

Special Paper

방송공학회논문지 제25권 제7호, 2020년 12월 (JBE Vol. 25, No. 7, December 2020)

<https://doi.org/10.5909/JBE.2020.25.7.1042>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

Interactive Pixel-unit AR Lip Makeup System Using RGB Camera

Hyeongil Nam^{a)}, Jeongeun Lee^{a)}, and Jong-Il Park^{a)†}

Abstract

In this paper, we propose an AR (Augmented Reality) lip makeup using bare hands interactively using an RGB camera. Unlike previous interactive makeup studies, this interactive lip makeup system is based on an RGB camera. Also, the system controls the makeup area in pixels, not in polygon-units. For pixel-unit controlling, the system also proposed a 'Rendering Map' that can store the relative position of the touched hand relative to the lip landmarks. With the map, the part to be changed in color can be specified in the current frame. And the lip color of the corresponding area is adjusted, even if the movement of the face changes in the next frame. Through user experiments, we compare quantitatively and qualitatively our makeup method with the conventional polygon-unit method. Experimental results demonstrate that the proposed method enhances the quality of makeup with a little sacrifice of computational complexity. It is confirmed that natural makeup similar to the actual lip makeup is possible by dividing the lip area into more detailed areas. Furthermore, the method can be applied to make the face makeup of other areas more realistic.

Keywords AR Makeup, pixel-unit make up, Rendering Map, Interactive Makeup

1. Introduction

Virtual makeup technology is developing with AR technology year after year. Commercialized products or studies takes the method of masking the entire area or part (lips, eye, etc.) to be made up or blending a certain mask [3, 4, 5, 9, 11, 12, 13, 17]. However, for a virtual makeup system like the actual makeup, the user should be able to make more detailed makeup even inside each area of the face

(lips, eyes, etc.). Specifically, it should be able to show the effect of makeup and repainting of smaller units.

Also, among the methods for interactive makeup, there is a method using Depth when using a makeup tool or a hand gesture [6, 7, 8, 16, 21, 23]. But if users can make up interactively without a depth camera, it will be applicable to a much wider range. In order to make makeup on face using hand movements with an RGB camera, landmarks of face can be detected by Dlib while Openpose is used for hand movement [1, 2]. More specifically, the system focuses on how to distinguish the area of the lips, the area of hand contact, and how the color of the lips changes. In this process, a Rendering Map was created to memorize the relative position where the finger was touched in real time and apply to the part for each frame. The Rendering Map contains the number of the polygon area and the ratio of x and y coordinates.

a) Department of Computer Science, Hanyang University, Seoul, Korea

† Corresponding Author : Jong-Il Park

E-mail: jipark@hanyang.ac.kr

Tel: +82-2-2220-0368

ORCID: <https://orcid.org/0000-0003-1000-4067>

※ This work was supported by the National Studies Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2019R1A4A1029800).

· Manuscript received September 25, 2020; Revised December 2, 2020; Accepted December 2, 2020.

dinates to be drawn in the area based on the landmark of the lip. By using this, the system can interactively store and show a part of the lips makeup where the hands are touched in real time. The method implemented in this way is compared and evaluated quantitatively and qualitatively with the makeup system of the previous study.

The contribution of this paper is as follows:

- Proposal of AR makeup system using Rendering Map interactively in real time
- Using only an RGB camera for interactive makeup without additional devices
- Comparison of this method with the existing interactive virtual makeup method.

II. Previous Related Work

1. Pixel-unit makeup

Existing makeup-related apps use a model-based or deep learning method to find facial features and cover images or colors with masks on the whole or parts of the face (lips, eyes, nose) [3, 4, 5, 9, 10, 17, 18]. Li et al. proposed makeup transfer using deep learning. They imposed a loss term, measuring the pixel-level differences between the reference image and the input image [3]. ModiFace creates a radial-gradient translucent colored mask and assigns colors to all pixels within the “lip box” according to the weighted-average value being either preset or chosen by a user

