A Smile/Laughter Recognition Mechanism for Smile-based Life Logging

Kurara Fukumoto
Faculty School of Engineering, Kobe University
1-1, Rokkodaicho, Nada, Kobe, Hyogo
657-8501, Japan
kurara@stu.kobe-u.ac.jp

Tsutomu Terada
Grad. School of Engineering, Kobe University
and JST PRESTO
1-1, Rokkodaicho, Nada, Kobe, Hyogo
657-8501, Japan
tsutomu@eedept.kobe-u.ac.jp

Masahiko Tsukamoto
Grad. School of Engineering, Kobe University
Institute for Clarity in Documentation
1-1, Rokkodaicho, Nada, Kobe, Hyogo
657-8501, Japan
tuka@kobe-u.ac.jp

ABSTRACT

Most situations that cause people to smile are important and treasured events that happen in front of the other people. In life-logging systems that record everything with wearable cameras and microphones, it is difficult to extract the important events from a large amount of recordings. In this research, we design and implement a smile-based life-logging system that focuses on smile / laughter for indexing the interesting / enjoyable events on a recorded video. Our system, features an original smile / laughter recognition device using photo interrupters that is comfortable enough for daily use and proposed an algorithm that detects smile / laughter separately by threshold-based clustering. The main challenge is that, since the reasons people smile and laugh are quite diverse, the system has to detect a smile / laughter as different events. Evaluation results showed that our mechanism achieved a 73% / 94% accuracy in detecting smile / laughter, while actual use of the system showed that it can accurately detect interesting scenes from a recorded life log.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous

General Terms

Human Factors

Keywords

Wearable Computing, Programming, Context Awareness

1. INTRODUCTION

The downsizing of portable computers has attracted a great deal of attention in the field of wearable computing [1]. Wearable computing can provide more effective and attractive services that conventional desktop / mobile devices. Moreover people's daily lives can be improved by wearable computing technologies because they contain extremely detailed information about their users [2, 3]. Of particular interest to the authors is the fact that we can record things 24 hours a day, 7 days a week with wearable video cameras that have a large capacity and high portability (in a process called life logging [4]). Although life log enables us to record amazing or sudden events that would not be possible without the wearable camera, it is difficult to isolate and retrieve particular events from enormous recordings. A method of indexing the recorded data is therefore required for efficient retrieval [5].

There are currently many indexing methods using various types of sensors. For example, the spread of smartphones has made obtaining GPS data much easier. There are many applications that index photos, notes, and videos with location information [6, 7, 8]. Kobayashi et al. implemented a life-logging system that indexes recorded data with the context of having a meal, going to the toilet, and smoking by using scent sensors [9]. Most of these indexing techniques are based on objective information related to biological or environmental data acquired from wearable sensors. Subjective information is also very important in life-logging systems, but it is extremely difficult to detect.

Most of these situations in which people smile occur during important and treasured events that happen in front of other people, such as participating in enjoyable activities and watching interesting things. In this research, we focus on smile / laughter for indexing interesting and enjoyable events. To the best of our knowledge, there is no research for indexing life-log data using smiles or similar indicators of positive emotions. We developed a smile / laughter recognition device that uses two photo interrupters built into glasses that can be worn with comfort in daily life since the sensors are non-contact and small enough to look natural. We also developed an algorithm that detects smile / laughter by simple threshold-based clustering. While there are many methods to detect smiles, including systems that use image processing [10] and electromyography (EMG) [11], these are not designed for use in daily life, and are typically uncomfortable or cannot be worn practically.

Since the reasons people smile and laugh are quite diverse,

AH'13, March 07 - 08 2013, Stuttgart, Germany

the system detects each as different indexes. Our system distinguishes them by the fact that laughter makes changes in both the eye and the cheek area while a smile makes changes in only the cheek area.

Prototype testing of the device showed that our system achieved 73% / 94% accuracy in detecting smile / laughter. We also implemented an application of video summarizing based on our approach. This actual-use scenario confirmed that our system could accurately detect the interesting scenes from a recorded life log.

