A method of micro facial expression recognition based on dense facial motion data

Yasuhiro AKAGI
Kagoshima University
Korimoto, 1-21-24
Kagoshima
890-8580, JAPAN
akagi@ibe.kagosima-u.ac.jp

Hiroshi KAWASAKI
Kagoshima University
Korimoto, 1-21-24
Kagoshima
890-8580, JAPAN
kawasaki@ibe.kagosima-u.ac.jp

ABSTRACT

In this paper, we propose a method for recognizing a micro expression which is a small motion appearing on a face by using a high density and high frame-rate 3D reconstruction method. Some studies report that the micro expressions are caused by the change of mental state. If we can recognize the micro expressions, this information could be useful for machines to understand the mental state of a human. With advancements of 3D reconstruction methods, methods have been proposed to reconstruct dynamic objects such as motions of a human's body in high accuracy with high frame rate. Based on the data obtained from the high quality shape reconstruction method, the proposed method recognizes the micro expressions. To detect a part of the face where the micro-expressions are appeared, we propose an experimental estimation of the part. We also report a recognition rate of the change of the mental state using the experimental system.

Keywords

Micro expression, Facial expression, AAM, Facial motion

1 INTRODUCTION

The human face displays much information of an internal (mental) state of a human. By using facial images, many studies have been proposed to estimate the human mental state based on a machine learning. In these facial expression recognition methods, they define 5-8 types of human expressions and detect these expressions at more than 90% recognition rate. This confirms in terms of study that human expressions have certain universality in which connects human emotions.

On the other hand, humans can purposefully control facial expressions different from their emotions. However, when experiments are performed to evaluate a facial expression recognition, whether or not it matches the recognized facial expression and the subject 's true emotion at this time is not distinguished, since it is excluded from this recognition problem. Ekman et. al. focuses on a facial expression when the mismatch occurs between the expression and emotion[1]. According to this work, they point out that when a person shows facial expressions different from their true intentions, there exists a part with small expression change and difference in time to form the expressions. They call the small change as "micro expression". In this study, we verify whether or not a person's change in mental state appears on a face and to figure out the part it is expressed on by focusing on small movements (micro expressions) on the face. To detect the micro expressions, a method which can capture facial motions with high accuracy and high frame rate are used. Then we capture the facial motions by using Kinect[5] (common accuracy and frame rate) to detect the micro expressions. By comparing the recognition results of the micro expressions between the two capturing method, we also show the required accuracy and frame rate to recognize the micro expressions.

2 RELATED WORKS

2.1 Facial expression recognition

There are many studies of the human understanding based on facial expressions, since the human face displays various information including different emotions. Among these, numerous experiments have been performed regarding human expression detection. As methods to detect expressions, using points that display the outline of parts such as a mouth (facial feature points), as well as dividing the face image into small

regions and using its uniqueness of the region with luminance gradient have been proposed. For method using facial feature points, there is a method called Active Appearance Model (AAM) that places feature points on the outlines of major facial parts such as eyes and nose, and using their changes as a feature[6]. Ekman et. al. first defined parts called Action Units (AUs) that serve as a index to capture facial movements, then extracted the movements from the image to detect expressions[7]. Kotsia et. al. studied 8 types of AU movements using Support Vector Machine (SVM) to reach over 95% expression detection rate[8].

On the other hand, for methods using uniqueness in small region, Shan et. al. extracted unique Local Binary Pattern as feature vector to be studied with SVM to detect expressions[9]. As a result, they were able to reach 91% identification rate for 7 different types of expressions. According to a survey article regarding expression detection, many methods have been proposed that successfully reached over 90% recognition rate[10], and general expression detection has been in practice. On another note, for the purpose of aiding speech detection from a voice, other methods have been proposed to detect speech from the shape of the mouth[11]. Many experiments have been performed to recognize the information displayed on the human face, by using facial feature points.

2.2 Researches focus on a micro expression

Ekman et. al. focused on the subtle movements displayed on the face[1]. In their research, they reported that when a person tries to display expression different from their emotions, the small movements appearing on their face (micro expression) becomes different from the expression with matched emotion. For the research that utilizes the change in micro expression, a method has been proposed that measures the amount of time it takes for the newborn child to recognize from the initial movement ($200ms$ from the beginning) to the time of expression change detection to identify their development disorder[12]. Su-Jing et. al. proposed a method that detects micro expression from an image[2, 3]. Here, 5 types of micro expressions were detected from a moving image captured at $60fps$ which resulted in about $30-47\%$ recognition ratio and concluded that it needs improvements. Studies about the micro expression recognition are still stages in-development and many studies researched about the method to detect the feature of the micro expressions[4]. Although mechanical detection of micro expressions is crucial for human understanding, its subtleness makes it difficult to detect.