[4]. Evangelista et al. detected face landmarks using a mobile face tracker and applied makeup colors to pixel coordinates in 3D based makeup maps [17]. Oztel et al. proposed a method based on image processing [18]. They detected a lip area and changed the lip color through color segmentation in the lip box. Li et al. decomposed an input image into albedo layer, diffuse shading layer and specular layer. They changed colors of the albedo layer after calculating them by Kubelka-Munk model [14]. Bernado et al. [15] proposed a method to render makeup textures and 2D mesh. The projection image is shown by matching a texture image and transformed vertex positions onto the projector’s image plane. These apps and studies use a mask to match the whole face and blend it or use deep learning to change the original face image to a style image. Even in commercial products, lip makeup is applied all over the lips with a one-touch interface [4, 5, 10]. However, for real virtual makeup, the users should be able to do makeup on the part of the face such as lips and eyes in more detail as desired by them. In other words, there is a need for a makeup system that can fine control the inner areas of the lips and eyes in smaller units. Borges’s studies proposed a system to do makeup by changing the color of the triangular area made from the feature point of the face [6]. This method has the advantage that the user can interactively check the process of the face makeup. However, there are some limitations by completely changing the color inside each polygon after dividing the face into polygons. First of all, in the middle of the makeup process, the makeup



Fig. 1. Comparison of (a) polygon-unit makeup and (b) pixel-unit makeup

part and the non-makeup part are separated in polygon-units, so it is expressed unnaturally. Also, it is difficult to apply detailed makeup styles within polygons already applied because the face area can be controlled only in the polygon-unit. The polygon-unit makeup refers to the makeup in polygon size as the previous works. Otherwise the pixel-unit makeup is the method managing makeup area in pixel size. In this paper, we compare these methods through experiment.

Figure 1 shows the difference between polygon-unit makeup and pixel-unit makeup as an example. Therefore, our method enables delicate makeup by dividing the lip area into more subdivided areas than previous studies.

2. Interactive Virtual makeup

It appeared as a mobile app of the existing product or studies to enable face makeup interactively using a touch device [4, 5, 19, 22]. Rahman's study used equipment for makeup. In the study, a smart mirror system was proposed and used to recognize makeup areas by using RFID and IR cameras as hardware [7]. And the study put a certain mark on the makeup equipment and Kinect recognized it

to make up the face. In the study, when the fingertip was recognized by Kinect and contacted within a certain area of the face, 3D coordinates were converted to 2D to identify the touched part and to change the entire polygon area. In addition, studies that have virtual makeup without a touch device have emerged [6, 7, 8, 16, 20, 23, 24]. And there was a study of a system using bare hands without any special equipment for the user [6].

However, for less equipment constraints and wider versatility, Interactive makeup is needed commonly used rather than interaction equipment using depth. Therefore, in this paper, makeup is possible by only bare hands and the RGB camera that can be encountered in daily life. We used Dlib to extract the landmarks of the face and OpenPose to detect the landmarks of the palms and fingers [1, 2]. OpenPose is suitable for detecting the whole hand and estimating the pose in that it can generate interaction using various hand gestures later.

III. AR Lip Makeup System

When a video frame comes in, it detects landmarks on

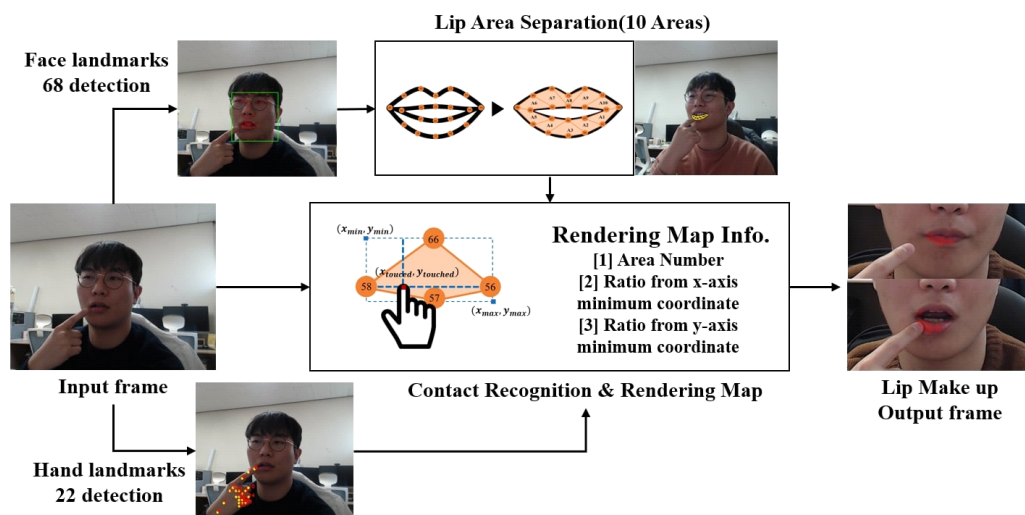


Fig. 2. Lip Makeup System Overview

the face and hands. The system detected 68 face landmarks across the face, considering not only the lip makeup but also the makeup of other facial areas with Dlib. Also, 22 landmarks were extracted to find the fingers and palms, considering the various uses of the hand by OpenPose.

