Cheek tracking interface for limb-disabled person

Noboru Murabayashi^{a,*}, Kenichi Yoshida^b

^aResearch group on information and social science University of Tsukuba, Tokyo, Japan

^bGraduate School of Business Science University of Tsukuba, Tokyo, Japan

*Corresponding Author: murabayashi@fw.ipsj.or.jp

Abstract

Computer interface for severely limb-disabled person is an important issue to investigate. Head, eye and mouse tracking are main sources of information for such interface. Since limb-disabled person cannot use under the neck, how to enrich information available is important. This paper tries to use mouse, more precisely cheek, information. Although recent progress of speech recognition technology realizes interfaces to control various machines accurately, the limb-disabled who cannot speak cannot use such speech recognition interface. Computer interfaces need to be developed according to their degrees of disability and must use the body part which they can move effectively. There are many handicapped persons whose degrees of disability are different. It is desirable to prepare various computer interfaces. Such interfaces must use various body parts to cope with various handicaps. From above viewpoints, this paper proposes the use of the cheek visual appearance changes using web camera. The experimental results with two layer convolutional deep learning processing show that the average recognition accuracy is 97%. In addition, the effects of the image size and deep learning network structure for the recognition performance are reported in this paper.

Keywords: cheek, visual appearance change, deep learning, face recognition, human machine interface.

1. Introduction

Due to the development of Information and Communication Technology (ICT), various human machine interfaces for operating household appliances, computer equipment or electric wheelchair have been developed for

many physically or visually handicapped persons and barrier-free is realized at various situations. Though recent progress of speech recognition technology realize man machine interfaces to control various machines accurately, the handicapped person who is not able to speak or is dysphasia may not use the interface.

Those man machine interfaces need to be developed according to their degrees of disability and used efficiently a body part which they can move. For example, if a physically handicapped person whose hand or feet is disable can move eye or blink, he may operate the PC, home appliances or electric wheelchair and communicate other persons by using his proper man machine interface. If another handicapped person whose eye or eye muscle is disabled can move mouth or tongue, he also may operate them by using the interface.

There are many handicapped persons whose degrees of disability vary and the parts of their body which they can move are different. In case it is hard work for the handicapped person to move his eye or tongue or to change his mouth open, he may feel it easy to puff out his cheeks. Furthermore, it is possible to combine the eye or head or tongue movement and the change of cheek visual appearance for the interface. It is desirable to prepare various man machine interfaces which use different body parts to cope with various physically or visually handicapped.

From the above viewpoint, this paper proposes a method to recognize the cheek visual appearance changes for human machine interface by using web camera. Although the image quality of the web camera is not good, it is easy to obtain at low cost and easy to process on a PC. Thus it is easy to use it in non-contact condition for various disabled people.

In addition, deep learning processing is applied to realize our proposed method. The effects of the image size and network structure of the deep learning processing for the recognition performances are also studied in this paper.

The rest of this paper is organized as follows. Section 2 summarizes related research. Section 3 explains the proposed recognition method and Section 4 presents the experimental results. Finally, Section 5 summarizes our findings.

2. Related Works

Computer interface for severely limb-disabled person is an important issue to investigate. Head, eye and mouse tracking is a main source of information used since they cannot use under the neck^[1]. For example, Song and Lin use head and mouse information^[2]. Several studies on detection of cheeks have been conducted. [3] studied on the method of the cheek detection based on the use of the thirteen essential face landmarks. [4] reports the study on the method for extracting the cheek region based on the lip chrome region, position of the mouth corner and geometric structure of the face. Several studies have been conducted on the interface using the cheek movement. For example, Takahashi et. al.^[5] proposed the human interface that uses the cheek swelling switch. They use two fiber units to detect a user's cheek movements. Choi^[6] proposed the wheelchair control system which used the electromyographic (EMG) signal based on the cheek movement and the electroencephalograph (EEG) signal. These conventional studies on the detection of the cheek movement use some kinds of mechanical system or electrical devices.

