温度画像処理による顔識別・表情認識

吉冨康成*, 宮浦智範*, 宮脇伸枝*, 冨田重幸*, 木村修一**

*: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電子システム科
〒880 宮崎市恒久 1-7-14

赤外線画像処理を用いた顔識別法と顔表情認識法を開発した。顔温度分布ヒストグラムと顔部分平均温度をニューラルネットワークの入力とし、ＢＰ法での学習を基に、未知の顔を識別する方法と顔形状因子（円形度, 縦横比）を基にした教師付き分類法の結果を統合することで、良好な顔識別が可能となった。更に、平均無表情温度画像と未知表情温度画像との局所的濃度差（温度差）をニューラルネットワークの入力とし、ＢＰ法での学習を基に、未知表情画像を解析することにより、無表情、幸福、驚き、悲しみを９０％の認識率で識別できた。

キーワード：画像解析, 顔識別, 表情認識, 赤外線, 温度画像, ニューラルネットワーク, 教師付き分類

Face Identification and Facial Expression Recognition Using Thermal Image Processing

Yasunari Yoshitomi*, Tomonori Miyaura*, Nobue Miyawaki*, Shigeyuki Tomita* and Shuichi Kimura**

*: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

A method for face identification has been developed by exploiting Infrared Rays Image Processing Technique. The method is based on 2-dimensional detection of temperature distribution of face. The temperature distribution and the locally averaged temperature are used as input for Neural Network, while the value of shape factors are used for supervised classification. By integrating information from Neural Network and supervised classification, the face was identified with excellent accuracy. Moreover, a method for recognition of facial expression has been developed with Infrared Rays Image Processing Technique. The local temperature-difference between the averaged neutral and the unknown expression faces is used as input for Neural Network. By the present method, neutral, happy, surprise and sad expressions were recognized with 90% accuracy.

Keywords:computer vision, image understanding, face identification, facial expression recognition, infrared rays, neural network, thermal image,supervised classification

1 Introduction

The present investigation is concerned with computer vision for identifying human face and detecting human feeling or inner mental situation through facial expression. For a future robot or computer system to work or live peacefully with human, it will be necessary to present capability for recognizing human face and human feeling to a robot. Although the procedure for recognizing face or facial expression has been received considerable attention in the course of computer vision research[1, 2, 3, 4, 5, 6, 7, 8, 9],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, 3, 4, 5, 6, 8, 9]. 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 at feature-boundaries on face.

Recently, several methods for face identification by visible rays and image analysis have been developed. However, it is essentially difficult that the face is identified under wide-ranged lighting condition. For this reason we try to use an image describing thermal distribution of face, with Infrared Rays (IR), instead of ordinary visible ray. Human face is easy to be extracted in the scene when the range of skin temperature is selected for producing the thermal image with IR, using the value of 1 as emissivity[10, 11].

For recognizing facial expression, we have been also using an image describing thermal distribution of face, and reported that neutral and smiling faces were distinguished with excellent accuracy[10, 11].

In the field of psychology, FACS(Facial Action Coding System)[12] to analyze facial expression was developed, resulting in that it became to be possible for computer graphic researcher to analyze facial expression as a subject. Some investigations for describing facial expression through feature deformation, mainly based on Action units (AUs:elements of expression in FACS) or other description, have been reported[4, 5, 8].

The goal of our current study is to present computer vision enabling identification of human face and detection of human feeling. For reaching the goal, it is necessary to collaborate with psychologists and construct data-bases for distinguishing one facial expression from others. However, since the investigation presented here is in the first stage for understanding human feeling with IR computer vision, image variety is limited for simplicity. The facial expressions selected here are neutral, happy, surprise and sad. The number of image-sequences is not big. The generality of the present procedure is discussed with a small speculation. In the course of IR image analysis for recognizing human information, in the present study, the thermal image processing, Neural Network (NN) and supervised classification are used for face identification. Moreover, the thermal image processing and NN are used for recognizing more kinds of facial expressions, generally.

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.

