Quantitative Assessment Support for the Improvement of Facial Appearance Around the Mouth Based on the Local Image Classification
（局所的な画像判別に基づく口唇周辺領域の顔貌改善のための定量評価支援）

２０２１年　３月

石 切 山　　順 一

Quantitative Assessment Support for the Improvement of Facial Appearance Around the Mouth Based on the Local Image Classification
（局所的な画像判別に基づく口唇周辺領域の顔貌改善のための定量評価支援）

石 切 山　　順 一

グローバル教育院

エンパワーメント情報学プログラム

筑波大学

２０２１年　３月

Declaration of Authorship

I, Junichi Ishikiriyama, declare that this thesis titled, “Quantitative Assessment Support for the Improvement of Facial Appearance Around the Mouth Based on the Local Image Classification” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: Junichi Ishikiriyama

Date: 18 February 2021

Abstract

Facial appearances are known to have considerable influence on interpersonal communication; when facial appearances are compromised, it may be detrimental to psychological and physical health. To remedy any possible disadvantages, medical treatment and livelihood support are provided to improve the appearance of the face through face-to-face assessment. However, such facial evaluations are easily influenced by individual subjectivity and is supported by specific individuals with expertise. These problems lead to a decrease in inter-rater reliability and the restriction of use in scenes of facial paralysis diagnosis and makeup confirmation for visually impaired persons, respectively.

We then propose artificial intelligence systems that can assess facial appearance, assist experts in their assessment, and perform face evaluation on behalf of experts. The first is makeup support for the visually impaired. Here, we propose an interface to confirm unintentional makeup such as makeup protrusion. The second is diagnosis support for facial nerve paralysis. Here, we discuss a quantification and classification method for facial movements to facilitate the diagnosis of paralysis level.

The research questions are as follows: 1) What local features affect the improvement of facial appearance represented by subregional face shapes and motions? 2) What are the relations between the expert's facial assessments and subregional facial features? 3) What are the functions of interfaces for use in uncontrolled environments?

We propose methodologies that can extract local features around the lip parts and represent the relation between them and the evaluator's non-verbal assessment factor as a mathematical model. We implement support interfaces that realize the use in actual situations and explore the answer to questions through experiments.

Depending on the results of performance evaluation and user studies, the proposed features in the local area were suggested to be related to the expert's assessment, and the output from the evaluation model showed values corresponding to the expert's decision. Furthermore, our proposed interface could perform appropriate assessment in uncontrolled environments, thereby demonstrating the feasibility of our proposal.

The findings herein prove that human facial assessment can be described by the integration of mathematical values that is culculated from local facial features. This knowledge provides a supportive technology for the expert's facial assessment fields, which have not been able to apply artificial intelligence technology in the past.

Acknowledgements

My greatest gratitude is for my supervisor Prof. Kenji Suzuki for his guidance and encouragement. Since I joined this laboratory, Prof. Suzuki has supported my research activities in various ways. Any achievements in this work had not been there without his existence.

My gratitude is also for Prof. Hideaki Kuzuoka, Prof. Itaru Kitahara, Prof. Yasushi Hada, and Prof. Ruriko Takano for their valuable supervision and inspiring comments on this work.

This stydy was supported in part by Shiseido Company, Limited, Special Needs Education School for the Visually Impaired, University of Tsukuba and Osaka Police Hospital. I would particularly like to thank Miho Ijima and Naoki Matsushiro for providing the opotunity of field works.

I would like to thank the administrative assistant of the artificial intelligence laboratory for their support, which helped me to carry out my work smoothly.

I would also likes to thanks Dr. Masakazu Hirokawa, Dr. Soichito Matsuda, Dr. Naomi Kuramoto, Dr. Taku Hachisu and Dr. Moder Hassan. Their opinions and the techniques learned from them are the important factor of this work.

Also, I would like to thank all the members in the Artificial Intelligence Laboratory for their great friendship and support. Special thanks to Jun Nishida, Yuta Kozaki and Kai Sasaki. The reason why I was able to embark on this path is because of them.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Research Backgrounds

The human face has a myriad of shapes and changes depending on individual differences and coordinated movements of facial muscles. The appearance is not only a part of interpersonal communication but also the confidence and quality of life [1, 2]. On the contrary, related research has reported that acquired facial disfigurement has a significant impact on the patient [3, 4]. Therefore, medical treatment and rehabilitation are being conducted in actual scenes to improve their appearance. We focus on two problems in the field of medical and nursing care, especially in the local area of facial appearance, and parallelly conduct research on them 1.1.

1.1.1 Confirmation of Makeup Finish by a Visually Impaired Person

Visually impaired persons consider applying makeup to be social etiquette and have the motivation to acquire the skills. They use common tools to apply cosmetics, checking for surface unevenness and the texture of the cosmetics with their fingertips. It is said that skilled people can apply makeup by themselves as well as healthy people do. However, this method requires delicate work without relying on visual information. They sometimes spoiled the facial impression by applying makeup to the wrong position (unintentional makeup). Therefore, many visually impaired persons tend to avoid makeup in their daily lives due to concerns and fears of unintentional makeup.

These problems are revealed in interviews that we conducted with our collaborators with visual impairments (including blindness, low vision, and visual field

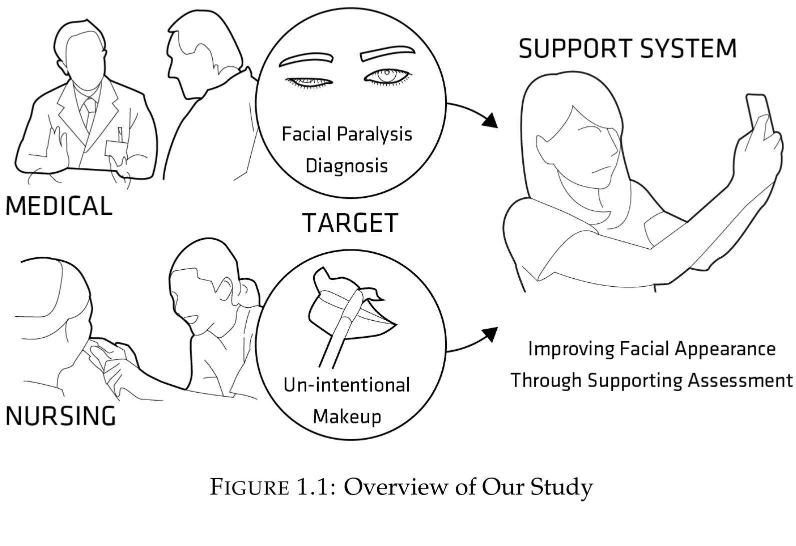

FIGURE 1.1: Overview of Our Study

narrowing), as well as with an expert on makeup for people with visual impairments (an instructor at a special school for the visually impaired) and its related study [5].

The above-mentioned work shows that there is a demand for support in building their confidence through confirmation of the makeup finish. In addition, the survey showed that people with visual impairments feel anxious about their makeup falling off when they go out in bad weather, and there is a need to provide an environment where they can fix their makeup on the spot.

Currently, makeup confirmation support has been conducted with a sighted assistant face-to-face. However, they have many situations where it is difficult to obtain cooperation, such as when they live alone or are at work. Therefore, there is a demand for support with regard to confirming the makeup finish by themselves.

1.1.2 Diagnosis of Facial Nerve Paralysis

Facial nerve paralysis is a disease wherein the facial nerve is damaged, resulting in the inability to move the facial muscles that control facial movements. The condition generally affects one side of the patient's face and significantly impairs facial expression. This concurrent eyelid and lip dyskinesia can lead to eye diseases, poor feeding, and adverse effects on face-to-face communication with others. Normally, these symptoms gradually resolve after a certain period of time, but poor prognosis can lead to sequelae that are difficult to cure.

To alleviate such symptoms and prevent sequelae, appropriate treatment and rehabilitation are required in medical institutions. In current medical practice, diagnoses based on visual measurements such as the Yanagihara's 40-point grading system (Yanagihara's method) and House and Brackmann's facial nerve grading system (H-B Method) are used to ascertain the level of paralysis in patients with facial paralysis and to provide them with treatment [6–8].

However, previous research has indicated that differences in diagnostic results tend to occur in the methodology between doctors without specializations [9]. The reason is that the evaluation is based on the doctor's observations of the facial small movement, which changes the complex mechanism over time. Therefore, objectivity with regard to the physician's diagnosis is needed.

1.2 Objective

We aim to solve the assessment problems in the fields of makeup of visually impaired persons and facial paralysis diagnosis through the support systems.

1.2.1 Makeup Confirmation Support for Visually Impaired Persons

We propose a methodology that allows visually impaired persons to confirm the makeup finish by themselves. We especially focus on lipstick application owing to the difficulty and effects on the overall impression of the face. Furthermore, we notice a mobile support interface owing to the increasing use of smart devices by visually impaired persons and implement it as an interactive application (Fig. 1.2).

We define "ideal makeup" as makeup applied by a third person (e.g., a makeup instructor) whom the visually impaired person trusts, and "unintentional makeup" as makeup with excessive lipstick protrusion on the lip contour. Through this study, we describe a methodology to improve unintentional makeup and bring it closer to the ideal makeup through warnings about the protrusion. As the relative survey revealed the low frequency of unintentional makeup, we consider that the "concern even if it is applied properly" is the majority of cases. Therefore, we place equal emphasis on providing support that communicates the absence of protrusions and relieves anxiety.

FIGURE 1.2: Makeup Support Interface for Visually Impaired Persons

FIGURE 1.3: A Wearable Camera Interface for Diagnosis Support of Facial Nerve Paralysis

Diagnosis Support of Facial Nerve Paralysis

We propose a methodology to facilitate the assessment of the paralysis level based on Yanagihara's method. We especially focus on the facial posture of "showing teeth" in the Yanagihara method, which has already been shown to be related to the other diagnosis [10, 11]. Furthermore, we consider the format for capturing appropriate diagnosis data such that it does not interfere with the doctor's operation (Fig. 1.3).

We define the results of the skilled doctor's diagnosis as desirable and aim to bring the non-skilled doctor's assessment closer to it. Moreover, by presenting the level of paralysis, indirect double verification of the diagnosis results may be realized. As an initial study for this purpose, we examined quantitative assessment indices contributing to the judgment of diagnosis.

In contrast, early and appropriate medical treatment has been reported to be effective in preventing sequelae of facial paralysis [12, 13]. Therefore, as an extension of support in the medical field, we consider the possibility of a simple diagnosis in the patient's living space.

1.3 Research Question

Here, we focus on facial assessment by experts in the fields of makeup and facial paralysis diagnosis for the visually impaired. In the realm of face recognition, expert judgments have been reported to consistently outperform those of the general persons, suggesting that with training and experience, reliable and accurate identification is possible [14]. Furthermore, expertise depends not on the ability to analyze local detailed features sequentially, but on the ability to analyze the features at ones. The report suggests a close relation between facial assessment and local detailed features [15, 16]. Based on this idea, we consider the following three questions to achieve our purposes.

1.3.1 Quantification of Local Features

Lip shapes and movements were assessed in the makeup confirmation and facial nerve paralysis diagnosis, respectively. Subsequently, what local features affect the improvement of facial appearance represented by subregional face shapes and motions? Related studies state some methodologies for quantifying lip shape and movements through image processing [17, 18]. However, the quantified values are difficult to use as explanatory variables for the severity of lipstick extrusion and paralysis. Therefore, we explore methodologies for calculating numerical feature values from the quantification of shape and motion in subregions of the face as variables that linearly explain the geometric shape change of facial images.

1.3.2 Formalization of Expert's Assessment

Human face evaluation is susceptible to individual subjectivity, but the problems herein assume quantitative evaluation [19, 20]. In facial nerve paralysis diagnosis and makeup confirmation, the doctor and the assistant assess the stage of the paralysis level and the presenting or absence of protrusion, respectively. Here, the former is an ordinal scale evaluation and the latter is a nominal scale assessment, which can be said to be a non-verbal factor for the assessment. This means that the face assessment model contains nonlinearities.

If such an evaluation can be translated into a quantitative model, a support system can be implemented. On the contrary, the report suggests a close relation between facial assessment and local detailed features [15]. Then, what are the relations between the expert's facial assessments and subregional facial attributes? We explore a mathematical model that describes an expert's assessment with the integration of local feature values.

1.3.3 Interaction for Facial Assessment Support

For facial image processing, appropriate facial images not including shading and keeping the frontal face are indispensable. In addition, a support interface is required to perform explanations which is easily understood. Then, what are the functions of interfaces for use in uncontrolled environments? Related researches use a desktop device and control facial posture using a chin mounter [5, 18]. However, the field of facial assessments are in daily life space or in examination room. The user may also have a disability or perform other medical examination duties. We explores a novel human-machine interaction for users that assumes disability and medical operation.

1.4 Outline

This paper is organized by different chapters shown below.

The next chapter introduces research in areas related to the findings provided by this study. Here, we clarify the problems of existing methods and findings, and explain the position of this research. In Chapter 3, we describe two implemented systems in the medical and nursing fields. We also state the experiments and results to verify the validity of the proposal. In Chapter 4, we consider whether the proposed method can provide the functions required in each field. We also provide an answer to the research question based on the findings. Chapter 5, we conclude this research and consider future prospects.

Chapter 2

Related Work

Here, we clarify our research's standing position from related research on makeup support for the visually impaired and diagnostic support for facial nerve palsy.

2.1 Facial Assessment Based on Mathematical Information

Since facial assessment includes subjectiveness, it is known that the results may be biased to some extent [19–21]. Therefore, attempts to realize objective evaluation by expressing non-verbal face assessment numerically have been continued.

Most reported face assessment flows on a computer are 1) quantification of facial features from facial images and 2) multivariate observation using the quantified values. In addition, methods that use 3) deep learning techniques to simultaneously quantify and observe have been actively proposed. We state the related works of these methodologies in this section.

2.2 Quantification of Local Features

Accurately measuring the shape of an object, not only a human face, is an important part of the object recognition task. In addition, quantifying the measured geometric descriptors as features has a significant impact on the difficulty of evaluating subsequent observations.

Because of the wide variety of facial shapes, not only among races and genders, but also among individuals, many studies are searching for a quantitative representation based on facial giometric mesurementvalues [22, 23]. Several studies are shown that holistic attributes such as facial averageness and symmetry are related to the evaluation of facial attractiveness [24, 25]. Other works report a automatic facial assessment method based on holistic attributes: the distance between the eyes, thickness of the lips, and symmetry of facial shape [26]. In contrast, local factors have also been found to be important in expert evaluations [15].

In this study, we assume that measurements of the thickness and movement of a part of the lips are local features, and that combinations of these features are holistic features. We do not discuss the strict categorization of local and holistic features according to size, but design features as descriptors of the problem.

Describing the lip shape quantification is important for focusing problems. In the shape measurement approach, related research is mainly used in the still image, which includes the frontal face. Here, there are a wide range of methods, including those based on positional relationships in the face [27], those based on color transformation and contour tracking [17], and those based on multiple features such as histograms of oriented gradients (HOG) features [28]. However, the reports design features to detect the degree of opening or closing of the mouth. An implementation that focus on lip contour deformation is proposed, however, the system are based on hue extraction, which is vulnerable to individual differences and shading [29]. Therefore, there is still room for debate on how to detect lip contour with high accuracy to detect the protrusion of makeup.

For the measurement of facial movements, previous researches propose methods that use facial feature points. First, a facial feature point detector is applied to each video frame and the positional coordinates on the image are measured. This method utilizes a regression model or trained neural network that tracks a given number of points (e.g., 48 or 68 points) [30–32]. However, since the method can only track contours of the local facial area, the method can not perfectly mesure the microtexture of facial surface. Second, a method using optical flow is proposed, in which motion on the face surface is used as a texture transition [18]. This method can measure transitions on the surface of the face, but it has problems toward individual differences and susceptibility to shading. Third, methods uses the local binary pattern are proposed, which is a feature describing the texture pattern [33–36]. Although these methods are robust against illumination changes, the feature is also difficult to use as a criteria because of the affectiveness toward the individual differences in human skin. Therefore, we focus on a method of sequential tracking of texture transitions in small regions in order to achieve both tolerance to individual differences and tracking of the face surface [37].

To provide the assessment, describing the measured values as features is also important. In the makeup assessment research, the primitive feature using the area of the aperture and the center point of the lips are proposed [29, 38]. Subsequently, the related srudies mention the symmetry and the before-after deformation which derived from the primitive feature [5, 29]. However, features expressed using reference points such as the center of gravity of a lip region tends to be affected by the shape change of lipstick protrusion. Therefore, features that assume contour changes in lip shape are necessary.

In mouth movement description approach, some works use the action units [39, 40]. This method is effective for observing the occurrence of facial expressions such as the presence or absence of paralysis, but it is unsuitable for measuring the intensity of expression creation. Other proposals use the symmetricity between the helthy side and the affected side [18, 35, 37, 41]. In addition to this, a proposal to include simple motion strength as a feature is made [36]. We took these suggestions, interviewed medical experts, and checked them against the essential criteria for the diagnosis of facial nerve palsy.

2.3 Formalization of Expert's Facial Assessment

We consider the representation of expert's assessment, which is an important part of solving the problems. Here, we focus in particular on attempts to implement the evaluation model on a computer.

