https://doi.org/10.2352/issn.2169-2629.2020.28.18

Development of Monitoring System for Facial Shape and Skin Color Using Depth Camera mounted on a Smartphone

Ikumi Nomura[1], Reimei Koike[1], Naoaki Rikihisa[2], Nobuyuki Mitsukawa[2], and Norimichi Tsumura[1]
1) Chiba University, Chiba, Japan, 2) Chiba University Hospital, Chiba, Japan

Abstract

Regular observation and recording of the changes in body appearance are essential for the process of the treatment of plastic surgery and dermatology, especially aesthetic surgery. Usually, physicians treat patients with medical interviews, pictures of the patient's faces before and after their treatment, anatomical data that including size, location, and color of the affected skin. However, it is difficult to capture the affected area under the same conditions every time because the captured range varies depending on the imaging angle and distance. There is a need to record three-dimensional shape of face parts such as cheek, nose, eye, and chin. Therefore, in this study, the face shape and the skin color were measured using the infrared depth camera and the RGB camera built in the smartphone three-dimensionally. We measured before and after modulating the shape and color of the face, and then, the change in volume and the change in skin pigment of skin color was calculated and visualized. This method makes it possible to analyze the skin shape and color independently of the viewing angle and the illumination direction. In this study, the depth sensor built in the smartphone showed the potential to monitor changes in facial shape and skin color. In the future, it is expected to contribute to the development of telemedicine, in which the patient measures their face at home and gets medical treatment consultation remotely.

1. Introduction

Regular observation and recording of the changes in body appearance are essential for the treatment of plastic surgery and dermatology, especially aesthetic surgery. In these departments, one of the main targets of treatment is the skin.

The physician primarily observes the condition of the affected area directly and uses measures and photographs as supports to record the status. Physical measurement and record should be made by the same physician, and diagnosis naturally depends on their experience because the measurement accuracy depends on the physician's experience. The diagnostic imaging is difficult to thoroughly photograph the same position of the affected part due to differences in angles, distances, and lighting conditions. That is, it is considered that both methods are imperfect in accuracy and reproducibility. Plastic surgery, dermatology, and aesthetic surgery often perform treatments on the face. Even a slight change in the face gives significant changes of impression, so especially strict measurement is needed for the face. In the current clinical practice, the measurement system in three dimensional based on a stereo image have been proposed. Obtaining a three-dimensional shape model eliminates the need to use a measure and physically measure the face. Also, mapping facial texture maps to the 3D model solves some of the problems with angles and distances. This makes it easier for both the doctor and the patient to understand the current condition and contributes to planning for future treatment.

Technological advances over the last several decades have replaced classic direct anthropometry (using rulers and calipers) and two-dimensional (2D) photography with non-invasive three-dimensional (3D) surface imaging methods[3]. One of the most common techniques for 3D facial surface imaging today is digital stereophotogrammetry, for example, such as the Vectra H1[2]. This technique involves capturing images of the facial surface from multiple cameras with overlapping fields of view and then using software, based on the MVS (Multi-View Stereo) [4] method, to merge these images into a single 3D model, with facial geometry represented as a dense point cloud and realistic facial skin texture[3]. Therefore, we are able to get three-dimensional shape data and a texture map representing detailed skin color information, use to evaluate the changes in volume and skin color of multiple 3D shapes quantitatively and quantitatively.

Masui et al. measured the edema in the lower leg using depth cameras such as Kinect v1 and Structure Sensor, and visualized its change in shape [5]. In this research, the face is measured using the infrared depth camera mounted on the smartphone, and the change of the face shape is analyzed as in the existing system by applying the method of Masui et al. to the face. In addition, a texture map is also calculated and its changes in skin color is analyzed. In addition, in order to confirm whether the system can be used practically as well as the existing system, as an experiment, we modulate the shape and skin color of the actual human face, and measure and visualize the changes using the proposed system.