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

,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 [13, 14]. In this study, however, the approximate value of 1 was used as ϵ for human skin. The values of ϵ for almost all substances except human skin are lower than that for human skin [13, 14]. 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. The input images were produced under the operating condition which presented lower-gray-level for colder part and higher-gray-level for hotter part.

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 surrounding.

Moreover, we have verified experimentally that the thermal image by our system is not influenced by the lighting condition[11]. Even at night, the face identification and the facial expression recognition by IR image processing were performed successfully without lighting, in the same way as that with lighting.

3 Recognition Algorithm

3.1 Face Identification

Figure 1 shows the flowchart for face identification.

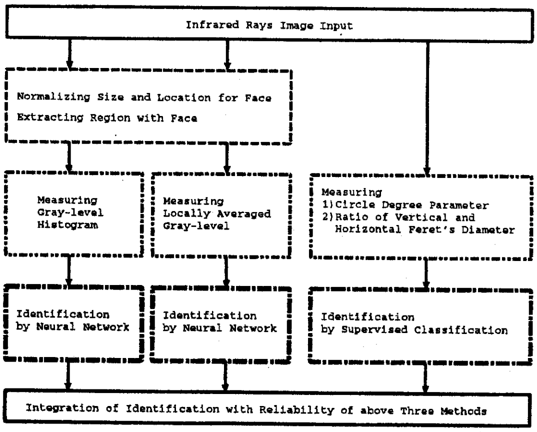

Figure 1: Flowchart for face identification.

The face is finally identified through integrating information from NN for gray-scale histogram of face, NN for mosaic image of face, and supervised classification for shape factors of face.

3.1.1 Face Identification with NN for Gray-scale Histogram

The flowchart for extracting the block image-region with the normalized face is shown in Figure 2. Firstly, the input image of front-view neutral face is normalized in terms of the size and the center of gravity for the face. Then, an image-region with the normalized face is extracted.

The gray-level histogram for the block image-region is measured for input to NN with three layers. With Back Propagation (BP) method[15], the face is recognized. The double value for total number of quantized gray-level in histogram is used as

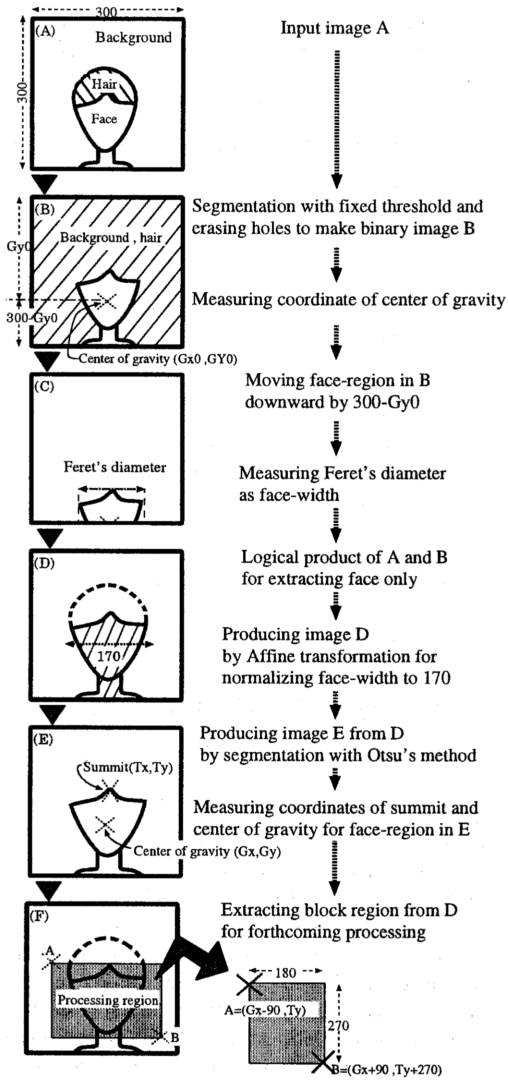

Figure 2: Flowchart for extracting the block image-region with the normalized face.

the unit number of input layer. The unit number of hidden layer is decided experimentally for improving the recognition accuracy for the face. The unit number of output layer is the number of the faces which should be recognized. The values of histogram frequency normalized by corresponding average value of learning data are transformed to 2 bits per data for input to NN. The conditions for making 0,1 data are decided experimentally for improving the recognition accuracy.