There also exist various studies on the recognition of the face expressions without using mechanical system. For example, Nishime et. al.^[7] applied the convolutional neural network (CNN) to recognize the human emotions, facial expressions. They studied image feature for the recognition. [8] uses the Lucas-Kanade optical flow tracker and random forest classifier. In the research of [9], facial expressions are recognized using Support Vector Machine (SVM). In the research of [10], facial expressions are judged using texture features, Gabor LBP (GLBP) features and Random Forest Classifier methods. As research on the facial gesture detection, for example, [11] proposed tongue-in-cheek non-contact facial gesture detection system which uses wireless signal to detect different facial gestures in four directions. It detects the movement of cheeks caused by

moving different parts of the mouth.

Several methods for man machine interface used the parts of the human body have been studied for a long time. For example, [12] studies to detect lip movement tracking and lip gesture recognition by using images taken with a web camera and cascade of boosted classifiers using Haar-like image features and recognition results are applied to the human computer interface. As the study on the recognition of the eye blinking, for example, [13] reports the method based on the blinking detection to assist the handicapped, who cannot move his body by himself freely, to communicate to another person. [14] uses illuminant markers to control the cursor movement in computer applications based on facial expressions such as cheek movement, eye brow, and mouth open. As the research to recognize the motion of the head, for example, [15] reports the method to operate the electric wheelchair with the interface that detects head gesture motion using depth sensor. The method of controlling the robot by detecting movements such as left and right rotation of the head with a web camera is studied in [16]. Other study is that [17] conducted the research on the man machine interface for the computer operation based on the detecting the direction of the face by using the web camera.

From those background reviews of the conventional studies, in this paper, we propose the recognition of the cheek visual appearance change for human machine interface only using a web camera and deep learning processing.

3. Recognition of the cheek appearance change

In this section, the proposed method for recognizing the cheek visual appearance change is described. Fig.3.1 shows the system configuration, i.e., the block diagram proposed method. The cheek visual appearance changes are recognized by the deep learning processing. A single and two-layer convolution deep learning processing are applied to evaluate recognition performance between different network structures.

The network structures are consisted of seven processing systems, that is, convolution, ReLU (rectified linear units), Max Pooling Tanh, affine, softMax, and categorical cross entropy as shown in Fig.3.1(1) and (2).

There are four types of the cheek visual appearance change, class 0: no change, class 1: puffed out in the left side cheek, class 2: puffed out in the right cheek side, class

3: puffed out in the both left and right cheek sides.

Figure 3.2 shows the outline of the method of setting the web camera. In this case, it is set on the same desk as the PC keyboard and the face image capturing camera.

(1) single-layer convolutional network

(2) two-layer convolutional network

Fig. 3.1. block diagram of the deep learning network

Fig. 3.2. web camera setting

4. Experimental Results

4.1 Training and testing data

To evaluate the performance of the proposed method, we have conducted experiments using the data shown in Fig.4.1 and Table 4.1.

Fig.4.1 shows the outline of the image data preparation.

Firstly, face image is captured as color video images by the web camera, image size (width $\times$ height) is 640×480 pixels. The color frame images are extracted from the video, and training and testing image data are made from them. Three different image sizes of the training and testing data are prepared to evaluate the effect of the image size for the recognition performance as shown in Fig.4.1. Three image sizes (width $\times$ height) are 128×128 pixels, 64×64 pixels and 32×32 pixels.

Table 4.1 also shows the four types changes of the example images captured by the web camera, and the training and testing data schemes. Deep learning processing was used to recognize cheek changes. In this experiment, Neural Network Console^[18] is applied as deep learning framework.