Infantile hemangiomas occur on the face, arms, and back in as many as 5% of infants, making them the most common benign tumor of infancy. Most Infantile hemangiomas are small, innocuous, self-resolving, and require no treatment. However, because of their size or location, a significant minority of infantile hemangiomas are potentially problematic. These include infantile hemangiomas that may cause permanent scarring and disfigurement, hepatic or airway infantile hemangiomas, and infantile hemangiomas with the potential for functional impairment, ulceration, and associated underlying abnormalities. Early intervention and/or referral (ideally by one month of age) is recommended for infants who have potentially problematic infantile hemangiomas. Surgery and/or laser treatment are most useful for the treatment of residual skin changes after involution and, less commonly, may be considered earlier to treat some infantile hemangiomas. [1]. If infantile hemangiomas occur on the face, prompt treatment is advisable because it can greatly affect the appearance of the face.

2. Proposed System

We introduce our system in three steps as below. 1) Measurement for 3D face, 2) Analysis of changes in 3D shape, 3) Analysis of changes in skin color.

2.1. Acquisition of 3D face models

In this study, a 3D shape of the face is measured by using an iPhone XR as a depth camera instead of a Kinect because it is easier to get the former one. The iPhone XR is a smartphone with a structured light depth sensor 30,000 dots of infrared light are projected onto the face, and it measures the shape by sensing the distortion [6]. The 3D facial model was reconstructed using Bellus3D software[7]. Bellus3D allows the iPhone XR to act as a 3D scanner in acquiring a point cloud, triangle meshes, and a texture map. Measurement and reconstruction of the 3D shape are completed in about 60 seconds. The 3D shape was obtained in .obj format, and the texture map was obtained in 4096×4096 in .jpg format. Figure 1 shows an example of the acquired 3D shape of the face. The texture map of Figure 1 is shown in Figure 2.

Figure 1. *Acquired 3D model.*

Figure 2. *Texture map.*

2.2. Visualization of 3D shape

In this section, we calculate the distance between the corresponding meshes after the 3D registration, and in order to visualize the volume change of the two 3D face models, we color them according to the distance. In this paper, we call the before or after treatment as the "earlier model" or the "later model" to distinguish between the two 3D shape models.

2.2.1. 3D registration

Firstly two 3D shape models are applied to the RANSAC and the ICP algorithm to register each position to match. Since the ICP algorithm, which is a general 3D registration method, strongly depends on the initial position between point clouds [8], the RANSAC method [9] is applied to remove the effect of outliers before the ICP algorithm. In this paper, the right cheek was filled with a cotton ball to modulate the facial shape to resemble an effect of surgery, as discussed in detail later. Figure 3 shows the results of 3D registration for the two point clouds. The cyan-colored points indicate the earlier model, and the magenta-colored points indicate the later model. After the above modulation, it is expected that two point clouds are matched in any of the parts except for the right cheek. Actually, according to Figure 3, it is seen that the RANSAC and ICP algorithm yielded an accurate 3D registration.

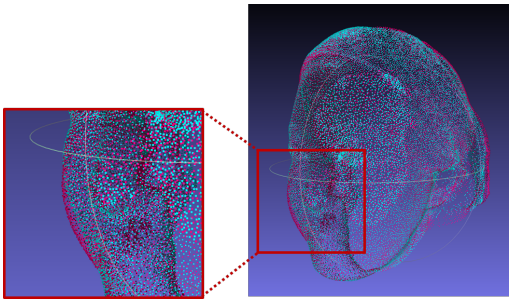


Figure 3. *3D model registration.*

2.2.2. Calculation of distances between two models

We compute the shape change from the earlier model to the later model. In this research, as in Masui et al.'s method, we apply the collision detection from the earlier model to the later model and calculate the distance between the two models. The Möller-Trumbore ray intersection algorithm [10] was used for the collision detection. An overview of the Möller-Trumbore ray intersection algorithm is shown in Figure 4. In three-dimensional