3.1.2 Face Identification with NN for Mosaic Image

The block image with the normalized face is transformed to a mosaic image made by locally averaged gray-level at $n \times m$ pixel elements. The locally averaged gray-level is measured for input to NN with three layers. With BP method, the face is also recognized. The double value for total number of blocks in the mosaic image is used as the unit number of input layer. The unit number of hidden layer and that of output layer are decided in the same way as mentioned in Face Identification with NN for Gray-scale Histogram. The values of locally averaged gray-level normalized by corresponding average value of learning data are transformed to 2 bits per data for input to NN. The conditions for making 0,1 data are decided experimentally for improving the recognition accuracy.

3.1.3 Face Identification with Supervised Classification for Shape Factor

The input image of the face is segmented with Otsu's method[16], followed by measuring SF1: the circle degree parameter calculated with area and perimeter, and SF2: the ratio of vertical and horizontal Feret's diameters. Using Euclidean distance as discriminant function where the element of characteristic vector is normalized by the corresponding average value of learning data, the face is recognized with supervised classification through comparing the test data with the averaged data of each learned class.

3.1.4 Integrated Face-Identification Method

The face is finally identified by integration of face-identification where the summation of recognition results of three different methods mentioned above is performed to select one face having the highest value among the summations for faces. Each recognition-result corresponds to the value of reliability of each method for identified face, 0 for other faces. The recognition accuracy in case of applying each method to learning data themselves is used as the value of reliability for each method.

3.2 Facial Expression Recognition

Front-view faces were used as the input images. Figure 3 shows the flowchart for recognizing facial expression. The algorithm for recognizing facial expression is as follows.

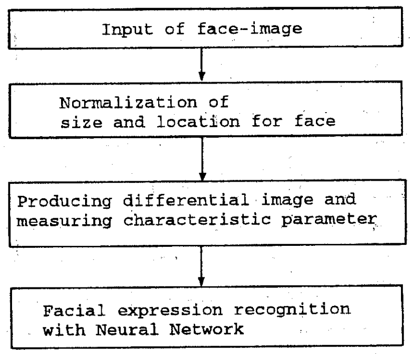

Figure 3: Flowchart for recognizing facial expression.

3.2.1 Normalization of Size and Location

The face-image is recorded under the natural condition without any special restriction for observed person. For this reason, as shown in Figure 4, the input image is normalized in terms of the size and the center of gravity for the face ,through performing segmentation with Otsu's method, measuring the horizontal Feret's diameter and the center of gravity for the binary image, and performing Affine transformation for the input image in order to normalize the horizontal Feret's diameter and the center of gravity for the binary image.

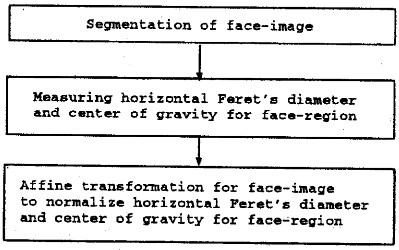

Figure 4: Flowchart for normalizing the size and location for the face.

3.2.2 Producing the Differential Image and Measuring Characteristic Parameter

Figure 5 shows the flowchart for producing the binary image to measure the characteristic parameter for recognizing facial expression. Beforehand, an averaged neutral face-image is produced as a reference image A. Then, with an unknown image B and an averaged neutral face-image A, the differential images A-B, B-A are made under the condition that the negative value made by the differential calculation is transformed to 0. Then, for both images A-B and B-A, a processing α of (1) smoothing with 3×3 median filter, and (2) segmentation with level 2 and level 3 as threshold, is performed, where level 2 and level 3 correspond to the double and triple values for thermal difference unit in IR apparatus, respectively. Moreover, for image A-B, a processing β of (1) contrast enhancement with gray-scale transformation, (2) smoothing 5 times with 3×3 median filter, (3) segmentation with Otsu's method ,and (4) erasing region with small area is performed.