Some of these studies in the machine learning field are regarding facial attractiveness assessment. In these studies, the face attributes is numerically represented by feature values: eye height, width, and thickness of lip. The methods realize human like evaluation by using classifiers such as the k-nearest neighbors method [42] and the support vector machine [43–46]. In contrast to these attempts, the problem in this study is not attractiveness, but trying to formalize the relations between undesirable facial impression and facial features.

On the other hand, it should be noted that our attempt is aimed at the expert's assessment: assessment that requires expertise. The related studies attempts to represent this expertise in a model using graph rules [47, 48]. Here, the rules are determined by specifying empirical values and learning from decision trees, but it is necessary to choose the appropriate method for each problem. The study which check makeup protrusions also use the rule based assessment model, and a threshold-based evaluation is proposed [5]. However, the study does not propose a learning method from the expert's assessment.

On the contrary, several assessment models have already been proposed for estimating the facial nerve paralysis level [18, 35–37, 41]. Here, single or multiple features, static or dynamic, are judged by a classifier such as a support vector machine.

2.4 Deep Learning Based Facial Assessment Method

Deep learning methods have been actively discussed in recent years. The methods perform feature extraction and classification in one-shot learning, which is particularly effective for problems where the domain knowledge is unknown.

Related studies on makeup recommendation tasks and facial paralysis diagnosis tasks have already been reported [8, 49–52]. These methods have advantages such as high recognition accuracy while not requiring feature design. On the contrary, they also have the disadvantage of poor interpretability due to the black box model. This point is especially a problem in the medical field where evidence is important. For this reason, we do not actively use it in the deep-learning-based implementation of this paper but use it only for detecting facial feature points.

2.5 Support System for Facial Assessment

In the makeup support and diagnosis support scene, capturing appropriate facial images that do not include shading and keep the frontal face into the picture are difficult. This is because the lighting position is rarely the same for each shot, and the positional relation between the camera and the face can also vary.

Therefore, the environment control through man–machine interaction. Furthermore, the usability of the interface needs to be used without assistance from others. Here, we notice the functions that enable both control of the environment and usability.

2.5.1 Support System for Visually Impaired Persons

Attempts to support visually impaired people's daily lives and facilitate their work by supplementing their vision have been made mainly through the support interface. Some proposals check faces and provide feedback through voice guidance. The technologies are split into two main types of recognition methods for image information. One of them is the method of recognition by human hands, represented by commercial applications such as aira (Aira Tech) and Be My Eyes (Be My Eyes) [53, 54]. In this method, the visually impaired person capture the image and send to a sighted assistant via the Internet. The remote assistant explains the contents of the image. Although this method can be expected to provide highly accurate and reliable feedback, applying the method to have an unspecified third party confirm the evaluation of a makeup applied face is not desirable.

On the other hand, automatic recognition methods that use machine learning and other techniques solve such problems. Commercial applications that perform object recognition have already been in the service, such as Tap Tap See (Cloudsight) and Viz Wiz [55–57]. In the research phase, an attempt is being made to use an eyeglass-type camera device to recognize other persons or objects and present them with voice guidance [58, 59]. However, these mobile applications do not take into account the usage toward the user's face, and have not been implemented to recognize cosmetic faces.

In the past, a number of interfaces designed to assist the visually impaired person's makeup which use the desktop computers and web cameras are proposed [5, 29, 60]. However, since these methods assume a controlled lighting environment and a desktop device, they are not suitable for use on the outside.

From above, we need an automatic assessment interface that can be used easily by the user without being affected by the convenience of others.

2.5.2 Diagnosis Support System on the Facial Nerve Paralysis

works use a movie wherein the patient performs a prescribed facial posture. Mainly, related works implement a human-machine interface that provides feedback on a set of data used to assess the disease. The reported system presents visual feedback of one or both of the predicted results of the diagnosis and the measurements of facial movements. In the method that feeds back the quantified values of facial movements, the shapes of the facial landmarks are measured using an image processing method, and primitive values used for judgments such as facial symmetry are presented [26, 61–63]. In the method of presenting predictive results for diagnosis, the above-mentioned measurements are used as features to estimate the level of paralysis using statistical or machine learning methods and subsequently presented [26, 37, 64]. The proposals are expected to improve the accuracy and simplify the difficulty of diagnosis by quantitatively presenting instantaneous difficult-to-grasp facial movements.

Previous facial movement tracking methods mainly use multiple markers on the face [65, 66]. However, in actual medical practice, doctors have no time to put tracking markers on patients' faces.

For methods without markers, many studies have attempted to apply trained detectors to each frame. However, since human facial expressions change rapidly, high-resolution and high-framerate video images or multi-view images are necessary [67, 68]. As the features used for diagnosis include very small facial movements, Therefore, the authors use large scale measurement equipment that is difficult to handle.

Although the method to implement the support system as a application is proposed, the photograph framework to take a patient's face is not considered [69].

Other implementations using mobile interfaces such as smartphones exist [70]. However, the method has limitations such as the inability to present the paralysis level as a function.

From the above, we need an assessment support interface that takes a facial movie in a format not interfering with the face-to-face diagnosis in the consultation room and perform advanced analysis such as an estimation of the level of paralysis.

Chapter 3

Methodology

Throughout this paper, we implement a novel support device in parallel on makeup support for visually impaired persons and facial paralysis diagnosis support. The system provides a solution to the research question, and we gain insight into the essentials to support an assessment for improving human facial appearance.

3.1 Overview

3.1.1 Quantification of Local Features

We propose the method to quantify the local features which are related in the irregular. In particular, we focus on defining numerical features that linearly correspond to geometric changes in the face shape.

In the study of makeup support, we make attempt to mesure the shape of the lips by suppressing the effects of ambient light and head posture by using image processing methods. We also propose the method for describing the lip thickness based on the lip slit line, which is specialized on the makeup protrusion detection.

In the study of diagnosis support, we propose the facial tracking methods based on the faceial feature point for mesureing the movement of facial surface. Subsequently, we quantify the mouth aperture and symmetry by by using the area ratio of small region pairs, which are less affected by traction on the healthy side, assuming left-right asymmetry.

3.1.2 Formalization of Expert ’ s Assessment

We investigate how to describe the nonverbal facial assessment of experts as a model. In particular, we aim to represent nonlinear facial assessment by numerical features and to clarify the relations between them.

In facial assessment requiring expertise, clarifying how the local features are related to the overall assessment is important.

While studying makeup support for visually impaired persons, we formalize the makeup protrusion assessment of a makeup teacher who is familiar with the makeup of the visually impaired persons. Here, we propose a simple learning method using a dataset including intentional makeup protrusion.

While studying support for facial nerve paralysis diagnosis, we propose the bivariate assessment model which is based on the results of interviews with medical specialists.

3.1.3 Interaction for Facial Assessment Support

We implement interfaces that can provide functions without special assistance, assuming that they are used in the actual field. In particular, we focus on the interaction to take the appropriate images and perform the comprehensible feedback.

While studying makeup support for visually impaired persons, we propose the selfie support and the protrusion place presentation. While studying support for facial nerve paralysis diagnosis, we implement the head mount camera interface to capture the facial video images without interfering with face-to-face diagnosis.

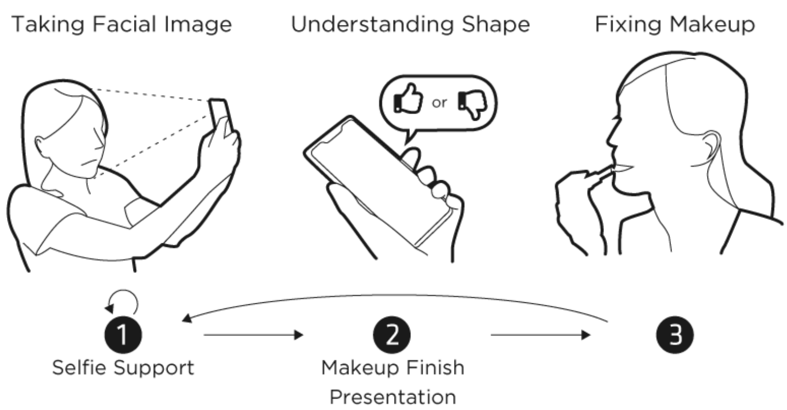

FIGURE 3.1: Outline of Interaction to Support Makeup Comfirmation

3.2 Makeup Support Interaction Design

Our proposed interface presents the state of the makeup finish through a three-step interaction (Fig. 3.1).

3.2.1 Taking Facial Image

To accurately check the finish of the makeup, an applied makeup image wherein the user's face is in the frontal position is indispensable. However, it is extremely difficult for visually impaired persons to take selfies using the camera. This is because maintaining the relative positions and angles of the camera and face without any visual information is necessary.

Therefore, we propose a support method to ensure appropriate head posture while capturing images. The interface uses haptic and audio feedback to present the head rotation and position, respectively. After proper head posture is maintained for a period of time, the system automatically captures a facial image and moves on to the next step.

3.2.2 Understanding Shape

The user is informed of the presence or absence of "unintended makeup" and its location through voice guidance. We focus on the difference between the ideal and current applied makeup shapes. and use this comparison as a criterion for detecting excessive protrusion. Through a simple learning method, the criteria for detecting partial protrusions are automatically decided. Our proposed learning method examines how an expert relates the degree and location to the protrusion. We propose a novel interaction wherein the presence or absence of protrusion is presented in each area around the lip contour by dividing it into five areas.

3.2.3 Fixing Makeup

The user uses a makeup remover to arrange the makeup shape. After fixing the makeup, the user takes an additional facial image and verifies the shape of the makeup again. Through this recursive process, the user's makeup is expected to gradually improve and approach the ideal shape.

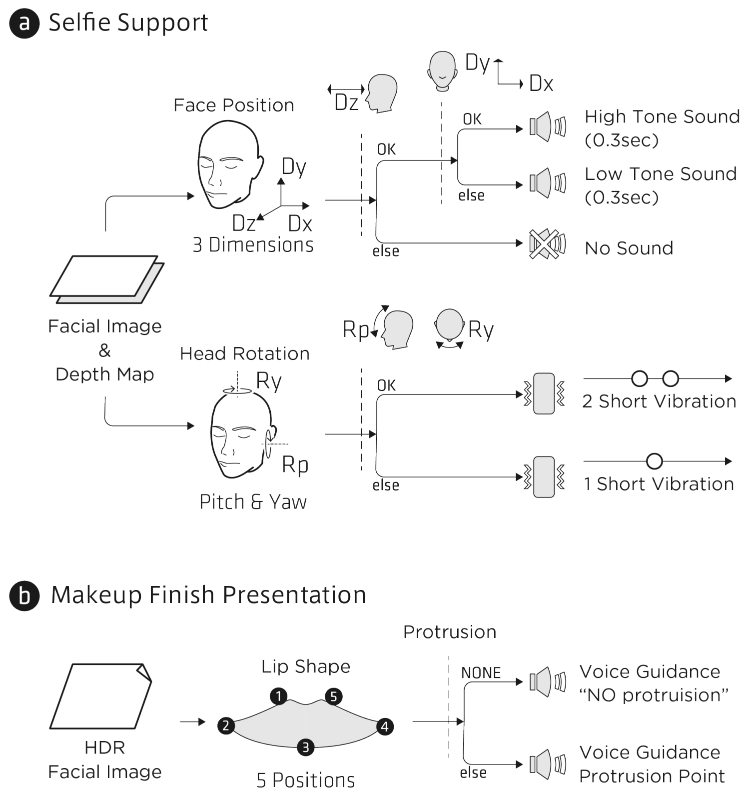

FIGURE 3.2: Outline of feedback. The System Make Two Types of Feedback to Realize the Proposed Interaction.

3.3 Implementation of Makeup Support System for Visually Impaired Persons

3.3.1 Selfie Support

The appropriate facial image is indispensable for accurately measuring the makeup shape. To capture an image whose face is arranged in the frontal position, the interface presents a difference between the ideal and current head positions (Fig. 3.2).

An RGB image (8-bit RGB, height: 4032 pixels, width: 3024 pixels) and a depth map (8-bit grayscale, height: 320 pixels, width: 240 pixels), indicating that the distance to the object is captured by the camera. Subsequently, the position of the face on the image plane and the head rotation are obtained using the face detection and rotation estimation methods. The system sequentially updates posture information: head rotation and face position, and provide feedback by vibration and sound, respectively. When the user maintains the appropriate posture for 2 s, the interface automatically takes a facial image and terminates the assistance. Throughout this implementation, the face detection method emphasizes the speed of computation and uses a hardware-embedded implementation [71]. We also used a deep learning-based method for head posture estimation, and the trained models used the 300W-LP dataset [72, 73].

Face Position

The distance between the ideal and current face positions is calculated with the following formula:

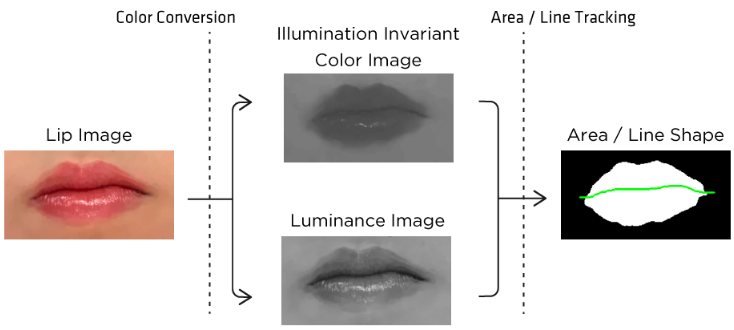

FIGURE 3.3: Outline of Lip Shape Extraction Method

$$D_x = |C_x - H_x| \quad (3.1)$$
$$D_y = |C_y - H_y| \quad (3.2)$$
$$D_z = |\frac{\alpha}{255}(\mathbb{D} - \beta)| \quad (3.3)$$

Here, D_x and D_y are the distances to the appropriate position on the X and Y coordinates: the center in the image plane. D_z is the distance to the appropriate position in the Z-axis direction perpendicular to the image plane. In addition, C_x and C_y are the X and Y coordinates of the center of the rectangular region surrounding the face position on the image plane obtained from the face detection method. Similarly, let H_x and H_y be the X and Y coordinates of the center of the image plane, respectively. In addition, $\mathbb{D}$ represents the pixel value of the center of the face on the depth map. The D_z is set to be 0.0 at a distance of 40 cm from the object and is changed by 1.0 every 20 cm. In this study, we set $T_x = 0.1W$, $T_y = 0.25H$, and $T_z = 0.5$ as empirical values. H_x and H_y denote the width and height of the RGB image, respectively.

If either D_x or D_y is out of the proper position, a low tone is presented. If D_z is out of the appropriate position, no audio feedback is given.

Head Rotation

The distance between the ideal and current head rotations is calculated with the following formula:

$$R_h = \sqrt{R_p^2 + R_y^2} \quad (3.4)$$

R_h is the difference up to the appropriate head rotation: both the pitch and yaw angles are 0 degrees. R_p and R_y are the pitch and yaw angles of the head, respectively, based on the degree method. The values were obtained by the head rotation estimation method.

When R_h falls below an empirically defined threshold T_r, the interface gives two short vibration feedbacks; otherwise, it gives one short vibration. In this study, we set $T_r = 7$.

3.3.2 Makeup Finish Presentation

The system presents the existence of protrusions in the applied makeup area by voice guidance (Fig. 3.2.b). The facial image (8-bit RGB, heigh : 4032 pixel, width : 3024 pixels) taken at the end of selfie support is performed for the detection of the protrusion. The interface takes a facial image as a high-dynamic-range image to reduce the ambient light effect. The imaging method uses a hardware-embedded implementation [74].

Subsequently, the protrusion detection is performed to the following procedure.

1. Correcting the rotation of facial image according to the head posture.
2. Cropping the partial image around the lip region.
3. Creating the lip area-emphasized image and the lip slit-emphasized image using the color conversion method (Fig. 3.3).
4. Extracting and tracking the lip area and lip slit line by the area extraction method, respectively (Fig. 3.3).
5. Detecting the protrusion position through comparison with a sample of the ideal shape (Fig. 3.5)

Rotation Correction and Image Cropping

To reduce the effect of rotation on the face roll angle, a rotation correction process is performed on a facial image. The correction is based on the face feature points inferred by the detector based on deep learning. We use the trained model based on the LS3D-W dataset [75].

The angle of the rotation movement is given by the following formula.

$$\theta = \tan^{-1}\left(\frac{Q_y - \mathbb{U}_y}{Q_x - \mathbb{U}_x}\right) + \frac{\pi}{2} \tag{3.5}$$

$$Q_x = \frac{\mathbb{R}_x + \mathbb{L}_x}{2} \tag{3.6}$$

$$Q_y = \frac{\mathbb{R}_y + \mathbb{L}_y}{2} \tag{3.7}$$

θ is the rotation angle. $\mathbb{R}_x$ and $\mathbb{R}_y$ are the X and Y coordinates of the right eye center in the image plane, respectively. Likewise, $\mathbb{L}_x$ and $\mathbb{L}_y$ and $\mathbb{U}_x$ and $mathbbU_y$ are the coordinates of the left eye and the lip, respectively. Finally, the corrected facial image is obtained by performing a rotation with the center coordinates of the lips as the rotation center in θ.

