The results showed that the classification accuracy is higher when the sad faces are excluded, although both cases have good recognition rates (above the 85% or recognition). As we can see, the classifier gives nice results when the number of hidden layers is above eleven. In the case of six faces classification, 11 layers is the best result obtained. In the first case, the best one

achieved has 14 hidden layers. Thus, we can state that the classifier using between 11 and 13 layers seems good performance in any case.

It is to note that the results show quite similar results independently of the number of training iterations. We only obtain some small deviations (more than 5% in the recognition rate) in the case of fourteen hidden layers with six categories and in the case of eight hidden layers with five categories. Thus, we can conclude that the number of training iterations is not sensitive in terms of the recognition accuracy.

Although our preliminary face expression classification experiment is a sort of initial step, we were able to confirm that the proposed system setting has a potential to get user feedbacks from the facial expressions. The use of facial recognition in videoframes can be very useful to capture immediate reactions along the time for specific events, such as visualization of a work of art, listening of musical pieces. For the users, this kind of recognition can be easier, as they can make natural movements, and not staying very quiet in front of a camera, which can also bias our final results.

KinectV2 sensor with HD face API seems very powerful tool to make such facial expression recognition system in a short development period. In this study, we did not use the facial motion, however, facial motion information seems very important to be able to detect the detailed expression information.

5 Conclusions

This paper described our preliminary study of facial expression recognition in order to extract user response information. Kinect V2 HD face API is applied to solve our problem, and animation units (AU) were used to detect facial expressions of the user to extract 6 categories (neutral, happiness, disgust, surprise, sadness, angry). To recognize the different faces, a machine learning using neural network MLP is designed and trained.

A total of 1,920 facial expression data sets were collected from 16 subjects. We performed a test using 80% of training data and 20% of validation data extracted from this images taking to the 16 individuals.

The recognition rate without “sadness” feature was around 90%, and the rate using every categories was around 80%. This lead us to conclude as a good system to recognize different facial expressions accurately.

As our next step, we plan to collect the facial expression data from the real audiences to be able to investigate the effective facial expression recognition process. Also, facial motion information needs to be tested to make the system more reliable.

References

1. Chang, Y.J., Chen, S.F., Huang, J.D.: A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities. Research in developmental disabilities **32**(6), 2566–2570 (2011)

2. Gupta, N., Rawal, A., Narasimhan, V.L., Shiwani, S.: Accuracy, sensitivity and specificity measurement of various classification techniques on healthcare data. IOSR J. Comput. Eng. (IOSR-JCE) **11**(5), 70–73 (2013)
3. Khoshelham, K., Elberink, S.O.: Accuracy and resolution of kinect depth data for indoor mapping applications. Sensors **12**(2), 1437–1454 (2012)
4. Puica, M., Florea, A.M.: Towards a computational model of emotions for enhanced agent performance. Phd Thesis. Romania: University of Bucharest (2013)
5. Vineetha, G.R., Sreeji, C., Lentin, J.: Face Expression Detection Using Microsoft Kinect with the Help of Artificial Neural Network. Trends in Innovative Computing (2012)
6. Youssef, A.E., Aly, S.F., Ibrahim, A.S., Abbott, A.L.: Auto-Optimized Multimodal Expression Recognition Framework Using 3D Kinect Data for ASD Therapeutic Aid. International Journal of Modeling and Optimization **3**(2), 112 (2013)
7. Zhang, Z.: Microsoft kinect sensor and its effect. IEEE MultiMedia **19**(2), 4–10 (2012)