The remainder of this paper is organized as follows. In Section 2, we discuss related work. In Section 3, we describes the design of the device and the algorithm to detect smile / laughter, and in Section 4 we explain our implementation of a prototype system. The experimental evaluation is descried in Section 5. Finally, we conclude and discuss future work in Section 6.

2. RELATED WORK

There have been several studies on detecting subjective emotions. Healey et al. analyzed four physiological signals — skin conductivity, blood volume pressure, respiration, and an electromyogram — on the masseter muscle [12]. They found that anger could easily be differentiated from peaceful emotions (90%–100%), that high and low arousal states could be distinguished (80%–88%), and that positive and negative valence states were difficult to distinguish (50%–82%). Subsets of three emotion states could be effectively separated (75%–87%), and characteristic patterns for single emotions were found. However, to the best of our knowledge, there is no research for indexing life-log data using smiles or similar positive emotional indicators.

There are several approaches for detecting smiles including algorithms that use image processing [10] or electromyography (EMG) [11]. In terms of the former, most current digital cameras have a smile shutter function that automatically takes a photo when it detects a smile [13]. As for the latter, Surakka et al. used the EMG to investigate facial and emotional reactions while viewing two different types of smiles and the relation of emotional empathy to these reactions [14]. They put several electrodes on a human face to detect the facial motions and interviewed participants to determine concentration during the smile. They reported two different types of smiles. One is a duchenne smile, which is a felt smile that moves the orbicularis oculi and zygomaticus, and the other is a non-duchenne smile, which is an unfelt smile that moves the orbicularis oculi but not the zygomaticus. Although both methods can detect smiles well, they are not suitable for everybody use because the former requires a camera in front of the face to capture it correctly and the latter requires the user to wear multiple uncomfortable devices. With our method, all the user needs to do is wear glasses that have two small photo interrupters. This fulfills the requirements for the daily use of the device.

3. SYSTEM DESIGN

Our objective with this research is to create a system that can detect smile / laughter for indexing the recorded data of a life log. We designed our system to meet the two requirements below.

(1) Detect smile and laughter separately

(a) Neutral

(b) Smile

(c) Laughter

Figure 1: Three types of facial expression.

Most previous research has treated a smile and laughter as the same human emotions. However, the meaning of a smile and laughter is quite different in daily life. A smile occurs when people are interested in what they like or when they want something in front of them. It occurs based on emotions: e.g., people think something is cute or lovely, or they have a desire for an object or event in front of them. In contrast, laughter occurs when people are obviously enjoying an event such as a conversation.

As described in Section 2, Surakka et al. reported that an unfelt smile moves the orbicularis oculi but not the zygomaticus (Fig. 1(b)) compared with an emotionless stare (Fig. 1(a)). In contrast, a felt smile moves both the orbicularis oculi and the zygomaticuszz (as shown in Fig. 1(c)) [14].

In the proposed system, we define laughter as a reaction of both the orbicularis oculi and the zygomaticus, and a smile as a reaction of only the zygomaticus. Our purpose is to classify the user state into one of three: laughter, smile, or other (such as neutral, eating, sneezing, or blinking).

(2) Be wearable for daily use

We have discussed how there are several methods to detect smiles by using image processing or EMG, but that they are unsuitable for use in daily life because the devices are uncomfortable to wear and they look unnatural.

There is a clear need for a device that can detect smile / laughter and also be worn comfortably. We designed a small device that uses two simple photo interrupters as sensors to detect smiles. These interrupters send and receive infrared light and measure depth by non-contact.

