温度画像処理による顔表情認識

吉冨康成*, 木村修一**, 平栄蔵**, 冨田重幸*

*:宮崎大学工学部情報工学科
〒889-21 宮崎市学園木花台西1-1
yoshi@cs.miyazaki-u.ac.jp, stomita@cs.miyazaki-u.ac.jp
**:宮崎県工業試験場機械部
〒880 宮崎市恒久1-7-14
kimura@miyazaki-iri.go.jp, hira@miyazaki-iri.go.jp

赤外線を用いた2次元温度分布画像の解析による顔表情認識手法を開発し、笑顔認識に適用した。額の高温領域の面積（S_0）と口とその周辺の高温領域の面積（S）の比（S/S_0）を画像処理により測定し、特徴量とした。Sは笑う際に増加するので、無表情の時の特徴量を1に規格化し、この規格化特徴量（SF）にしきい値を設定して笑顔を認識することとした。対象2名の計114枚の正面顔画像（無表情と笑顔）に対して実験を行ない、笑顔認識率100%を達成できた。この認識率は、対象者、画像収録日時、眼鏡の有無にほとんど影響されなかった。本法は原理的に照明条件や肌の色に対してrobustであり、感性認識にも活用される可能性をもつと考えられる。本法の任意表情認識への一般化についても見通しを得ている。

キーワード：画像解析, 表情認識, 赤外線

Facial Expression Recognition Using Thermal Image Processing

Yasunari Yoshitomi*, Shuichi Kimura**, Eizou Hira** and Shigeyuki Tomita*

*:Department of Computer Science and Systems Engineering, Faculty of Engineering, Miyazaki University
1-1 Gakuen Kibanadai Nishi Miyazaki 889-21 Japan
yoshi@cs.miyazaki-u.ac.jp, stomita@cs.miyazaki-u.ac.jp
**: Mechanics Department, Miyazaki Prefectural Industrial Research Institute
1-7-14 Tsunehisa Miyazaki 880 Japan
kimura@miyazaki-iri.go.jp, hira@miyazaki-iri.go.jp

A method is presented for recognition of facial expression. The method is based on 2-dimensional detection of temperature distribution of face, using Infrared Rays. The area ratio of hotter region consisting of mouth and its surrounding is measured to identify smiling face, using image processing. The area ratio becomes bigger in smiling. 114 image-subjects describing smiling or neutral face were used for demonstrating the feasibility of present method. The method can be applied to both cases with and without glasses. The recognition accuracy for smiling face reached to 100%. Moreover, the recognition accuracy hardly depended on both person and scene. Based on the present method, a general recognition method for any facial expression is being developed favorably, in order to provide possibility toward future human-interface systems to detect feeling.

Keywords:computer vision, image understanding, facial expression, infrared rays

1 Introduction

The present investigation is concerned with computer vision for detecting human feeling or inner mental situation through facial expression.

One target of computer vision research is to present humanlike visual ability to computer. For the target, computer should be able to visually communicate with human, mainly using recognition capability for human facial expression as a major medium for non-verbal communication[1, 2, 3, 4]. Facial expression plays an important role in communication in our daily life. The problem of facial expression recognition is one of the most attractive and tough ones in the field of pattern recognition.

For a future robot or computer system to work or live peacefully with human,it will be necessary to present capability for recognizing human feeling to a robot. Although the procedure for the characterization or identification of human faces has been received considerable attention in the course of computer vision research[1, 2, 3, 4, 5, 6],the present stage is far from goal of humanlike capability,especially from the point of understanding human feeling or mind. One reason for the difficulty is that input image from ordinary camera has a slight difference between neutral and smiling faces, for example, from the point of gray level distribution. Some researchers have tackled the issue of understanding human feeling through facial expression[1, 2, 3, 4]. However,the trials seem to be tough jobs because the gray level distribution of input image from ordinary camera has a just slight difference among various facial expressions, in addition to low contrast of facial feature-boundaries. For this reason we try to use an image describing thermal distribution of face, with Infrared Rays (IR), instead of ordinary visible ray[7].