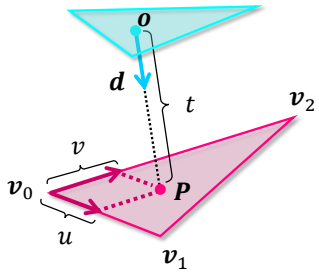


Figure 4. *Möller–Trumbore intersection algorithm.*

space, a point $\boldsymbol{R}$ on a ray whose length is $t \geq 0$ in the direction of unit vector $\boldsymbol{d}$ from the ray starting point $\boldsymbol{o}$ is represented by Equation 1.

$$\boldsymbol{P}(t) = \boldsymbol{o} + t\boldsymbol{d} \tag{1}$$

When a ray hits a point $\boldsymbol{T}$ on a mesh consisting of points $\boldsymbol{v}_0$, $\boldsymbol{v}_1$, and $\boldsymbol{v}_2$, the colliding points are expressed by the center of mass coordinate system as Equation 2.

$$\boldsymbol{T}(u,v) = (1-u-v)\boldsymbol{v}_0 + u\boldsymbol{v}_1 + v\boldsymbol{v}_2 \tag{2}$$

where, $u \geq 0, v \geq 0, u+v \leq 1$. When a ray and a mesh collide at a point P on the mesh, the point R on the ray in Equation 1 and the point $\boldsymbol{T}$ on the mesh in Equation 2 have the same coordinates, and thus are expressed by Equation 3.

$$\boldsymbol{R}(t) = \boldsymbol{T}(u,v) \tag{3}$$

By computing Equation 3, the ray and mesh intersection problem becomes a simple ternary linear equation problem solving three variables (t,u,v). The above equations can be solved by using Cramer's formula. And if three variables (t,u,v) are existing, it means that the ray and the mesh are intersected. Thus, we obtained the distance t between the corresponding meshes for the earlier and later models.

2.2.3. Visualization of distances

Based on the distance t between the corresponding meshes, the change in shape was visualized by coloring the meshes according to the distance. The color scale shown in Figure 5 was used for coloring. Red indicates an increase in volume, and blue indicates a decrease in volume. The range of the scale, 5 mm, was determined empirically from the results.

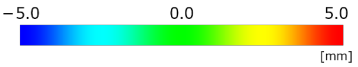


Figure 5. *Color scale.*

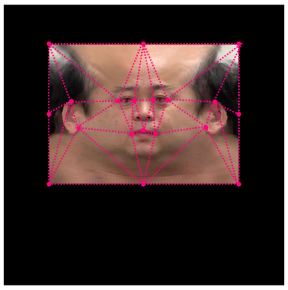
Visualization of skin texture

Next, we visualize the change in skin color in three dimensions. The texture map obtained in section 2.1 has different (u, v) coordinates for each 3D shape model. In other words, the facial landmarks in the texture map have a different coordinate for each measurement. Therefore, if we simply overlay the texture maps, the size and position of the faces are different, even on the same person. In this study, the texture map was converted by an affine transformation based on the facial landmarks. This transformation makes it possible to compare and calculate pixel values for skin color. Next, we apply the separation method for skin pigment [12] to the texture map to obtain a map of skin pigment on the face that excludes the unevenness of lighting. Then, by calculating the change in concentration of skin pigment between corresponding pixels, the change is calculated and visualized. In this section, we will explain in the above step.