Totally, five binary images are produced with above procedure. Then, for the binary images made by the processing α, the measured area fraction in each block demonstrated in Figure 6 (α) is divided by the corresponding average value of learning data. Then, for the binary image made by the processing β, the smaller area fraction in two area fractions of blocks demonstrated in Figure 6 (β) is divided by the corresponding average value of learning data. The 17 normalized area fractions are used as the values of characteristic parameters for facial expression.

Each block demonstrated in Figure 6 (α) is decided, based on the studies in the fields of psychology, especially FACS-AU[12], and computer vision

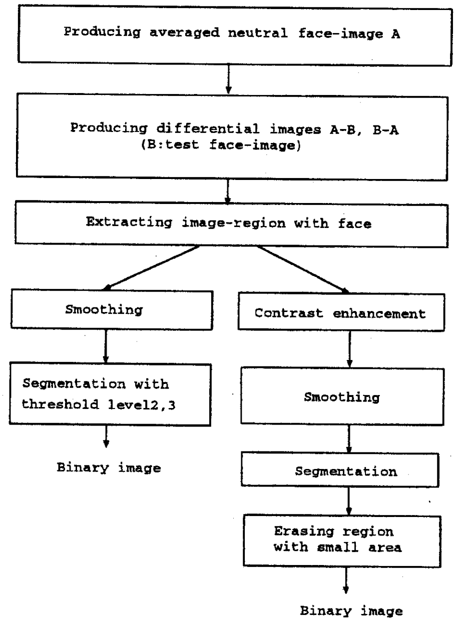

Figure 5: Flowchart for producing the binary image α(Left), β(Right),to measure the characteristic parameter for recognizing facial expression.

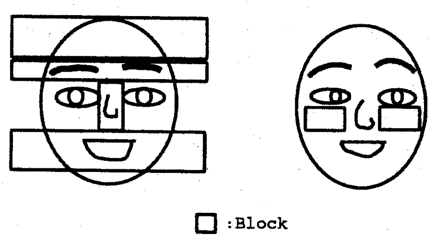

Figure 6: Blocks for measuring characteristic parameters.(α):Left,(β):Right.

with visible ray[4, 6, 9], while each block demonstrated in Figure 6 (β) is decided experimentally for improving the recognition accuracy.

3.2.3 Recognition with Neural Network

The 16 values of normalized area fraction from the processing α are transformed to 2 bits per data. The value of normalized area fraction from the processing β is also transformed to 2 bits. The condition for making 0,1 data for the characteristic parameters is decided experimentally for improving the recognition accuracy.

The 34 bits data which are transformed from the 17 values of characteristic parameters are used as input data for NN with three layers. With BP method, the facial expression is recognized. The unit number of input layer is 34. The unit number of hidden layer is decided experimentally for improving the recognition accuracy for facial expression. The unit number of output layer is the number of facial expressions which should be recognized.

4 Experiment and Discussion

4.1 Face Identification

For evaluating the recognition accuracy of the present method, face-image sequences of man A, man A with glasses, man B and woman A were collected. The total number of images was 40 for learning data, 80 for test data. Since the recognition accuracy of the present method might be influenced by room temperature, the two conditions for room temperature were collected. The room temperature in recording all learning-images and the half test-images was about 302K ,while that for the other half test-images was about 285K. For extracting face-image successfully, the detected temperature range by IR was selected as 303 to 307.5K for the room temperature of 302K, 298.5 to 303K for that of 285K. The unit number of input, hidden and output layer of NN for gray-scale histogram made by 10 gray-scale pitch was selected as 52, 22 and 4, respectively. On the other hand, the unit number of input, hidden and output layer of NN for mosaic image made by locally averaged gray-scale at 30 $\times$ 30 pixel elements was 108, 54, 4, respectively. The condition for transforming the characteristic parameters to 2 bits per data is that the value smaller than 0.5 is transformed to 00, the value from 0.5 to smaller than 1.0 is transformed to 01, the value from 1.0 to smaller than 1.5 is transformed to 10, the value bigger than 1.5 is transformed to 11. Figure 7 shows each example for 4 face-images.