First, the lip part is divided into polygons based on facial landmarks (10 polygons having four angles in the implementation). This divided polygon area is a map for storing coordinates information that will be adjusted in real time when the hand (in the implementation, using the index fingertip) touches the lip polygons. Based on the Rendering Map, the lip makeup is adjusted to the next frame and displayed to the user as an output. (Fig 2).

1. Lip Area Separation

Among 68 landmarks of the face, 20 key points related to lips were divided into small polygon-units. The divided polygons are used as the management unit of the Rendering Map for dynamically rendering the pixel color for each real time frame. Those are division units for indicating the position of the lips after the finger is in contact with them. The area of the lip can be divided into triangles, squares or more polygons.

When polygons are divided into too large units, the number of areas to be managed is reduced, but the amount of information to be managed per area may be too large. Otherwise, when those are divided into too small units, the number of areas to be managed increases, but the amount

of information to be managed per area is relatively reduced. Therefore, in this paper, the user's lips color can be changed by empirically dividing the lips into 10 polygon areas with four angles so that the system quickly can find x, y coordinates on 2D for each frame. As shown in Fig. 3, the lip area is drawn and well distinguished (Fig 3).

2. Contact Recognition

After dividing the area of the lip into polygons as described above, it is necessary to determine whether a landmark of the hand is inside the polygon so that the lip makeup can be adjusted only in the specific area. In detail, when a point is located inside the polygon, we can determine by the feature of that point. As shown in Fig. 3, when a half straight line is drawn from the target point (in Fig 4) to be identified, if the number of points that meet the polygon is an odd number, it is determined to be an internal point. On the contrary, a straight half line is drawn at the point missed point (in Fig 4) to be discriminated, and if the number of intersection points is even, the point is determined to be outside. To do this we must consider that the half line is always a horizontal line parallel to the x axis. Then the system checks for every line of the polygon that each line has an intersection with a half line. For implementation, the y-coordinate of the target point should be between the y-coordinates of two vertices as below conditions, and the x-coordinate of the intersection point between the horizontal line and the line segment passing

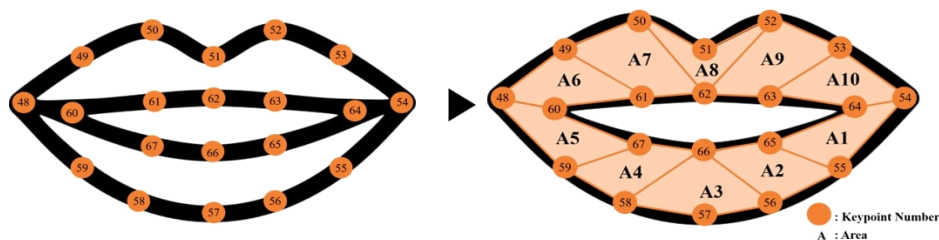


Fig. 3. Lip Area Separation (10 Areas)

through the target point is larger than the x-coordinate of the target point.

Conditions (If N is last, the N + 1 vertex is the first vertex),

$$y_N > y_{Target} > y_{N+1} \quad (1)$$

- y_N, y_{N+1} : y-coord. of N th and (N+1) th vertex of the polygon

- y_{Target} : y-coord. of the target point

$$x_{Target} < x_{Intersection} \quad (2)$$

- x_{Target} : x-coord. of the target point

- $x_{Intersection}$: x-coord. of the intersection of the horizontal line and the line

(Nth vertex, (N + 1) th vertex)

■ Intersection

● P: Point inside polygon

● P': Point outside polygon

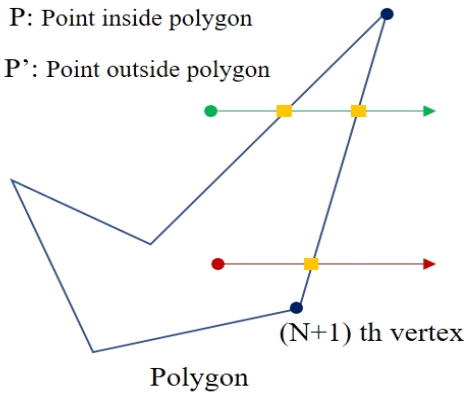


Fig. 4. For explaining conditional expressions to determine if a point is inside any polygon