2.3.1. Conversion of texture map

In this method, the texture map is transformed by performing the affine transformation for each triangle divided by the facial landmarks so that the coordinates of the source matches the target, taking into account the shape of the face. We first consider the facial landmarks for the transformation. In the acquired texture map, it is difficult to obtain the landmarks on the contour because the contour and the neck are drawn continuously, as shown in Figure 2. Therefore, following a face morphing system such as FUTON [11], we set a total of 19 facial landmarks for the eyes, mouth, ears, etc., and define a triangular patch with these landmarks as the vertices, so that the texture map can be transformed by the affine transformation of that triangular patches. Figure 6 shows the results of the transformation of Figure 2. Figure 6(a) is the same texture map as Figure 2. The 19 points in the figure represent manually set facial landmarks, and the dotted lines represent the triangular patches created by each point. The texture map is transformed by using an affine transformation matrix in which the triangular patches in Figure 6(a) match the coordinates of Figure 6(b), which is the target. The affine transformation matrix includes scaling, translation and rotation of the image. The transformed texture map is shown in the figure. 6(c). The texture in Figure 6(c) is the same as the source (Figure 6(a)), but the feature point coordinates are identical to those of the target (Figure 6(b)). It can be seen that the texture map can be transformed to match landmarks with arbitrary coordinates.

(a) Source. (b) Target.



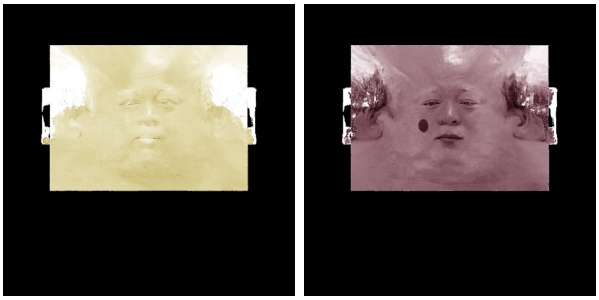
(c) Converted source.

Figure 6. *Conversion of texture map with 19 facial landmarks.*

2.3.2. Estimation of skin pigment map

By applying the pigment separation method based on the independent component analysis proposed by Tsumura et al [12] to the texture map image, the skin pigment maps of melanin and hemoglobin components, which constitute the skin color, are extracted from the texture map image. The pigment separation method assumes that melanin pigments in the epidermis and hemoglobin pigments in the dermis are spatially distributed independently to form the skin tone. We estimate a map of the skin pigment component by applying independent component analysis to the pixel values of the skin image. Changes in the concentration of pigment components are visualized by calculating the differences in pigment concentration between the corresponding texture maps and coloring them.

In this study, we measured and analyzed the skin replicating the hemangioma, as the details will be described later. The acquired texture map is the same, as shown in Figure 2. The results of the skin pigment separation [12] in Figure 2 are shown in Figure 7. Figure 7 is colored using melanin and hemoglobin pigment vectors for separation. Based on the skin pigment separation method, shades were mapped in the direction of the shade vector, and the concentration distributions of melanin and hemoglobin pigment components were obtained less affected by the unevenness of illumination. The hemangioma reproduced on the right cheek of the original image shown in Figure 2 was almost separated as a hemoglobin pigment component by Figure 7(b) and had almost no effect on the melanin pigment component shown in Figure 7(a), although its outline was slightly visible.

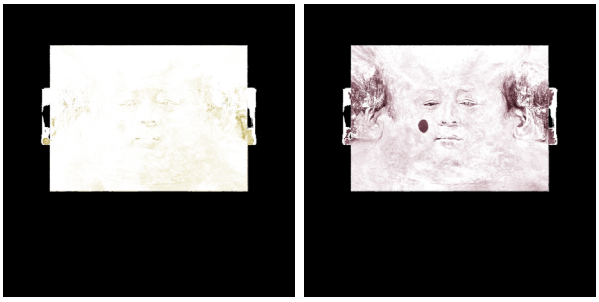


(a) Melanin map. (b) Hemoglobin map.

Figure 7. *Separation of skin pigment.*

2.3.3. Visualization of differences in skin pigment

The obtained concentrations of the skin pigments are subtracted in each pixels to calculate the amount of its change, and then colored with the vector of the skin color used in the separation to visualize the amount of its change. Figure 8 shows the visualization results of the change in skin pigment concentration. Whereas the melanin pigment component is unchanged in Figure 8(a), the entire face region appears white, the hemoglobin pigment component shown in Figure 8(b) shows the presence of a pigmented spot on the right cheek. The changes in skin color were visualized by the above procedure. In the actual system, this differential skin pigment map is texture mapped to a 3D shape model and displayed.



