Spatial normalization of facial thermal images using facial landmarks

Kent Nagumo¹ · Kosuke Oiwa¹ · Akio Nozawa¹

Received: 15 April 2021 / Accepted: 21 September 2021 / Published online: 9 October 2021
© The Author(s) 2021

Abstract

Human–computer interaction (HCI) is an interaction for mutual communication between humans and computers. HCI needs to recognize the human state quantitatively and in real-time. Although it is possible to quantitatively evaluate the human condition by measuring biological signals, the challenge is that it often requires physical constraints. There is an increasing interest in a non-contact method of estimating physiological and psychological states by measuring facial skin temperature using infrared thermography. However, due to individual differences in face shape, the accuracy of physiological and psychological state estimation using facial thermal images was sometimes low. To solve this problem, we hypothesized that spatial normalization of facial thermal image (SN-FTI) could reduce the effect of individual differences in facial shape. The objective of this study is to develop a method for SN-FTI and to evaluate the effect of SN-FTI on the estimation of physiological and psychological states. First, we attempted spatial normalization using facial features. The results suggested that SN-FTI would result in the same face shape among individuals. Since there are individual differences in facial skin temperature distribution, the inter-individual correlation coefficient is suggested to be lower than the intra-individual correlation coefficient. Next, we modeled the estimated drowsiness level using SN-FTIs and compared it with Normal. The results showed that SN-FTI slightly improved the discrimination rate of drowsiness level. SN-FTIs were suggested to reduce the effect of individual differences in facial structure on the estimation of physiological and psychological states.

Keywords Facial thermal image · Spatial normalization · Drowsiness estimation · Infrared thermography

1 Introduction

Human–computer interaction (HCI) is an interaction for mutual communication between humans and computers and is expected to be used in various fields such as medicine, welfare, and industry. HCI requires quantitative and real-time recognition of the human condition. Physiological indices with these characteristics are useful as indices for computers to recognize the human state. The measurement of biological signals often requires physical restraint, and the measure itself may cause mental and physical stress to the subject [16]. Conversely, infrared thermography can measure skin temperature non-contact and non-invasively

with high sensitivity, accuracy, and reproducibility [2, 12, 19, 25]. There have been many physiological and psychological studies based on time-series changes in the facial skin temperature [9, 13]. For example, physiological and psychological states such as respiration rate [17], heart rate [11], emotion [7, 21], drowsiness [5], and mental stress [8, 23] have been estimated. In these studies, biometric information has been estimated from the amount of change in temperature values within regions of interest, such as the nose, mouth, and cheeks.

Studies using the spatial characteristics of facial skin temperature distribution have also been conducted. Adachi et al. attempted to construct a model to estimate the level of drowsiness from a single facial thermal image (FTI) using CNN, a form of deep learning [1]. The individual model had a 70–90% discrimination rate for estimating the three levels of drowsiness. However, because of the individual differences in facial contours, a general model for estimating drowsiness levels could not be constructed. Furthermore, there was an issue that FTI was affected by the angle at which the thermal image was taken.

This work was presented in part at the 26th International Symposium on Artificial Life and Robotics (Online, January 21–23, 2021).

✉ Kent Nagumo
nagumo@ieee.org

¹ Aoyama Gakuin University, 5-10-1 Fuchinobe, Chuo-ku, Sagami-hara, Kanagawa, Japan

In a study of the human brain atlas, spatial normalization is performed on PET and MRI images. The human brain varies in size and shape, making it challenging to compare regions between subjects. Spatial normalization is performed to transform the subject's brain scan image into regions of the template brain scan image to reduce individual brain differences [3, 4, 6, 10, 18, 22, 24]. It was hypothesized that the spatial normalization of the facial thermal image (SN-FTI) would reduce the effect of individual differences in face shape. The objective of this study is to develop a method for SN-FTI and to evaluate the effect of SN-FTI on the estimation of physiological and psychological states. In this paper, first, spatial normalization of facial thermal images using facial landmarks was attempted. Next, the SN-FTI was used to estimate the level of drowsiness and compared to the normal FTI. To the best of our knowledge, this paper proposes the first attempt at spatial normalization focused on facial thermal images.