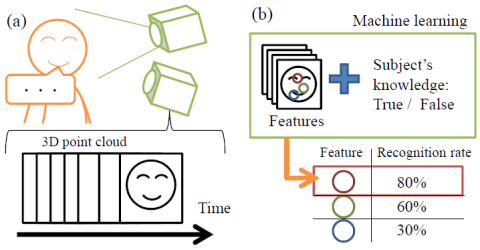

Figure 1: Overview of the proposed system. (a)Measurement of a face during speech. (b)Detection of a part where a micro-expression appears.

2.3 Methods of feature detection of a face

As mentioned in 2.1, in order to detect expressions based on facial feature points, it is necessary to extract high quality feature points from the image. Dollar et. al. achieved error difference of 3 pixel by studying the uniqueness on the image using Random fearn[13]. Also, Cao et. al.[14] was able to track feature points in real time based on AAM, from the depth map obtained from devices such as Kinect[5]. To detect the micro expressions, it requires highly precise measurement with high quality face shape data, we propose a method that utilizes the detailed 3D shape data with subtle changes displayed on the face. In our research, by using the highly precise measurement of the 3D shape data, we examine which movement related to human mental state is being displayed on the face. From this result, we experimentally created a system that estimates a simple human mental state, and tested its detection accuracy.

3 OVERVIEW OF THE RECOGNITION TARGET AND ITS PROCESS

Our experiments are performed under the following hypothesis: if the micro expressions are appeared according with the changing of a person 's mental state, by learning the pair of a facial motion and a mental state of a person, a machine learning method can find the relevances between the facial motion and the mental state (fig. 1). This hypothesis agrees with the approach of a facial expression recognition method. In the following sections, we will explain the overview of the proposed method.

3.1 Experimental method and the state of the subject to be detected

In this research, facial motions under different mental states are required to recognize the micro expressions. To collect these motions, we assign the following subjective experiments to capture facial motions (fig. 2). The subject will see one vocabulary every three seconds on the display, and is to read them out loud. After the experiment, the subject is asked whether or not

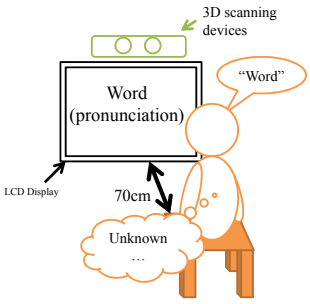

Figure 2: Experiment environment of the task.

they knew the vocabulary asked. Then if it was the first time a subject pronounced a word, even if the subject could assume the meaning of the word from the word, the word is categorized as "unknown". From this experiment, we can record two types of conditions where one knows the vocabulary already and their vocal state is normal; the other is where the subject did not know the vocabulary and speaks under pressure of thought. The reason to choose this method is that the recognition problem becomes simple with only two states (knows/don 't know), and the subject can correctly answer the mental state after the experiment. A dataset[15] was used for the vocabulary set.

3.2 Measurement method of 3D shape of a face

In our experiment, we measure the 3D shape of the subject 's face from the front while they are answering the questions. For measurement, we use the two types of 3D reconstruction methods; one is that allows high resolution and high frame rate, another is a simple measurement method using Kinect. The facial motion data from the first measurement method is to determine if there are changes in the face when the subject experiences an unknown word, and to detect the region with changes. The second measurement method was done to determine if a simple measurement system can accommodate the proposed detection problem. Table 1 and fig. 3 show the difference between the two measurement methods in respect to measurement preciseness and the difference in the number of measured points. The Table 1 shows the numbers of points when the reconstruction system (camera) is located in the minimum distance from a face. The next section will explain the detection method of the facial area where the micro-expression is appeared from an input point cloud.

3.3 Estimation of the site of micro expression using precise 3D facial motions

According to the research of Ekman et. al. mentioned above[1]., the human face has parts that are easy to actively control and parts that are not so easy. It

Table 1: Comparisons of specifications of each measurement method.