3. Rendering Map

When a finger touches over a divided area, makeup should be made at the coordinates of the lip area where the finger is in contact. makeup should be done on the coordinates of the lips area which the finger is touched. However, landmarks of the lips are detected differently each frame, it is necessary to remember the relative position of the parts where makeup should be affected. Therefore, in this paper, the Rendering map is proposed and stores landmark information of the polygon and ratio information about the detailed location in the polygon. For example, if there are 4 different landmarks in a polygon as Landmark1 to Landmark4 in regular order, ratio information should be calculated as the pixel position in contact with the intersection of a line (Landmark 1 to Landmark 3) and the other (Landmark 2 to Landmark 4).

More specifically, as shown in the Fig 5, once a contact is recognized, a bounding box is created based on the landmark to which the contacted area belongs, and the relative positions of the x and y coordinates are calculated. And the system stores those in the Rendering Map along with the area number. Then, when the corresponding lip region landmark can be detected in the subsequent frame, the makeup effect is reflected in the corresponding position by calculating the position to be drawn using the ratio in the Rendering Map. In this paper, the position can be specified

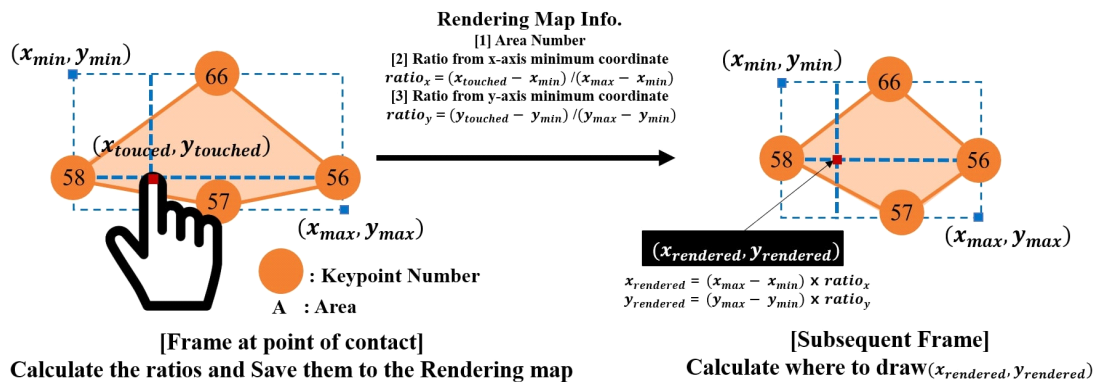


Fig. 5. Process for embedding information in Rendering Map

by the coordinates, and the makeup effect is applied to 2-4 pixels up and down from the position. This essential process enables smaller lip area makeup.

4. Lip Makeup

In the case of pixel color changing, the blending method is applied for using the shade of lips naturally reflected in the user's existing lighting environment, rather than the masking method that covers the whole lips with a certain color. Although multiple colors can be artificially adjusted at once based on where they are contacted, we adjusted the pixel values to emphasize only the red color of the lip from the existing color for the sake of naturalness. In this way, the system can selectively apply the effect of the area being overlaid. Also, the color would be highlighted when the same position is selected several times like actual makeup.

As shown in Fig. 6, when the red color is mixed at a constant weight ratio (original: 0.75, red: 0.25), it is compared to the case of blending once and changed pixel in three times (Fig 6). In this way, we will be able to selectively apply effects to the system that have been painted once and at multiple times with a certain ratio.

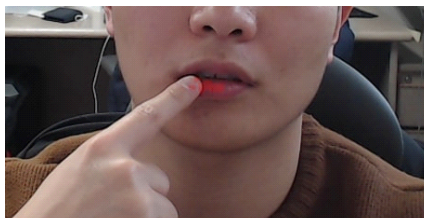
IV. Experiment and evaluation

1. Implementation and experiment

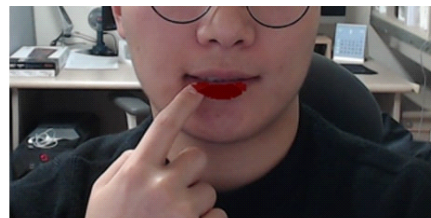
In order to find the feature points of the face and hands

in the RGB camera, we basically used the open source system. To find the feature points of the face, we could find the face landmarks using the Dlib library and use the deep learning based OpenPose to detect the hand landmarks. The experiment received input images with the Logitech 960's Camera with a frame size of 1024 x 1024. Also, when a pixel is specified in the Rendering Map, makeup effect is commonly reflected by 2 pixels up and down based on this. And the distance between the camera and the experiment participants in the x-axis direction was measured at 0.45-0.5m. Participants were 11 campus students aged 23-31 years (mean 27.4 years). Two experiments were conducted for each participant. First, six locations (under lips (5 areas) and upper left part (1 area)) were randomly specified in advance, and one makeup was performed per area.