(a) Difference map in melanin. (b) Difference map in hemoglobin.

Figure 8. *Differences of skin pigment.*

3. Experiment

In this study, we aim to propose a system that can measure the three-dimensional shape of a face and compute and visualize the minute changes in shape and skin color in two different states. Therefore, in this chapter, in order to confirm the behavior of the proposed system on real subjects, we measured faces with different shapes and skin tones, and analyzed them by the proposed system. Section 3.1 describes the measurement procedure of this experiment, and section 3.2 describes the measurement data and analysis results.

3.1. Experimental setup

In this study, we used the iPhone XR as a proposed method to measure facial shape and skin texture. The faces were not fixed and measured in a real-environment under a white LED light source. Four subjects were measured: subject 1: female in 20s, subject 2: male in 20s, subject 3: male in 50s, and subject 4: male in 50s. The subjects were all Japanese. The following manipulations were performed with the help of a physician to reproduce the modulated face shape and skin color. For the shape, two or three medical cotton balls were stuffed into the right cheek to create a puffy state of the cheek. For color, we applied artificial hemoglobin to reproduce hemangiomas or acne-like redness.

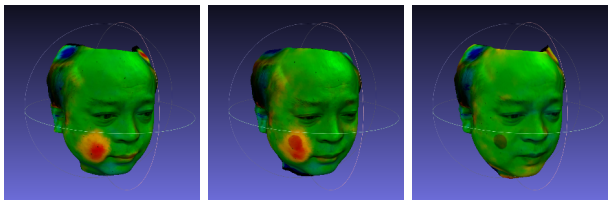
Each state and measurement procedure is explained. First, as a state (1), we measure the initial state. Next, as a state (2), the right cheek is filled with a cotton ball and the shape modulation state is measured. Then, as a state (3), by applying artificial hemoglobin on the puffy skin of the right cheek, we measure the modulated shape and color. Finally, as a state (4), the cotton ball of the right cheek is removed and the color modulation state is measured.

3.2. Experimental result

Four states of the face were measured for four subjects using Bellus3D on the iPhone XR. The visualization results of the changes in the face between the four states are shown on the 3D geometry. The results are described in Sections 3.2.1 and 3.2.2, respectively.

3.2.1. Results of changes in the shape of the 3D model

In this section, we visualize the shape change in each state of the four subjects, as an example, subject 4: male in the 50s. The initial state (1) is used as the target for 3D registration. The result of the 3D registration is shown in Figure 9. The facial shape was modulated when changed from the state (1) to (2) and from the state (1) to (3). In fact, according to both Figures 9(a) and 9(a), the right cheek is colored in red as meaning of bulged. That is, the 3D registration and intersection algorithms are done correctly. The results showed a similar trend in all four subjects.



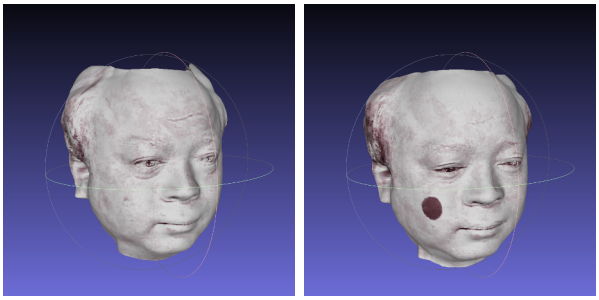
(a) From the state (1) to (2). (b) From the state (1) to (3). (c) From the state (1) to (4).

