https://doi.org/10.2352/issn.2169-2629.2019.27.7
Reprinted from
Journal of Imaging Science and Technology® 63(4): 040406-1–040406-9, 2019.

Detecting Wetness on Skin using RGB Camera

Mihiro Uchida
Graduate School of Science and Engineering, Chiba University, Chiba, Japan
E-mail: Mihiro.uchida@gmail.com

Norimichi Tsumura▲
Graduate School of Engineering, Chiba University, Chiba, Japan

Abstract. *In this study, we propose a method to detect wetness on the surface of human skin and skin phantoms using an RGB camera. Recent research on affect analysis has addressed the non-contact multi-modal analysis of affect aimed at such applications as automated questionnaires. New modalities are needed to develop a more accurate system for analyzing affects than the current system. Thus we focus on emotional sweating, which is among the most reliable modalities in contact methods for affect analysis. However, sweat detection on the human skin has not been achieved by other researchers, and thus it is unclear whether their feature values are useful. The proposed method is based on feature values of color and glossiness obtained from images. In tests of this method, the error rate was approximately 6.5% on a skin phantom and at least approximately 12.7% on human skin. This research will help to develop non-contact affect analysis.* © *2019 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.2019.63.4.040406]

1. INTRODUCTION

Many recent studies have examined non-contact methods for affect analysis [1–5]. Such methods can be useful, for instance, for automating and improving questionnaires while subjects watch a video, by using physiological signals [3–5]. This is because respondents of questionnaires do not always answer honestly. The conventionally used contact method to analyze affect may prevent people from concentrating on the relevant stimuli because of discomfort due to devices attached to them. Thus, questionnaires that are automated using non-contact analysis for affect can be useful in such situations.

The system of non-contact affect analysis is based on analyses of facial expression, voice, and pulse waves using a microphone or a video. It thus may not be suitable for people who are mute, "poker-faced," or have fewer cardiac outputs than the average. To improve the system to be applicable for a wider class of people, new modalities of non-contact affect analysis are needed.

Therefore, we focus on emotional sweating, which is among the most reliable modalities in contact-based affect analysis. For example, sweat appears on your palms when you feel nervous or anxious even though it is not hot. This is emotional sweating, and is conventionally evaluated using skin conductance and skin potential. These indices are obtained by measuring the electric state of the skin through electrodes attached to it. These electric methods are not suitable for non-contact analysis of affect. On the contrary, Uchida et al. analyzed the difference between dry and wet palms based on color features [6]. However, their database did not consider wetted palms due to sweat, and could not detect wetness on the surface of the palm. Therefore, it remains unknown whether the features they proposed are useful for detecting wetness, especially due to sweating. Moreover, the skewness of the Y component, which is related to glossiness [7], showed no significant trend. Thus, glossiness could not be used as a feature. They also assumed a change in internal reflection. However, images that they processed contained internal reflection as well as surface reflection, which can appear as noise in the analysis.

In this study, we use both color and glossiness as features to detect wetness on the skin, including through sweat. The remainder of this paper is structured as follows and an overview is provided in Figure 1: We construct databases of images of skin phantoms and human skin as described in Section 2. Some image processing methods were applied to captured images in the databases, and this is described in Section 3. In Section 4, we describe the training of classifiers to estimate whether skin in a given image is wet based on the feature values using a support vector machine (SVM). In Section 5, we detail our results and discuss them. Finally, we summarize this study and offer directions for future work in Section 6. All codes in this study were written in MATLAB.

2. DATABASE CONSTRUCTION

In this section, we describe the procedure to construct our databases. We constructed them using images of skin phantoms (Beaulax) and human skin.

2.1 *Database of Skin Phantoms*

We used five skin phantoms made by Beaulax with different anisotropic textures copied on the cheeks of Japanese females. All skin phantoms were identical in color, like the color of the skin of Mongoloids. An example of the skin phantoms is shown in Figure 2, and Figure 3 shows our experimental setup to construct the database of skin phantoms. Their geometry was empirically determined as

▲IS&T Member.

Received Mar. 13, 2019; accepted for publication June 8, 2019; published online Aug. 22, 2019. Associate Editor: Jean-Baptiste Thomas.
1062-3701/2019/63(4)/040406/9/$25.00

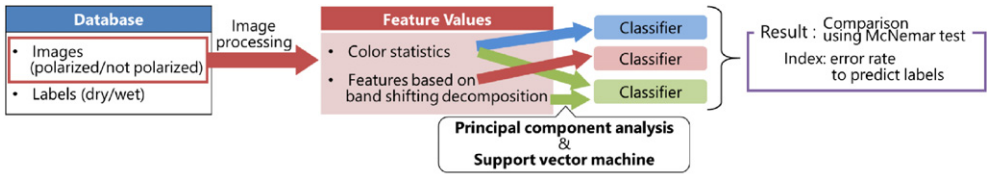


Figure 1. The procedure of our study.