In order to show the detailed makeup process of the lip area in the experiment, it is necessary to consider the situation of makeup the part of the lips rather than the entire lips. And the baseline of completion was based on the time measured when the makeup of the lip region was finished (Fig. 7). While the polygon-unit method fills the color of a polygon with a single touch, the pixel-unit method has to be performed several times to fill one polygon. In consideration of this, only the case where the color is filled in more than 80% of the area was allowed. In addition, we evaluated qualitatively, including interviews and user satisfaction surveyed on the Likert 5-point scale, by showing the user's makeup on their lips. Second, they randomly specified 4 locations in advance, and reapplied makeup



a. From A3 to A4: 1 time



b. From A2 to A4: 3 times

Fig. 6. Blending Effects with RGB-Red (0, 0, 255), Each pixel weight: Original 0.75, Red 0.25

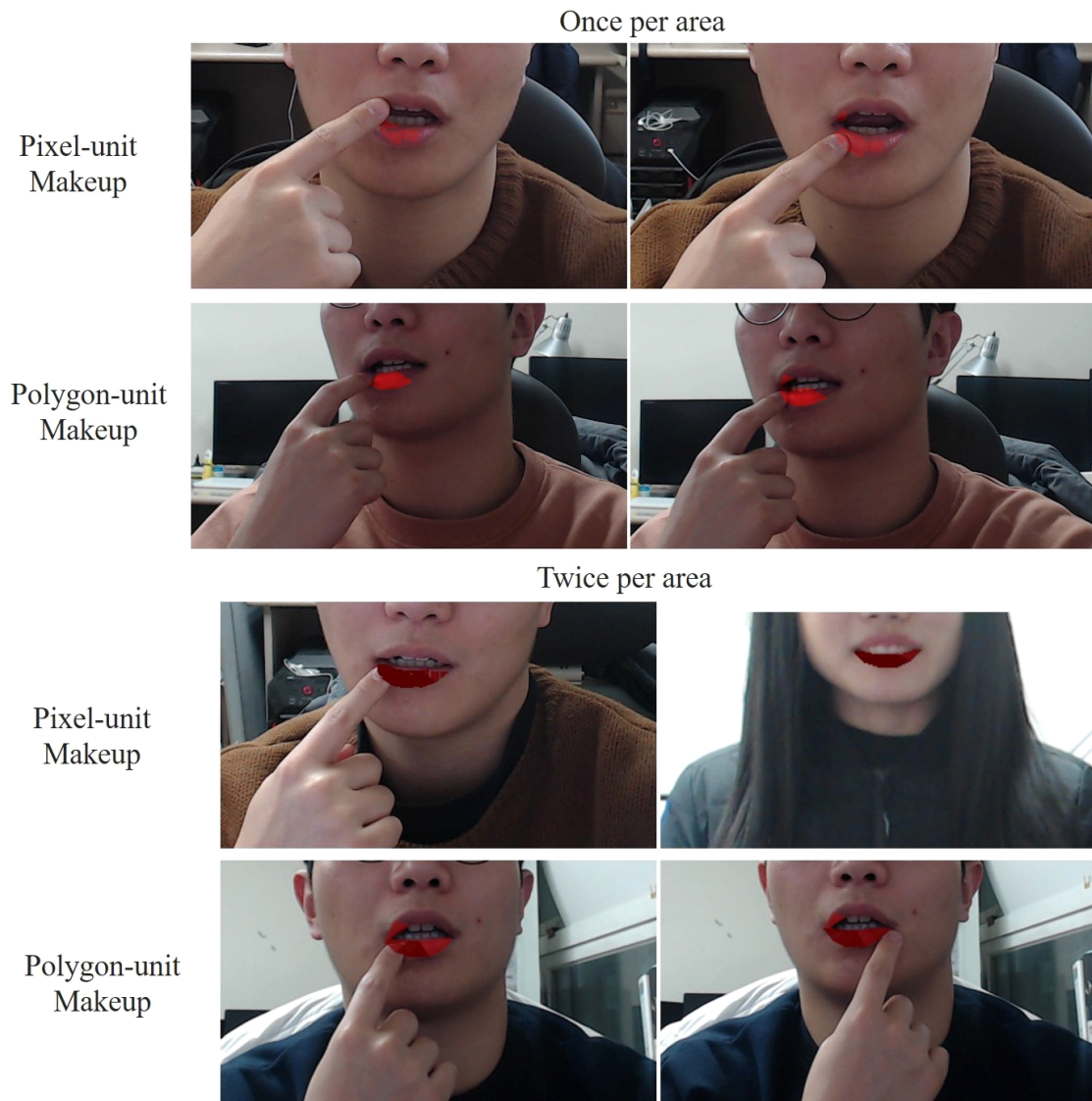


Fig. 7. Lip Makeup Comparison: Ours (pixel-unit) and previous work (polygon-unit)

more than twice. Like the previous experiment, the makeup of the pixel-unit is allowed only if more than 80% of the area is filled. Also, the qualitative part was evaluated by including Likert 5-point scale and interview.

2. Results

First, when one makeup was performed per area, the pol-

YGON-unit method was performed quickly within 1.4 seconds on average, and the proposed method in the pixel-unit performed also within 2.1 seconds on average. This is because the makeup method in polygon-units fills one area at once, whereas the pixel-unit method requires several hand movements to fill one area, which takes more time.

There is a small difference in the average time of doing makeup between twice and once per area (Table 1).

Table 1. Average time(std.) for lip makeup(second)

	once per area	twice per area
polygon-unit makeup	1.381 (0.421)	2.368 (0.458)
pixel-unit makeup	2.04 (0.396)	2.706 (0.474)

However, the method of pixel-units can provide detailed makeup expression. Natural and detailed expressions are important in makeup. Even considering the time, our method is appropriate. The qualitative evaluation showed that the makeup method (4.18) of the pixel-unit was higher than the average satisfaction (2.81) of the polygon-divided part when one makeup was made per area. Most of the participants expressed the opinion that the boundary divided by polygons was unnatural, and the opinion that the makeup method of the pixel-unit expressing the touched part of the finger was natural, and explained that they scored based on these opinions (Fig 8).

Next, when two or more makeup were performed per area, the time in the polygon-unit method was less than 2.64 seconds on average, and the pixel-unit method was less than 3.76 seconds on average. However, the satisfaction test showed a larger gap than when one makeup was performed per area (pixel-unit: 4.27, polygon-unit: 2.36). This is the same reason as the mentioned above, but the majority said that the degree of naturalness differed more clearly. (7 out of 11).

V. Conclusion

We proposed a system that could make users do virtual lip makeup with hand gestures interactively in real-time using only an RGB camera. To do this, we created a Rendering Map that separates the area of the lip and dynamically enabled the system to draw the makeup position for each frame. When the finger touches the area of the separated lips, the relative position is calculated as a ratio based on the separated landmarks and stored in the Rendering Map. Using the stored position, when landmarks are captured again in a later frame, the makeup part can be calculated and rendered at the position. We compared our method of pixel-units with the previous studies which are the method of simply dividing landmarks into polygons and verified their effectiveness through time and satisfaction level. Compared to the makeup of the pixel-unit, a polygon is filled with makeup color by one touch in the makeup of the polygon-unit, therefore the execution time is slower than the makeup of the polygon-unit. However, the makeup method of the pixel-unit using Rendering Map is much more natural and detailed control than the makeup method of a polygon-unit without the unnatural border between the polygons or the parts with and without makeup. In addition, since the method uses only an RGB camera, the parameters to be considered are effectively reduced