We focused on the difference of facial motion using data from the photo interrupters. The sensors are above the eye (tail of the eye) and cheek areas. Two sensors are fixed to a pair of glasses, as shown in Fig. 2. It looks natural because the attached devices are fairly small, so these glasses can be

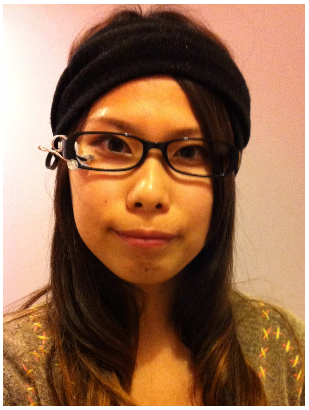
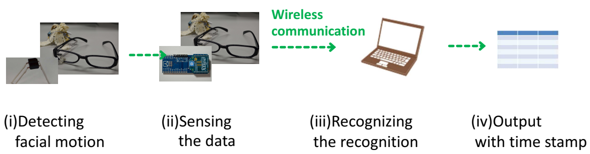
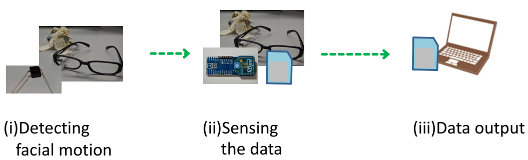


Figure 2: Snapshot of user wearing our device.



(a) the PC based system



(b) A stand alone system

Figure 3: System structure of smile-detecting device.

discretely worn in daily life.

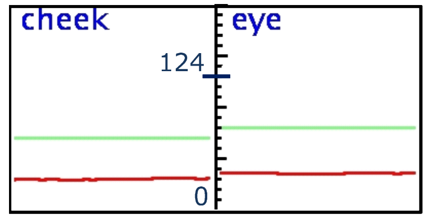
3.1 System structure

We designed two types of data logger. One is a PC-based system in which the device sends information on smiling / laughter to the PC in real time, and the other is a stand alone system in which the device stores the information in its memory and then sends the stored data in bulk when the device connects to a PC.

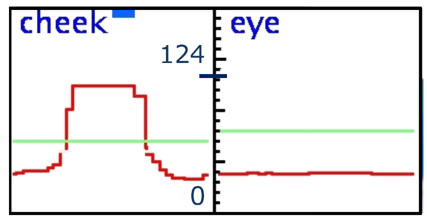
Figure 3 shows the structure of the PC-based system. It consists of the glasses equipped with two depth sensors, a microcontroller, and a PC. The two sensors detect facial motion and the microcontroller analyzes the data after it receives the information from the sensors. The sensed data are sent to the PC via ZigBee wireless communication. The recognition software on the PC recognizes a smile or laughter by referencing the sensor data and then records its results with a time stamp.

3.2 Recognition method

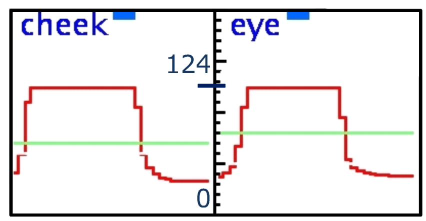
This section describes the detailed algorithm of detecting smile and laughter. Our method not only detects these two emotional indexes — it also cleans the noises that happen in everyday life. As described above, we define a smile on the basis of orbicularis oculi and zygomaticus movement and laughter on the basis of zygomaticus movement.



(a) Neutral face



(b) smile



(c) laughter

Figure 4: The depth for each facial expression.

Figure 4 shows the sequence of depth data for the neutral, smile, and laughter states. The left part of each graph shows the data from the sensor above the cheek area(in red), and the right part shows the data from the sensor above the eye area (also in red). Green lines indicate the tentative threshold to detect smile / laughter. The vertical axis indicates depth and the horizontal axis indicates time. When the interrupter gets closer than 1 mm, the depth value becomes 124, and when it gets further than 10 mm, the depth value becomes 0. The time duration shown in each graph is about three seconds.

Looking at the waveforms for smile, the value of the cheek clearly exceeds the threshold when a smile starts. Then, when the user ends the smile, the value decreases immediately. However, the value of the eye area does not change. Next, both waveforms for laughter exceed the threshold when the user starts laughing. Then, when the user stops laughing, both values decrease immediately with the same timing. There is almost no change in the status of a neutral face, which indicates that smile and laughter can be detected using the two thresholds for cheek area and eye area. To put it simply: the system recognizes a smile when only the value of the cheek exceeds the threshold, it recognizes laughter when the values of both cheek and eye exceed their thresholds, and it recognizes “other" when both values are under their thresholds.