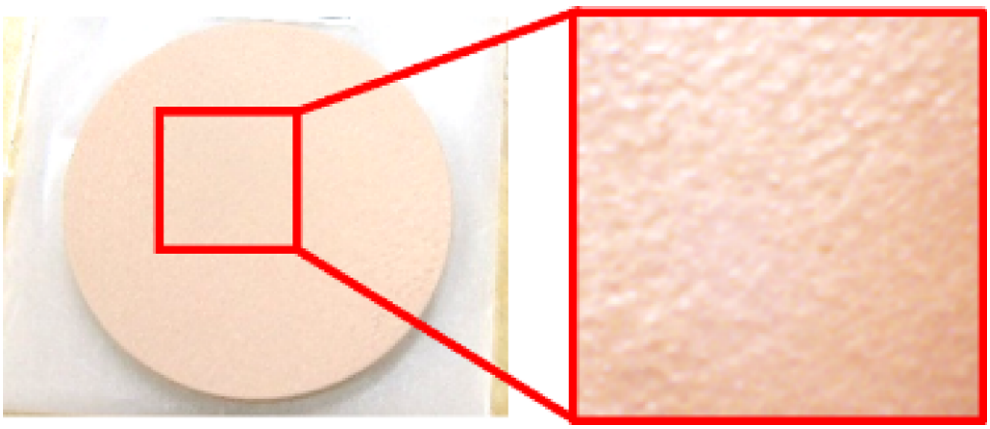


Figure 2. An image of skin phantom.

the image contained many highlights. We captured images of the skin phantoms with a polarizing camera (Lucid, PHX050S-Q, whose sensor is Sony IMX250 MYR) under artificial solar light (Seric, XC-100AF) in a dim room. We used the polarizing camera because that is compact and not expensive, and because that helps to simultaneously remove surface reflection, which is the cause of gloss [8]. We developed software for capturing and imaging using SDK provided by Lucid. One [px] in the captured image corresponded to approximately 0.2 [mm] of physical space. A polarizing plate was attached in front of the artificial solar light. Images of different polarizing angles could be simultaneously obtained, and we used 0° images as ones containing gloss and 90° images as images without gloss. The direction of polarizing plate was empirically decided so that 90° images do not contain gloss. We assumed that spectral response is the same for each polarizing bands because the difference between each bands is based on only the direction of polarizer not any spectral factors. Shutter speed was 1.0 [s], and was set to reduce color changes by the pulse wave when capturing images of the human skin. We noted that the texture of the skin phantom changed after slight rotation because of its anisotropy. We thus captured images of skin phantoms by rotating them to increase the number of samples. We then trimmed the images and obtained 128 × 128 [px] images, such as the red frame in Figure 2. We labeled phantoms whose surface was wiped with a dry tissue as "dry," and labeled phantoms whose surface was wiped with a wet tissue as "wet." We obtained 200 data items in total, 100 images each labeled "dry" and "wet."

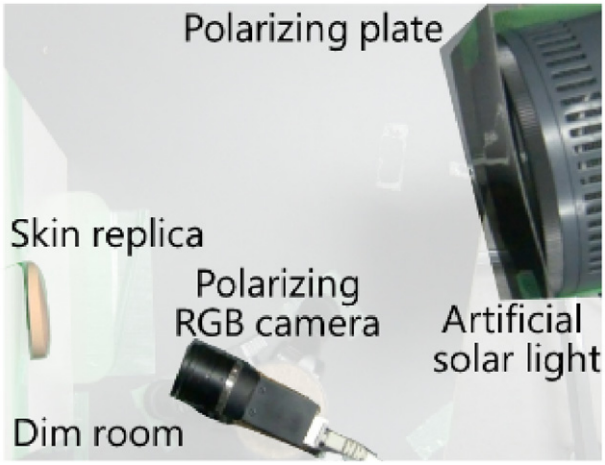


Figure 3. Our experimental setting for skin phantoms.

2.2 *Database of Human Skin*

We now explain the construction of the database of images of human skin. The subjects were four Japanese males and a Japanese female of ages ranging from 22 to 25 years. We captured images of their left palms. Figure 4 shows our experimental setup to construct the database of human skin. The experimental settings were nearly identical to those described in Section 2.2 except that the subject's hand was placed on a table to prevent movement. We gave the subjects five stimuli as follows: "relaxing," "mental calculation," "softly gripping a computer mouse for 10 minutes," "imagining nervous experiences," and "wetting the surface of the palm by tissue." "Mental calculation," and "imagining nervous experiences" are emotional stimuli, and

Figure 4. Our experimental setting for constructing a database of images of human skin.

we expected that the subjects' palms would become wet during these tasks due to the activation of the sympathetic nerves. "Softly gripping a computer mouse for 10 minutes" caused sweat to appear on the palm. "Relaxing," "'mental calculation," and "imagining nervous experiences" lasted for 2 [min], and we captured images of the palm at 30 [s], 60 [s], and 90 [s]. We did not use images captured at 120 [s] because the affect might not have lasted this long and the shutter speed was high to keep the palm wet. We also labeled samples based on the sense of touch using one evaluator. We then trimmed images and obtained 128×128 [px] images, such as the red frame in Figure 5. We labeled the skin assessed to be dry by the evaluator as "dry" and that determined to be wet as "wet." We obtained 55 data items in total: 33 images of dry palms and 22 images of wet ones.

3. FEATURE EXTRACTION

In this section, we describe the proposed method to obtain feature values. We explain the change caused by wetness. First, light re-enters the skin by total reflection at the water–air interface if the surface of the skin is covered with water. As a result, more light is absorbed by the chromophore in the skin so that luminance decreases and saturation increases [9]. Second, the rough surface of the skin becomes smooth because it is covered with water. Therefore, sharp highlights appear rather than dull ones to increase glossiness as shown in Figure 6. (a) shows dry skin and (b) shows wet skin. We apply two methods to obtain values to represent these features. We first explain the color statistics [10] and then describe feature values based on band-sifting decomposition [11].