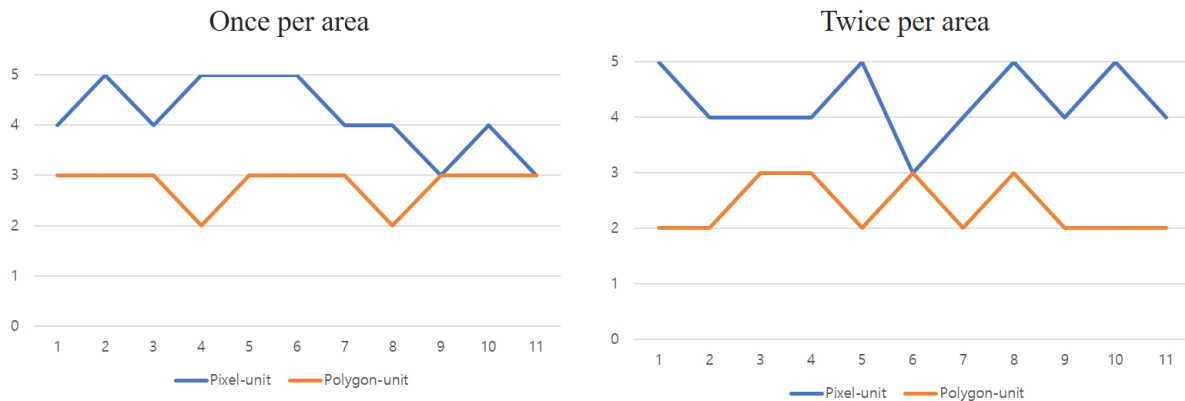


Fig. 8. Satisfaction Level Comparison: Ours (pixel-unit) and Previous work (polygon-unit) in case of makeup once and twice per areas

compared to the previous studies that require further adjustment of depth information for interaction. In the future, the blending ratio between the original color and makeup material color can be varied in the blending section, and the virtual makeup system will be much more natural if it reflects all the effects of the repaint. In addition, it can be exploited to various makeup productions such as smudging or contouring makeup. Furthermore pixel-unit makeup can be applied to a range of sizes of makeup material. And we will apply the method not only to the lips but also to other parts of the face.

References

- [1] Dlib C++ Library. <http://dlib.net/> (accessed Sep 20, 2020).
- [2] Z. Cao, G. Hidalgo, T. Simon, S. Wei and Y. Sheikh, "Open-Pose: real-time multi-person 2D pose estimation using Part Affinity Fields", *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Hawaii, USA, pp.7291-7299, July 2017.
- [3] T. Li, R. Qian, C. Dong, S. Liu, Q. Yan, W. Zhu and L. Lin, "BeautyGAN: Instance-level Facial Makeup Transfer with Deep Generative Adversarial Network," *Proceedings of ACM Multimedia Conference(MM '18)*. pp 645-653. Seoul, Republic of Korea, October 2018.
- [4] P. Aarabi, "Method and system for simulated product evaluation via personalizing advertisements based on portrait images," US Patent 9,275,400 B2, to MODIFACE INC., Patent and Trademark Office, Washington D.C., 2016.
- [5] SNOW application. <https://www.snowcorp.com/ko/> (accessed Sep 20, 2020)
- [6] A. Borges and C. Morimoto "A Virtual Makeup Augmented Reality System," *Proceeding of 21st Symposium on Virtual and Augmented Reality (SVR)*, pp. 34-42. Rio de Janeiro, Brazil, 2019.
- [7] B. Treepong, P. Wibulpolprasert and S. Hasegawa, "The Development of an Augmented Virtuality for Interactive Face Makeup System," *Proceeding of Advances in Computer Entertainment Technology (ACE)*, London, UK, pp. 614-625. 2017.
- [8] A. Rahman, T. Tran, S. Hossain, "Augmented Rendering of Makeup Features in a Smart Interactive Mirror System for Decision Support in Cosmetic Products Selection." *Proceeding of the IEEE/ACM 14th International Symposium on Distributed Simulation and Real Time Applications*, Fairfax, Virginia, USA, pp203-206, 2010.
- [9] R. Bao, H. Yu, S. Li and B. LI, "Automatic makeup based on generative adversarial nets," *Proceeding of the International Conference on Internet Multimedia Computing and Service*, New York, USA, pp.1-5, 2018.
- [10] Business Apps & Solutions | PERFECT Application. <https://www.perfectcorp.com/business>. (accessed Sep 20 2020).