	Measurement method	
	High accuracy	Kinect
Resolution(mm)	0.3	2
No. of points	30,000	5,000
Frame rate	200	20

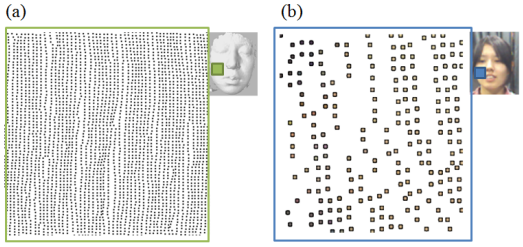

Figure 3: Comparison of point clouds captured by both measurement methods. (a)Points scanned by the high accuracy method. (b)Points scanned by Kinect.

is reported as one example, the characteristic region around the mouth is easily controlled, spaces between eye brows are difficult to control, where the true emotion are often expressed. Since facial movement is likely to have different uniqueness between different parts, the face is divided into mouth region, nose region, and the eye brow region and determined which region best contains uniqueness regarding detection of words used in this experiment. For each region, points are selected randomly from the starting frame of the face with speech, and variation in the depth direction at those positions is recorded. Then, using the feature vector for this variation and by placing supervised learning labels as the unknown words, the difference in the recognition rate is compared (fig. 1(b)). From the result, in the parts with high identification rate, the movement related to the unknown word is considered to be present. The detailed method is stated in section 4.

3.4 Recognition of unknown word with Kinect

As method explained in section 3.3, by using a large number of feature points with high-precision sensor, it is possible to measure the small changes in the face and although it is advantageous for the recognition of subtle movements of the face, it requires special equipment for capturing and also too costly. On the other hand, its measurement accuracy is low and when using sensor with low frame rate, it is possible to not be able to find the subtle movement changes. For this problem, since it is possible to limit the part with unique deformation as mentioned in section 3.3, the change in the recognition rate is examined for the narrow measurement range. Cheaper 3D measurement device, Kinect is

used for this experiment. Ideas of measurement method using the Kinect are described in section 5.

4 STUDY OF SPEECH FEATURE BASED ON HIGH-PRECISION 3D SHAPE DATA

This section experimentally detects which site appears to uniquely show the difference in the shape by recognition of the words during pronunciation using the 3D shape data of the face taken on high speed and high precision.

4.1 Measurement of high-precision facial movement data

In the study presented in this section, it is necessary to extract the movement unaffected by the difference of the words spoken by the subject, and only the movements originating from movements from unknown words. Furthermore, since there is a possibility that a micro expression might have a ‘ habit ’ specific to the subject, two patterns of study are performed for the case of extracting individual features for each subject, and for the case of extracting regardless of the subject. For this study, multiple speakers are asked to read the words multiple times, and the face shapes are measured individually. The measurement was performed continuously until 14 known and 14 unknown words appear to be used to be further evaluated. Whether the words were known or unknown is reported from the subject after the measurement (after speech).

4.2 Study of movement data and the data used for the comparison survey

Since there are facial parts that are easy to control and parts that are not so easy[1] we divide a face into three regions: mouth, nose, and eyes for defining the feature vectors of facial movements. This study was done using the shape deformation data in the region of radius 2.5cm from the center position, for the three parts: mouth, nose, and between the eyebrows (fig. 4). We given the center positions of each region manually and 200 points are selected randomly in each region. We define the feature of the facial movement by using the depth directional changes of 100 frames (0.5 sec) at each point. By applying the facial motion tracking method[16], we track the motion of each points. The following measurement was performed for the 14 known and unknown words respectively, with each dataset being represented by 100 reference points. We define the feature vector of each point as Z-axis movements of the point in 100 frames (100 dimension vectors). Also, each feature vector was given a label whether or not the words were known or unknown. Based on these feature vectors and labels, even when given with multi-dimensional data with numerous instructing data, mechanical study is performed that can efficiently learn using two class classifying Local Deep Kernel Learning (LDKL)[17].

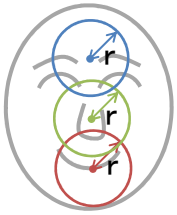

Figure 4: Comparison areas of a motion feature.