Next, the rectangular area cropping around the lips is calculated as follows.

$$U_h = \chi \cdot |\mathbb{M} - \mathbb{N}| \tag{3.8}$$

$$U_w = \chi \cdot |\mathbb{R} - \mathbb{L}| \tag{3.9}$$

Here, U_h and U_w denote the height and width of the rectangular region around the lip, respectively. The $\mathbb{N}$, $\mathbb{M}$, $\mathbb{R}$, and $\mathbb{L}$ are 2D coordinates at the base of the nose, center of the lips, center of the right eye, and center of the left eye, respectively. χ is a coefficient for the scale adjustment. We set to $\chi = 1.5$, where the lip contour is within the image.

Color Conversion and Lip Shape Extraction

A color conversion method is used to enhance the lip area in the peripheral image, and the area is then extracted by segmentation (Fig. 3.3). First, we use a specific color transformation to obtain lip area-enhanced images with robust color invariants to disturbed light. Next, we use the Chan–Vese level-set method for performing segmentation [76]. We define a closed curve that minimizes the energy function as shown below.

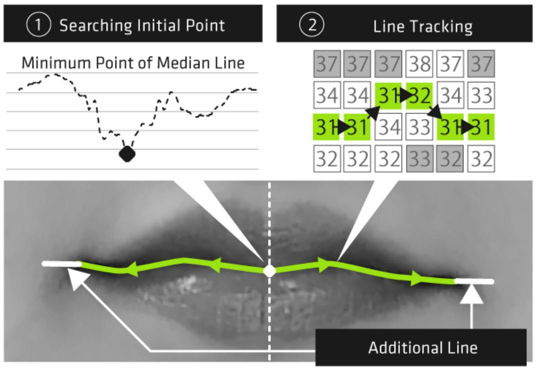

FIGURE 3.4: Outline of Lip Slit Line Tracking

$$
\begin{aligned}
E(c_i, c_o, O) =& \kappa \cdot l_O \\
&+ \int_{O_i} |V_{(x,y)} - c_i|^2 dxdy \\
&+ \int_{O_o} |V_{(x,y)} - c_o|^2 dxdy
\end{aligned} \tag{3.10}
$$

Here, E is the energy function and O is the closed curve that performs segmentation. c_i and c_o are the average pixel values of the inner and outer regions of the closed curve, respectively. Similarly, O_i and O_o are groups of pixels in the inner and outer regions of the closed curve, respectively. l_O is the length of the closed curve, and $V(x, y)$ is the pixel value at coordinates (x, y) in the image. The κ is an adjustment term and we herein set $\kappa = 0.1$. For the initial segmentation, a rhombus shape with a width two-fifths of the length and width of the lip area enhancement image is used. The number of updates of the closed curve is empirically set to 500.

Lip Slit Line Tracking

A color transformation method is used to enhance the lip slit: areas between the upper and lower lips. Subsequently, the line is calculated by the simple tracking method (Fig. 3.3). First, we perform a transformation of the L-color components in the laboratory chromatic system [77]. Next, we perform a luminance normalization transformation to make the color distribution uniform across the entire image [38]. Furthermore, to reduce the effect of specular reflection of the lips, a 3 x 3 size average filter is applied to the image.

Finally, we apply the following procedure to the processed image (Fig. 3.4).

1. Setting the initial point at the coordinate of the pixel with the lowest value on a line bisecting the X-axis vertically in the image.
2. Selecting the next tracking point, which is the minimum luminance point within a left and right window (width: 1 pixel, height: 3 pixels).
3. Repeating the process of 2 for the left and right mouth corners until the pixel value of the tracking point exceeds the threshold T_l. Empirically, we set the average pixel value of the image to 0.7.
4. Adding a line of length $\mathbb{L}$ parallel to the X-axis to the endpoints of the tracking points. We set $\mathbb{L}$ to a value 0.1 times the width of the image.

Lip Shape Quantification

We obtain the quantified lip shape features from the obtained lip area and the lip slit line. The procedures are robust to the user's posture. First, we quantify the lip

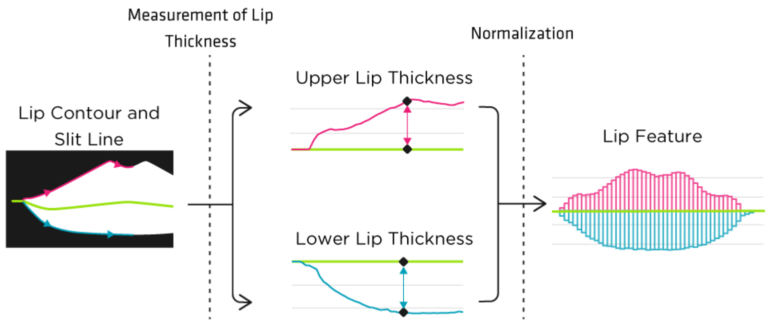

FIGURE 3.5: Quantification of the Lip Thickness

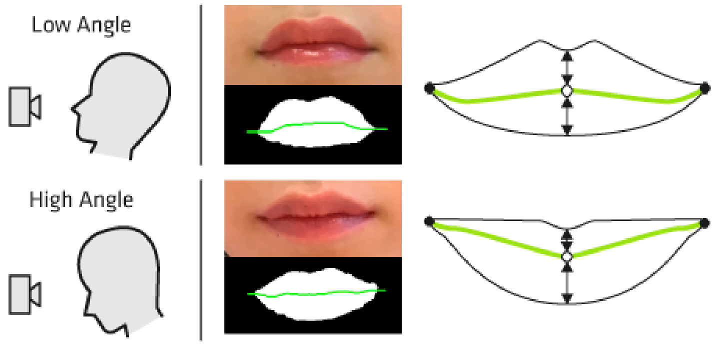

FIGURE 3.6: Relations Between Lip Contour Shape and Facial Pitch Angle

thickness, and the distance between the lip contour and the slit line are separately calculated in the upper and lower lip regions. For each pixel point on the lip slit line, we compute the set of distances to the endpoints of the upper and lower lips in the Y-axis direction on the image plane (if there is no endpoint in the lip area, the distance is set to 0). The features of the lip contour shape are calculated by normalizing the number of dimensions to the obtained distance set calculated as follows:

$$G_u(n) = \frac{1}{\mathbb{S}} \sum_{k=n\cdot\mathbb{S}}^{n\cdot(\mathbb{S}+1)} \frac{\phi_u(k)}{\mathbb{H}} \quad (n \in [0, \epsilon)) \tag{3.11}$$

$$G_b(n) = \frac{1}{\mathbb{S}} \sum_{k=n\cdot\mathbb{S}}^{n\cdot(\mathbb{S}+1)} \frac{\phi_b(k)}{\mathbb{H}} \quad (n \in [0, \epsilon)) \tag{3.12}$$

$$\mathbb{S} = \frac{K}{\epsilon} \tag{3.13}$$

Here, G_u and G_b denote quantified features of the shape of the lip contour of the upper and lower lips, expressed as a set of G_u and G_b, respectively. ϕ_u and ϕ_b are the set of distances (lip thickness) in the upper and lower lip areas, respectively. K is the number of dimensions of the distance set and K is the number of dimensions after normalization. We set K as 60 in this study. In addition, $\mathbb{H}$ is a term for scale normalization; herein, the height of the cropped images is used.

The obtained feature values of the lip contour shape are affected by a small pitch angular rotation (Fig. 3.6). The magnitude of the face's pitch angular rotation can be estimated based on the shape of the lip slit lines. Therefore, we compensate for the shape magnitude between the upper and lower lip shapes; the lower lip feature increases and the upper lip feature decreases if the lip slit line arcs downward. If the line arcs upward, the upper lip and lower lip features increase.

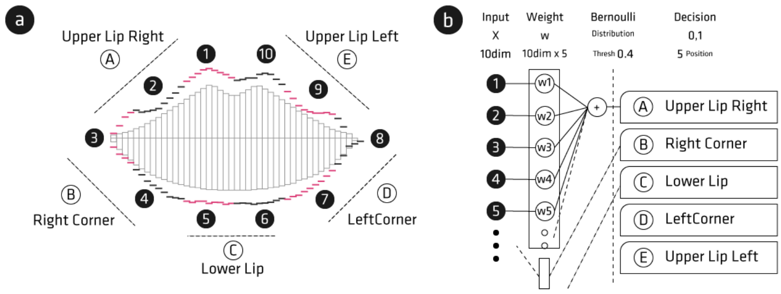

FIGURE 3.7: a : The Locations of Protrusion Warnings and the Division of Lip Features. b : The Decision Model of Protrusion

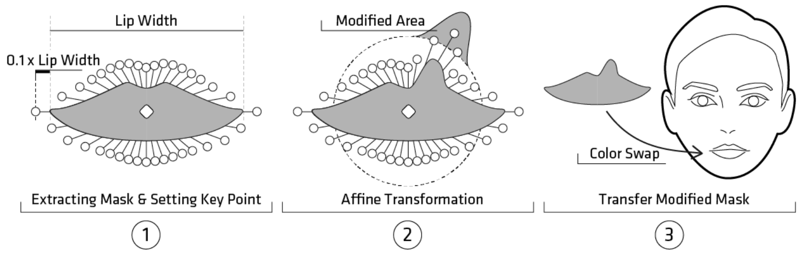

FIGURE 3.8: Outline of the Makeup Protrusion Simulator

$$F_u = G_u(1 - \frac{\mathbb{P}}{\mathbb{H}}) \tag{3.14}$$

$$F_b = G_b(1 + \frac{\mathbb{P}}{\mathbb{H}}) \tag{3.15}$$

$$\mathbb{P} = P_m - \frac{P_r + P_l}{2} \tag{3.16}$$

Here, F_u and F_b are the corrected feature values of the upper and lower lips, respectively. P_m, P_r, and P_l are the Y-axis coordinates of the center of the lip slit line, right angle of the mouth, and left angle of the mouth, respectively. The $\mathbb{H}$ is the height of the cropped image.

Protrusion Detection

The interface automatically detects the position of protrusion from the lip features. In this detection, setting the appropriate criteria is important: the makeup applied position and the amount. For the former, I decided to present the protrusion in the following five areas: 1) right upper lip, 2) right corner, 3) lower lip, 4) left corner, and 5) left upper lip through interviews with a special support school teacher (Fig. 3.7.a). For the latter, we propose a simple evaluation model to construct a system reflecting the opinions of human evaluators (Fig. 3.7.b).

Prior to the implementation of the evaluation model, we created a dataset including several makeup protrusion images. We use the expert's decision, which is evaluated from the visual stimulation, as labels of the protrusion existence. Evaluators viewed a simulated image that intentionally applied a protrusion on the display and labeled five areas around the lips (Fig. 3.9).

We subsequently train the evaluation model from the obtained dataset. We used a simple Gaussian linear regression for the training model, assuming that the extrusion decision in each domain is independent. We define an evaluation model from the lip feature inputs and correct answer labels obtained from the face images in the dataset as follows:

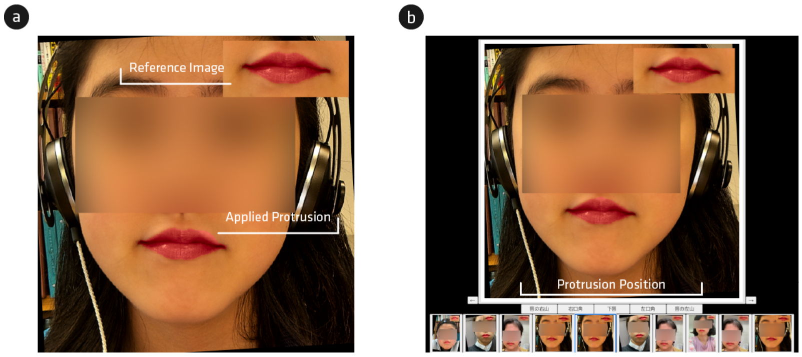

FIGURE 3.9: a : Makeup Protrusion Image. b : Feedback Application UI

$$y = \frac{1}{N}\sum_{n=1}^{N} w_n x_n \tag{3.17}$$

y is a five-dimensional number taking the value from 0 to 1, representing the existence of protrusion in the five lip regions. x is the difference between the features of the ideal makeup shape sample and the current lip features. A sample of ideal makeup shapes was applied by a teacher at a special needs school. Moreover, we used the averages over the 10 intervals around the lip (Fig. 3.7.a).

w is an NxM constant matrix, sampled from a half-normal distribution with covariance 0.1. Here, N denotes the number of dimensions of the input, 10, and M is the number of labels in the model: 5. The evaluation model is trained on the MCMC method (sample size: 2000; burn-in interval: 1000) as computing the predictive distribution is difficult.

The overhang is determined by thresholding the average of the model's output after 1000 samples. In this study, the threshold value is 0.4, depending on the accuracy and reproducibility of the model.

The dataset comprises images simulating makeup protrusion in accordance with the following procedure.

1. Extracting the lip area from the image as a mask of the lip shape. Setting key points on the radiation from the center of gravity of the mask; the number of key points was set to 72. The key points are defined as points at a distance 0.1 times the lip width from the lip contour.

2. Moving the key points out of the lip partially. According to the modified points, the mask is deformed by the affine transformation. The moving is calculated as follows:

$$K_y(k) = sin(2\pi k/\mathbb{K})(l_k + \xi(k)) + \mathbb{C}_y \tag{3.18}$$

$$K_x(k) = cos(2\pi k/\mathbb{K})(l_k + \xi(k)) + \mathbb{C}_x \tag{3.19}$$

$$\xi(k) = \gamma\, W_k \frac{N_k - min(N_k)}{max(N_k)} \quad (k \in [0, \mathbb{K})) \tag{3.20}$$

$K_y(k)$ and $K_x(k)$ are the values of the X-axis and Y-axis at the k-th keypoint, respectively. $\mathbb{C}_y$ and $\mathbb{C}_x$ are the values of the X-axis and Y-axis of the center of gravity point, respectively. And $\mathbb{K}$ is the number of key points. γ is a constant that determines the scale of the protrusion. W_k is the distance from the center of gravity to the kth point. N is a value using the normal distribution on the circumference (Von-Mises distribution) and is defined by the following equation:

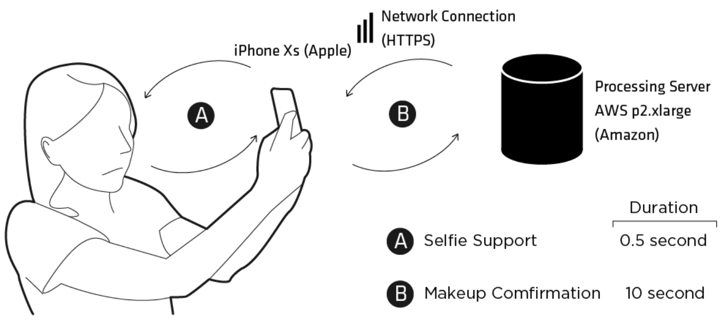

FIGURE 3.10: Outline of Hardware Setting

$$N_k = \frac{exp(\upsilon\ cos(\theta_k - \mu_k))}{2\pi I_0(\upsilon)} \tag{3.21}$$

Here, N_k is the value at the k-th key point and θk is the direction in which the keypoint is located in terms of the center of gravity. we set the value of $\theta k = 2\pi k/\mathbb{K}$. μ_k and υ are constants that determine the scale and direction of the protrusion, respectively. I_0 is a first-class variant Bessel function of order 0.

In the above variant, the shape of the protrusion is controlled by three variables: γ, μ_k and υ. We set three values of $\gamma = [0.4, 0.6, 0.8]$. Also, μ_k is set to indicate the direction of one of the 10 equal positions of the lip contour. υ is set to 0.5. We get various shape types (3 size, 10 positions) of protrusion with applying above process for each images.

3. Color transformation is performed on the obtained mask region in the face image. We select the transfer color (8-bit RGB color. R: 209, G: 67, B: 65), which can be clearly seen on the display. In addition, the target facial image and the target color are projected onto the LAB space, and only the mask region is transcribed with A and B color components.

Finally, I apply the above procedure to eight portraits taken from different models to obtain 240 portraits, including the various scale and direction of the protrusion. We also obtained a 480-image dataset with left–right flipping because the lip shape evaluation appears to be symmetrical.

Presentation of the Status of Makeup Finish

According to the results obtained, two types of Japanese voice guidance are provided by the interface. If no protrusion is detected in any area segment, the interface feedback is “There is no protrusion”. On the contrary, if an overhang is detected, the interface feedback is ‘ ‘ There is a protrusion in <position>”. The direction of presenting to the user is the side seen from the user’s position.

3.3.3 Hardware and Software Setting

We implement the interface as a smartphone application (iPhone Xs, Apple). In this paper, the presentation of vibrations and sounds, as well as the capture of face images and depth maps, uses built-in hardware functions. The interface light user’s face by the built-in flash. The voice presentation uses a framework (AVFoundation) built into the hardware. The speed of voice presentation is set at 1.7 times speed based on the hearing conducted in advance.