This is the manner in which we set the basic algorithm to detect smile / laughter. We implemented an application

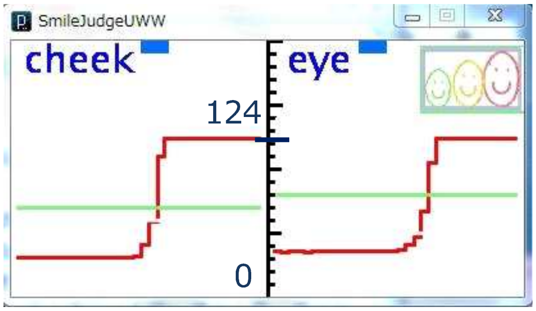


Figure 5: Application that showing waves and recognitions.

that shows the current values of the sensors and the detected facial expression in icons, as shown in Fig. 5. We used this application to investigate the appropriate position of the sensors and an algorithm to achieve high true-negative accuracy in natural situations.

3.3 Sensor placement

We used several test subjects to investigate the positions at which each sensor had the most sensitive reaction. Result showed that the appropriate position at the cheek is around the top of the cheekbone, which lifts up with a smile, and that at the eye is around the middle point of the corner of the eye, which droops with a smile. We therefore set the criterion of correct sensor placement such that each depth value for a neutral face should stay between 10 and 30, the cheek depth value of smile should be over 90, and the eye depth value of laughter should be over 90. Under this criterion, these facial expressions are clearly detected. Here, we define laughter as a loud laughing sound or obvious laughing, and a smile as little sounds or obvious smiling. The sensor placement can be adjusted to different head size and shape, body weight, etc. by using bolts. We conducted an experiment to determine if different physical characteristics among users resulted in different depth data. Eight male and four female university students took part in this experiment using the prototype device. The results showed that there is not much difference in the sequence of depth data between a male who has a big head and a female who is below average weight. This experiment demonstrated that our system can clearly and effectively identify smile / laughter regardless of user.

3.4 Padding of preliminary motion

In our original detection method, a smile is frequently detected before laughter because the value of the cheek usually exceeds the threshold before that of the cheek exceeds the threshold. However, this detection of a smile is just a preliminary motion and not an actual smile as defined above. Therefore, we add a process to change the unnecessary smile with laughter when it is a preliminary motion.

Figure 6 shows an example of someone laughing. The upper graphs are depth values for the cheek and eye areas, and the lower part shows the recognition results. As you can see in the figure, when laughter occurs after a neutral face, at first, the cheek value goes over the threshold. This causes the system to incorrectly output a smile. To prevent this, the system modifies the recognition results: if the duration of a smile between neutral and laughter is shorter than a specific time, the duration of the smile is changed to the duration of laughter.

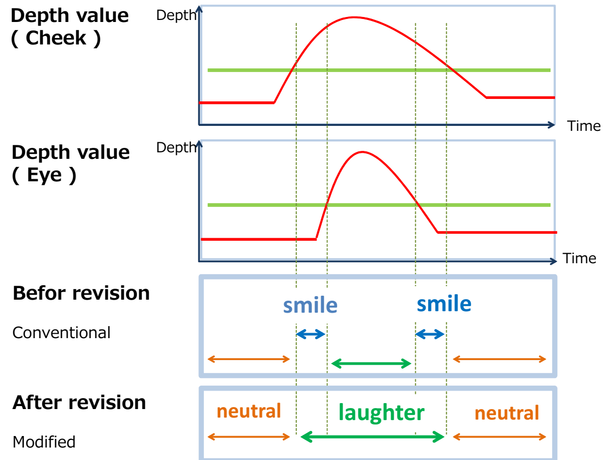


Figure 6: At beginning and end of laughter.