The training processing is conducted with 1022 color images and the image sizes are 128×128 pixels, 64×64 pixels and 32×32 pixels color images. When the image size is large, it is easy to understand state change of the cheek. However, if the size is too large, depending on the PC performance, memory processing may freeze. The testing processing for evaluating the recognition is carried out with 300 image data which are not included within the training image data.

In video image, the four kinds of cheek appearance are changed randomly in the experiments.

Fig. 4.1. outline of the data set preparation

Table 4.1 Training and testing data scheme and example

class	0	1	2	3
image				
direction of change	no change	left	right	left and right
number of training data (total : 1022)	469	205	175	173
number of testing data (total : 300)	137	57	57	49

4.2 Recognition performance

The experimental results of the confusion matrix, precision, recall and F-measure are shown in Table 4.2 (1) - (6). Summary of the recognition results based on Table 4.2 are shown in Table 4.3 (1) (2) (3).

Average accuracy, average Precision, average recall and average F-measure are defined as follows respectively.

$$\text{Average accuracy} = \sum_{n=0}^3 (S_n / S_t) \quad (1)$$

S_n : the number of correctly recognized n-th class data

S_t : the number of total testing data (= 300)

$$\text{Average precision} = \sum_{n=0}^3 (P_n / N_t) \quad (2)$$

P_n : precision of the n-th class data

N_t : the number of total class (= 4)

$$\text{Average recall} = \sum_{n=0}^3 (R_n / N_t) \quad (3)$$

R_n : recall of the n-th class data

$$\text{Average F-measure} = \sum_{n=0}^3 (F_n / N_t) \quad (4)$$

F_n : F-Measure of the n-th class data

Where;

F-measure =

$$(2 \times \text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (5)$$

The recognition performances of the two layer convolutional network on image size of 128×128 pixels are slightly better than that of the single layer convolutional network as shown in Table 4.2 (1), (2) and Table 4.3 (1). Average accuracy, average precision, average recall and average F-measure are more than 97% in two layer convolutional network.

The experimental results tendency of the image size 64×64 pixels and 32×32 pixels are similar to that of image size 128×128 mentioned above. That is, the recognition performances on the two layer convolution network of both image sizes are superior to that on the single layer convolutional network.

Considering the effect of the image size to recognition performances, there are performance difference results on the two layer convolutional network and the single layer convolutional network, there are no significant differences in recognition performances between any three image sizes as shown in Table.4.3 (1) (2) (3).

Table 4.2 Results of the experiments

(1) single layer convolution network (128×128 pixels)

label	0	1	2	3	Recall
0	136	1	0	0	0.993
1	7	50	0	0	0.877
2	10	0	47	0	0.825
3	6	1	0	42	0.857
Precision	0.855	0.962	1.00	1.00	
F-Measure	0.919	0.917	0.904	0.923	

(2) two layer convolution network (128×128 pixels)

label	0	1	2	3	Recall
0	135	2	0	0	0.985
1	2	55	0	0	0.965
2	0	0	57	0	1.00
3	3	0	0	46	0.939
Precision	0.964	0.965	1.00	1.00	
F-Measure	0.975	0.965	1.00	0.968	

(3) single layer convolution network (64×64 pixels)

label	0	1	2	3	Recall
0	136	1	0	0	0.993
1	3	53	0	1	0.930
2	0	0	57	0	1.00
3	6	0	0	43	0.878
Precision	0.938	0.981	1.00	0.977	
F-Measure	0.965	0.955	1.00	0.925	

(4) two layer convolution network (64×64 pixels)

label	0	1	2	3	Recall
0	136	0	0	1	0.993
1	2	55	0	0	0.965
2	2	0	55	0	0.965
3	1	0	0	48	0.980
Precision	0.965	1.00	1.00	0.980	
F-Measure	0.978	0.982	0.982	0.980	

(5) single layer convolution network (32×32 pixels)

label	0	1	2	3	Recall
0	137	0	0	0	1.00
1	3	51	0	3	0.895
2	2	0	55	0	0.965
3	5	0	0	44	0.898
Precision	0.932	1.00	1.00	0.936	
F-Measure	0.965	0.944	0.982	0.917	