Since the investigation presented here is the first step for understanding human feeling using computer vision with IR,image variety is limited for simplicity. The facial expression selected here is neutral or smile. The number of image-sequences is not big. The generality of the present procedure is discussed with a small speculation. Although the goal of our current study is to present computer vision enabling detection of human feeling, the future system will be also able to distinguish more precisely one person from others because thermal image of face has more fruitful information than ordinary image with visible ray.

2 Image Acquisition and Analysis System

In this study,thermal images of human faces were produced by Thermal Video System (Nippon Avionics Co.,Ltd,TVS-3500) with IR. The principle of thermal image generation comes from well-known law by Stefan and Boltzmann, which is expressed by the following equation.

$$W = \epsilon \cdot \sigma \cdot T^4 \qquad (1)$$

,where W is radiant emittance (W/cm^2),ϵ is emissivity,σ is Stefan-Boltzmann constant ($=5.6705\times10^{-12}W/cm^2K^4$),$T$ is Temperature (K).

For human skin, ϵ is estimated at 0.98 to 0.99 [8, 9]. In this study, however, the approximate value of 1 was used as ϵ for human skin, since the relative temperature distribution of face was powerful enough to present information for recognizing facial expression. The values of ϵ for almost all substances except human skin are lower than that for human skin [8, 9]. Accordingly, human face is easy to be extracted in the scene when the range of skin temperature is selected for producing the thermal image, using the value of 1 for ϵ. Face-images were observed by a monitor through an IR thermal image system with 16 (4 bits) as thermal level and recorded in 8-mm-video and then digitized with 256 levels (8 bits) per pixel. These digitized images were stored on fixed disks having a spatial resolution of 300×300 pixel elements and processed by a personal computer (NEC PC-H98S model 8).

In principle, the temperature measurement by IR dose not depend on skin color, darkness and lighting condition, resulting in that face and its characteristics are easier to be extracted in the input image containing face and its surroundings having non-humanlike temperature. Therefore, the present method for recognizing facial expression should be more general and useful than that using ordinary visible ray.

Moreover, we have verified experimentally that the thermal image by our system is not influenced by the lighting condition. Even at night, the facial expression recognition by IR image processing was performed successfully without lighting, in the same way as that with lighting.

3 Approach

3.1 Input Images

The final target in a series of our investigations is directed to development of computer capability for recognizing human feeling. However,the observed facial images were influenced by not only person, but also its face location and orientation,in addition to its mental or inner situation. In the beginning, therefore,male front-view faces with expression of smile or neutral were selected as the input images. The influences of subjects and glasses on recognition accuracy for facial expression were also investigated. The input images were produced under the operating condition which presented lower-gray-level for colder part and higher-gray-level for hotter part. Fig.1 shows typical examples of input images of smiling and neutral faces. In Fig.1, brighter point means hottter skin. Each face had thermal range by 5 to 10 K with each spatial distribution. As shown in Fig.1, it was the most marked that the area ratio of hotter region consisting of mouth and its surrounding became bigger in smiling.

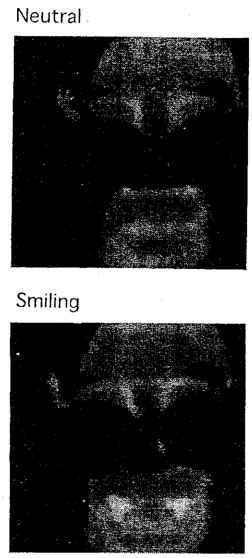

Figure 1: Typical examples of neutral and smiling faces.