The response duration of each support interaction is 0.3 s for selfie support and 10 s for makeup confirmation. We use external servers (AWS, Amazon) to process the makeup confirmation owing to the high computational burden of quantifying the lip contour shape.

Throughout this implementation, we used part of the OpenCV implementation for the image manipulation process [78]. The core part of the program was written in Python 3 [79].

3.4 Experiments of Makeup Support Interface for Visually Impaired Persons

We conducted the basic performance of the proposed method via a quantitative evaluation considering two aspects: the quantification of the lip shape and the detection of makeup protrusion. Subsequently, we also clarify the feasibility and usability of use cases in actual makeup situations.

3.4.1 Accuracy of Lip Shape Quantification

The experiment describes the effects of the proposed features on the degree of intentional makeup protrusion and head rotation. We compared the quantified feature value between a normal makeup applied face and a protrusion applied face.

In this experiment, we recruited one participant (female) and captured facial images with intentionally applying makeup protrusions in two situations.

- Three levels of protrusion were applied—none, 2-mm, and 4-mm—to three areas: right upper lip. right corner and lower lip in frontal face position (Fig. 3.11).
- Applying one level of protrusion, 2 mm, to one area: right upper lip, in 5 head poses: frontal, pitch +8 to +10 deg and -8 to -10 deg, roll +8 to +10 deg, and -8 to -10 deg (Fig. 3.12).

In this experiment, we extracted lip shape features from the facial images and verified whether detecting protrusion in each area of the predetermined segments was possible.

In this experiment, we defined five segments: right upper lip, right corner, lower lip, left corner, and left upper lip, through interviews with a medical specialist (Fig. 3.13). The actual distinction between the regions was based on the dimension i of the features F_u and F_b (Formula 3.14–3.16) of the upper and lower lips as follows:

① **Right Upper Lip** $F_u: 15 \leq i < 30$

② **Right Corner** $i < 15$

③ **Lower Lip** $F_b: 15 \leq i < 45$

④ **Left Corner** $45 \leq i$

⑤ **Left Lower Lip** $F_u: 30 \leq i < 45$

We used a smartphone (iPhone Xs, Apple) to captured facial images from a distance of 40 cm. To evaluate only the performance of the image processing procedure, the interface did not perform a selfie support interaction. The position and angle of the participants' faces were controlled using a chin rest in case 1. The head rotation of the images was controlled with real-time monitoring in case 2 [75]. The protrusion was applied using a lip liner (MAQUILLAGE Smooth Stay Lip Liner N PK210, SHISEIDO) because it required a delicate application technique.

Through the results, we verify that the features quantified by our proposed method have usable accuracy for detecting lip makeup stick-out.

Result for the Various Protrusion Setting

Figure 3.14 denotes the quantified upper and lower lip shapes, with the horizontal axis representing the dimension of the lip contour features and the vertical axis representing the feature values. Here, the values at each location increase in a stepwise manner with the 2-mm and 4-mm protrusions.

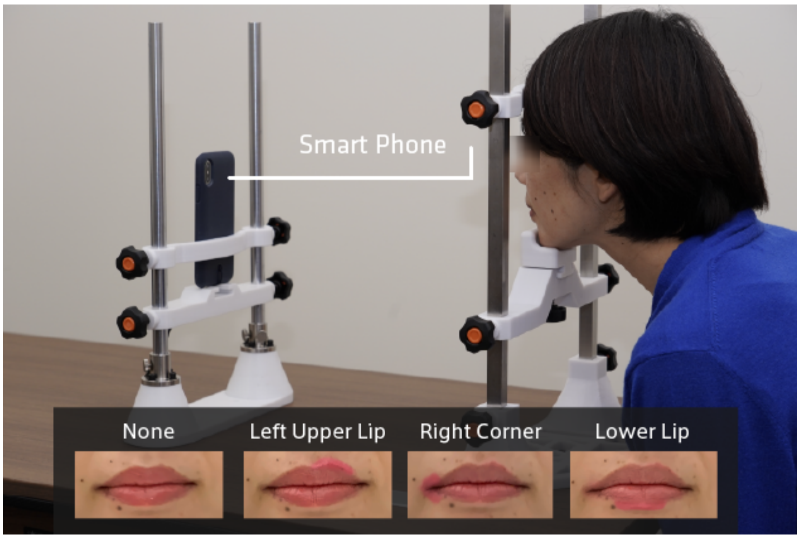

FIGURE 3.11: Overview and Taken Images in Various Protrusion Condition

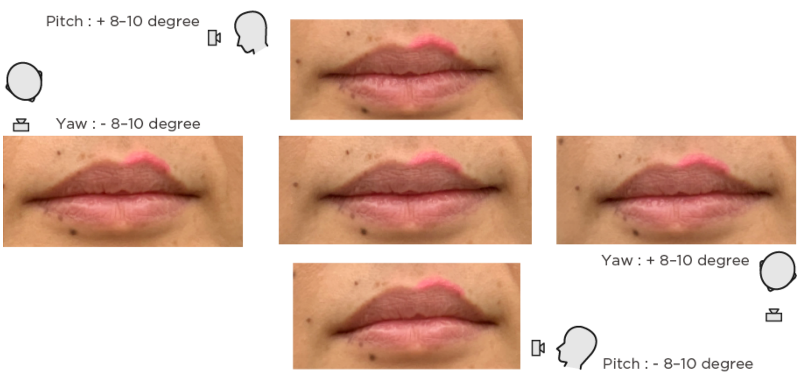

FIGURE 3.12: Taken Images in Various Head Rotation Condition

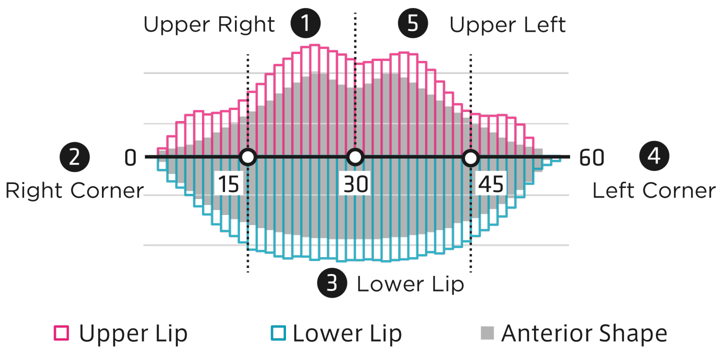

FIGURE 3.13: Regional Segmentation of the Lip Contour

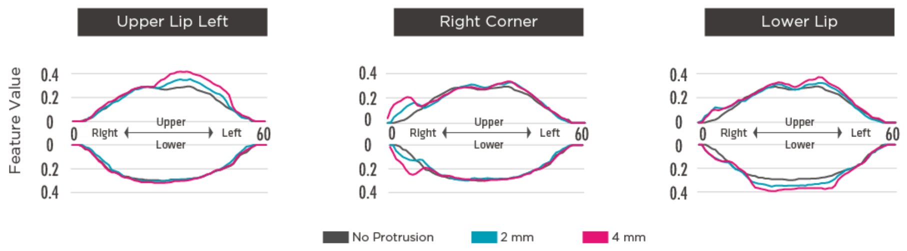

FIGURE 3.14: Protrusion Level and Quantified Feature Value

FIGURE 3.15: Comparison Between Lip Areas

Figure 3.15 shows the difference in feature values between with and without a 2-mm or 4-mm protrusion. The horizontal axis represents the regional segmentation of the lip contour, and the vertical axis represents the difference value. Here, a statistically significant difference ($p < 0.05$) was observed between the protruded area and other areas, au, using the Mann–Whitney U test. In contrast, the feature value, regarding the right corner, that is, in the condition of lower lip 2-mm protrusion, does not show a significant difference ($p = 0.20$).

Result for the Various Head Rotation Setting

Figure 3.16 denotes the quantified upper and lower lip shapes for the various head rotations. The horizontal and vertical axes represent the dimensions of the lip contour features and the feature values, respectively. In the right figure, where the yaw angle is varied, the features drift in the direction opposite to the face is tilted. On the contrary, in the left figure, where the pitch angle is varied, the feature value decreases in the protruded area when the image is captured looking up.

Figure 3.17 shows the difference in feature values between with and without a protrusion in each head pose. The horizontal axis represents the regional segmentation of the lip contour, and the vertical axis represents the difference value. Here, a statistically significant difference ($p < 0.05$) was observed between the protruded area and other areas, au, using the Mann–Whitney U test. On the contrary, the feature value for the right upper lip case, in which the yaw angle setting is 8 deg, does not show a significant difference ($p = 0.06$).

3.4.2 Feasibility Study in the Actual Makeup Scene

To verify the feasibility of the proposed method for assistive interaction in actual usage environments, we measured the facial posture of the user in the selfies and evaluated the estimation of makeup protrusion.

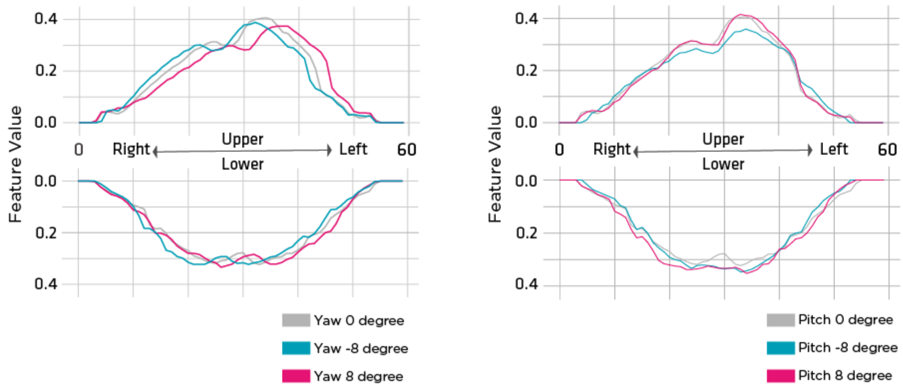

FIGURE 3.16: Overview of Quantified Feature Value in Various Head Rotation

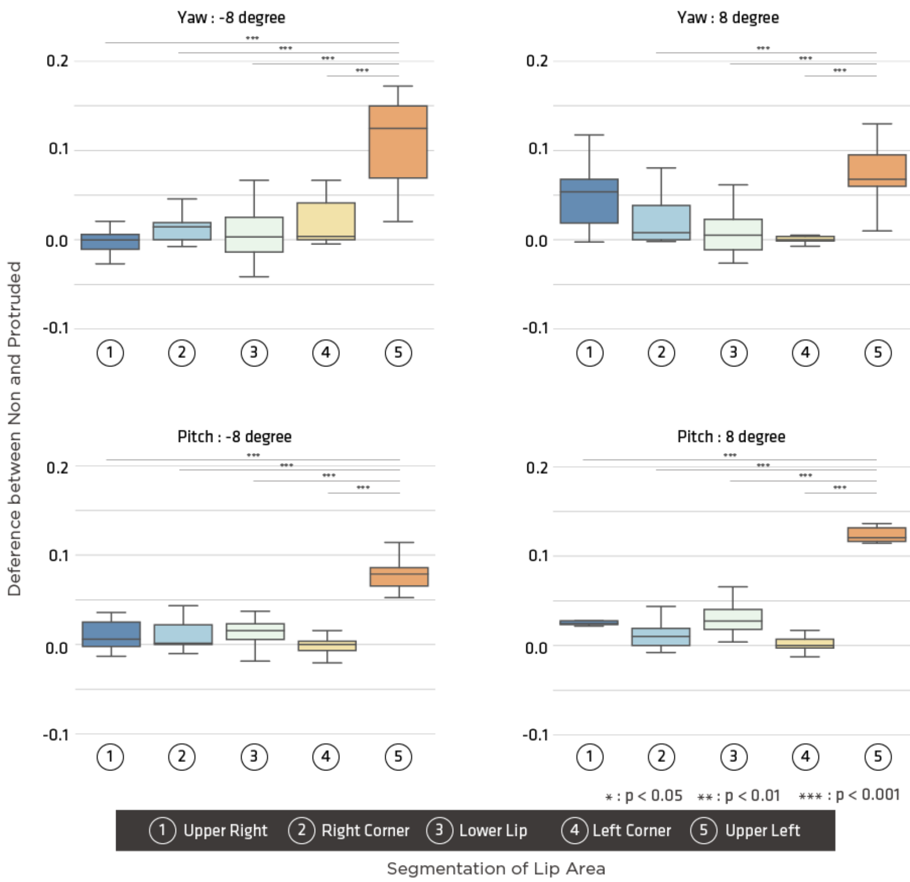

FIGURE 3.17: Comparison Between Different Head Rotation

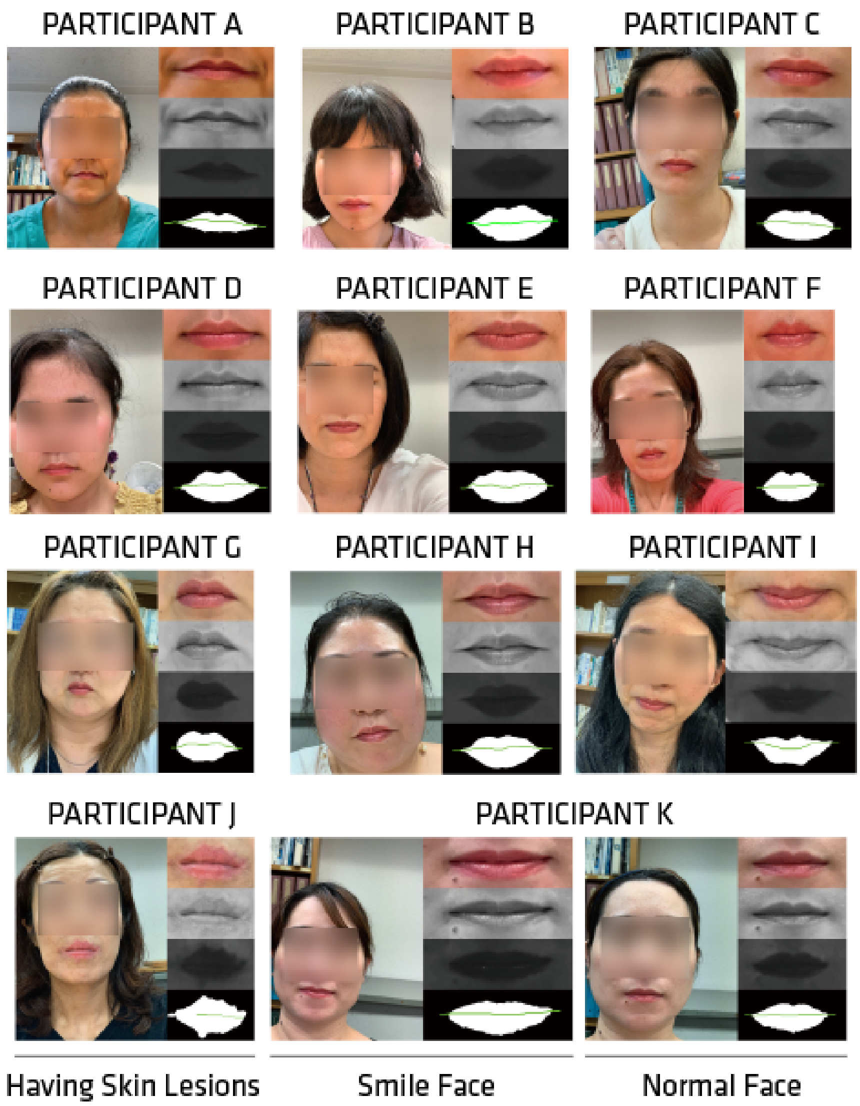

FIGURE 3.18: Selfies Taken Throughout the Experiment

In this experiment, we recruited 11 visually impaired participants (including blind, low vision, and visual field narrowing, including blind, low vision, and visual field narrowing) who wore makeup on a daily basis. Subsequently, we also requested an expert in the visually impaired person's makeup (a lecturer at a special needs school) to judge the makeup protrusion as an evaluator. We conducted an experiment using the following procedure.

First, we conducted 15-min guidance to the participants to familiarize them with the interface operation. Second, the evaluator applied the makeup to the participants. Subsequently, the participants captured a picture using the interface and wiped off the makeup. I used the taken image as a sample of an ideal makeup shape (non-protruded makeup shape). Third, the participants applied lipstick by themselves. Afterward, the evaluator judged the makeup protrusion as to whether there was an overflow on five steps (1: non-protruded, 5: protruded significantly). Finally, participants took selfies by using the interface and received audio guidance on the cosmetic shape of the lip area.

Through this experiment, the same type of lipstick (MAQUILLAGE Dramatic Rouge P RD366, SHISEIDO) was used.