3.1 *Color Statistics*

In this section, we describe the method to obtain feature values related to color from images with gloss and images without gloss, respectively. RGB values are first converted into CIEXYZ coordinates [12] under D55 light source using MATLAB. We then obtain two feature values, the mean and skewness of the Y component, known as the luminance component. The mean is related to a decrease in luminance and the skewness to that in glossiness [7]. Then, $u'v'$ coordinates are calculated from the XYZ coordinates: [13]

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} \dfrac{4X}{X+15Y+3Z} \\ \dfrac{9Y}{X+15Y+3Z} \end{pmatrix}, \tag{1}$$

where (u', v') are the $u'v'$ coordinates, and X, Y, and Z are the CIEXYZ coordinates. Following this, we obtain saturation and hue by the following formula:

$$S = 13\sqrt{(u'-u'_n)^2-(v'-v'_n)^2} \tag{2}$$

$$H = \tan^{-1}\left(\frac{v'-v'_n}{u'-u'_n}\right), \tag{3}$$

where S is saturation, H is hue, (u', v') are the $u'v'$ coordinates of the object, and (u'_n, v'_n) are the uv' coordinates of the white point. We obtain two features: mean saturation related to an increase in saturation, and the entropy of hue. These four features are related to the human perception of wetness. Finally, we obtained the feature values as follows: The mean of the Y component was obtained from images with gloss, the skewness of the Y component was obtained from images with gloss, the mean of saturation was obtained from images with gloss, the entropy of hue was obtained from images with gloss, the mean of the Y component was obtained from images without gloss, the skewness of the Y component was obtained from images without gloss, the mean of saturation was obtained from images without gloss, and the entropy of hue was obtained from images without gloss. Fig. 7 shows the example of each color components before obtaining feature values. Figure 7(a) shows the original image, (b) shows the Y component image, (c) shows the saturation image, and (d) shows the hue component image. We obtained some statistics from these components.

3.2 *Feature Values based on Band-sifting Decomposition*

In this section, we describe the method to obtain feature values related to glossiness based on band-sifting decomposition [11], which uses the appearance of a given material, such as highlights.

We first extracted the L^* component of CIELAB color space from the original RGB image with gloss [14]. The L^* image was then decomposed in terms of frequency. This is achieved with pyramid structure such like Laplacian pyramid. The overview is shown in Figure 8. The bottom image is the original L^* image. The second image from the bottom is obtained by smoothing the original image. The third image from the bottom is obtained by smoothing the

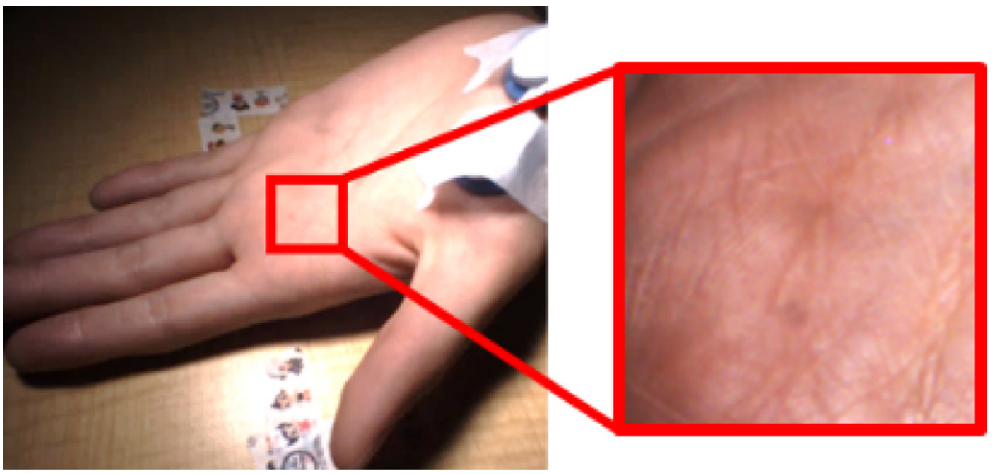
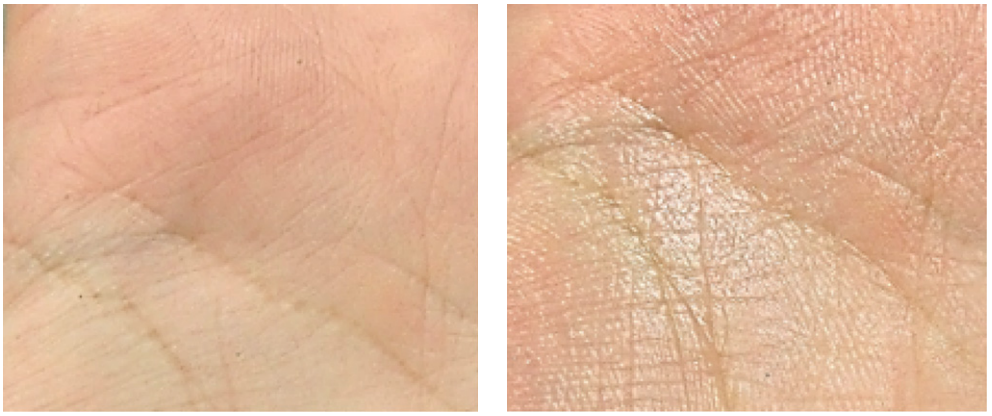
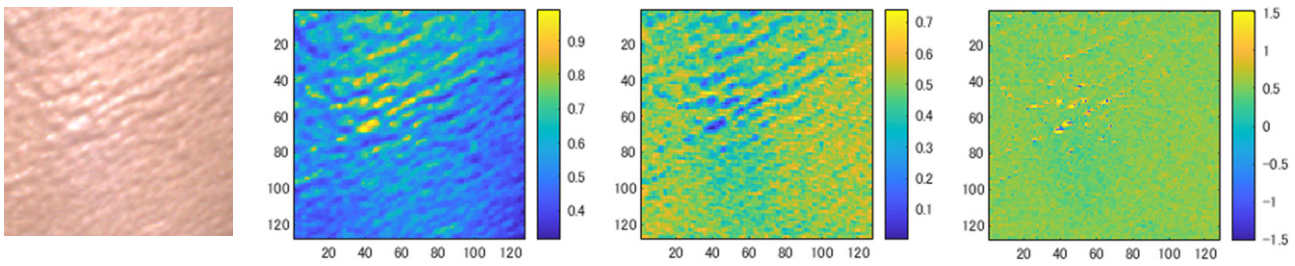


Figure 5. Example of the image of human skin.