Figure 9. *Results of changes in shape.*

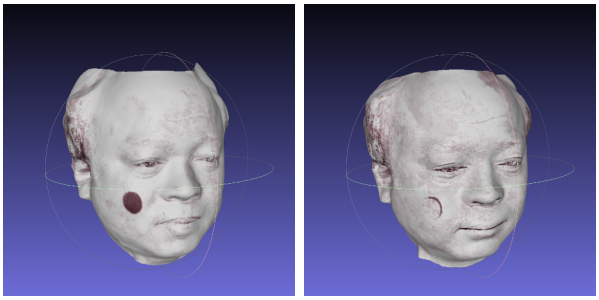
3.2.2. Results of changes in texture map of 3D model

In this section, we visualize the color change in each state in the same subject (subject 4: male in 50s). The melanin and hemoglobin components are obtained by applying the skin pigment separation method to the texture map images, and the amount of change will be visualized. The results are shown in Figure 10. In comparison with the results shown in Figure 7 in Section 2.3.2, we caught less change for melanin in this time, so we explain the results only for hemoglobin. When comparing before and after application of artificial hemoglobin (from states 1 to 3, and 1 to 4), Figure 10(a) and Figure 10(b) show that a circular difference in hemoglobin dye concentration appeared at the applied site. On the other hand, it was confirmed that positional misalign-

ment occurred between texture maps when artificial hemoglobin was applied with different shapes, such as the changes in states 3 to 4 shown in Figure 10(d). These results showed a similar trend in all four subjects.



(a) From the state (1) to (2). (b) From the state (1) to (3).



(c) From the state (1) to (4). (d) From the state (4) to (3).

Figure 10. *Results of changes in hemoglobin.*

4. Discussion

In this study, we proposed a system to visualize shape and color changes in the treatment of skin diseases on the face. A 3D shape model and texture map of the face were obtained by using a structured optical depth sensor on the iPhone XR. The distance of the shape modulation part is calculated by performing the 3D registration and intersection algorithm for the 3D shapes of multiple faces. We successfully visualized the shape modulation part by coloring the mesh according to the distance. In addition, it was possible to compare the texture maps of each pixel by transforming the texture maps so that the (u, v) coordinates are the same.

First, we discuss the visualization results of the change in the shape. The results of Figure 9 show that the entire face was colored green, which means no change in shape, and only the right cheek area, where the shape was modulated, was colored red as a change of less than 5 mm. The red-colored area corresponds to the area where the cotton ball was inserted, and it can be said that the shape change could be visualized as predicted.

Next, we discuss the visualization results of the change in skin color. When comparing before and after color modulation, we found that the modulation was visualized in the right cheek, mainly for the hemoglobin pigment component, from Figure 7 and Figure 10. However, as the change from the state (4) to (3) shown in the Figure 10(d), when the shape is modulated between states with pigmented spots, it is confirmed that the pigmented spots do not completely overlap when the texture map is transformed, resulting in a slight deviation during subtraction. The proposed transformation of the texture map using the landmarks is just only a two-dimensional image processing of the texture map, which represents the three-dimensional view in two dimensions. In other words, it may be due to the fact that the transformation is not strictly based on the facial 3D shape. Therefore, it is suggested that a transformation based on the shape modulation of the actually acquired 3D shape model is necessary to transform with more strict shape considerations. The separation was not perfect for the melanin pigment component in Figure 7, according to the results shown in Figure 7. Since the skin pigment separation was originally an estimation assumed a two-layer model of the skin [12], it probably did not work correctly when the skin was exposed to the surface, as in the present experiment.

This transformation made it possible to compute changes in skin pigment concentration and visualize skin color modulation. The results were similar when performed on four subjects. Using a depth camera in a smartphone, it was confirmed that the changes in face shape and color could be easily analyzed. In the past, measurements were mainly made using very expensive specialized medical equipment. In the future, however, it is expected to become possible to easily measure the state of the face in three dimensions by using a measurement system like the one described above.