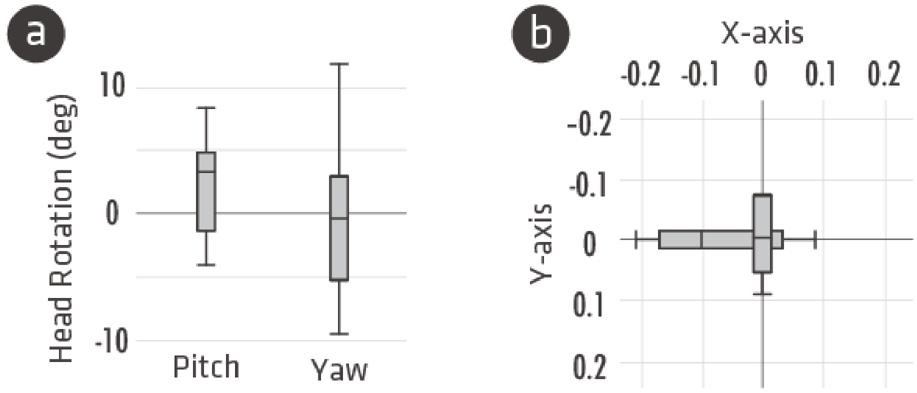

FIGURE 3.19: Head Posture in the Selfies
a:Head Rotation (deg), b:Head Position

TABLE 3.1: Confusion Matrix Between Interface's Protrusion Inference and Expert's Evaluation

(Unit: Number of persons)		Interface's Inference	
		Not Protruded	Protruded
Expert's Evaluation	Not Protruded	7	1
	Protruded	0	1

Results for the Head Posture

In the selfie session, all 11 participants were able to capture pictures of their own makeup faces without assistance. Figure (Fig. 3.18).a shows the head rotation around the pitch and yaw angle in the images captured throughout the experiment. As a result, both the pitch angle and angle were within ± 10 degrees. Figure 3.19.b denotes the difference between the face center coordinate and image center of the selfie image. Here, the center coordinate of the face is normalized such that the origin $(0, 0)$ is the height and width of the face image. The head coordinates parallel to the X-axis are ± 0.1 of the origin. On the contrary, the head coordinates parallel to the Y-axis were in the range of -0.15 to $+0.1$.

The interface presented the state of makeup finish for 9 of the 11 participants. Two participants were excluded from this evaluation because their lips were quantified abnormally. Here, as shown in the table, the interface presented no protrusion for seven of the nine participants. For eight of the nine participants, the evaluator rated less than 2 (no noticeable protrusion) in all lip regions. The precision of interface inference is 1.0, the recall rate is 0.875, and the F-value is 0.933.

Subsequently, we show a figure that indicates the lip shapes evaluated by the evaluator as protruded. The image in the left column is rated by the evaluator as not protruded in all lip regions (rating 1), while the interface provided feedback as having makeup protrusion in the upper lip right. The image in the right column is rated by the evaluator as protruded in the left and right upper lip (rating 5), while the interface provided feedback as having makeup protrusion in the right corner.

TABLE 3.2: Accuracy of Result Presentation of Non-protruded Condition

Precision	Recall	F
1.000	0.875	0.933

TABLE 3.3: SUS Questionnaire (English Version)

1	I think that I would like to use this application frequently.
2	I found the application unnecessarily complex.
3	I thought the application was easy to use.
4	I think that I would need the support of a technical person to be able to use this application.
5	I found that the various functions in this application were well integrated.
6	I thought there was too much inconsistency in this application.
7	I would imagine that most people would learn to use this application very quickly.
8	I found the application very cumbersome to use.
9	I felt very confident using the application.
10	I needed to learn many things before I could get going with this application.

3.4.3 User Study

We conducted a qualitative evaluation of usability to verify whether the proposed method can provide makeup support in daily life from the viewpoint of users.

In this experiment, we recruited eight visually impaired participants and divided them into two groups.

Short-Term Group
A group of four beginners (all blind). They were loaned at the proposed interface for 3 days. Facial rotation pitch and yaw angles were limited to 7 deg. In addition, direct face-to-face guidance was provided before the start of the experiment.

Long-Term Group
A group of four persons who applied makeup on a daily basis (including blindness, low vision, and visual field defects). They were loaned at the proposed interface for two weeks. Facial rotation pitch and yaw angles were limited to 3 deg. In addition, non-directive guidance was remotely provided before the start of the experiment.

To examine only the usage of the application, we targeted those who use smartphones on a daily basis in both groups. We requested the participants to use the interface at least once a day, with no upper limit on the number of times. We used a makeup shape applied under the guidance of an expert in the visually impaired person's makeup (a lecturer at a special needs school) was used as a sample of an ideal makeup shape.

We validated whether short-term proficiency with the equipment is possible in the former group. We also examine whether the interface can be used in a more practical setting in the latter group. Through a comparison of the two groups, we also discuss the effect of the restriction of facial posture on usability.

We conducted the SUS and open-ended questionnaire after the trial period [80]. SUS had been used to evaluate the usability of various systems. Participants answered 10 questions on a 5-point ordinal scale (0: Strongly disagree, 4: Strongly agree). The results of the individual responses are summed and scored between and 0–100, and the distribution of the total mean score is known to be approximately 70 [81]. We also focus on the individual questions used for basic usability research. The content of the question in Japanese is based on the already proposed text [82] (Table 3.3).

After the trial period, we also mesured the time required from the start of using interface to the successful completion of the selfie in the former group. Thtough this experiment, the same type of lipstick (MAQUILLAGE Dramatic Rouge P RD366, SHISEIDO) was used.

FIGURE 3.20: Answers of the Part of SUS and Comments, a: Short-term Group, b: Long-term Group

TABLE 3.4: Number of Selfies Per Participant and Time Required for Selfies

Participant	A	B	C	D
Selfie	7	5	3	10
Duration (Second)	25	7	150	20

Results of the Short-Term Group

Table 3.4 shows the participants' use of the interface throughout the experiment (Participants A and B have acquired visual impairment, and participants C and D have congenital visual impairment). The selfie values are the total number of times the face image was taken, and feedback was returned from the interface during the experiment. The duration values are the time to the completion of capturing selfies. In this experiment, we recorded the first use of the interface at the end of the experiment because this measurement result was significantly reduced by repeated use of the interface.

The average SUS score given by the participants is 73.75. Figure 3.20.a shows the answers of the part of SUS and additional comments. In the part of SUS questions, all participants responded with a value of 4 or higher for the question "I would like to use the application repeatedly in the future" On the contrary, the responses to "The application was easy to use" and "I never felt uneasy using the application" tended to differ among participants. Participants A and D for the former question, and participants A and B for the latter question answered 2 or lower value: it was difficult to use.

In the free-response questionnaire, participant C answered "It took me some time to be able to take a selfie smoothly". On the contrary, participants (A and B)

TABLE 3.5: Answer of Questionnaire: The Time It Took you Until the Taking a First Selfie

Participant	Time
E	It took me more than three hours at first, but I got used to it and was able to shoot in 5-10 minutes
F	several hours
G	About 1 hour
H	After about 2 days of practice

TABLE 3.6: Answer of Questionnaire: Fequency of Use

Participant	Fequency of Use
E	2-3 times a week
F	Every day
G	About once every 2 days
H	About once a day (each shot often took more than 10 minutes, and once I was able to shoot once, I was done)

answered that the interaction was effective in supporting selfies as stated "In had said that I had protrusion, even through I had fixed alarted position".

Results of the Long-Term Group

Table 3.5 and Table 3.6 shows the participants' use of the interface throughout the experiment (Participant E was blind, F and G have low vision, and H has intensively low vision).

The average SUS score given by the participants is 70.65. Figure 3.20.b shows the answers of the part of SUS and additional comments. In the part of the SUS questions, participants F and G responded with a value of 4 or higher, Participant H answered 2 for the question "I would like to use the application repeatedly in the future" On the contrary, for the question "The application was easy to use", participants F and G answered 4 or higher and the other two participants answered two or lower values.

In the open-ended questionnaire, Participant E answered "I'd like to see the shooting range (recognition range) expanded to make it easier to shoot." Participant H answered "I'd like to see the audio speed of the explanation of the shooting results slowed down a bit". Participant F answered "The taking selfie was very fun".

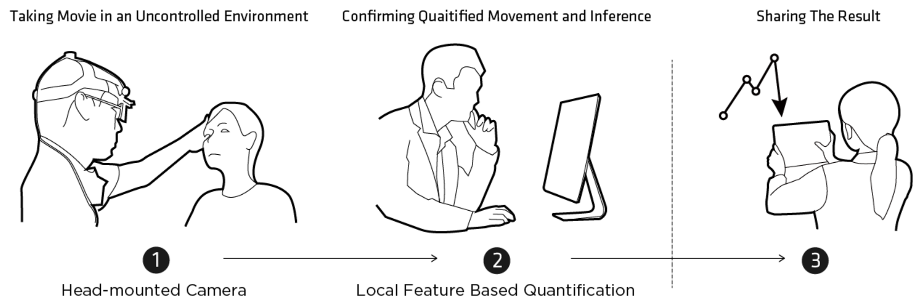

FIGURE 3.21: Facial Nerve Paralysis Diagnostic Support System

3.5 Support Systems for Facial Nerve Paralysis Diagnosis

We propose an interventional support system for the diagnosis of facial paralysis. Our proposed system aims to enhance the diagnosis by supporting the following procedure.

3.5.1 Taking Movie in an Uncontrolled Environment

The doctor takes the facial video himself/herself in the facial nerve paralysis diagnosis. In this paper, we aim to acquire an appropriate facial video that does not interfere with facial paralysis diagnosis in an actual environment. We focus on the face-to-face format of the Yanagihara 40-point method and propose a video captured using a head-mounted camera. This format allows the doctor to capture a frontal facial movie while giving posture instructions to patients.

3.5.2 Confirming Quantified Movement and Inference

The doctor reviews the results of the quantification of facial movements and estimates the paralysis levels from the system. Here, the system evaluates the lip posture by showing teeth (i-) that have already been statistically correlated with other facial nerve palsy diagnosis methods. For facial movements, the system visually presents time-series changes in the facial local area. On the contrary, for the inference of the facial paralysis level, the system presented a five-point scale from 0 to 4. The results confirm that the system supplements the doctor's decision criteria and reduces diagnostic errors in individuals.

3.5.3 Sharing the Result

This system is not for only the time of diagnosis but also to ensure consistency. Sequentially recording the patient's condition in numerical aspects provides the same index for intermittent diagnoses. That is, it is expected to minimize errors at different times and by different doctors.

3.6 Implementation of the Diagnosis Support Interface

We describe the implementation of the proposed system, an imaging device for clinical use, and an image processing methodology. This is intended for use in the clinical field and minimizes disturbances caused by environmental changes that contribute to accurate validation.

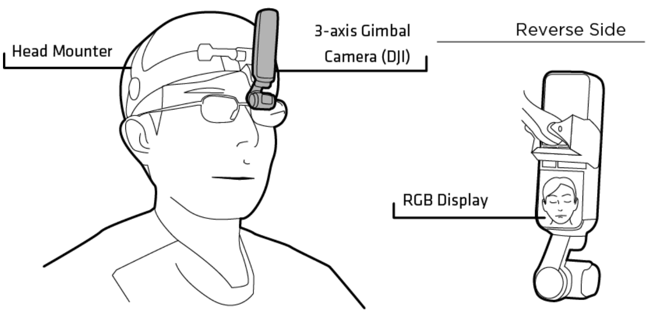

FIGURE 3.22: Head Mount Camera for Facial Paralysis Diagnosis

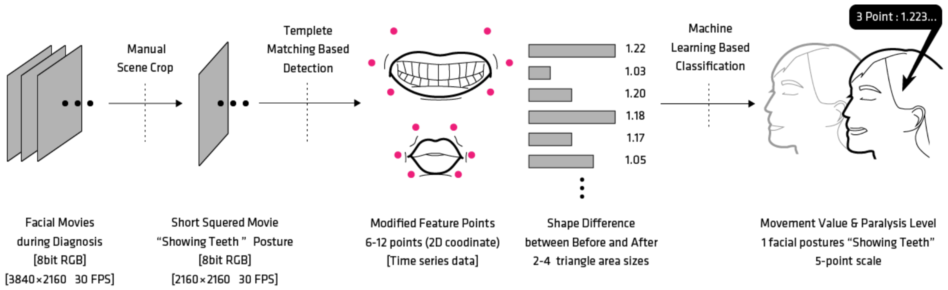

FIGURE 3.23: Outline of Processing

3.6.1 Head-mounted Camera

Visual comparison of the paralyzed and healthy sides is crucial for the diagnosis of facial nerve paralysis. Therefore, the facial images used for validation must include the frontal face of the patient. Subsequently, a high-resolution movie is required to track the texture of the human face in each frame as the minute facial movements are also included in the important criteria.

In this study, a camera with a three-axis gimbal stabilizer (OSMO Pocket, DJI) mounted on a head mounter at the same height as the doctor's eye level was used (Fig. 3.22). The doctor confirmed that the patient face is in the frame center with reference to the LCD display on the back of the camera.

3.6.2 Facial Texture Movement Tracking

The captured movie (width: 3840; height: 2160; 8-bit RGB; 30 FPS) goes through several steps and are output as facial movement features and estimated paralysis levels (Fig. 3.23). Here, we describe a method for quantifying the movement of a person's local facial surface.

First, we crop a square (width: 2160; height: 2160) area in the frame of the facial movie from the center. Then, we also manually cropped the section of the "showing teeth" face posture. Second, the system performs adaptive histogram normalization on the facial movie as a pre-processing step [83]. Here, the system computes the neighborhood cumulative distribution function for each image sub-region separately. In this implementation, we empirically set the size of the small region to be an 8×8 rectangle and the clip limit of the histogram amplification to 2.0.

The system estimates feature points based on template matching to detect small movements of the facial surface. The feature points are culculated by the following procedure.

1. Detecting facial feature points in the first frame. The system performs face detection using the histogram of oriented gradients' features [84]. Subsequently, within the rectangular region, the system performs facial feature point detection based on the inter-pixel error features in a regression tree

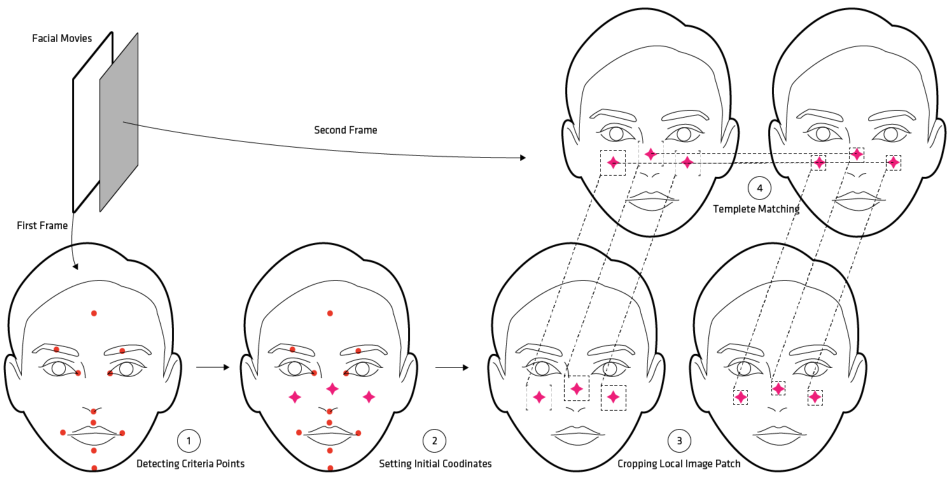

FIGURE 3.24: Outline of Facial Texture Movement Tracking

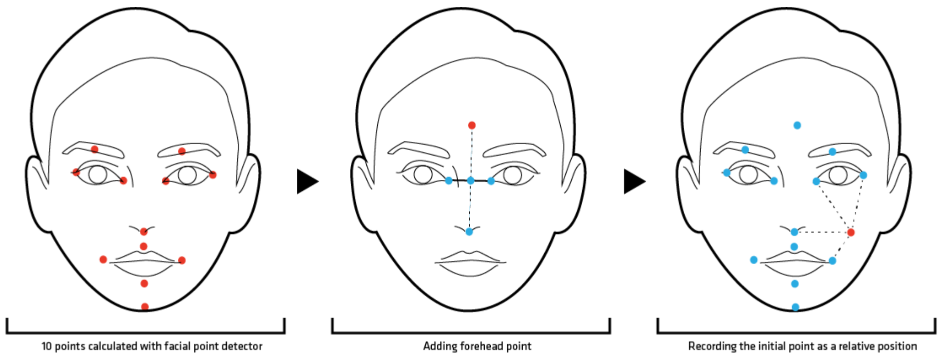

FIGURE 3.25: Tracking Initialize Procedure

[31, 85]. We selected 13 facial points: forehead, right eyebrow, left eye brow, right eye inner corner, left eye inner corner, right eye corner, left eye corner, nose septum, mouth top, mouth right, mouth left, mouth bottom, and chin. Their forehead points are calculated by the point of the inner corner of the eye and the nose septum. We set the center of each eye's inner corners as the reference point. We set the forehead point as calculated by extending from the reference point to the forehead region. The length to the region is the same as that of the bridge of the nose septum (Fig. 3.25).

2. Setting the initial tracking points to the first frame. These feature points are arbitrary coordinates on the facial surface. To share the position of the setting point between different patients, the system obtains the coordinates relative to the reference point. All two-dimensional coordinate data treated below are normalized with the X-axis as the width of the face (length between the left and right cheeks) and the Y-axis as the height from the eye line to the tip of the chin (Fig. 3.25).

$$P_x(i) = T_x - J_x(i) \quad (i \in [0, d)) \tag{3.22}$$

$$P_y(i) = T_y - J_y(i) \quad (i \in [0, d)) \tag{3.23}$$

Here, $P_x(i)$ and $P_y(i)$ are the x and y coordinates of the i-th encoded tracking points, respectively. T_x and T_y are the X-axis and Y-axis coordinates of a tracking point. $J_x(i)$ and $J_y(i)$ are the X-axis and Y-axis coordinates of the i-th reference points, respectively. d is the number of reference points. In this paper setting, it is 13.