2 Spatial normalization of facial thermal images (SN-FTI)

2.1 Methods

The SN-FTI was performed by the following three steps: (1) defining the coordinates of the template facial landmarks, (2) defining the regions to be transformed according to the Delaunay diagram using the facial feature points, and (3) affine transformation of each region defined in (2) to the coordinates of the reference facial landmarks.

In step (1), the FTIs were measured from the front side using infrared thermography (FLIR A615—Model: A615, FOV 45°, FLIR Systems, Oregon). Nineteen subjects between the ages of 21 and 24 years participated in the measurement of FTIs. They were fully informed about the experiment and the purpose of the study before their participation. All participants signed a consent form. As shown in Fig. 1, the measured FTIs were manually annotated with the 68 facial landmarks used in the study by Nagumo et al. The coordinates of each facial landmark were scaled so that the height of the feature point was 130 pixels. The coordinates of the template facial landmarks were taken as the average of the scaled coordinates of transformed facial landmarks for all subjects. The facial landmarks were extracted using the face alignment method [20].

In step (2), the coordinates of the template's facial feature points were used to create a Delaunay diagram, as shown in Fig. 2. The triangular region in the Delaunay diagram is the region to be transformed by spatial normalization.

In step (3), for each of the regions defined in step (2), an affine transformation to the coordinates of the template's facial feature points was performed, and spatial normalization was

Fig. 1 Example of 68 landmarks in a facial thermal image

Fig. 2 Delaunay diagram based on the template facial landmarks

performed. The affine transformation can rotate, scale, and translate the image. The equation for affine transformation is shown as follows:

$$\begin{bmatrix} y' \\ x' \end{bmatrix} = A \begin{bmatrix} y \\ x \end{bmatrix} + b \quad (1)$$

$$= \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \quad (2)$$

where x, y are the coordinates before the transformation, and x', y' are the coordinates after the transformation. The values of the matrix in A are parameters related to the rotation and scaling of the image, and the values of the vector in B are parameters related to the translation of the image. The coordinates of the three vertices of the triangular region defined in step (2) were used to calculate A and B .

2.2 Evaluation

2.2.1 Evaluation methods

To evaluate the accuracy of SN-FTI, an experiment was conducted to measure the FTI from two different angles. Seven subjects (five males and two females) aged 22–24 years participated in the experiment. Figure 3 shows the experimental system. They were fully informed about the experiment and the purpose of the study before their participation. All participants signed a consent form. Infrared thermography (FLIR A615—Model: A615, FOV 45°, FLIR Systems, Oregon) was used to capture the FTIs. The distance between the subject and the infrared thermography was set at 60 cm. The thermography resolution was 640×480 pixels, and the infrared emissivity of the skin was set to 0.98. The

Fig. 3 Experimental system

experiment consisted of two measurement sections (Small and Large). The subjects were asked to turn their heads in nine directions (center, top center, top right, center right, bottom right, bottom center, bottom left, center left, and top left). Subjects were asked to move their head angles to 20° and 45° in the Small and Large conditions, respectively.

The intra-individual correlation coefficients and inter-individual correlation coefficients were used to evaluate the accuracy of SN-FTIs. Spatial normalization was performed on the FTIs measured in the above experiments. When calculating the intra-individual correlation coefficients, the correlation coefficients between the center SN-FTIs and SN-FTIs of the other eight poses were calculated for each condition (Small and Large) and each subject's data. When calculating the inter-individual correlation coefficients, the correlation coefficients for each subject's center SN-FTIs were calculated. Correlation coefficients were calculated using Pearson's product ratio correlation coefficient.

2.2.2 Results and discussion

This section describes the results of the experiments in Sect. 2.2.1. Examples of SN-FTIs for Small and Large conditions are shown in Fig. 4. Figure 5 shows the average intra-individual correlation coefficient, which was 0.77 for the Small condition and 0.61 for the Large condition. It is suggested that the further the face angle is from the center, the less accurate the spatial normalization becomes. Table 1 shows the inter-individual correlation coefficients for the center SN-FTIs. All inter-individual correlation coefficients

Fig. 4 Example of SN-FTI

Fig. 5 Mean intra-individual correlation coefficients of SN-FTIs ($N = 7$). Error bars represent the standard error

calculated for the frontal images of each subject were above 0.38, indicating a moderate correlation. Therefore, it was suggested that SN-FTI would result in the same face shape among individuals. Since there are individual differences in facial skin temperature distribution, the inter-individual correlation coefficient is suggested to be lower than the intra-individual correlation coefficient.