(a) (b)

Figure 6. The difference in surface reflection between dry skin and wet skin: (a) dry skin, (b) wet skin.



(a) (b) (c) (d)

Figure 7. Examples of color component: (a) the original image, (b) the Y component image, (c) saturation components, and (d) hue components. The bar beside images represents the value of the components.

second image from the bottom. After that, we calculated the difference between nth image from the bottom and n+1th image from the bottom to decompose L^* component by frequency. We apply a guided filter [15], which can preserve edges although it is a smoothing filter. Then, the L^* image was decomposed into high-frequency components (described as H) and low-frequency components (described as L). The components were then decomposed in terms of amplitude, where a decomposed value higher than 1.2 σ was considered a high-amplitude component (described as H) and that lower than 0.8σ was considered a low-amplitude component (described as L). σ represents standard deviation. The interval between 0.8σ and 1.2σ was linearly connected to modulate the appearance naturally, as in a previous study [11]. The values were also decomposed in terms of sign, such as positive (described as P) and negative components (described as N).

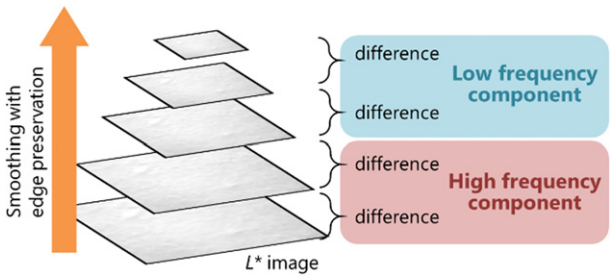


Figure 8. Overview of the decomposition in terms of frequency.

We obtained feature values as the mean of each representative operator. The operator is a finely decomposed component, for example, high frequency, high amplitude, and positive. We selected the following nine operators mentioned in a previous study: [11]. (frequency, amplitude, sign) = (H, H, P), (H, H, N), (H, H, A), (H, L, P), (H, L, N), (H, L, A), (L, A, P), (L, A, N), (L, A, A). "A" means "all components." Figure 9 shows the examples of images before obtaining feature values. Fig. 9(a) shows the original image. Fig. 9(b) shows the (H, H, P) component image, and Fig. 9(c) shows the (H, H, N) component image. The bar beside images shows the value of the component. Pixels corresponding highlights in Fig. 9(a) are emphasized in Fig. 9(b), pixels corresponding skin grooves in Figure 8(a) are emphasized in Fig. 8(c). Thus, the band-shifting components are related to appearance. We obtained feature values as the mean of each components.

3.3 *Results of Feature Values*

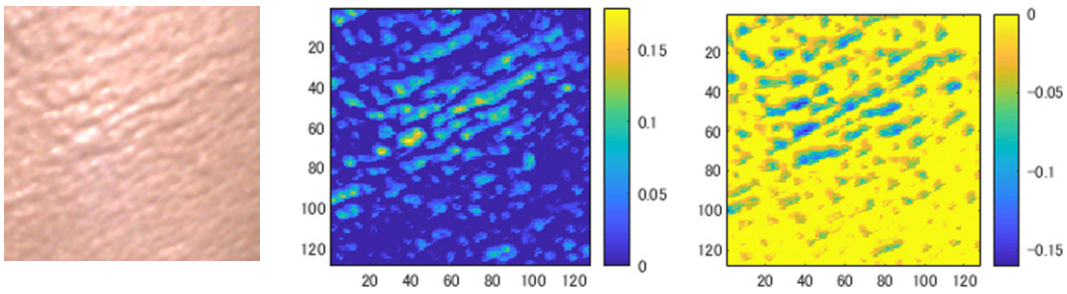
In this section, we show the results of feature values expected to have significant differences between dry and wet skins. We normalized the scale of feature values into [0, 1] to display and tested them using the student's t-test. The null hypothesis was that the mean of each group was the same.

Figure 10 shows the feature values of the skin phantoms. (a) shows the result of "mean of Y component" obtained from images without gloss, (b) shows the result of "skewness of Y component" obtained from images with gloss, (c) shows those of "mean of saturation" obtained from images without gloss, (d) shows the result of "entropy of hue" obtained from images with gloss, (e) shows those of "mean of HHP operator," and (f) shows those of "mean of LAP operator." The vertical axis of each graph shows the normalized value of each feature and the horizontal axis its label, such as "dry" and "wet."

In Fig. 10(a), the mean of the wet group is slightly larger but there is no significant difference ($p = 0.0772$). We expected the mean of the wet group to have been significantly smaller than that of the dry group because of the change in the path of light caused by wetness. However, there was no significant visual difference possibly because of the differences between skin phantoms. In Fig. 10(b), the mean of the wet group is significantly larger than that of the dry group ($p = 7.60 \times 10^{-18}$). This result is as expected because the skewness of the Y component is related to glossiness, and wetness increases glossiness on the surface. In Fig. 10(c), the mean of the wet group is significantly larger ($p = 2.62 \times 10^{-10}$). This result is as expected because of the change in the path of light caused by wetness. In Fig. 10(d), the mean of the wet group is slightly larger but there is no significant difference ($p = 0.824$). We expected the mean of the wet group to have been significantly larger than that of the dry group as Sawayama et al. have shown [10]. However, this is the same result as obtained by Uchida et al. [6]. This may be related to the fact that the skin showed no significant color change. In Fig. 9(e), the mean of the wet group is significantly larger than that of the dry group ($p = 4.78 \times 10^{-3}$). This is the expected result because the HHP operator is related to sharp highlights [11], and wetness increases sharp highlights on the surface. In Fig. 10(f), the mean of the wet group is significantly larger than that of the dry group ($p = 1.15 \times 10^{-4}$). However, we expected that the mean of the wet group would be significantly smaller than that of the dry group because the LAP operator is related to dull highlights [11], and wetness reduces them on the surface. This may be because the threshold of the band-sifting decomposition was not optimized to the settings of our experiment. As a result, we obtained both color and glossy features.