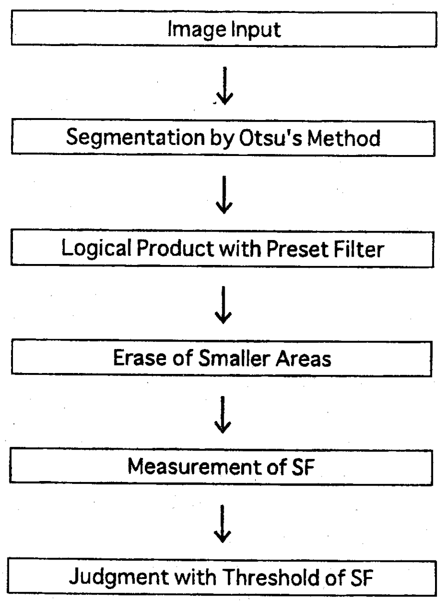

Figure 2: Flow chart for recognizing neutral and smiling faces.

3.2 Recognition Algorithm

The algorithm for recognizing smiling and neutral faces is shown in Fig.2. The procedure of image processing is as follows.

1) A filter for extracting upper reference region and lower characteristic region for a typical set of smiling and neutral faces is produced with manual operation. The filter, which is a binary image with 1 for upper reference region and lower characteristic region, 0 for other region, is basically used for every subject of recognition. The upper skin region which has little chance between smiling and neutral faces is selected as a reference. On the other hand, the lower skin region containing mouth and its surrounding whose area becomes bigger in smiling is also selected as a characteristic part for recognizing smiling face. The central region containing nose and eyes is included in neither reference nor characteristic regions since some persons wear glasses which prevent IR from penetrating them.

2) The input-gray-image is transformed into binary image with Otsu's segmentation method [10].

3)The logical product for the binary image and the filter is performed,followed by erasing smaller areas which have no meaning for recognizing smiling face.

4)The relative area (S/S_0) of the characteristic part, where S, S_0 are the area of the characteristic part, the reference part respectively, is measured, followed by the normalization that the relative area (S/S_0) is multiplied by a coefficient (C) which transforms the smallest relative area for each male neutral face in a series of scenes into unit, viz., 1. The normalization is very useful and powerful for comparing one scene with another or one person with another, since the positioning of face is not so strict, besides the variety caused by individuality on not only feature but also facial expression change in smiling. The normalized value $C(S/S_0)$ is hereinafter referred as SF (Smiling Factor).

5)The measured SF is compared with a threshold. The threshold which is the lowest SF value judged to be from smiling face is decided as an experimental parameter. Therefore, the measured SF is judged to be from smiling face, if it is not smaller than the threshold. On the other hand, it is judged to be from neutral face if it is smaller than the threshold.

Fig.3 shows examles of the filter, original, binary and extracted images. The extracted regions consist of the reference and characteristic parts which are described with white and gray, respectively.

3.3 Experimental

Image sequences of smiling or neutral faces were collected. We assembled a file of 114 pictures. Since our initial goal was to demonstrate the feasibility of present method, we made our efforts to create a relatively simple population, namely, front-view-pictures of one male face or another. The image sequences of one main male were collected for 2 days. The first day presented 3 sequences with each 12 images. The other day presented 4 sequences with 18,18,14,14 images respectively. The main male usually wears glasses. However, some sequences provided the faces without glasses intentionally. Images of another male were also collected for 1 day with 1 sequence consisting of 14 images. He did not wear glasses as usual. Beyond this, no other selection was used. The faces of the males were roughly positioned in the center of the image. However, since the restriction was not strict like mechanical positioning, some images had partial lack of face, for example. The position error of face in input image might cause deterioration of recognition accuracy for smiling and neutral faces. The sequences of images were acquired off line with 8-mm-video for IR image, followed by storing them as image files for further processing.

First, the filter described in Recognition Algorithm was produced with manual operation, based on a typical pair of smiling and neutral faces of the main male. The filter was also used for another male face. Secondly, the series of processings mentioned in Recognition Algorithm were performed for each image.