3 Estimation of drowsiness level using SN-FTI

In this study, a drowsiness induction experiment was conducted to collect facial thermal images at various drowsiness levels.

3.1 Experimental methods

3.1.1 Experimental system

In this experiment, FTIs and visible images were acquired while inducing drowsiness in the subject. The visible images of the face were acquired to evaluate drowsiness from facial expressions. The measurement system consisted of infrared thermography (TVS-200EX, Japan Avionics Co., Ltd., Tokyo) and a near-infrared camera (DC-NCR300U, Hanwha

Q-Cells Japan, Inc., Tokyo). In this study, the infrared thermography and the near-infrared camera were placed at 100 and 70 cm in front of the subject, respectively. FTIs were recorded at a resolution of 320×240 pixels at a sampling rate of 1 fps. The infrared emissivity of the skin was 0.98. The face's visible image was recorded at a resolution of 640×480 pixels at a sampling rate of 60 fps.

3.1.2 Procedure and conditions

Seven subjects, aged 21–23 years, participated in the experiment. They were fully informed about the experiment and the purpose of the study before their participation. All participants signed a consent form. The experiment was conducted during the day to control for potential circadian rhythm effects. Subjects were asked not to eat, drink caffeine, or smoke for two hours before the experiment. The experiment was conducted three times and started after the participants had been in the room for at least 30 min to get used to the laboratory's temperature. The laboratory's temperature was 23.8 ± 0.1 °C. The experiment consisted of a 1-min resting state segment and a 15-min drowsiness induction segment; the two resting-state segments were provided before and after the drowsiness induction segment. During each resting-state segment, subjects were instructed to sit with their eyes closed. In the drowsiness-inducing segment, the subject was instructed to gaze at the image using only eye movements to induce drowsiness. The images for drowsiness induction were displayed on a liquid crystal monitor placed in front of the subject's eyes. Then, the circular moving target was moved in a circular orbit; one cycle was set to 2 s, and the lights were turned off to induce drowsiness further.

3.1.3 Drowsiness level

Eight evaluators, who were different from the subjects who participated in the drowsiness induction experiment, objectively evaluated the drowsiness level from the facial expressions in the drowsiness induction segment every 20 s. The drowsiness level based on facial expressions [15] was evaluated according to the evaluation criteria shown

Table 1 Inter-individual correlation coefficients of SN-FTIs in the center

	Subject A	Subject B	Subject C	Subject D	Subject E	Subject F	Subject G
Subject A	–						
Subject B	0.66	–					
Subject C	0.43	0.44	–				
Subject D	0.60	0.64	0.41	–			
Subject E	0.57	0.53	0.50	0.53	–		
Subject F	0.60	0.46	0.38	0.77	0.45	–	
Subject G	0.69	0.64	0.40	0.78	0.46	0.79	–

Table 2 Drowsiness level based on facial expression

Drowsiness level	Example of facial expression and motions
1. Awake	Eye movements are quick and frequent Blink cycles are stable at approximately two per two s and are active Body motions are active
2. Slightly drowsy	Lips are parted Eye movements are slow
3. Drowsy	Blinks are slow and frequent Repositions body on seat Touches hand to face
4. Very drowsy	Blinks assumed to occur consciously Shakes head Frequently yawns
5. Extremely drowsy	Eyelids closing Leans head back and forth

Table 3 Definition of two levels of drowsiness based on drowsiness evaluation values

	Evaluation value				
	1	2	3	4	5
Pattern A	Low	High			
Pattern B	Low		High		
Pattern C	Low			High	

Table 4 Definition of three levels of drowsiness based on drowsiness evaluation values

	Evaluation value				
	1	2	3	4	5
Pattern D	Low	Medium	High		
Pattern E	Low	Medium		High	
Pattern F	Low		Medium	High	

in Table 2. The average of the values of the eight evaluators was then used as the final evaluation value. In this study, the drowsiness levels were divided into two and three levels, considering the distinction between normal and abnormal from the viewpoint of preventive safety when considering the final application: “high level” and “low level” for the two levels of drowsiness (Table 3), and

“high level,” “medium level,” and “low level” for the three levels of drowsiness (Table 4).