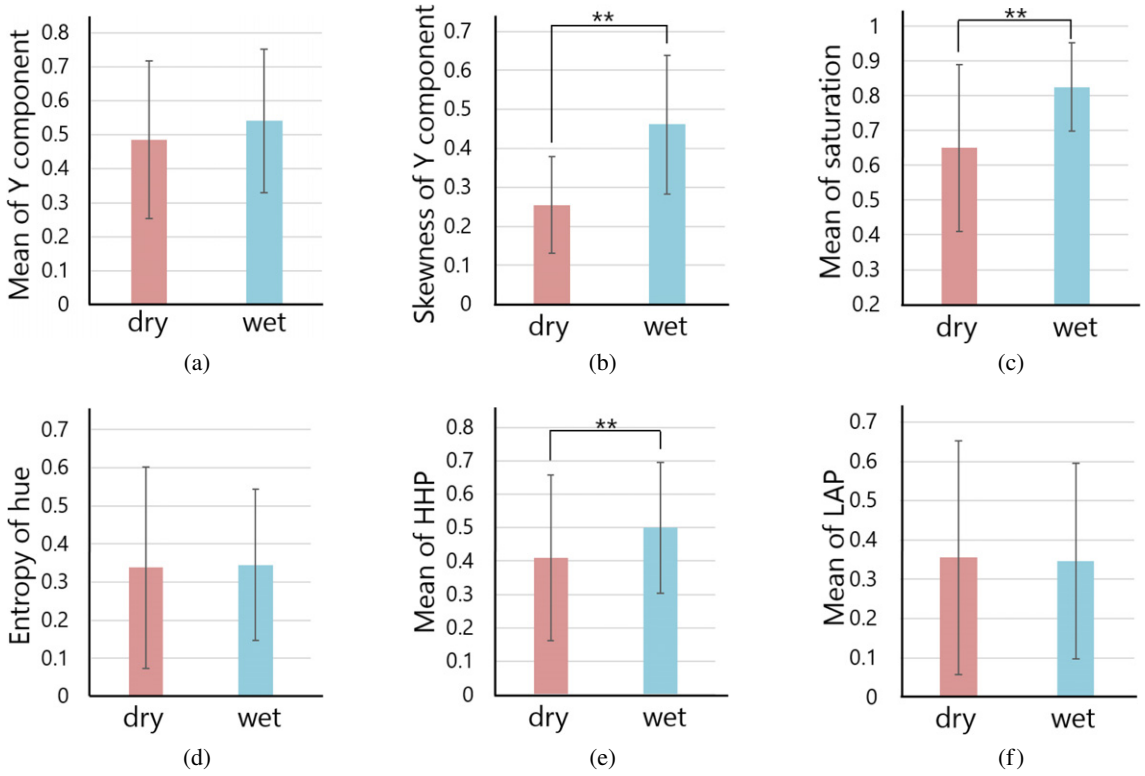
Figure 11 shows the feature values of images of the human skin. (a) shows the result of the "mean of Y component" obtained from images without gloss, (b) shows that of the "skewness of Y component" obtained from images with gloss, (c) shows that of the "mean of saturation" obtained from images without gloss, (d) shows the result of "entropy of hue" obtained from images with gloss, and (e) shows the result of "mean of HHP operator." The vertical axis of each graph shows the normalized value of each feature and the horizontal axis shows the label, "dry" or "wet."

In Fig. 11(a), the mean of the wet group is slightly larger but there is no significant difference ($p = 0.853$). We expected the mean of the wet group to be significantly smaller than that of the dry group because of the change in the path of light caused by wetness. In Fig. 11(b), the mean of the wet group is slightly smaller but there is no significant difference. ($p = 0.620$). However, we expected it to be significantly larger than that of dry group because the skewness of the Y component is related to glossiness, and wetness increases glossiness on the surface. In Fig. 11(c), the mean of the wet group is slightly larger but there is no significant difference ($p = 0.502$). We expected it to be significantly large than that of the dry group. In Fig. 11(d), the mean of the wet group is slightly smaller but there is no significant difference ($p = 0.313$). We expected it to be significantly larger than that of the dry group as a previous study has shown [9]. In Fig. 11(e), the mean of the wet group is slightly smaller but higher than that of the dry group ($p = 0.634$). However, we expected it to be significantly larger because the HHP operator is related to sharp highlights [10], and wetness increases sharp highlights on the surface. In Fig. 11(f), the mean of the wet group is significantly larger than that of the dry group ($p = 0.913$), but we expected it to be significantly smaller because the



(a) (b) (c)

Figure 9. Example of components: (a) the original image, (b) (H, H, P) components, and (c) (H, H, N) components. The bar beside images represents the value of the components.



(a) (b) (c) (d) (e) (f)

Figure 10. The differences between dry and wet skin phantoms on feature values expected to have significant differences. (a) Mean of Y component in images without gloss of skin phantoms. (b) Skewness of Y component in images with gloss of skin phantoms. (c) Mean of saturation in images with gloss of skin phantoms. (d) Entropy of hue in images with gloss of skin phantoms. (e) Mean of HHP operator of skin phantoms. (f) Mean of LAP operator of skin phantoms.