(6) two layer convolution network (32×32 pixels)

label	0	1	2	3	Recall
0	135	2	0	0	0.985
1	2	55	0	0	0.965
2	0	0	57	0	1.00
3	0	0	0	49	1.00
Precision	0.985	0.965	1.00	1.00	
F-Measure	0.985	0.965	1.00	1.00	

Table 4.3 Summary of the recognition results

(1) (image size : 128 x 128 pixels)

network	single layer	two layer
Average accuracy	0.917	0.977
Average precision	0.954	0.982
Average recall	0.888	0.972
Average F-measure	0.916	0.977

(2) (image size : 64 x 64 pixels)

network	single layer	two layer
Average accuracy	0.963	0.980
Average precision	0.974	0.986
Average recall	0.950	0.976
Average F-measure	0.961	0.981

(3) (image size : 32 x 32 pixels)

network	single layer	two layer
Average accuracy	0.957	0.987
Average precision	0.967	0.988
Average recall	0.939	0.988
Average F-measure	0.952	0.988

4.3 Learning characteristics

Fig.4.2 shows the learning characteristics of the experiments. There are more perturbations of the results on two layer convolution network compared with that on the single layer convolution network. In addition, characteristics of image size of 64×64 pixels and 32×32 pixels are changing largely compared with that of 128×128 pixels. That is, learning process of image size 128×128 pixels is conducted more smoothly than that of 64×64 pixels and 32×32 pixels on both single and two layer convolution network.

Relation between the network structure and the learning characteristics for improving the recognition performance requires further study.

(1)

(2)

(3)

Fig. 4.2 Learning characteristics

5. Conclusion and Future works

In this paper, we proposed a method to recognize the cheek visual appearance change for human machine interface. In addition, the effects of the image size and deep learning network structure for the recognition performance are studied. The experiments which use the image captured by the web camera reveal that:

For the recognition performance;

- 1) The experimental results with two layer convolutional deep learning processing show that the average recognition accuracy is 97%.
- 2) The recognition accuracy of the two layer convolutional deep learning processing is superior to that of the single layer convolutional network.
- 3) There is no significant effect of the image size on the recognition accuracy within the experiments.

For the learning characteristics;

- 1) There are more perturbations of the results on two layer convolution networks compared with that on the single layer convolution network.
- 2) The learning process of image size 128×128 pixels is conducted more smoothly than that of 64×64 pixels and 32×32 pixels on both single and two layer networks.

In this paper, only the recognition performance of the visual appearance change is analyzed. The evaluation of the actual human machine interface for the physically handicapped person is not conducted. In actual application, it might be different depending on the detection performance, the way of use and various situations such as disabled people, web camera setting method, surrounding brightness etc.

The cheek visual appearance changes within a specific period are considered as a time series data. Handling it as time series data makes it possible to use more information than the four types cheek visual changes that were recognized in the experiments. The actual evaluation in the various situations and the applications of those processing as mentioned above to our proposed method remain as future issues.