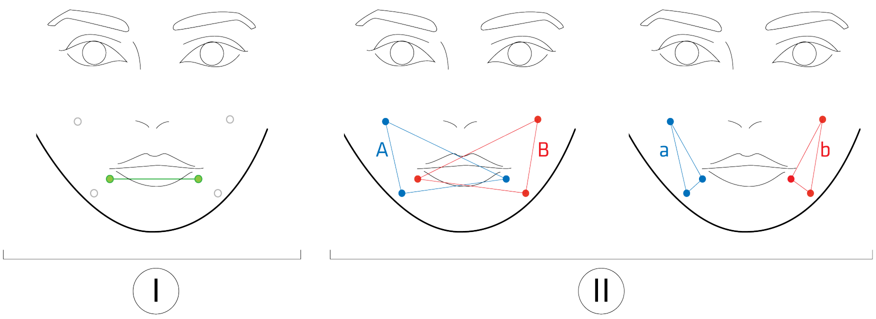

FIGURE 3.26: Tracking Point and Area

On the contrary, to apply the tracking points to different patients, the system use the above encoded tracking coordinates as follows.

$$Q_x = \sum_{j=0}^{d} S(j)(E_x(j) - P_x(j)) \quad (3.24)$$

$$Q_y = \sum_{j=0}^{d} S(j)(E_y(j) - P_y(j)) \quad (3.25)$$

$$S(j) = \frac{I(j)^2}{\sum_{s=0}^{d} I(s)^2} \quad (j \in [0, d)) \quad (3.26)$$

$$I(j) = \sqrt{P_x(j)^2 + P_y(j)^2} \quad (j \in [0, d)) \quad (3.27)$$

$Q_x(j)$ and $Q_y(j)$ are the x and y coordinates of the j-th tracking point, respectively. $E_x(j)$ and $E_y(j)$ are the j-th reference point's coordinates of the target patient's face. $P_x(i)$ and $P_y(i)$ are the x and y coordinates of the j-th encoded tracking points, respectively. d is the number of reference points. In this paper setting, it is 13.

3. Performing a template matching on the updated frame [86]. The system tracks the point located on the face based on the similarity comparison between the anterior and posterior local regions. In addition, we partly apply the Lucas–Kanade algorithm to increase robustness because the facial points do not move significantly before and after the frame [37, 87, 88].

   First, local images are cropped in the square, which is centered at the current tracking point, from the source and resized image. The window needs to be fine-tuned according to the size of the imaged face; however, we herein assumed a face with a height of 1400 pixels and set a rectangular area of 20 × 20 in the image. The local images below are called L1 and L2, respectively.

   Next, in the resized updated image, the temporal tracking point is set at the center point of the region, which has the highest similarity to L1 for each of the points existing in the circular region around the current tracking point. We set the circular region, or the search region, to a 50 × 50 square. Furthermore, the system computes the similarity to L2 for each of the points existing in a circular region around the temporal point in the updated image. The system uses the coodinates of similarity-minimized point as the updated coordinate. Throughout the above process, the scale transformation is assumed to be 0.1 times, and the matching similarity index is a zero-means normalized cross-correlation [89].

3.6.3 Quantification of Facial Movement

In this paper, we propose the following two types of features based on interviews with specialists in facial nerve paralysis. The system tracks 6 facial points on the left and right sides; the point on the zygomatic muscle, at the lower jaw and at the lower corner of the mouth (Fig. 3.26).

Maximum Mouth Aperture

We quantify a mouth aperture that describe the movement strength of the entire lip, including the affected and healthy sides. The feature is the maximum distance between the left and right lip corners of the mouth in the time series (Fig. 3.26.I). Here, the mouth aperture is normalized by the value at the rest posture.

Average Symmetry

The average symmetry feature describe the movement symmetricity between the affected and healthy sides (Fig. 3.26.II). The feature value is calculated areas composed of the points as below.

$$F_d = |F_h - F_a| \tag{3.28}$$

$$F_h = \frac{A}{q} \tag{3.29}$$

$$F_a = \frac{B}{(1-q)} \tag{3.30}$$

$$q = \frac{a}{a+b} \tag{3.31}$$

Here, F_d is the area feature difference between the affected and healthy sides. F_a and F_h indicate the area feature values of the affected and healthy sides, respectively. A and B are the areas from the side of the left and right lips to the opposite corner of the mouth, a and b are the lateral areas of the left and right corners of the mouth, respectively. Each values are normalized by these values at the rest posture.

Since the length of the patient's movements varies from diagnosis to diagnosis, we cut out the interval to the rising edge and use the average value of F_d among the frames as the feature value. We cut out the frames based on the mouth aperture, and the interval from 0.1 to 1.0 are cropped when the maximum value was normalized to 1.0 (t_R in Fig. 3.27.a).

Here, the values of the feature values F_a and F_h are larger on the healthy side and smaller on the affected side (Fig. 3.27.b). In other words, the difference of the integrated value of each feature in the cropped interval is the average symmetry feature.

3.6.4 Simple Classification on Facial Paralysis

The system estimates the palsy level based on the Yanagihara 40-point method. We used a support vector machine (SVM) with a linear kernel for classification [43]. In support vector classification, we use the L2 penalty for the penalization and the value of 0.0001 as the tolerance for stopping criteria.

3.6.5 Facial Diagnosis Dataset

We create a dataset for validating the proposed methodology. The dataset contains the diagnosis data: facial movies during "showing teeth" ' facial movement, extracted feature values, and paralysis level labels (Fig 3.28). Here, the movies are captured by the proposed head-mounted camera in an uncontrolled environment with ambient light and patient movement. The dataset contains 42 movies, which has duplicate the 19 patients. The labels of the paralysis level are based on the

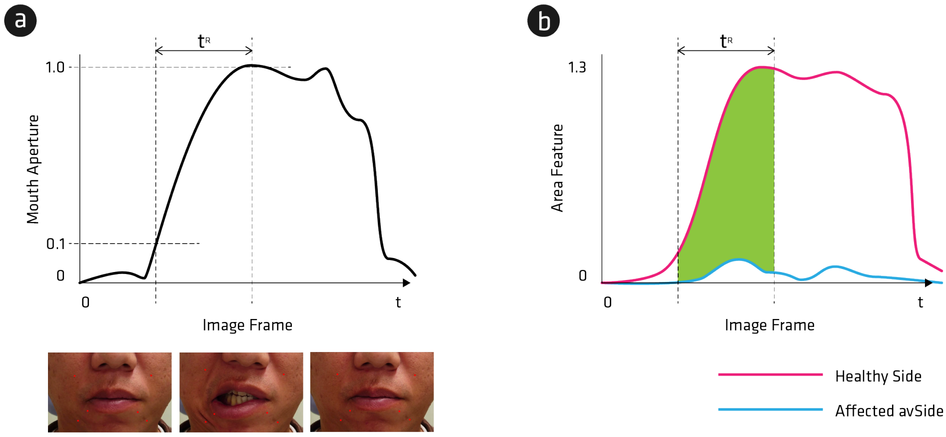

FIGURE 3.27: Transition of Proposed Feature Value, a: Mouth Aperture Through Image Frames, b: Area Feature Through Image Frames

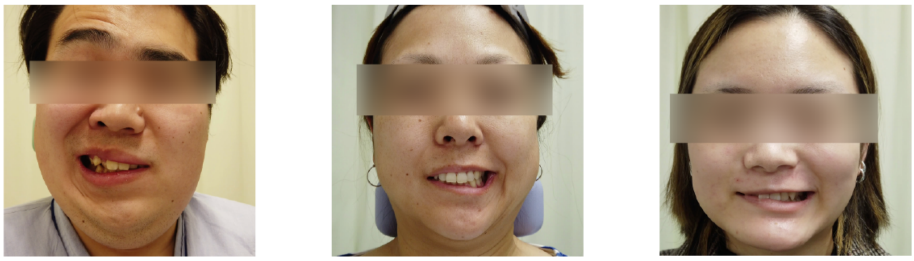

FIGURE 3.28: Examples of Movie Frame in Dataset

actual diagnosis of an expert doctor of the Yanagihara's grading method. The ratio of the paralysis level included in the dataset is not controlled (Level 4 : 3 people, Level 3 : 6 people, Level 2 : 7 people, Level 1 : 15 people, Level 0 : 11 people).

3.6.6 Hardware and Software Settings

We used a separately prepared computer to process the images captured by the head-mounted camera (OSMO Pocket, DJI). Throughout this implementation, we used part of the OpenCV implementation for the image manipulation process [78]. We also used scikit-learn to implement the support vector machine [90]. The core part of the program was written in Python 3 [79].

TABLE 3.7: Result of OLS Regression

Explanatory Variable	Coefficient	T Value	p Value
Maximum Mouth Aparture	11.477	-2.762	>0.000
Average Symmetry	-2.554	8.703	0.009

3.7 Experiment of Support in Facial Nerve Paralysis Diagnosis

3.7.1 Performance Evaluation

The experiment examines the relation between the proposed feature value and the specialist's diagnosis.

Here, we extracted the proposed features: maximum mouth aparture and average symmetry from movies in the facial diagnosis dataset: sections showing teeth facial movement (42 movies). We assumed that the proposed feature has a linear relation to the paralysis level. Therefore, we performed multiple regression analysis on each feature based on ordinary least squares (OLS) regression.

Result of Regression

The maximum lip width and the average similarity had coefficients of 11.477 and -2.554, respectively. The T-values were -2.762 and 8.703, which showed significant p-values ($p<0.05$) for both explanatory variables. The degree of freedom adjusted R-squared value was 0.676, and the F-value was 38.

Figure 3.29 shows the distribution of explanatory variables. The X-axis and Y-axis denote the values of the average symmetry and maximum mouth aparture, respectively. Subsequently, the color of each point denotes the paralysis levels. From this figure and the results of the OLS regression analysis, the paralysis level increases the change with feature values, decrease in average symmetry, and increase in maximum mouth aparture.

3.7.2 Feasibility Study of Paralysis Level Estimation

The experiment examines the feasibility of the paralysis level estimation based on Yanagihara's method.

We compared the inference results based on the machine learning method and the paralysis level. The video data for learning and the correct answer labels were obtained from the facial paralysis diagnosis database, and the validation was based on leave-one-out validation. For inference, the extracted features of maximum mouth aparture and average left-right symmetry were used. Each feature was normalized using the z-score method so that the mean was 0 and the standard deviation was 1. Additionally, we used an SVM based on the linear kernel as the machine learning method.

Result of Inference

The SVM-based estimator had a correct rate of 0.57 and an error variance of 0.55.

Figure 3.30 shows the decision plane of the inference result and correct answers. The X-axis and Y-axis denote the value of normalized feature values. The segmented background planes correspond to the respective paralysis levels. In this figure, the evaluation of the paralysis level increases from the lower right to the upper left. On the contrary, the plane at level 2 is drawn outside the correct answer plot.

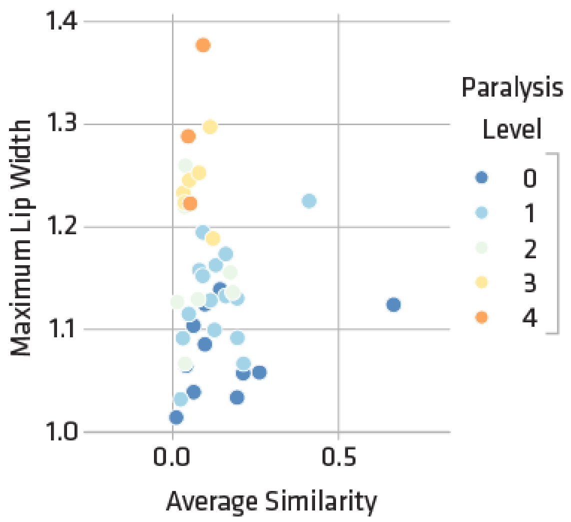

FIGURE 3.29: Distribution of Explanatory Variables by Paralysis Levels

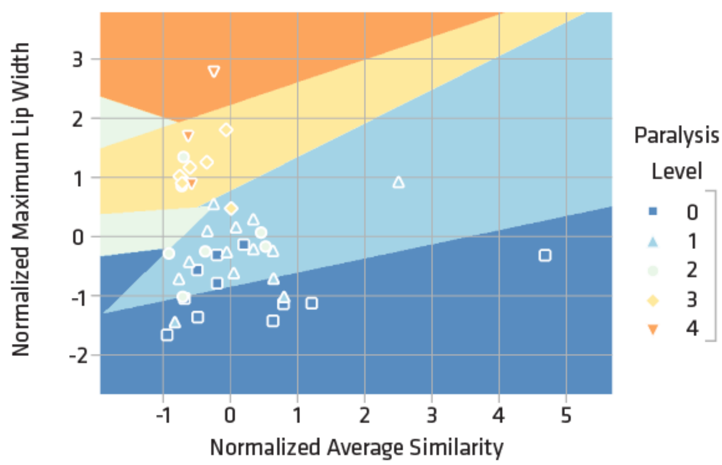

FIGURE 3.30: Decision Plane of Support Vector Machine

Chapter 4

Discussion

4.1 Makeup Support Interface for Visually Impaired Persons

4.1.1 Accuracy of the Lip Shape Quantification

Various Protrusion Setting

The feature values showed high values in the intentional protruded lip area. Subsequently, the accuracy was high enough to detect the overflow of 2 mm, which was the detection target of this study. On the other hand, the condition in which the protrusion was applied to the lower lip did not show any significant difference from the right corner of the mouth. This was because only the lower lip region occupied a larger lip area of the frontal surface than the other regions so that a more extensive protrusion is necessary to obtain significant differences. However, it may also be due to the effect of slight head rotation, which will be discussed below.

Various Head Rotation Setting

The proposed feature values drifted with the head rotation in the results. However, the difference values toward the normal shape was not change significantly. Therefore, proposed interface will able to describe the protrusion when the head rotation angle is within 8 degrees.

On the other hand, the value in the intentional protruded area with +8 degree head rotation did not have significant difference toward the value in right upper lip area. Although it is only a finding, the difference value tended to be less than that of the other protruded areas when the head rotation was with -8 degrees around the pitch angle. Therefore, the feature value based on the proposed method may have a case that judge as protruding due to the contour shape distortion regarding the head rotation.

4.1.2 User Studies in Actual Makeup Scene

Selfie Support

According to the results, all selfie images photographed the frontal face near the center. This shows that our method is capable of teaching the appropriate head posture for visually impaired persons.

Makeup Finish Confirmation

The results show that the system did not process the two participants (corresponding to participants J and K in the figure 3.14) correctly. One of them had skin lesions and erythema scattered around the lips, which made the appearance of a makeup shape excessively large. Another participant had a laughing face different from

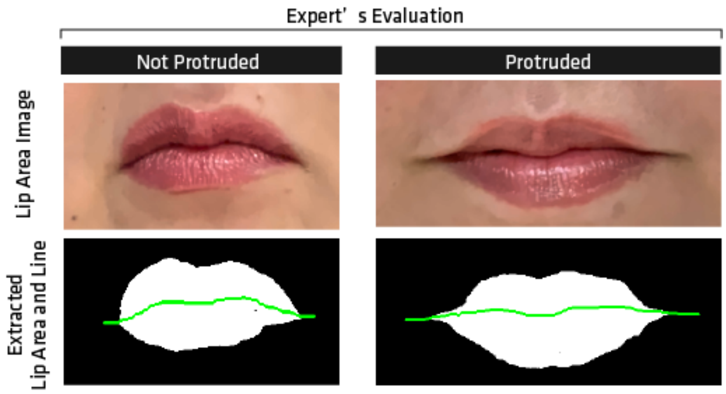

FIGURE 4.1: Lip Images Evaluated as Protruded

the resting face in the ideal shape sample of the experiment. To avoid such unconscious stiffening of the lips, one should consider providing voice guidance to encourage relaxation during selfies.

Table 3.2 indicates that the precision rate of the makeup confirmation is 1.0, which is generally consistent with the expert's evaluation. On the contrary, for one participant, the result did not match with the expert's evaluation in terms of the existence of protrusion. This participant (figure 4.1, left column) applied a lipstick thickly from the right corner to the right upper lip, which improved the symmetry of the lip shape visually. This might have been related to the expert's evaluation without protrusion. Therefore, improving the validity by proposing a model that includes the application balance of the entire lip would be possible.

Subsequently, one participant (figure 4.1, right column) received the result not matching the expert's evaluation in the protrusion area discrimination. This participant had a lip shape wherein the shadows at the corners of the mouth were long on both sides, and the lip line shape was easily changed by the rigidity of the corners of the mouth. Our proposed method is significantly affected by the quantification based on the lip line. Therefore, introducing a correction method that is more robust against changes in the lip corner before and after taking selfies is necessary.