LAP operator is related to dull highlights [10], and wetness decreases them on the surface. There was no significant difference possibly because differences between individuals were larger than that between dry and wet samples. However, this is a trivial problem for learning-based classifiers because we used a non-linear method.

4. TRAINING A CLASSIFIER AS SUPPORT VECTOR MACHINE

To detect wetness on skin, we trained and compared three classifiers: (A) a classifier based on only the color statistics described in Section 3.1, (B) a classifier based on feature values described in Section 3.2, and (C) a classifier based on all feature values described in Section 3. In this section, we describe the method to construct these classifiers.

In this study, there was a correlation between feature values as shown in Figure 12. (a) shows the correlation in the case of skin phantoms and (b) shows it in the case of human skin. However, if feature values are correlated, the response of the classifier may not be stable because of multicollinearity. Therefore, we applied principal component analysis (PCA) to the feature values to avoid multicollinearity. PCA can be used to obtain vectors orthogonal to one another [16].

The first principal component vector is defined as a vector representing the direction in which variance is

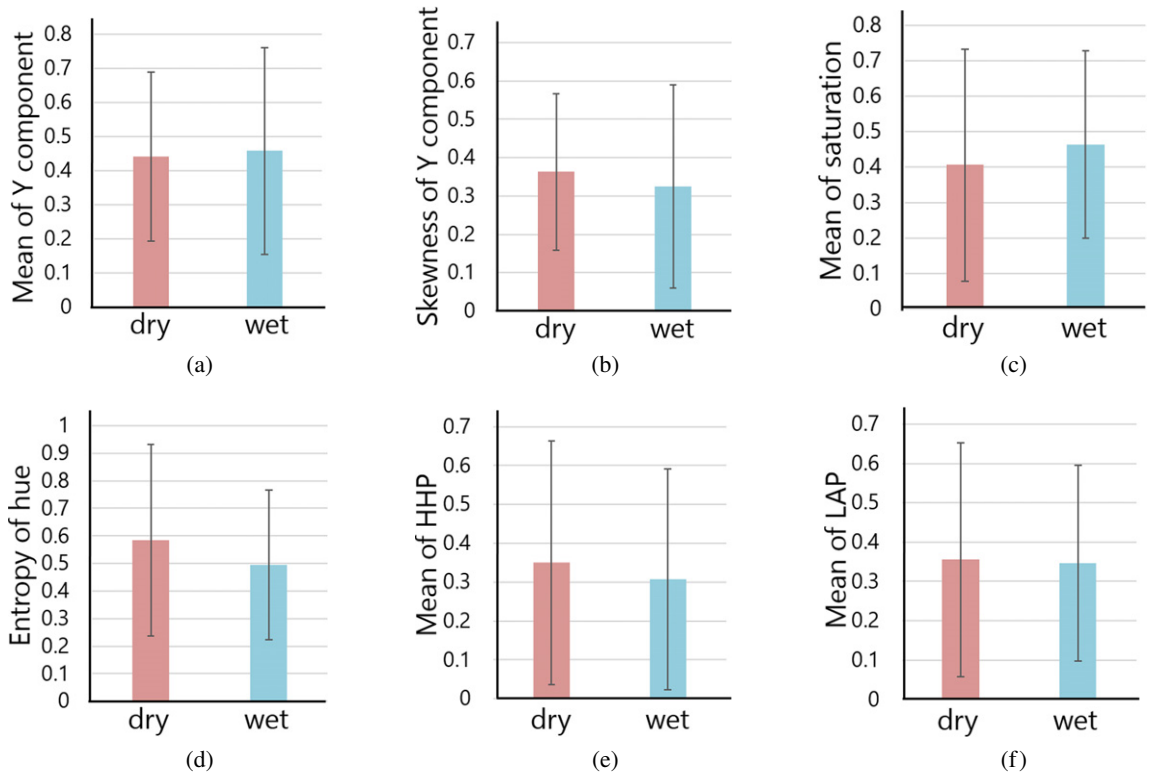


Figure 11. The effect of differences between dry and wet human skin on the feature values expected to have significant difference. (a) Mean of Y component in images without gloss of human skin. (b) Skewness of Y component in images with gloss of human skin. (c) Mean of saturation in images with gloss of human skin. (d) Entropy of hue in images with gloss of human skin. (e) Mean of HHP operator of human skin. (f) Mean of LAP operator of human skin.

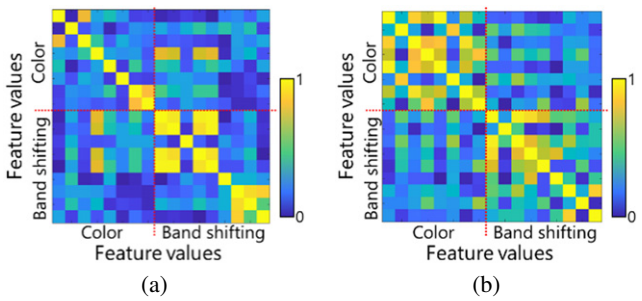


Figure 12. The absolute value of correlation between feature values: (a) correlation in the case of skin phantoms, (b) correlation in the case of human skin. The colors correspond to the absolute value of the correlation.

the maximum in the data space. The second principal component is defined as a vector orthogonal to the first principal component, and representing the direction in which the variance is the second maximum in the data space. The lth principal component is defined as a vector orthogonal to the first $\sim l$-1th principal components, and representing the direction in which the variance is the lth maximum in the data space. This analysis is achieved with the eigenvalue decomposition of the variance–covariance matrix of the feature values through the following formula:

$$z_l = \sum_{i=1}^{n} w_{il} x_i, \quad \text{where } (\boldsymbol{V} - \lambda \boldsymbol{I})\, \boldsymbol{w}_l = \boldsymbol{0},\ l = 1, \ldots, L, \tag{4}$$

where z_l is the lth principal component, $\boldsymbol{w}_l = (w_{1l}, \ldots, w_{nl})^T$ is the weight of lth principal component, $\boldsymbol{x}$ represents coordinates in the data space, $\boldsymbol{V}$ is the variance–covariance matrix, and λ is the eigenvalue. n is the number of dimensions of the data space and L is the total number of principal components.