- [11] S. Kristina, R. Tobias, H. Matthias, T. Thorsten, B. Volker and S. Hans-Peter, "Computer-Suggested Facial Makeup," *Computer Graphics Forum*, Vol. 30, No.2, pp.485-492, April 2011.
- [12] D. Guo and T. Sim, "Digital face makeup by example," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, Miami, USA, pp. 73 - 79, 2009.
- [13] A. Nishimura and I. Siio, "iMake: computer-aided eye makeup," *Proceedings of the 5th Augmented Human International Conference*, Kobe, Japan, pp.1-2, 2014.
- [14] C. Li, K. Zhou and S. Lin, "Simulating makeup through physics-based manipulation of intrinsic image layers." *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, pp.4621-4629, 2015.
- [15] A. Bermano, M. Billeter, D. Iwai and A. Grundhöfer, "Makeup Lamps: Live Augmentation of Human Faces via Projection," *Computer Graphics Forum*, Vol.36, No.2, pp.311-323, May 2017.
- [16] E. Iwabuchi, M. Nakagawa and I. Siio, "Smart Makeup Mirror: Computer-Augmented Mirror to Aid Makeup Application," *Proceeding of Human-Computer Interaction. Interacting in Various Application Domains (HCI)*, San Diego, CA, USA, pp.495-503, 2009.
- [17] B. Evangelista, H. Meshkin, H. Kim, A. Aburto, B. Rubinstein and A. Ho, "Realistic AR makeup over diverse skin tones on mobile," *Proceeding of SIGGRAPH Asia*, Tokyo, Japan, pp.1-2, 2018.
- [18] G. Oztel and S. Kazan, "Virtual Makeup Application Using Image Processing Methods," *International Journal of Scientific Engineering and Applied Science*, Vol.1, No.5, pp.401-404, August 2015.
- [19] D. Almeida, P. Guedes, M. Silva, A. Silva, J. Lima and V. Teichrieb, "Interactive Makeup Tutorial Using Face Tracking and Augmented Reality on Mobile Devices," *Proceeding of 2015 XVII Symposium on Virtual and Augmented Reality*, Sao Paulo, Brazil, pp.220-226, 2015.
- [20] B. Treepong, H. Mitake and S. Hasegawa. "Makeup Creativity Enhancement with an Augmented Reality Face Makeup System," *Computers in Entertainment*, Vol.16, No.4. pp.1-17, November 2018.
- [21] Z. Ye et al., "Method of virtual makeup achieved by facial tracking," US Patent 9,224,248 B2, to ULSee Inc., Patent and Trademark Office, Washington D.C., 2015.
- [22] M. Choe et al., "Makeup supporting methods for creating and applying a makeup guide content to makeup user's face on a real-time basis," US Patent 10,083,345 B2 to Myongsu CHOE, Patent and Trademark Office, Washington D.C., 2018.
- [23] Y. Goto, "Makeup simulation system, makeup simulation apparatus, makeup simulation method, and makeup simulation program," US Patent 8,107,672 B2, to Shiseido Company, Ltd., Patent and Trademark Office, Washington D.C., 2012.
- [24] C. Cheng et al., "Systems and methods for interactive virtual makeup experience," US Patent 2018/0165855 A1, to Perfect Corp., Patent and Trademark Office, Washington D.C., 2018.