However, there are many problems in terms of accuracy when aiming for practical use as a medical device. We have not been able to verify the accuracy of the acquired shapes. Therefore, it is necessary to verify the accuracy of the 3D shape measured by the iPhone XR by comparing it with the ground truth data obtained using a 3D scanner with guaranteed accuracy. Color correction was not taken into account because the measurements were made in the same environment under a white light source. In the future, it is necessary to measure the face with a color chart attached to the face under various light sources, evaluate the color reproducibility of each light source, and perform color correction. For the transformation of texture maps, we consider the use of mesh-to-mesh correspondences obtained by the intersection algorithm in order to more strictly consider the 3D geometry.The purpose of this study was to visualize the skin color change in the texture map. In the future, we will consider using different skin models for the regions of interest to cover specific skin structures, such as hemangiomas. In the future, we will prove that it is possible to put this into practice in clinical practice.

5. Conclusion

In this study, the face shape and the skin color were measured using the infrared depth sensor and the RGB sensor built in the smartphone three-dimensionally. We measured before and after modulating the shape and color of the face and calculated, and then, the change in volume and the change in skin pigment of skin color was calculated and visualized. This method makes it possible to analyze the shape and skin color independent of the viewing angle and the direction of the illumination light. In this study, the depth camera built in the smartphone showed the potential to monitor changes in facial shape and skin color. In order to contribute to the development of telemedicine, in which the patient measures their face at home and gets medical treatment consultation remotely, we need to confirm the accuracy and the reproducibility of our system strictly by using the 3D scanner for the 3D shape and skin color.

References

[1] Krowchuk, Daniel P., et al. ”Clinical practice guideline for the management of infantile hemangiomas.” Pediatrics 143.1 (2019): e20183475.

[2] Canfield Scientific, Inc., ”VECTRA H1 3D Imaging System”, `https://www.canfieldsci.com/imaging-systems/vectra-h1-3d-imaging-system/`, (accessed 27th March 2020).

[3] Camison, Liliana, et al. ”Validation of the Vectra H1 portable three-dimensional photogrammetry system for facial imaging.” International journal of oral and maxillofacial surgery 47.3 (2018): 403-410.

[4] Seitz, Steven M., et al. ”A comparison and evaluation of multi-view stereo reconstruction algorithms.” 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06). Vol. 1. IEEE, 2006.

[5] Kenta Masui, Kaoru Kiyomitsu, Keiko Ogawa-Ochiai, Takashi Komuro, and Norimichi Tsumura. ”Technology for visualizing the local change in shape of edema using a depth camera”. Artificial Life and Robotics. 24. 10.1007/s10015-019-00541-1.

[6] Apple Inc., ”iPhone”, `https://www.apple.com/iphone/`, (accessed 27th March 2020).

[7] Bellus3D, Inc., ”Bellus3D: High-quality 3D face scanning”, `https://www.bellus3d.com/`, (accessed 27th March 2020).

[8] Besl, Paul J., and Neil D. McKay. ”Method for registration of 3-D shapes.” Sensor fusion IV: control paradigms and data structures. Vol. 1611. International Society for Optics and Photonics, 1992.

[9] Fischler, M. A., and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24(6), 381-395.

[10] Möller, Tomas, and Ben Trumbore. ”Fast, minimum storage ray-triangle intersection.”, Journal of graphics tools, 2.1 (1997): 21-28.

[11] Mukaida, Shigeru, et al. ”Facial image synthesis system: FUTON-evaluation as tools for cognitive research on face processing.” Trans. IEICE 85.10 (2002): 1126-1137.

[12] Norimichi Tsumura, et al. ”Image-based skin color and texture analysis / synthesis by extracting hemoglobin and melanin information in the skin.” ACM Transactions on Graphics (TOG) 22.3 (2003): 770-779.

Author Biography

Ms. Ikumi Nomura was received the B. Eng degrees in computer science from Chiba University, Chiba, Japan, in 2018. She is now a master course student of the same university. Her research interest includes skin color and facial tracking it in long-term.