This same situation can also happen after laughter (also shown in the figure). The same modification can be applied at the end of laughter.

Here, laughter usually has its own reverberation and can be detect as smile. In such a case, the reverberation is objectively a smile, and the system does not modify the recognition results since the duration of the smile is longer than the threshold.

3.5 Padding of repetition in smiles

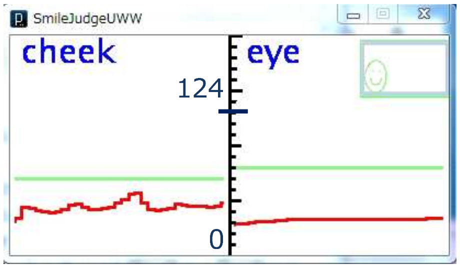
If smiles and laughter are recognized again and again at short intervals, the system recognizes that these smiles and laughter have occurred separately. However, objectively speaking, these smile and laughter stem from the same event. In other words, the same facial expression is maintained for a long time.

The system therefore clusters these smiles and laughter into one group. If smiles occur continually at short intervals, the system recognizes these smiles as belonging to one scene of an event. In concrete terms, if the duration from the end of the previous smile to the beginning of the current smile is shorter than a specific time, the system adds a smile between these two smiles. This modification is effective for extracting interesting conversational topics from a life log.

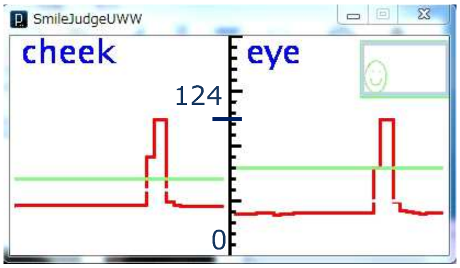
3.6 Recognition in various situations

For our method to be effective in daily life, it must have not only a high true-positive rate but also a low false-positive rate. Therefore, we conducted a preliminary experiment to estimate the ratio of false-positives. We made sure to include various motions that could be mis-identified as a smile / laughter. These include talking, winking, eating, narrowing the eyes, yawning, blinking, shaking the head, coughing, and walking. Figure 7 shows the results of actions that might be detected as a smile / laughter. With talking, the value of the cheek changed when the user pronounced "i" strongly, as shown in Fig. 7(a). In such a situation, although the value occasionally exceeds the threshold, the duration over the threshold is much shorter than that of a smile / laughter.

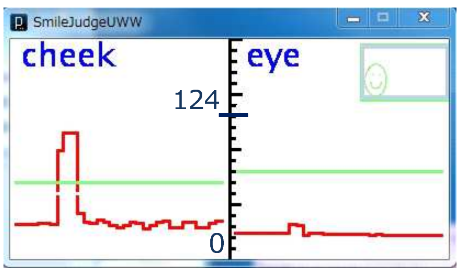
When strongly winking, there is a case of detecting a smile / laughter of very short duration, as shown in Fig. 7(b).



(a) Talking



(b) Winking



(c) Eating

Figure 7: Factors of noises.

However, the duration over the threshold is also quite short. As for eating, in most cases the value did not increase to the threshold. When users cram their mouths with food, the value of the cheek does rise, as shown in Fig. 7(c). The duration over the threshold in the case is also short.

For the other actions, there were no cases in which the value exceeded the threshold. Since smiles and laughter (our target) occur due to emotion, they tend to go on for several seconds. In contrast, the noises mentioned above only last 0.1 to 0.8 sec. Our system disregards such short durations over the threshold because it recognizes them as a noise from daily life.