4.3 Investigation of the recognition rate when considering individuality of the micro expression

We collect the motion data of 28 words (14 known, 14 unknown) from 6 subjects. The 24 words (12 known, 12 unknown) ware used for machine learning and 4 words ware used for evaluation. Table 2 shows the recognition ratio for 6 individuals. In any subject ’s case, the result from the motion features of the area around the nose had the highest detection rate. In particular, 75% or more detection rate was shown for the knowledge of words, and is highly likely that the nose area displays a feature movement. On the other hand, subject B ’s recognition ratio of unknown words was 46%, which shows that movement feature did not occur. Further, the recognition rate for the mouth region indicates 54 – 65%, but it cannot be said so when it is considered to have two type classification, thus motion resulting from the unknown words were found at the mouth region. This is also true for the region between eye brows.

4.4 Investigation of the recognition rate when not considering individuality of the micro expression

Next, we investigate the recognition ratio using the data without distinguishing characteristic movements for all subjects. Evaluation was done using the set of vocabulary not used in the learning process for each individual. The result is shown on Table 3.

Although the identification rate of the result has increased here, when it is compared with the data used for the learning process in section 4.3, it has decreased. This result shows that micro expression has some levels of individual differences. This is connected to the research mentioned by Ekman et. al. that ability to hide emotions such as a startle is different from subject to subject[1]. Therefore, when trying to make an identification device that uses multiple subjects, a solution is needed to overcome individual differences.

Table 2: Recognition rate of unknown words based on the motion data of the same person（%）. T:Known, F:Unknown words.

	Subject A		Subject B		Subject C		Subject D		Subject E		Subject F		Total	
	T	F	T	F	T	F	T	F	T	F	T	F	T	F
Mouth	62	65	54	65	66	45	77	45	70	13	29	62	60	49
Nose	79	77	77	46	75	78	70	69	78	57	73	74	75	67
Brow	44	5	72	34	41	19	55	18	85	51	71	21	61	34

Table 3: Recognition rate of unknown words based on the all motion data（%）. T:Known, F:Unknown words.

	Subject A		Subject B		Subject C		Subject D		Subject E		Subject F		Total	
	T	F	T	F	T	F	T	F	T	F	T	F	T	F
Mouth	45	54	69	62	52	75	36	54	53	70	24	68	46	64
Nose	72	45	62	73	73	79	75	76	87	31	73	86	74	65
Brow	58	4	50	61	53	15	45	21	84	76	43	30	55	35

5 DETECTION WITH A SIMPLE SYSTEM USING KINECT

From the result of section 4, when identifying the nose area, information to help identify the cognitive word can be obtained. Based on this knowledge, the change of detection rate using Kinect, which has more noise than the measurement method explained in section 4, is researched.

5.1 Measurement data and pretreatment

In addition to the target measurement of the 3D shape, 64 points of feature points can be obtained from the Kinect as the facial feature points[5].. Out of them, by using the 12 points that make up the outline of the nose, it is used as the detection and the study of the depth direction of the point as its feature vector. However, since there is a possibility that measurement errors and noise are contained in the Kinect, by pre-analyzing the motion information for frequency and also removing the signal component of 10Hz or more, variation of the time series have been smoothed out. In order to gather data from all of the seven subjects involved, facial data was measured for the full 60-words vocalization dataset. The known and unknown word ratio is 215:145. The ratio of the known words is structured to be slightly higher.

5.2 Recognition based on motion characteristics using Kinect

As similar to the experiment done in section 4.3, the study and its recognition is done using the same subject. From the 60-word worth of vocalization data gathered from one subject, 12 points that make up the nose and 40 frames (2.0sec) were obtained. Also, 45 words during the study process, and 15 words for evaluation were used. The result is as shown on Table 4.

From this result, although it can recognize the information regarding the unknown words from a partial subject, there are some subjects that couldn ’t capture any feature points. As stated in section 4.4, for subjects that shows little micro expressions, lowering the measurement accuracy makes it difficult to capture the feature points.

Table 4 also shows the results obtained when using only motion feature vectors for the mouth and eye brow area. As with the identification result based on accurate data, as compared to the nose area, recognition rate has decreased. Within the result, there are some results from the mouth data from subject A and C, subject B ’s eye brow data that has more than 80% accuracy. Since this is a two class problem, if the detection device tends to put the input data more in one class, one class will end up having higher recognition rate than the other. It is likely that the similar phenomenon occurred in our experiment, since the other ratio was at 43%.