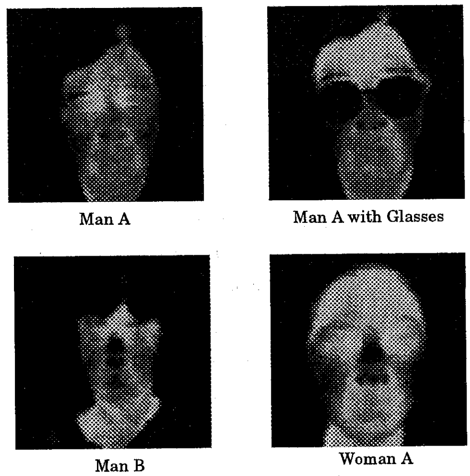

Figure 7: Examples of faces.

The value of reliability of face-identification-method was 1 for every three method. When the room temperature for test-image was the same as that for learning-image, the integrated accuracy for identifying 40 faces was 100%. On the other hand, when the room temperature for test-image was different from that for learning-image by 17 K, the integrated accuracy for identifying 40 faces was 87.5%. To improve the integrated accuracy of the second case, the input gray-scale histogram for the 40 test-image was normalized in terms of the same average of the histgram as that for learning-data, resulting in that the integrated accuracy for face identification was improved to 97.5%. The poor recognition with 2.5% was happened by abnormally hotter face of man A, caused by local heating especially at forehead. The exception is considered to be caused by rare feeling transition.

The present result was for four kinds of face-images. The present recognition accuracy for four kinds of face-images was excellent. When the number of persons which should be recognized is increased, the modification of image processing, NN and supervised classification might be necessary for improving the recognition accuracy.

4.2 Facial Expression Recognition

For evaluating the recognition accuracy of the present method, image sequences of neutral, happy, surprise and sad faces of one female were collected. We assembled 10 images per facial expression as learning data and 20 images per facial expression as test data. The total number of images was 40 for learning data, 80 for test data. Then, the averaged face-image produced from 10 neutral faces was used as a reference image A. The room temperature in recording all learning-images and all test-images was about 302K. For extracting face-image successfully, the detected temperature range by IR was selected as 303 to 307.5K. Facial expression was made by intentional action and the judgement as right answer for the facial expression was from the observation for the image by herself. The unit number of input, hidden and output layer of NN was selected as 34, 17 and 4, respectively. Figure 8 shows each image example for 4 facial expressions.

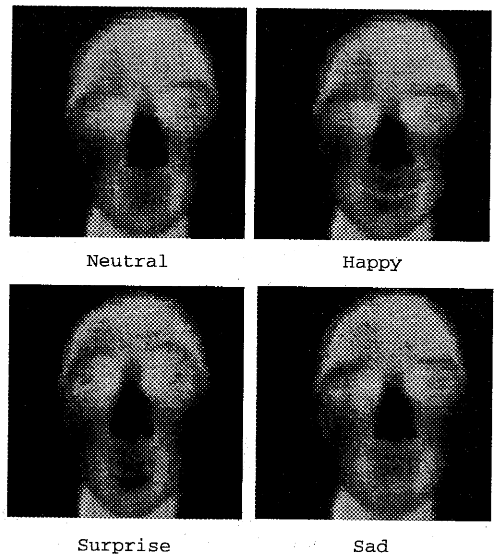

Figure 8: Examples of facial expressions.