4. IMPLEMENTATION

We implemented a prototype device that detects smile / laughter, shown in Figure 8. Two photo interrupters are attached to the frame of a pair of the glasses. The location of the sensors is changeable since the best position varies depending on the user. Detailed snapshots of the parts used to attach the sensors are shown in Figure 9. The sensor for the eye area is fixed to the arm via a bolt through a hole in the side of the glasses frame. The sensor for the cheek is fixed to the front of the glasses frame. The sensor for the eye area

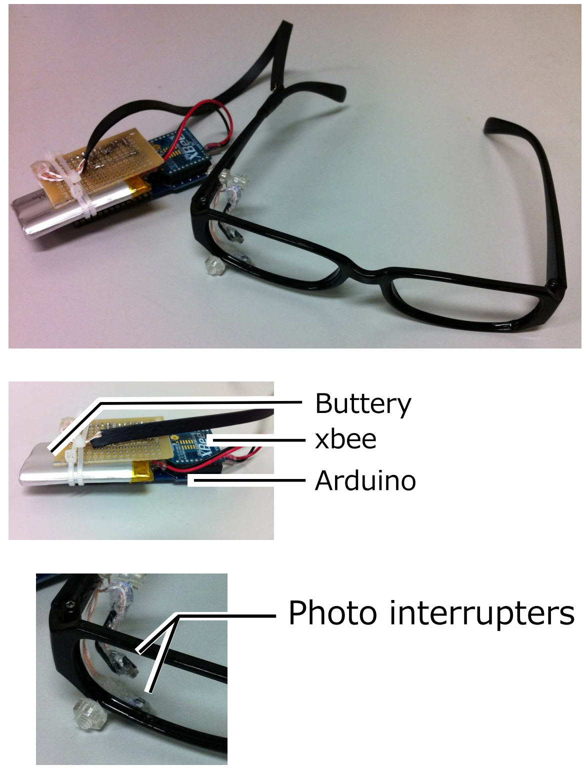


Figure 8: Prototype of device.

can be adjusted by moving it vertically and horizontally to the eye. The sensor for the cheek area can also be adjusted by moving it horizontally to the cheek. In order to adjust the depth of the sensor to users' facial shapes, we prepared two different heights of the bolt that fixes the sensor of the cheek part.

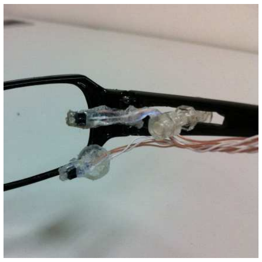
The microcontroller and communication device is attached to a hair band or other kind of accessory. Depth data from the sensors are analyzed by Arduino Fio and stored on an SD card or sent to a PC via ZigBee wireless communication using an xBee unit. The prototype has a 1100–mAh lithium polymer battery. Recognition software on the PC was implemented using Processing on Windows 7. Processing was also used to implement a video viewer for the life log video. It reads facial expression data from the recognition results and video data and then outputs the script for AviSynth. When users run the script, they can see the picked up scenes from the life log video. Those parts of the video that contain a smile / laughter duration within a 5–seconds margin before these duration are shown as the default. Users can select either a smile-based or laughter-based video and watch it to re-experience interesting events.

For using these application in daily life, users can wear a wearable camera and other sensors such as a GPS or accelerometers.

5. EVALUATION

5.1 Evaluative experiment 1

We conducted an evaluative experiment to investigate the recognition ratio. In this evaluation, laughter is defined as a smile with lifted cheeks and drooped eyes, and a smile is de-



(a) Side view (inside)

(b) Side view (outside)

(c) Front view

(d) Top view

Figure 9: The part of sensing

(a) Neutral

(b) Smile

(c) Laughter

Figure 10: Correct data (Intuitive recognition)

fined as a smile with lifted cheeks without drooped eyes. We analyzed spontaneous smiling. Participants sat on a chair and watched a video for about 10 min. The videos, which were chosen the participants themselves with the expectation they would choose something they thought would make them smile, included parts of variety shows, cats frisking about, and so on. The smiles were caused by external factors and not done on purpose. Each participant did this experiment in a classroom in which there were many other people. In this evaluation, we also consider unexpected smile / laugh that may come up in conversation. Participants performed additional actions, including eating food, talking with someone, and winking 10 times for each action so as to provide additional data.