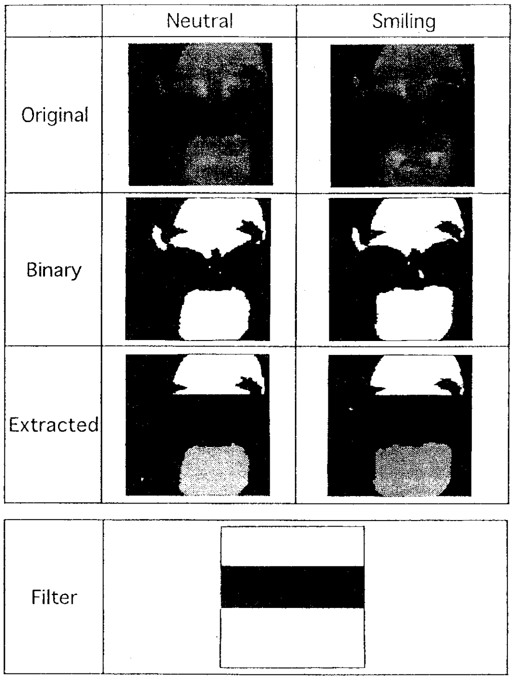

Figure 3: Examples of the filter, original, binary and extracted images of the main male.

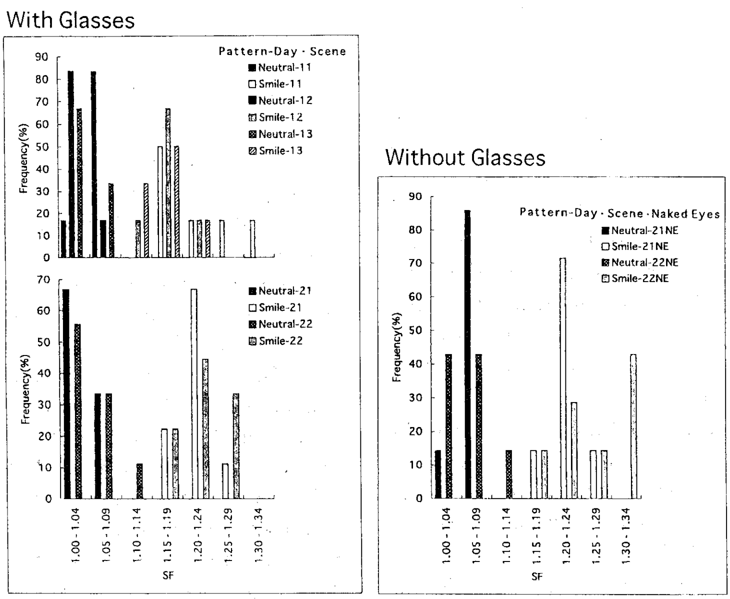

Figure 4: Distributions of SF values for the main male.

3.4 Results and Discussion

Fig.4 shows the histograms of SF value for the main male. His smiling face was perfectly recognized in terms of separate distributions of SF for smiling and neutral faces. The recognition accuracy of smiling and neutral faces hardly depended on days, scenes and glasses.

Fig.5 shows the relationship between the threshold of SF value and the recognition accuracy of smiling and neutral faces. Here,the threshold of SF means the lowest SF value judged to be from smiling face. The best threshold for recognizing smiling and neutral faces hardly depended on glasses. However,the observation day influenced the relationship between threshold of SF and recognition accuracy of smiling and neutral faces. It might be mainly caused by change of mental situation. Nevertheless,the highest recognition rate with each best threshold of SF was 100% for every sequence of image with or without glasses. Moreover, the recognition rate with the best threshold of SF, 1.10 to 1.125, was higher than 96% for all images.

Fig.6 shows examples of extracted areas, the reference (white) and characteristic (gray) regions, in face, with use of the filter produced for the other person. In this case, the facial expression of second male was analyzed, using the filter produced for the main male.

Fig.7 shows the histograms of SF value for the second male. Using the filter for the main male, the recognition rate for facial expression of second male could not reach 100% in terms of small overlap in SF distributions for smiling and neutral faces.