The 16 values of normalized area fraction from the processing α are transformed to 2 bits per data under the condition that the value smaller than 0.5 is transformed to 00, the value from 0.5 to smaller than 1.0 is transformed to 01, the value from 1.0 to smaller than 2.0 is transformed to 10, the value bigger than 2.0 is transformed to 11. Then, the value of normalized area fraction from the processing β is transformed to 2 bits under the condition that the value smaller than 0.5 is transformed to 00, the value from 0.5 to smaller than 0.8 is transformed to 01, the value from 0.8 to smaller than 1.2 is transformed to 10, the value bigger than 1.2 is transformed to 11. Table 1 shows the recognition accuracy of the present experiment. For happy and surprise, we got the excellent accuracy of 95 to 100 %, while, for neutral and sad, we got the good accuracy of 80 to 85 %. The totally averaged accuracy was 90 %. Moreover, through consideration on the poorly recognized data, it was found that the poorly recognized face had different characteristics from the normal input face in terms of thermal distribution. Since the input image was made by intentional action, the poorly recognized data might be caused by imperfect action. In addition, since the facial expression judgement as right answer was made by herself, the poor recognized data might be caused by her judgement with some ambiguity. However, the recognition accuracy of 90 % might show that the present method for recognizing facial expression has the potential almost equal to human.

		Input emotion			
		Neutral	Happy	Surprise	Sad
Recognized emotion	Neutral	80 %			15 %
	Happy		95 %		
	Surprise			100 %	
	Sad	20 %	5 %		85 %

Table 1: Recognition accuracy for facial expression.

The present result was for one person. However, since the face identification method with IR image analysis has been developed, the data base of characteristic parameter for facial expression which can be made for each person will be available. Namely, after identification of the person, the facial expression can be recognized with the data base.

In the present method, the supervised learning was used. For the person and/or facial expression without learning data, the averaged face and/or facial expression can be useful for recognizing facial expression. Moreover, the averaged learning data or characteristic data can be also useful for recognizing facial expression.

Moreover, using NN, the intensity of facial expression will be defined for recognizing facial expression in detail.

5 Conclusion

A method for face identification has been developed by exploiting Thermal Image Processing Technique. The method is based on 2-dimensional detection of temperature distribution of face, using IR. The front-view face in input image is normalized in terms of location and size, followed by measuring the temperature distribution, the locally averaged temperature and the shape factor of face. The measured temperature distribution and the locally averaged temperature are used as input data for NN, while the value of shape factors are used for supervised classification. By integrating information from NN and supervised classification, the face was identified with excellent accuracy.

A method for recognition of facial expression has been also developed with Thermal Image Processing Technique. The method is also based on 2-dimensional detection of temperature distribution of face. The front-view face in input image is normalized in terms of size and location, followed by measuring the local temperature-difference between the averaged neutral and the unknown expression faces. The measured local temperature-difference caused by the rearrangement of face muscle and the inner temperature change is used as input data for NN. Neutral, happy, surprise and sad expressions were recognised with 90% accuracy.

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] 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).

[3] 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).

[4] Harashima H.,Choi C.S. and Takebe T.:3-D Model-Based Synthesis of Facial Expressions and Shape Deformation, *Human Interface*,Vol. 4, pp. 157-166 (1989). (in Japanese).

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

[6] Mase K.:Recognition of Facial Expression from Optical Flow,*Trans.IEICE*, Vol. E74, No. 10, pp. 3474–3483 (1991).

[7] 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).

[8] 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).

[9] Kobayashi H. and Hara F.:Analysis of Neural Network Recognition Characteristics of 6 Basic Facial Expressions,*Proc. of IEEE Internat. Workshop on Robot and Human Communication*, pp. 222–227 (1994).

[10] 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).

[11] Yoshitomi Y.,Kimura S.,Hira E. and Tomita S.:Facial Expression Recognition Using Thermal Image Processing,*IPSJ SIG Notes* , CVIM103-3, pp. 17–24 (1997).

[12] Ekman P. and Friesen W.V.: Facial Action Coding System, Consulting Psychologists Press,Inc.,San Francisco,CA (1978).

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

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

[15] Rumelhart D.E., Hinton G.E., Williams R.J.:Learning internal representations by error propagation, *Parallel Distributed Processing:Explorations in the Microstructures of Cognition*, Vol. 1, MIT Press, pp. 318-362 (1986).

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