Accurate data is based on the recording data of participants' faces and is judged objectively as a smile, laughter, or "other" every second. Figure 10 shows a snapshot of the accurate data in this experiment. We spontaneously compared the accurate data with recognition data to determine whether the smile value would be the same or not. Four university students took part in this evaluation.

The recognition ratio shown in Table 1 was calculated on the basis of the true-positive ratio, which is the recognition ratio of smiling. Results of the evaluation showed that this system identified the context of a smile with 73% accuracy and laughter with 94% accuracy. A false negative occurred while participants were laughter and smile. This occurred due to smiles that were difficult to understand such as rapid or small smiles. In this reason, these mis-detection can be considered as allowance limits of error.

Several false positive occurred when a participant was eating. This occurred because the participant opened his mouth wider than usual.

All of the obvious smiles and laughter were correctly recognized. The duration of smiles that occurred when watching a pleasant scene (in which a cute child was surprised by mischief) and of laughter when watching a funny scene (which was an important point of a comedy) were much longer than the others. This indicates a link between the importance of a scene and the duration of a smile / laughter.

All participants said that the device did not impede vision and was not uncomfortable to wear, so it is practical for use in daily life. The result of this evaluation demonstrate that our prototype is comfortable enough for daily use and useful for detecting the interesting scenes from a recorded life log.

5.2 Evaluative experiment 2

We conducted an evaluative experiment to investigate the usefulness of the proposed system.

The tester was the first author of this paper. She put on our device and a wearable camera, as shown in Fig. 11. We assumed a situation of going out with friends and took data indoors (at home and at a restaurant), outdoors (at day and at night), and on a bus. The actions performed during the experiment included eating, talking, playing a video game, walking, taking the bus, and so on.

Figures 12 and 13 shows a snapshot of the experiment. Our system made it possible to pick out the interesting or funny scenes that made the user smile, as well as topics that were not so interesting but were important for the user.

Table 1: Recognition ratio using the proposed device [%].

		Recognition data			Accuracy [%]
		Laughter	Smile	Other	
Correct data	Laughter	44	1	2	93.6
	Smile	7	47	10	73.4
	Other	0	5	117	95.9
					Total accuracy : 89.2 [%]

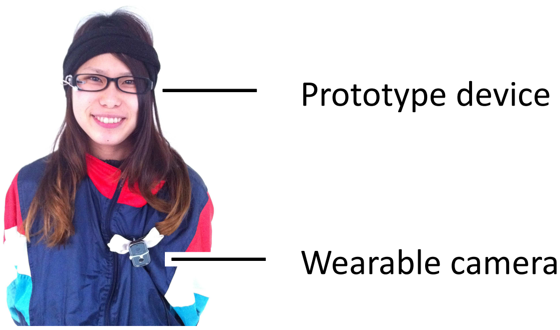


Figure 11: The user wearing our device and wearable camera

(a) Talking with a friend

(b) Talking with friends

(c) Looking at favorite thing

(d) Eating delicious food

Figure 12: Indexed as smile

Smiling continued when talking with people face to face, so only topics that were long enough to understand were picked out. There were a few short scenes picked out that were not long enough for other people to understand the conversation, but it was enough for the user because the key words of topics were picked out and she was able to remember what she had said and felt at the time. About a fifth of the most amusing things discussed during the experiment were picked out. There were some scenes in which the system could not understand the situation; however, three fourths of the scenes that had the meaning were picked out. Important scenes with long durations were usually picked out, so we can estimate the importance of a scene based on the length.

The event of greeting the bus driver was picked out, so the picked-out video seems to contains the event of meeting someone for first time. Smiling when walking or eating was picked out. A cute scene in which a cat was fawning on the user was also picked out. Nobody seemed to notice or stare at the prototype device, so it could be implemented naturally.