3.2 Analysis methods

To evaluate the impact of SN-FTI on the estimation of physiological and psychological states, a neural network constructed with the same layers as in the previous study [1] was used for supervised learning to estimate the level of drowsiness. In this previous study, the drowsiness level was estimated when the input images were facial region thermal images that had not been preprocessed. In this study, the input images of the network were compared under two conditions: facial thermal images without preprocessing (Normal) or SN-FTI. Drowsiness levels were estimated for patterns A–F. The number of thermal images was randomly subtracted so that the number of thermal images for each drowsiness level would be the same. The total number of thermal images for each pattern is shown in Table 5. For each condition, the input image was a single channel with a size of 130×130 pixels.

The CNN consisted of three convolutional layers, three pooling layers, and one fully connected layer (Table 6). The number of output classes was two for two-class drowsiness level estimation and three for three-class drowsiness level estimation. The learning rate was 0.001, the learning algorithm rule was backpropagation, the optimization method was Adam [14], and the loss function was the cross-entropy error. If the verification loss did not improve in 5 epochs, the training was stopped.

To perform leave-one-out cross-validation (LOOCV), the data of six subjects were split into training data and the data of the remaining subjects into test data, and this was repeated until all subjects' data were used as the test data. This is equivalent to k -fold cross-validation with $k = 7$. The number of images in the training and test data is different for each subject in the test data, but the number of images in the test data is more than 300 for all data sets. The conclusive discrimination rate was the average of the rates obtained by LOOCV.

3.3 Results and discussion

This section describes the results of the experiments in Sect. 3.1. Figures 6 and 7 show the discrimination rates of the estimated two- and three-level drowsiness levels. In all

Table 5 Total number of input images for each pattern

	Pattern A	Pattern B	Pattern C	Pattern D	Pattern E	Pattern G
Number of input images	2282	6538	3514	3423	3423	5271

Table 6 Construction of CNN

Layers	Filter size	Stride
Convolution 1	$40 \times 6 \times 6$	2
Max pooling 1	2×2	1
Convolution 2	$40 \times 4 \times 4$	1
Max pooling 2	2×2	1
Convolution 3	$40 \times 4 \times 4$	1
Max pooling 3	2×2	1
Fully connected

Fig. 6 Discrimination rate of drowsiness levels using CNN ($N = 7$). Error bars represent the standard error. (Two levels of drowsiness) **Fig. 7** Discrimination rate of drowsiness levels using CNN ($N = 7$). Error bars represent the standard error. (Three levels of drowsiness)

patterns, our method slightly improved the discrimination rate of drowsiness levels over Normal. This suggests that SN-FTIs reduce individual differences in the facial structure on drowsiness estimation. However, the discrimination rate is still low, and it is necessary to construct an optimal network for SN-FTIs.

4 Conclusions

As mentioned in the introduction, the objective of this study is to develop a method for SN-FTI and to evaluate the effect of SN-FTI on the estimation of physiological and psychological states. First, we attempted spatial normalization using facial features. The results suggested that SN-FTI would result in the same face shape among individuals. Since there are individual differences in facial skin temperature distribution, the inter-individual correlation coefficient is suggested to be lower than the intra-individual correlation coefficient. Next, we modeled the estimated drowsiness level using SN-FTIs and compared it with Normal. The results showed that SN-FTI slightly improved the discrimination rate of drowsiness level. SN-FTIs were suggested to reduce the effect of individual differences in facial structure on the estimation of physiological and psychological states.