Fig.8 shows the relationship between the threshold of SF value and the recognition accuracy of smiling and neutral faces, related to Fig.7. The highest recognition rate with the best threshold of SF was 86%. Even when using the filter produced for the second male, the best record could not reach 100%. Therefore, the imperfect recognition is considered to be mainly caused by smaller difference between smiling and neutral faces of the second male, compared with the main male. However, as you can see from Fig.5 and Fig.8, the recognition rate with the best threshold of SF, 1.10, was higher than 86% for all images of both the main and second males.

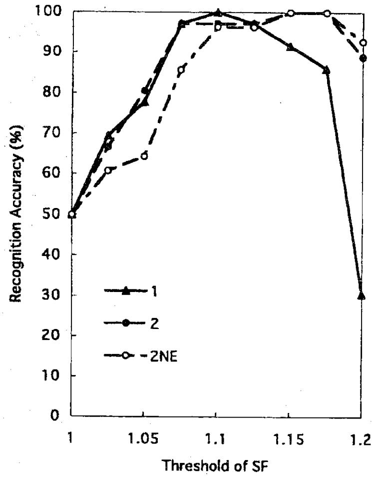

Figure 5: Relationship between the threshold of SF value and the recognition accuracy of neutral and smiling faces for the main male. 1,2 and 2NE indicate the first, second and second (naked eyes) image-recording-day, respectively.

Based on the present method, a general recognition method for any facial expression is being developed favorably, in order to provide possibility toward future human-interface systems to detect feeling. The method consists of the two parts. The first part is the face identification by IR image analysis with both Back Propagation Neural Network and conventional pattern recognition. The second part is the facial expression recognition by IR image analysis with Back Propagation Neural Network on the assumption that the face has been identified. In a forthcoming paper, a fruitful detail of the study will be presented. As the next step, it is necessary to collaborate with psychologists and construct data-bases for recognizing human feeling more generally through facial expression described with IR.

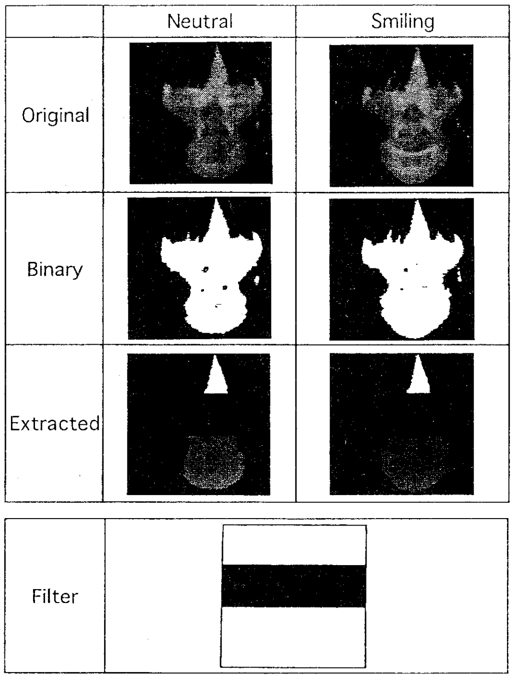

Figure 6: Examples of the original, binary and extracted images , using the filter produced for the other person.

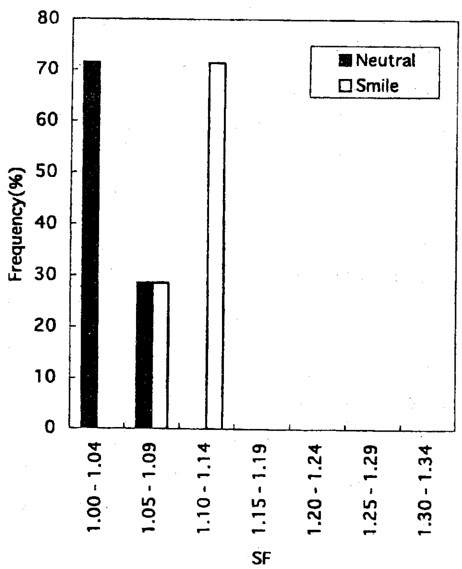

Figure 7: Distributions of SF values for the second male.