These results indicate that the proposed interface is limited in its ability to support "improving makeup through protrusion warnings" However, our interface is capable of supporting "correctly communicating protrusion when it does not exist"

4.1.3 User Study

Short-term Group

The SUS score of 73.73 is average with regard to score distribution in previous researches, corresponds to "Good" in adjective form. In the SUS responses, all users answered affirmatively to the continued use (Fig. 3.20.a). Table 3.4 shows that the three participants had taken more selfies than the three times required at the beginning of the experiment. When the participants were asked about this at the end of the experiment, they answered that they actively tried to use this as a makeup training opportunity. These results indicate that the proposed interface has usability satisfying the requirements of a visually impaired person's makeup confirmation.

On the contrary, interface operability was differently evaluated by each participant in the SUS and the open-ended questionnaire. The time required to take the selfie as shown in the table 3.4 also differed among the participants. These results denote that the operation difficulty of proposed interface has individual differences.

Although this is only a reference, participants with congenital visual impairment tended to take significantly more time to learn the operation at first. In the future, we will consider periodically providing detailed face positions with voice guidance; "Your face is facing up to the right" to ease the operation.

Throughout the experiment, participant A and B claimed that they was still warned even through they fixed the makeup. The same participants reported that they felt uneasy about using the application in SUS as well, and such mechanical errors during use may have led to a decrease in reliability. The proposed interface could not feedback the amount of protrusion. Therefore, the ideal support may be difficult in some cases, such as when makeup correction is incomplete.

Long-term Group

The SUS score of 70.65 is average in the score distribution in previous researches, corresponds to "Good" in the adjective form.

However, this group was divided into two groups: positive and negative responses for continuous use (Fig. 3.20.b). Subsequently, the answers to the question about the difficulty of operation were the same. The result indicates that the long-term group had a different impression about the usability between participants.

In the comments, participant E answered that the angle of view of the camera was too narrow, and participant H answered that the playback speed of audio feedback was too fast. For the latter, we found that same participant took a long time to photograph in Table 3.6. The two participants are who rated the low value for continuous use and difficulty of operation in the SUS, respectively. In addition to the reasons given by the participants in their comments, the reason might be the increase in the difficulty of selfies due to the severe head angle limitation.

The another reason of the result might be that the explanation of the use of the equipment was done remotely. Unlike face-to-face guidance which can instruct by moving the participant's hands directly, explaining the physical positional relationships between the face and the interface is difficult in the remote guidance. Therefore, In Table3.5, participant E and H took significantly longer time than the other participants in the time between the first time the device was turned on and the first selfie was taken. In other words, both of them are expected to have more difficulty in grasping the data than the other participants.

These results suggest that it is necessary to expand the angle of view in which face shots can be taken and to further devise instructions for taking selfies, in addition to applying a feedback speed that is appropriate for each individual. The feedback speed can be changed by adding functions. On the other hand, the angle of view depends on the hardware, and the instruction at the beginning of use must be covered by additional information, which is a issues in the future.

It is noteworthy, that the selfie skill had leveled off, even though the participants was used for a period longer than the short-term group. As mentioned above, our proposed method is affected by head rotation. However, the experimental results show that the restriction of head rotation may lead to a decrease in usability. To provide optimal feedback in this trade-off relationship, further investigation is necessary. However, participants F and G with mild visual impairment responded positively to the comments and SUS. The result suggests that participants with low vision can use the proposed interface, even if it is a remote explanation without physical interaction.

4.2 Discussion on the Makeup Support Interface for Visually Impaired Persons

4.2.1 Performance Evaluation

The results show that the maximum mouth aparture and the average symmetry indicated a significant difference in the palsy level. Here, the maximum lip width

TABLE 4.1: Expert Doctor's Assessment Criteria

Palsy Level	0	1	2	3	4
Palsy Side Movement	Not Moving	Slightly Moving	Obviously Moving	Moving	Moving
Movement Symmetry	Assimmetry	Asymmetry	Asymmetry	Slightly Asymmetry	Symmetry

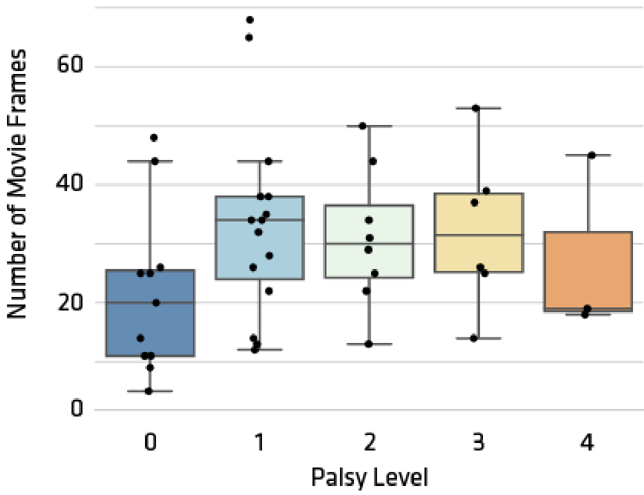

FIGURE 4.2: Boxplot of Number of Frames in Dataset Movies

and average similarity were positive and negative coefficients, respectively. The result shows that the expert's diagnosis result is highly evaluated when there is a large change in the mouth aparture and a small difference between the left and right sides of the facial movement. Figure 3.29 that plots the distribution of each feature shows the same result. Based on this, the patient having significant paralysis are considered to not be able to make obvious lip movement.

The maximum mouth aparture feature showed a smaller p-value than the average similarity feature, indicating that it better reflected the differences in diagnostic results. This is consistent with the report observations that changes in a lip width are more significant than changes in lip height [91].

On the other hand, the adjusted R-squared value was 0.676, which did not reach the target value of 0.8. This indicates that it is difficult to completely explain the expert's evaluation with only two proposed features. One of the reasons for this is that the proposed method is verified based on two-dimensional images, but the experts diagnose by directly looking at three-dimensional faces, and it is thought that features are truncated in the process. For reference, in the interview, the expert doctor answered that he paid attention to the movements of the whole face to observe the after effect [92]. Therefore, Measuring the movements of other parts of the mouth might be worth to consider.

4.2.2 Feasibility Study of Paralysis Level Estimation

The results showed that the paralysis level estimation based on the proposed features agreed with 56% with the correct answer. As mentioned earlier, the level estimation used in this study is a five-level estimation, which is a special setting for this study. As the error variance of 0.55 is lower than 1.0, it is expected to be accurate enough for estimation in an actual diagnosis scene.

In this experiment, the decision plane in the figure 3.30 shows that the palsy-level distribution gradually increases from the right side of the X-axis to the upper side of the Y-axis. While this is the same as described in Experiment 1, in this section, we refer to the decision plane in the figure and conduct the investigation in detail.

TABLE 4.2: Comparison of the Estimation Accuracy

Methodology	Accuracy(%)	Average Error
Ngo 2014, GBLBP–B [35]	55.2	0.64
Ngo 2016, Tracking [37]	72.5	0.31
Proposed	57.0	0.55

First, in the group judged as level 0 in the decision plane, most cases had no lip movement. On the contrary, the 1 case, indicating the average similarity, is judged to have a low palsy score even when the lips move slightly.

In the group judged as Level 1, many cases in which the correct label is 0-2 are mixed in. The reason for this is that the database used in this experiment has a large number of Level 1 cases, so the range of this level was expanded. Subsequently, the different facial movement times of the patients in the database might have an effect. Figure 4.2 shows the distribution of the number of frames for each case in the dataset. Here, the X-axis is the correct palsy level and the Y-axis denotes the number of frames from the start to the end of the movement. The figure shows that the number of frames ranges from 10 to 50, especially in the cases of Level 1-2. The number of frames varied between 10 and 50, in the cases of Levels 1-2, indicating that the onset duration of the facial posture varied by more than 1 s. This is a matter that affects facial impressions, such as whether there is a slight expression or not. Standardizing the facial motion duration or finding features not affected by the duration is necessary [67].

The group that was judged as Level 3 contained all but one correct Level 3 case. However, the group judged as Level 4 included two cases that may have been affected by the longer duration of onset, mild sequelae, and asymmetry at rest.

Although a simple comparison is not possible due to the different datasets, Table 4.2 shows a comparison of the accuracy of paralysis level estimation with related studies on the 5-scale Yanagihara's grading method. Here, the proposed method is more accurate than the method using local binary patterns and Gabor filters, and the estimation results are inferior to the state of the art method [37]. The difference in the estimation results is thought to be related to the presence or absence of control of the shooting environment, such as head rotation, in the target image data.

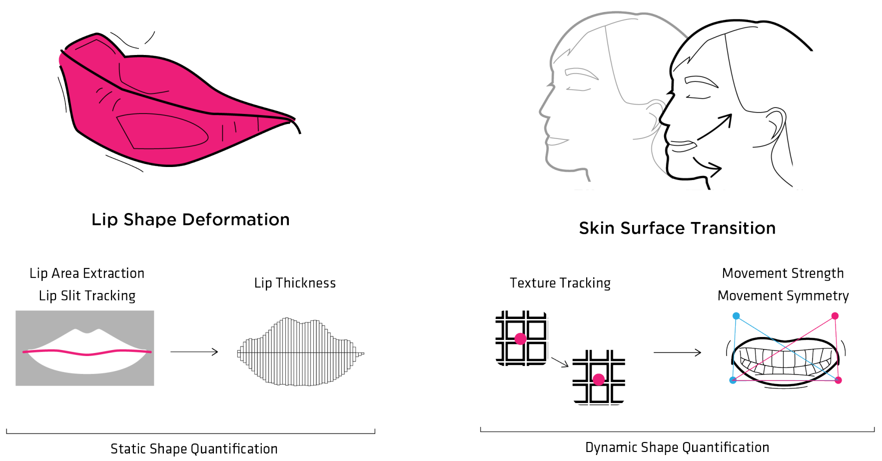

FIGURE 4.3: Proposed Quantification Methods

4.3 Overall Discussion

To address the research questions of our study, we discuss the implementation and experimental results.

4.3.1 Proposed Quantification Methods to Extract Local Features

"What local features affect the improvement of facial appearance represented by subregional face shapes and motions?"

We proposed methodologies that extract local feature around the mouth (Fig.4.3). While studying makeup support for the visually impaired persons, the proposed system quantifies the lip thickness by applying a color transformation enhancing the lip area and lip lines. Here, we quantify the deformation of lip shape based on the comparison between the ideal lip shape and the current lip shape. While studying support for facial nerve paralysis diagnosis, the proposed system tracks local texture transition on the facial surface. Since these measurements are time series data with a very large number of dimensions, we design two features: movement strength and similarity between left and right movements. We quantify these features from the width of lip and combination of local area size, respectively.

We designed the two features to linearly increase or decrease in geometric shape change of facial images. From the experiment results of both studies, the relations between the proposed features and expert's assessment, decision of "Presence or absence of makeup protrusion" and "Aperture and symmetry of lip area", were confirmed. Therefore, we have partially developed a method for numerically observing the process of shape change in localized areas that will contribute to the improvement of facial appearance.

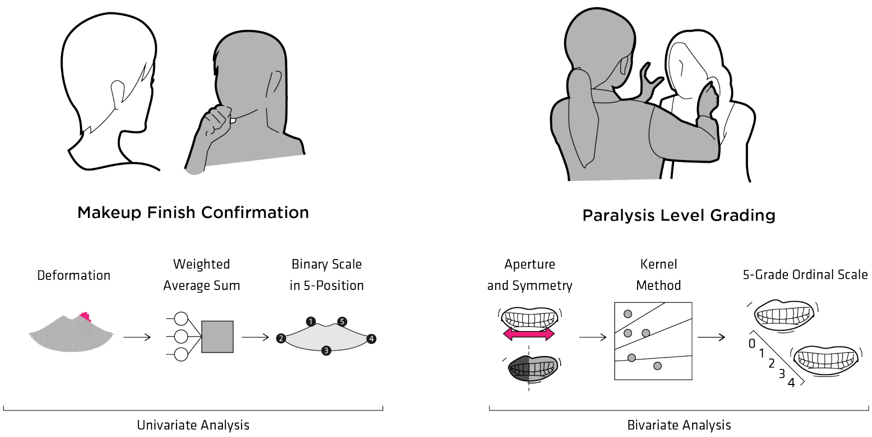

FIGURE 4.4: Proposed Formalization Methods

4.3.2 Interaction for Facial Assessment Support

"What are the relations between the expert's facial assessments and subregional facial features?"

We proposed methodologies that formalized the assessment criteria of experts based on the quantified local feature value (Fig.4.4).

While studying makeup support for visually impaired persons, we implement the system that comfirms the presence or absence of lip makeup protrusions at 5-positions. The system performs a classification based on a protrusion assessment value that is derived by the weighted average sum. We can think that the assessment model is one of the univariate analysis from the single attribute that is calculated from multiple feature values.

While studying for facial nerve paralysis diagnosis, we proposed the assessment model that grades on a 5-step scale by the kernel method based classification. Here, we used two features, lip aperture and symmetry, as variables. In other words, the assessment model is a bivariate analysis.

The experiment results showed that our method is capable of performing evaluations that reflects the experts' evaluations. We designed the models to interpret the non-verbal factors of the expert's assessment as mathematical description in both studies. Our proposed models integrate local features, allowing us to specify how they work as assessment criteria for individual experts.

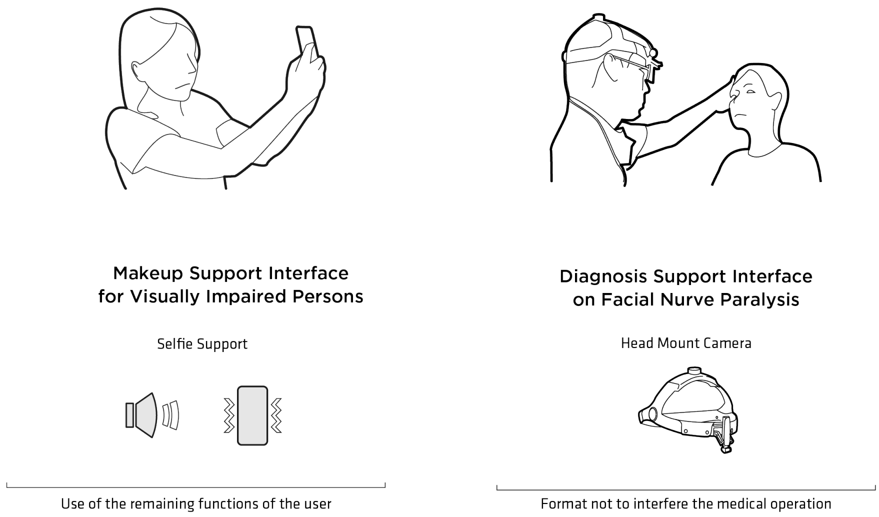

FIGURE 4.5: Proposed Interfaces

4.3.3 Assessment Support in Actual Use

"What are the functions of interfaces for use in uncontrolled environments?"

We proposed methodologies to capture facial video images and provide feedback on the validation results (Fig.4.5). Here, we mentioned about the user's use or disability-related restriction in an uncontrolled environment.

As for feedback presentation, in the makeup support research, the positional relation between the face and the camera was presented to a visually impaired person by voice and vibration. The interface implemented in the diagnosis support study was designed such that the user can take movies while checking the patient's face position without using their hands.

Through the experiments and the database imaging, participants and physicians were able to take photographs that captured the frontal face. In summary, the proposed interactions allow us to grasp the difference between the current position and the appropriate position, using the user's remaining functions or not interfering with their operation.

Chapter 5

Conclusion

This chapter wraps up our research, summarizing it and exploring the future direction. We clarify the academic contribution of this paper, and its possibilities are mentioned.

5.1 Makeup Support Interface for Visually Impaired Persons

We propose a human–machine interaction to confirm applied makeup for visually impaired persons. The interaction is structured in two parts: selfie support and makeup status feedback through the mobile interface. In the selfie support phase, the proposed interface presents the facial posture by multimodal feedback, making the visually impaired user capture an appropriate facial image by themselves. In another phase, the proposed interface alerts the makeup protrusion around the lip contour, which is designed to understand the position even if visually impaired. To realize this interaction, we implement a novel application that includes accurate lip shape quantification methods and protrusion detection methods.

Through performance evaluation experiments of lip shape quantification, the result showed that the proposed image processing method is accurate enough to detect lip overflow as large as 2 mm. Next, through a user study with a visually impaired person, the results showed that our proposed method can capture the face image facing the camera and can correctly feedback the existence of the makeup protrusion. In addition, we evaluated the usability of the system through continuous use. The questionnaires showed that the functions of the interface fulfill the requirements of visually impaired persons to confirm the completion of makeup.

5.2 Facial Nerve Paralysis Diagnosis Support

We explore and implement a diagnosis support for facial paralysis in a format that does not interfere with medical operations. The system includes a wearable platform that assists in taking facial movies in the face-to-face diagnosis scene. In the image processing aspect, our method enables highly accurate visual inspection in an uncontrolled environment by the time-series micro-motion measurement at specific points on the facial surface. We also derived an objective symmetry index that is independent of the patient's paralysis status by extracting features specific to facial posture based on the Yanagihara method.