We then obtained the principal component scores, which are shown in Figure 13, with coordinates in the space represented by the principal components. These scores were normalized to the same scale.

Following this, we trained the non-linear classifiers to detect wetness on the skin. We used an SVM [17, 18] to obtain the classifiers. An overview is shown in Figure 14. The normalized principal component scores were mapped to high-dimensional space by a radial-basis function kernel:

$$k(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\left\|\boldsymbol{x} - \boldsymbol{x}'\right\|^2\right), \tag{5}$$

where $\boldsymbol{x}$ and $\boldsymbol{x}'$ are factors of the dataset. We calculated a classifier that can linearly separate data in high-dimensional space to maximize the margin of the error. This is described

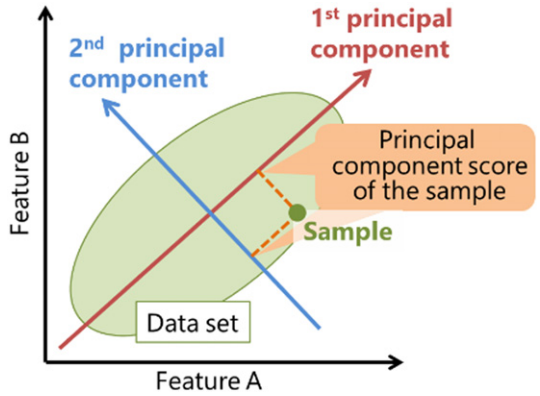


Figure 13. Overview of principal component analysis and principal component scores.

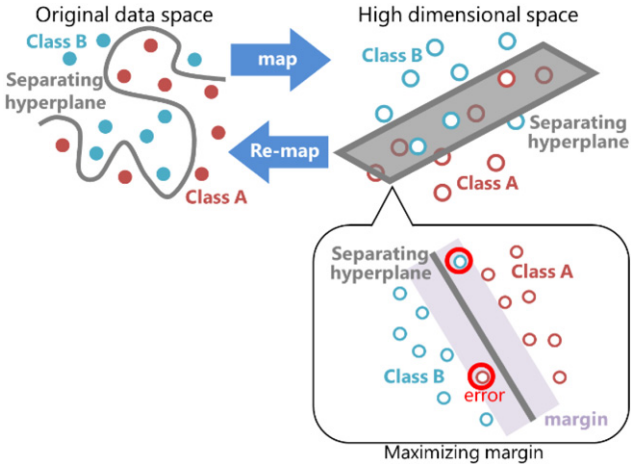


Figure 14. Overview of support vector machine.

as follows:

$$\min_{\boldsymbol{w}_{svm},\gamma,\boldsymbol{\xi}} \left[\frac{1}{2} \|\boldsymbol{w}_{svm}\|^2 + C \sum_{i=1}^{n} \xi_i \right] \text{subject to } (\boldsymbol{w}_{svm}^T \boldsymbol{x}_i + \gamma) \geq 1 - \xi_i, \xi_i \geq 0, \forall i = 1, \ldots, n, \tag{6}$$

where $\boldsymbol{w}_{svm}$ is the weight for the SVM, C is a parameter to control the acceptance of the error, $\boldsymbol{\xi}$ is the error, and γ is the bias.

5. RESULT AND DISCUSSION

We evaluated accuracy of the classifiers by the leave-one-out error rate. It is a cross-validation method where the classifier is trained with all samples except one sample as rotating the left one. We then statistically tested the difference in accuracy with the McNemar test [19]. This test focuses on labeling the data; the labels are predicted correctly by one classifier but not by another. The statistical value is described as follows:

$$t^* = \min(n_{12}, n_{21}), \tag{7}$$

where n_{12} is the number of samples whose labels were correctly predicted by model A but not by model B and n_{21} is the number of samples whose labels were correctly predicted by model B but not by model A. The critical region is represented as follows:

$$n_{12+n_{21}} C_{t^*} 0.5^{n_{12}+n_{21}} + \sum_{t=0}^{t^*-1} {}_{n_{12}+n_{21}-1} C_t 0.5^{n_{12}+n_{21}-1} < \frac{a}{2}, \tag{8}$$

where a is the level of significance empirically determined to be $a = 0.05$. The null hypothesis was that "the accuracy of both classifiers is the same."

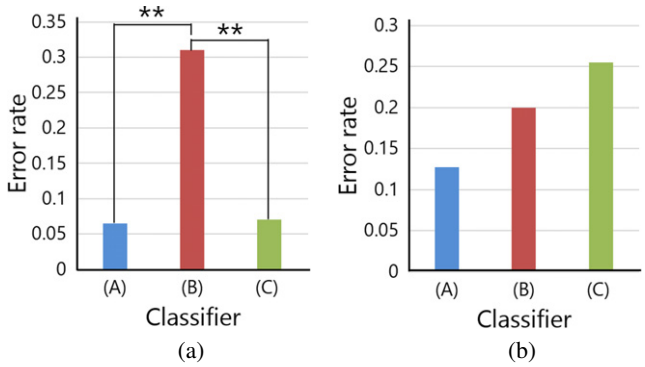


Figure 15. The results of error rates of the classifiers: (a) the case of skin phantoms, (b) the case of human skin.