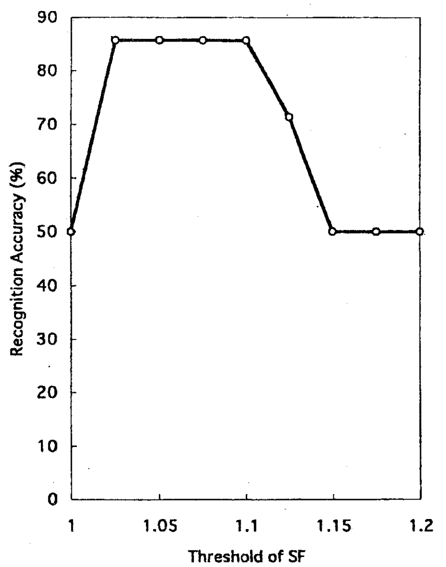

Figure 8: Relationship between the threshold of SF value and the recognition accuracy of neutral and smiling faces for the second male.

4 Conclusion

A method is presented for recognition of facial expression. The method is based on 2-dimensional detection of temperature distribution of face, using Infrared Rays. The area ratio of hotter region consisting of mouth and its surrounding is measured to identify smiling face, using image processing. The area ratio becomes bigger in smiling. 114 image-subjects describing smiling or neutral face of the main and second males were used for demonstrating the feasibility of present method. The method can be applied to both cases with and without glasses. The recognition accuracy for smiling and neutral faces reached to 100% in the case with the best threshold of SF value for the main male. The recognition rate with the best threshold of SF value was higher than 86% for all images of both the main and second males. Moreover, the right recognition rate hardly depended on scene,despite of rough face positioning in image.

References

[1] Sakai K.,Nagao M. and Kanade T. :Computer Analysis and Classification of Photographs of Human Faces,*In First USA-JAPAN Computer Conference*,volume session 2–7 (1972).

[2] Yuille A.L.,Cohen D. S. and Hallinan P. W :Feature Extraction from Faces Using Deformable Templates,*Proc. of IEEE Conf. on Computer Vision and Pattern Recognition* ,pp. 104–109 (1989).

[3] Mase K.:An Application of Optical Flow - Extraction of Facial Expression, *IAPR Workshop on Machine Vision and Application*, pp. 195–198 (1990).

[4] Matsuno K.,Lee C.and Tsuji S. :Recognition of Facial Expressions Using Potential Net and KL expansion, *Trans. IEICE*, Vol. J77-D-II, No. 8, pp. 1591–1600 (1994). (in Japanese).

[5] Sirovich L. and Kirby M.:Low Dimensional Procedure for the Characterization of Human Faces, *J. Optical Society of America* ,Vol. 4, No. 3, pp. 519–524 (1981).

[6] Turk M.A. and Pentland A.P.:Face Recognition Using Eigenfaces,*Proc. of IEEE Conf. on Computer Vision and Pattern Recognition* , pp. 586–591 (1991).

[7] Yoshitomi Y.,Kimura S.,Hira E. and Tomita S.:Facial Expression Recognition Using Infrared Rays Image Processing,*Proc. of the Annual Convention IPS Japan*, No. 2, pp. 339–340 (1996.9).

[8] Kuno H.:Sekigaisen Kougaku,pp. 22,IEICE,Tokyo (1994). (in Japanese).

[9] Kuno H.:Sekigaisen Kougaku,pp. 45,IEICE,Tokyo (1994). (in Japanese).

[10] Otsu N.:A Threshold Selection Method from Gray-Level Histograms, Trans. IEEE, Vol. SMC-9, No. 1, pp. 62–66 (1979).