6 CONCLUSIONS

In this paper, by using the dense, accurate facial shape measurement, we propose a method to find subtle yet specific moving parts called micro expression. In order to verify the set of methods to identify the site of micro expression and its accuracy, as a simple recognition problem, vocalizing unknown words was assigned to be further analyzed mechanically. As a result, for the task given to the subject, by using highly accurate data of the facial movement, we were able to identify feature parts that matches with the human mental state, and able to estimate at about 75% accuracy. On the other hand, if the accuracy of the face shape measurement was decreased by few mm and lower the frame rate to $20fps$, recognition of the micro expression became troublesome. Further research will improve the experimental method to create micro expression, as well as

Table 4: Recognition rate（%） based on feature vectors around the mouth, nose and brow. T:Known, F:Unknown words.

		Mouth			Brow			Brow		
		T	F	Total	T	F	Total	T	F	Total
Subject	A	87	43	67	64	78	70	62	0	33
	B	67	33	53	56	60	57	88	14	53
	C	80	40	53	100	50	80	60	50	53
	D	36	55	44	67	67	67	46	42	44
	E	56	50	53	33	67	47	25	29	27
	F	72	0	53	88	29	60	83	22	47
	G	25	73	60	75	54	60	20	70	53

by performing on larger scale subject number, we aim to structure feature movements based on micro expressions.

ACKNOWLEDGMENT

This work was supported in part by NEXT program No.LR030 and KAKENHI No.25870570 in Japan.

7 REFERENCES

[1] Ekman, P.: Facial expression and emotion. American Psychologist **48** (1993) 384–392

[2] Wang, S., Chen, H.L., Yan, W.J., Chen, Y.H., Fu, X.: Face recognition and micro-expression recognition based on discriminant tensor subspace analysis plus extreme learning machine. Neural Processing Letters **39** (2014) 25–43

[3] Yan, W.J., Wu, Q., Liu, Y.J., Wang, S.J., Fu, X.: Casme database: A dataset of spontaneous micro-expressions collected from neutralized faces. In: Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on. (2013) 1–7

[4] Polikovsky, S., Kameda, Y., Ohta, Y.: Facial micro-expressions recognition using high speed camera and 3d-gradient descriptor. In: Crime Detection and Prevention (ICDP 2009), 3rd International Conference on. (2009) 1–6

[5] Zhang, Z.: Microsoft kinect sensor and its effect. MultiMedia, IEEE **19** (2012) 4–10

[6] Cootes, T., Edwards, G., Taylor, C.: Active appearance models. In: Proceedings of the 5th European Conference on Computer Vision, ECCV 98. Volume 1407 of Lecture Notes in Computer Science., Springer Berlin Heidelberg (1998) 484–498

[7] Ekman, P., Friesen, W.V.: Facial action coding system. A Technique for the Measurement of Facial Movement. (1978)

[8] Kotsia, I., Pitas, I.: Facial expression recognition in image sequences using geometric deformation features and support vector machines. Trans. Img. Proc. **16** (2007) 172–187

[9] Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on local binary patterns: A comprehensive study. Image and Vision Computing **27** (2009) 803 – 816

[10] Bettadapura, V.: Face expression recognition and analysis: The state of theart. CoRR **abs/1203.6722** (2012)

[11] KOMAI, Y., MIYAMOTO, C., TAKIGUCHI, T., ARIKI, Y.: Phoneme analysis of image feature in utterance recognition using. MIRU 2010 **IS3** (2010) 1771–1778

[12] ICHIKAWA, H., YAMAGUCHI, M.K.: Dynamic subtle facial expression can be recognized by 6- to 7-month-old infants. Japanese Psychological Research **56** (2014) 15–23

[13] Dollár, P., Welinder, P., Perona, P.: Cascaded pose regression. In: CVPR. (2010)

[14] Cao, Z., Yin, Q., Tang, X., Sun, J.: Face recognition with learning-based descriptor. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010. (2010) 2707–2714

[15] Kazuyo, T., Hayamizu, S., Ohta, K.: The etl speech database for speech analysis and recognition research. First International Conference on Spoken Language Processing (1990)

[16] Akagi, Y., Furukawa, R., Sagawa, R., Ogawara, K., Kawasaki, H.: A facial tracking and transfer method with a key point refinement. In: SIGGRAPH Posters. (2013) 79

[17] Jose, C., Goyal, P., Aggrwal, P., Varma, M.: Local deep kernel learning for efficient non-linear svm prediction. In: Proceedings of the 30th International Conference on Machine Learning (ICML-13). Volume 28. (2013) 486–494