The comparative result is shown in Figure 15. (a) shows results of skin phantoms and (b) shows those for the human skin. * means $p < 0.025$, and ** means $p < 0.005$ because we applied Bonferroni correction. In Fig. 15(a), the error in (B) is significantly higher than that in the others. The p-value between (A) and (B) was 6.81×10^{-10}, and between (C) and (B) was 7.64×10^{-11}. In Fig. 15(b), there is no significant difference and the p-value was 3.85×10^{-2} between (C) and (A). The error rate of (C) was the largest and that of (A) was the smallest of the three.

We expected (C) to deliver the highest accuracy because it contained more feature values than the other classifiers. However, the accuracy of (A) tended to be the highest. This suggests that feature values related to glossiness were noisy because glossiness is very sensitive to the experimental geometry, whereas the color change is robust against it. Moreover, this might have obtained because the surface was not wet enough to recognize the difference in gloss.

Trends in the case of skin phantoms and human skin were different. Error rates in the case of human skin were higher than those in the case of skin phantoms, possibly because of inter- and intra-class differences, such as difference in the phase of pulse wave, difference in skin color, and difference in shape. Moreover, we trained the classifiers on a small dataset so that they could not learn to classify beyond inter- and intra-class differences. This problem can be solved by normalizing the differences, such as by introducing a baseline or by training the classifiers on a large dataset.

6. CONCLUSION AND FUTURE WORK

In this study, we proposed a method to detect wetness on the skin using an RGB camera. We first constructed databases of skin phantoms and human skin. We obtained feature values, eight color statistics, and nine means of the operator using band-sifting decomposition. Following this, we removed the correlation among the feature values using PCA. We then trained support vector classifiers based on normalized principal component scores to detect wetness on the skin. Finally, we compared the accuracy of classifiers based on only color statistics, on feature values based on band-shifting decomposition, and a classifier based on all feature values. As a result, we found that color statistics are useful for constructing a classifier that can detect wetness on the skin.

It is necessary to improve the system to apply it. It needs to be rendered robust against the amount of sweat, the geometry to capture, and skin color. It is also important to detect changes in the volume of sweat in the sweat ducts. These problems can be solved with infrared and near-infrared imaging because some bands therein have very large water absorbance, and because near-infrared light and infrared light invade deep into the skin. This will form part of our future work in the area.

ACKNOWLEDGMENT

We thank Saad Anis, PhD, from Edanz Group (www.edanzediting.com/ac) for editing a draft of this manuscript.

REFERENCES

[1] M. Mansoorizadeh and N. M. Charkari, "Hybrid feature and decision level fusion of face and speech information for bimodal emotion recognition," CSICC **2009**, 652–657 (2009).

[2] D. McDuff, R. E. Kaliouby, J. F. Cohn, and R. W. Picard, "Predicting ad liking and purchase intent: Large-scale analysis of facial responses to ads," IEEE Trans. Affect. Comput. **6**, 223–235 (2015).

[3] H. Monkaresi, N. Bosch, R. A Calvo, and S. K. D'Mello, "Automated detection of engagement using video-based estimation of facial expressions and heart rate," IEEE Trans. Affect. Comput. **8**, 15–28 (2017).

[4] G. Okada, K. Masui, and N. Tsumura, "Advertisement effectiveness estimation based on crowdsourced multimodal affective responses," *Proc IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops* (2018).

[5] P. R. Chakraborty, D. W. Tjondronegoro, L. Zhang, and V. Chandran, "Towards generic modelling of viewer interest using facial expression and heart rate features," IEEE Access **6**, 62490–62502 (2018).

[6] M. Uchida, I. Nomura, and N. Tsumura, "Color-based non-contact analysis of skin changed by sweating for emotion estimation," CIC **1**, 253–258 (2018).

[7] I. Motoyoshi, S. Nishida, L. Sharan, and E. H. Adelson, "Image statistics and the perception of surface qualities," Nature **447**, 206 (2007).

[8] N. Ojima, T. Minami, and M. Kawai, "Transmittance measurement of cosmetic layer applied on skin by using image processing," *Proc. 3rd Scientific Conf. of the Asian Societies of Cosmetic Scientists* (1997), Vol. 114.

[9] J. Lekner and M. C. Dorf, "Why some things are getting darker when wet," Appl. Opt. **27**, 1278–1280 (1988).

[10] M. Sawayama, E. H. Adelson, and S. Y. Nishida, "Visual wetness perception based on image color statistics," J. Vis. **17**, 7 (2017).

[11] I. Boyadzhiev, K. Bal, S. Paris, and E. H. Adelson, "Band-sifting decomposition for image-based material editing," ACM Trans. Grap. **34**, 163 (2015).

[12] T. Smith and J. Guild, "The C.I.E. colorimetric standards and their use," TrOS **33**, 73–134 (1931-32).

[13] C. Poynton, *Digital Video and HDTV* (Morgan-Kaufmann Publishers, Burlington, MA, 2003), p. 226.

[14] K. McLaren, "XIII–The development of the CIE 1976 ($L^*a^*b^*$) uniform colour space and colour–difference formula," J. Soc. Dyers Colour. **92**, 338–341 (1976).

[15] K. He, J. Sun, and X. Tang, "Guided image filtering," IEEE Trans. Pattern Anal. Mach. Intell. **35**, 1397–1409 (2013).

[16] I. Jolliffe, *Principal Component Analysis* (Springer, Berlin, Heidelberg, 2011).

[17] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," *Proc. COLT '92* (1992), p. 144.

[18] C. Cortes and V. N. Vapnik, "Support-vector networks," Mach. Learn. **20**, 273–297 (1995).

[19] Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," Psychometrika **12**, 153–157 (1947).