In this study, we used the same network as in the previous study to evaluate the SN-FTI for estimating drowsiness. In the future, we plan to attempt to build an optimal model of the SN-FTI. We also plan to use the SN-FTI for the evaluation of indices other than drowsiness.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adachi H, Oiwa K, Nozawa A (2019) Drowsiness level modeling based on facial skin temperature distribution using a convolutional neural network. *IEEJ Trans Electr Electron Eng* 14(6):870–876
- Ahmadi N, McQuilkin GL, Akhtar MW, Hajsadeghi F, Kleis SJ, Hecht H, Naghavi M, Budoff M (2011) Reproducibility and variability of digital thermal monitoring of vascular reactivity. *Clin Physiol Funct Imaging* 31(6):422–428
- Ashburner J, Friston KJ (1999) Nonlinear spatial normalization using basis functions. *Hum Brain Mapp* 7(4):254–266
- Ashburner J, Neelin P, Collins D, Evans A, Friston K (1997) Incorporating prior knowledge into image registration. *Neuroimage* 6(4):344–352
- Bando S, Oiwa K, Nozawa A (2017) Evaluation of dynamics of forehead skin temperature under induced drowsiness. *IEEJ Trans Electr Electron Eng* 12:S104–S109

6. Collins DL, Neelin P, Peters TM, Evans AC (1994) Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J Comput Assist Tomogr* 18(2):192–205
7. Ebisch SJ, Aureli T, Bafunno D, Cardone D, Romani GL, Merla A (2012) Mother and child in synchrony: thermal facial imprints of autonomic contagion. *Biol Psychol* 89(1):123–129
8. Engert V, Merla A, Grant JA, Cardone D, Tusche A, Singer T (2014) Exploring the use of thermal infrared imaging in human stress research. *PLoS ONE* 9(3):e90782
9. Fernández-Cuevas I, Marins JCB, Lastras JA, Carmona PMG, Cano SP, García-Concepción MÁ, Sillero-Quintana M (2015) Classification of factors influencing the use of infrared thermography in humans: a review. *Infrared Phys Technol* 71:28–55
10. Friston KJ, Ashburner J, Frith CD, Poline JB, Heather JD, Frackowiak RS (1995) Spatial registration and normalization of images. *Hum Brain Mapp* 3(3):165–189
11. Gault T, Farag A (2013) A fully automatic method to extract the heart rate from thermal video. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp 336–341
12. van den Heuvel CJ, Ferguson SA, Dawson D, Gilbert SS (2003) Comparison of digital infrared thermal imaging (DITI) with contact thermometry: pilot data from a sleep research laboratory. *Physiol Meas* 24(3):717
13. Ioannou S, Gallese V, Merla A (2014) Thermal infrared imaging in psychophysiology: potentialities and limits. *Psychophysiology* 51(10):951–963
14. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. *arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980)*
15. Kitajima H (1997) Prediction of automobile driver sleepiness. *Jpn Soc Mech Eng Collect Pap (C book)* 63(613):93–100
16. Kranjec J, Beguš S, Geršak G, Drnovšek J (2014) Non-contact heart rate and heart rate variability measurements: a review. *Biomed Signal Process Control* 13:102–112
17. Lewis GF, Gatto RG, Porges SW (2011) A novel method for extracting respiration rate and relative tidal volume from infrared thermography. *Psychophysiology* 48(7):877–887
18. Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, Woods R, Paus T, Simpson G, Pike B et al (2001) A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philos Trans R Soc Lond Ser B Biol Sci* 356(1412):1293–1322
19. McCoy M, Campbell I, Stone P, Fedorchuk C, Wijayawardana S, Easley K (2011) Intra-examiner and inter-examiner reproducibility of paraspinal thermography. *PLoS ONE* 6(2):e16535
20. Nagumo K, Kobayashi T, Oiwa K, Nozawa A (2021) Face alignment in thermal infrared images using cascaded shape regression. *Int J Environ Res Public Health* 18(4):1776
21. Nakanishi R, Imai-Matsumura K (2008) Facial skin temperature decreases in infants with joyful expression. *Infant Behav Dev* 31(1):137–144
22. Rorden C, Bonilha L, Fridriksson J, Bender B, Karnath HO (2012) Age-specific CT and MRI templates for spatial normalization. *Neuroimage* 61(4):957–965
23. Shastri D, Merla A, Tsiamyrtzis P, Pavlidis I (2008) Imaging facial signs of neurophysiological responses. *IEEE Trans Biomed Eng* 56(2):477–484
24. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15(1):273–289
25. Zaproudina N, Varmavuo V, Airaksinen O, Närhi M (2008) Reproducibility of infrared thermography measurements in healthy individuals. *Physiol Meas* 29(4):515

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.