References

- [1] Amer Al-Rahayfeh, Miad Faezipour, "Eye Tracking and Head Movement Detection: A State-of-Art Survey", IEEE Journal of Translational Engineering in Health and Medicine, Vol.1, DOI: 10.1109/JTEHM.2013.2289879, 2013
- [2] Y. Song Y. Luo J. Lin, "Detection of movements of head and mouth to provide computer access for disabled", Proc. Int. Conf. TAAI, pp. 223-226., 2011
- [3] M. Frackiewicz, H. Palus & K. Radlak, "Automatic cheek detection in digital images", Computational Vision and Medical Image Processing V, pp49-55, 2016
- [4] Yuncong Yang, Jing Zhang, Li Zhuo, "Cheek region extraction method for face diagnosis of Traditional Chinese Medicine", Signal Processing (ICSP), 2012 IEEE 11th International Conference on , Volume 3, pp. 1663-1667, 2012
- [5] Yoshihiko Takahashi, Shinichiro Suzukawa, "Easy Human Interface for Severely Handicapped Persons and Application for Eating Assist Robot", Proc. of IEEE International Conference on Mechatronics, Pages:225-229, 2006
- [6] Kyuwan Choi, "Control of a vehicle with EEG signals in real-time and system evaluation", European Journal of Applied Physiology, Volume 112, Issue 2, pp 755-766 (February 2012)
- [7] Nishime, Taiki, Endo, Satoshi, Toma, Naruaki, Yamada, Koji, Akamine, Yuhei, "Feature Acquisition and Analysis for Facial Expression Recognition Using Convolutional Neural Networks", Transactions of the Japanese Society for Artificial Intelligence, 32(5), pp.1-8, 2017-09-01
- [8] Xiaorong Pu, Ke Fanb, Xiong Chen, Luping Ji, Zhihu Zhou, "Facial expression recognition from image sequences using two fold random forest classifier", Neurocomputing 168(2015) pp.1173-1180,
- [9] J. Kamalakumari, Dr. M. Vanitha, "Image Sequences Based Facial Expression Recognition Using Support Vector Machine", International Journal of Engineering and Technology, Vol 9 No 5 Oct-Nov 2017 pp.3605-3609 J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.
- [10] M. Arfan Jaffar, "Facial Expression Recognition using Hybrid Texture Features based Ensemble Classifier", International Journal of Advanced Computer Science and Applications, Vol. 8 , No.6, pp.449-453, 2017
- [11] Mayank Goel, Chen Zhao, Ruth Vinisha, Shwetak N. Patel, "Tongue-in-Cheek: Using wireless Signals to Enable Non-Intrusive and Flexible Facial Gestures Detection", Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI 2015), Pages 255-258 (2015)
- [12] Piotr Dalka and Andrzej Czyzewski, "Lip movement and gesture recognition for a multimodal human-computer interface", Proc. Computer Science and Information Technology, 2009. IMCSIT '09. International Multiconference on 12-14 Oct. 2009, pp365-369
- [13] Takahito Niwa, Ippei Torii, Naohiro Ishii, "Development of Communication Tool for Physically Handicapped with Involuntary Movements by Line-of-Sight Detection", Applied Computing and Information Technology/3rd Intl Conf on Computational Science/Intelligence and Applied Informatics/1st Intl Conf on Big Data, Cloud Computing, Data Science & Engineering (ACIT-CSII-BCD), 2016 4th Intl Conf on 12-14 Dec. 2016
- [14] M. Vasanthan, M. Murugappan, R. Nagarajan, "Facial expression based computer cursor control system for assisting physically disabled person", Communication, Networks and Satellite (ComNetSat), 2012 IEEE International Conference on, July 2012 , pp. 172-176.
- [15] Hirofumi Ohtsuka, Tatsuya Kato, Koki Shibasaki, "Non-contact head gesture maneuvering system for electric wheelchair using a depth sensor", Sensing Technology (ICST), 2015 9th International Conference on, 8-10 Dec. 2015
- [16] P.B. Alappanavar, Kishor Gardi, Shubham Raut, Swapnil Ghorpade, Sameer Waghmare, "Head gesture recognition for hands-free control of robot movements", International Journal Of Engineering And Computer Science ISSN:2319-7242, Volume 4 Issue 3, Page No. 11014-11017, March 2015
- [17] Takehiko Tomikawa, Toshiaki Yamanouchi, and Hiromitsu Nishimura, "An Adaptability of Head Motion as Computer Input Device", Journal of Automation and Control Engineering Vol. 4, No. 2, pp166-170, April 2016
- [18] <https://dl.sony.com/> (accessed:2018.03.08)