Introduction Authors



Hyeongil Nam

- 2018 : B.S. degree in Business Administration, Hanyang University, Seoul, Korea
- 2018 ~ present : Combined Master and Ph.D student, Department of Computer Science, Hanyang University, Seoul Korea
- ORCID : <https://orcid.org/0000-0002-6017-8869>
- Research Interest : Virtual and augmented reality, Computer Vision/Graphics, Realistic Rendering, Human-Computer Interaction



Jeongeun Lee

- 2015 : B.S. degree in Informaion Systems, Woosong University, Dajeon, Korea
- 2016 ~ 2019 : Researcher, Ucon System Corporation, Daejeon, Korea
- 2019 ~ present : Master Student, Department of Computer Science, Hanyang University, Seoul, Korea
- ORCID : <https://orcid.org/0000-0002-1478-442X>
- Research Interest : Virtual and augmented reality, Computer Vision, Realistic Rendering



Jong-II Park

- 1987 : B.S. degree in Electronics Engineering, Seoul National University, Seoul, Korea
- 1989 : M.S. degree in Electronics Engineering, Seoul National University, Seoul, Korea
- 1995 : Ph.D. degree in Electronics Engineering, Seoul National University, Seoul, Korea
- 1992 ~ 1994 : Visiting Researcher, NHK Science and Technology Research Laboratories, Tokyo, Japan
- 1995 ~ 1996 : Researcher, Korean Broadcasting Institute, Seoul, Korea
- 1996 ~ 1999 : Researcher, ATR Media Integration and Communication Research Laboratories, Kyoto, Japan
- 1999 ~ present : Professor, Department of Computer Science, Hanyang University, Seoul, Korea
- ORCID : <https://orcid.org/0000-0003-1000-4067>
- Research Interest : Virtual and augmented reality, Computational Imaging and display, Computer Vision/Graphics, Human-Computer Interaction