6. CONCLUSION

In this research, we designed and implemented a smile-based life-logging system focusing on smile / laughter for indexing the interesting / enjoyable events on a recorded video. Although systems of detecting smiles already exist, there are not wearable and are therefore difficult to use in daily life. To solve this problem, we developed a smile / laughter recognition device using photo interrupters that is comfortable enough for daily use and proposed an algorithm to detect smile / laughter separately by threshold-based clustering. We also proposed methods for modifying the recognition to correspond to objective recognition by people, distinguishing noises from smiles, and clustering several smiles together as one group. Evaluation results showed that our mechanism achieved 73% / 94% accuracy for detecting smile / laughter, and an actual use of the system confirmed that it can correctly detect interesting scenes from a recorded life log.

For future work, we plan to design additional types of data loggers and improve the recognition accuracy.

As we mentioned in Section 2, the PC-based system of data loggers that we implemented in this study requires wireless PC communication. We intend to implement a stand alone system to solve this problem. As for the recognition, we distinguished three types of face: smiling, laughing, and neutral. If we equip the device with another type of sensor, the recognition range will increase. For example, if using sound data based on a microphone, the device will be able to recognize laughing with a voice. We need to modify the algorithm accordingly.

7. ACKNOWLEDGMENTS

This section was eliminated for blind review.

8. REFERENCES

[1] M. Miyamae, T. Terada, M. Tsukamoto, and S. Nishio: Design and Implementation of an Extensible Rule Processing System for Wearable Computing, Proc. of IEEE International Conference on Mobile and Ubiquitous Systems: Networking and Services (MobiQuitous 2004), pp. 392–400 (2004).

[2] A. Cassinelli, C. Reynolds, and M. Ishikawa: Augmenting Spatial Awareness with Haptic Radar, Proc. of the 10th IEEE International Conference on Wearable Computers (ISWC 2006), pp. 61–64 (2006).

(a) Being Talked a funny story

(b) Enjoying talking with a firend

(c) Watching at funny movie

(d) Having to an unusual event

Figure 13: Indexed as laughter

[3] K. Ouchi, T. Suzuki, and M. Doi: LifeMinder: A Wearable Healthcare Support System Using User ’ s Context, Proc. of the 2nd International Workshop on Smart Appliances and Wearable Computing (IWSAWC 2002), pp. 791–792 (2002).

[4] J. Gemmell, G. Bell and R. Lueder: MyLifeBits: A Personal Database for Everything, Communications of the ACM, vol. 49, Issue 1, pp. 88–95 (2006).

[5] T. Kawamura, Y. Kono, M. Kidode: Wearable Interfaces for a Video Diary: Towards Memory Retrieval, Exchange, and Transportation, Sixth International Symposium on Wearable Computers (ISWC’02), pp. 31–33 (2002).

[6] GPS Log, http://gpslogapp.com/.

[7] NIKON IMAGE SPACE, http://nikonimagespace.com/.

[8] T. Ueoka, T. Kawamura, Y. Kono, and M. Kidode: I ’ m Here!: a Wearable Object Remembrance Support System, Proc. of the 5th International Symposium on Human Computer Interaction with Mobile Devices and Services (MobileHCI ’ 03), pp. 422–427 (2003).

[9] Y. Kobayashi, T. Terada, and M. Tsukamoto: A Context Aware System Based on Scent, Proc. of the 15th International Symposium on Wearable Computers (ISWC’11), pp. 47–50 (2011).

[10] R.Brunelli, and T. Poggio: Face recognition: features versus templates, Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 15, Issue. 10, pp. 1042–1052 (1993)

[11] A. Gruebler, and K. Suzuki: Measurement of distal EMG signals using a wearable device for reading facial expressions, Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE, pp. 4594–4597 (2010).

[12] J. Healey, R. Picard, Rm E and The Media Laboratory: Digital Processing of Affective Signals (1997).

[13] Smile Shutter, sony, http://www.sony.co.uk/.

[14] V. Surakka and J. K. Hietanen: Facial and Emotional Reactions to Duchenne and Non-Duchenne Smiles, International Journal of Psychophysiology, Vol. 29, No. 1, pp. 22–33 (1998).