In accordance with the performance evaluation of the proposed feature quantification, the quantified feature values based showed significant relations with the expert's diagnosis result. Subsequently, the inference result of the paralysis level in the experiment showed that the accuracy of the prediction was sufficient for use in normal diagnosis cases.

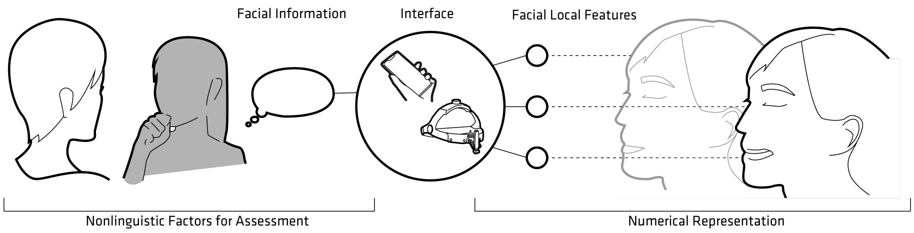

FIGURE 5.1: Our Proposed Interface Bridge Between Non-verbal Facial Assessment and Numerical Facial Features

5.3 General Conclusion

We summarize this research as a whole, and describe the contribution to human informatics. My definition of human informatics is "The academic that aims at further understanding of human and utilization of systems by quantitatively representing people themselves".

In this study, we investigate the realization of an interface that quantitatively measures and assesses human faces in actual scenes. In particular, this study focuses on an expert's facial assessment, which requires expertise and experience. We proposed quantification methods to clarify the relation between the local facial features and the expert's assessment. Here, the feature values are designed to have linearity with expert ratings. Subsequently, we considered the formalization of the assessment model of experts which is translate the mathematical feature value to the non-verbal knowledge. In the procedure, we made an attempt to represent nonlinear models numerically.

In accordance with the results of performance evaluation and user studies, the proposed features in local facial areas were suggested to be related to the expert's assessment, and the output from the evaluation model showed values corresponding to the expert's decision. Subsequently, the interface using our proposed method can be used smoothly in uncontrolled environments, demonstrating the feasibility of our proposal.

I propose a method for a facial assessment system in which there are individual differences between the evaluated person and the evaluating person. Here, I make it possible to make representations based on the integration of measured numerical descriptors on the local facial surface.

In other words, this study provides the knowledge that "Non-verbal and nonlinear facial assessments can be represented by integrating numerical features that have a linear relationship with face shape" (Fig. 5.1). I believe that it is possible to understand person's preferences and lifestyles using facial information as an extension of our works.

5.4 Future Prospect

We herein proposed facial evaluation frameworks by looking at a person's facial appearance from a microscopic point of view, which requires expertise that could not be achieved in the past. The proposed framework for face support is expected to indirectly support the improvement of facial appearance in people's living and working scenes. On the contrary, this study has revealed some problems that require further investigation.

Based on these findings, we summarized the direction of future development. In this study, we have dealt with problems such as makeup and facial movements in the lip region. However, our proposal is expected to be applied to a wide range of problems in the entire face. We discuss the potential of the proposed method, which is not limited to the problems in this study.

5.4.1 Quantifying the Original Face

In our proposed method, completely verifying the facial attributes assigned to the original face is difficult. For example, the rough skin has similar color elements to the lip contour. Through that, human can accurately distinguish between such noisy information and original surface. This is because we have already imaged the plausible shape of a face just by taking a glance at it. This can be seen in that in the diagnosis of facial paralysis, doctors sometimes predict the face before the onset of paralysis from the face that has already been paralyzed. Although deep learning has been proposed as a technique for this type of face restoration, restoring the original face to some extent by using facial symmetry seems possible [93].

Thus, if the shape of the original face is known, both the accuracy of quantification and the effort required for use should be reduced. However, it is suggested that this method can be applied to identity verification using the difference between a face and its real face. A report on the danger of makeup impersonation in identity verification based on face images is currently reported. Our technology may be a potential solution to this problem [94].

In relation to the above, the relativity and absoluteness of face assessment should be examined. In this study, in both the makeup support and diagnosis support studies, the former and the latter were normalized by the size of the lips on the image, and by the facial features at rest, respectively. This means that we are applying an evaluation index that varies relative to the face shape of the individual. However, we do not think that the fixed assessment indices such as 2 mm makeup protrusion or 1 cm mouth movement are unnecessary. Therefore, we should consider the relations between the normalization and the assessment results.

5.4.2 Considering the Influence of the Facial Features and Scales

This study did not examine the influence of features other than those in local regions on facial assessment. Already, attempts to incorporate texture features into face evaluators have been proposed [33, 34, 36]. The local binary pattern is one of them, and the symmetry can be easily obtained by left-right comparison using histogram intersection, which is worth considering [24, 25].

A related work reports features that inclease the accracy of facial recognition tasks [95]. Subsequently, we consider that the system keeps an individual's log data, including the facial change transition, might be able to extract the individual preferences [96]. Other research conducts the examination of hidden evaluation criteria of experts [97]. This is a powerful evaluation model method, although it needs to be reconciled with already established diagnostic criteria. Such an optimisation of multiple features, not limited to the proposed features, will bring the proposed system closer to the actual assessment.

With this realization, healthcare applications that make use of the face will be attained in the future. The report that a person's face contains information about his or her physical and mental health has already been indicated in related research [98]. In the field of nursing, where face-to-face health confirmation is important, and in the education field, where mental care is important, this will be one of the ways for staff to quantitatively verify the transition.

5.4.3 Ease of Interface Use Based on 3D Scanning

The proposed interface has a trade-off relation between high accuracy of facial assessment and difficulty of operation. In the proposed interface, there is a trade-off between the high accuracy of face evaluation and the difficulty of operation. This was due to the difficulty in controlling head rotation, which can be solved by manipulation based on 3D facial measurements.

Among the related research, we have focused on the implementation using deep learning [99]. Although the problem with this related technique is that the results are reconstructed close to the average face at this time, the technique of restoring a 2D image to a 3D image can be easily applied to the proposed method herein. On the contrary, proposed quantification tends to be affected by the depth image acquisition capability of the imaging equipment [100]. Therefore, the development of hardware is the most welcome, and future applications will correspondingly have broad use.

References

[1] Nezlek J. Wheeler L. Reis H. T. "Physical attractiveness in social interaction". In: *Journal of Personality and Social Psychology* 38.4 (1980), pp. 604–617.

[2] Dennis B. W. Litner J. A. and P. A. Adamson. "Impact of cosmetic facial surgery on satisfaction with appearance and quality of life. Archives of facial plastic surgery". In: *Archives of facial plastic surgery* 10.2 (2008), 79–83.

[3] McKenna S. P. Siddhi K. McGrouther-D. A. Brown B. C. and A. Bayat. "The hidden cost of skin scars: quality of life after skin scarring". In: *Journal of Plastic, Reconstructive & Aesthetic Surgery* 61.9 (2008), pp. 1049–1058.

[4] A. De Sousa. "Psychological issues in acquired facial trauma". In: *Indian journal of plastic surgery: official publication of the Association of Plastic Surgeons of India* 43.2 (2010), p. 200.

[5] Terada S. Miki Y. Sasagawa C. Hanafusa A. and T. Ikeda. "Makeup support system for visually impaired persons: Overview of system functions". In: *International Conference on Computers for Handicapped Persons (ICCHP)*. 2010, pp. 338–345.

[6] Koike Y.. Togawa K. Mori H. Nishimura-H. Hazama H. Tamachi H. Tomita H. Okuda Y. Kobayashi T. Naitou J. Yanagihara N. and H. Hosomi. "On Standardised Documentation of Facial Palsy". In: *Journal of Otolaryngology of Japan* 80.8 (1997). in Japanese, pp. 799 –805.

[7] J. W. House and D. E. Brackmann. "Facial nerve grading system". In: *Otolaryngology–head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery* 93.2 (1985), 146 – 147.

[8] Lee H. R. Jeong J. H. Kim J. and W. S. Lee. "Features of facial asymmetry following incomplete recovery from facial paralysis". In: *Yonsei medical journal* 51.6 (2010), pp. 943–948.

[9] N. Matsushiro. "Differences in the evaluation of facial palsy (Yanagihara grading system) among 52 ENT doctors at Osaka University". In: *Facial Nerve Research Japan* 29 ().

[10] O-Uchi T. Yoshihara S. Inoue Y.-Ogata A. Mashino H Satoh Y. and J. Kanzaki. "Evaluation of Facial Nerve Function by Yanagihara's 40 Points Grading System and House and Brackmann's Facial Nerve Grading System". In: *Otology Japan* 3.2 (1993). in Japanese, pp. 241–246.

[11] Nakazato-H. Abiko Y. Sugihara M. Ikeda M. and K. Hiroshige. "Facial Paralysis Evaluations Using the Yanagihara Method, the House-Brackmann Method, and Self-evaluation by Patients". In: *Journal of Otolaryngology of Japan* 104.7 (2001). in Japanese, pp. 735–743.

[12] Aoyanagi-Y. Suzuki H. Koike Y. and J. Watanabe. "Early prognosis and diagnosis of peripheral facial nerve palsy". In: ().

[13] A. Ahmed. "When is facial paralysis Bell palsy? Current diagnosis and treatment". In: *Cleveland Clinic journal of medicine* 72.5 (2005), pp. 398–405.

[14] C. Wilkinson and R. Evans. "Are facial image analysis experts any better than the general public at identifying individuals from CCTV images?" In: *Science and justice : journal of the Forensic Science Society* 49.3 (2009).

[15] De Graef P. Verfaillie K. Van Belle G. and Busigny T. "Whole not hole: expert face recognition requires holistic perception". In: *Neuropsychologia* 48.9 (2010), pp. 2620–2629.

[16] Gold J. M. Arcurio L. R. and T. W. James. “The response of face-selective cortex with single face parts and part combinations”. In: *Neuropsychologia* 50.10 (2012), pp. 2454–2459.

[17] Rubin D. M. Gritzman A. D. and A. Pantanowitz. “Comparison of colour transforms used in lip segmentation algorithms”. In: *SIViP* 9 (2015), 947–957.

[18] Kurata K. Tanaka H. Kawamoto M. Isono M and H. Azuma. “An Objective Evaluation Method for Facial Mimic Motion”. In: *Otolaryngology - Head and Neck Surgery* 114.1 (1996), pp. 27–31.

[19] N. N. Oosterhof and A. Todorov. “The functional basis of face evaluation”. In: 105.32 (2008), pp. 11087–11092.

[20] Amico A. Ibáñez-Berganza M. and V. Loreto. “Subjectivity and complexity of facial attractiveness”. In: *Scientific reports* 9.1 (2019), pp. 1–12.

[21] N. Epley and E. Whitchurch. “Mirror, mirror on the wall: Enhancement in self-recognition”. In: *Personality and Social Psychology Bulletin* 34.9 (2008), pp. 1159–1170.

[22] Katic M. J. Farkas L. G. and C. R. Forrest. “International anthropometric study of facial morphology in various ethnic groups/races”. In: *Journal of Craniofacial Surgery* 16.4 (2005), pp. 615–646.

[23] Gaia G. Miriam B. Claudia D. Marcio-D. M. Chiarella S. and F. F. Virgilio. “Age- and sex-related changes in three-dimensional lip morphology”. In: *Forensic Science International* 200.12 ().

[24] M. J. Swain and D. H. Ballard. “Color indexing”. In: *International journal of computer vision* 7.1 (1991), pp. 11–32.

[25] R. Thornhill and S. W. Gangestad. “Human facial beauty”. In: *Human nature* 4.3 (1993), pp. 237–269.

[26] Dusseldorp J. Hadlock T. A. Guarin D. L. and Jowett N. “A Machine Learning Approach for Automated Facial Measurements in Facial Palsy”. In: *Neuropsychologia* 20.4 (2018), pp. 335–337.

[27] Cheng S. Chuang C. F. Shih F. Y. and P. S. P. Wang. “Comparison of colour transforms used in lip segmentation algorithms”. In: *International Journal of Pattern Recognition and Artificial Intelligence* 22.03 (2008), pp. 515–534.

[28] Mauro M. Khan K. and R. Leonardi. “Multi-class semantic segmentation of faces”. In: *2015 IEEE International Conference on Image Processing (ICIP)*. 2015.

[29] Kuraya N. Hirayama M.J. and Y Komachi. “A Dialog Based Speech User Interface of a Makeup Support System for Visually Impaired Persons”. In: *International conference on Human Interface and the Management of Information*. 2010, pp. 261–268.

[30] Zafeiriou S. Cheng S. Asthana A. and M. Pantic. “Incremental Face Alignment in the Wild”. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 2014, pp. 1859–1866.

[31] V. Kazemi and J. Sullivan. “One millisecond face alignment with an ensemble of regression trees”. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 2014, pp. 1867–1874.

[32] Ali N. Dar S. H. Iqbal R. N. Butt A. R. Sajid M., Baig M. J. A. Riaz I. Zafar B. S. T., and S. Baig. “Data augmentation-assisted makeup-invariant face recognition”. In: *Mathematical Problems in Engineering* 2018 (2018).

[33] Pietikainen M. Ojala T. and D. Harwood. “Performance evaluation of texture measures with classification based on Kullback discrimination of distributions”. In: *Proceedings of 12th international conference on pattern recognition*. Vol. 1. IEEE. 1994, pp. 582–585.

[34] Shan C. Ardabilian M. Wang Y. Huang D. and L. Chen. “Local binary patterns and its application to facial image analysis: a survey”. In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 41.6 (2011), pp. 765–781.

[35] Seo M. Ngo T. H., Y. W. Chen, and N. Matsushiro. “Quantitative assessment of facial paralysis using local binary patterns and Gabor filters”. In: *Proceedings of the Fifth Symposium on Information and Communication Technology*. 2014, pp. 155–161.

[36] Zhang S. Dong J. Liu L. Wang T. and H. Yu. “Automatic evaluation of the degree of facial nerve paralysis”. In: *Multimedia Tools and Applications* 75.19 (2016), pp. 1573–7721.

[37] Chen Y. W. Matsushiro N. Ngo T. H. and M. Seo. “Quantitative Assessment of Facial Paralysis Based on Spatiotemporal Features”. In: *IEICE Transactions on Information and Systems* E99.D.1 (2016), pp. 187–196.

[38] L. Yuanyao and L. Qingqing. “Lip segmentation using automatic selected initial contours based on localized active contour model”. In: *EURASIP Journal on Image and Video Processing* 2018.7 (2018).

[39] P. Ekman and W. V. Friesen. “Measuring facial movement”. In: *Environmental psychology and nonverbal behavior* 1.1 (1976), pp. 56–75.

[40] Minnigerode L Volk G. F. Denzler J. Haase D. and O. Guntinas-Lichius. “Automated and objective action coding of facial expressions in patients with acute facial palsy”. In: *European Archives of Oto-Rhino-Laryngology* 272.5 (2015), pp. 1259–1267.

[41] C. J. Linstrom. “Objective facial motion analysis in patients with facial nerve dysfunction”. In: *The Laryngoscope* 112.7 Pt 1 (2002), 1129–1147.

[42] T. Bailey and A. Jain. “A Note on Distance-Weighted k-Nearest Neighbor Rules”. In: *IEEE Transactions on Systems, Man, and Cybernetics* 8.4 (1978), pp. 311–313.

[43] V. Vapnik. *The nature of statistical learning theory*. Springer science & business media.

[44] Wichmann F. A. Bülthoff H. H. Graf A. B. and B. Schölkopf. “Classification of faces in man and machine”. In: *Neural Computation* 18.1 (2006), pp. 143–165.

[45] Dror G. Eisenthal Y. and E. Ruppin. “Facial attractiveness: beauty and the machine”. In: *Neural computation* 18.1 (2006), pp. 119–142.

[46] Dror G. Leyvand T. Meilijson I. Cohen-Or D. Kagian A. and E. Ruppin. “A machine learning predictor of facial attractiveness revealing human-like psychophysical biases”. In: *Vision research* 48.2 (2008), pp. 235–243.

[47] Xu H. Xing J. Liu S. Zhou-X. Liu L. and S. Yan. “Wow You Are so Beautiful Today”. In: 2013.

[48] Jiang S. Wang S. Alashkar T. and Y. Fu. “Rule-Based Facial Makeup Recommendation System”. In: *2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017)*. 2017, pp. 325–330.

[49] Jiang S. Wang S. Alashkar T. and Y. Fu. “Examples-Rules Guided Deep Neural Network for Makeup Recommendation”. In: 2017.

[50] Shafique T. Baig M. J. A. Riaz-I. Amin S. Sajid M. and S. Manzoor. “Automatic grading of palsy using asymmetrical facial features: A study complemented by new solutions”. In: *Symmetry* 10.7 (2018), p. 242.

[51] Seo M. Ngo T. H. Matsushiro N. Yoshihara H. and Y. W. Chen. “Automatic feature point detection using deep convolutional networks for quantitative evaluation of facial paralysis”. In: *2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. 2016, pp. 811–814.

[52] Zhang S. Liu L. Wu G. Wang T. and J. Dong. “Automatic Facial Paralysis Evaluation Augmented by a Cascaded Encoder Network Structure”. In: *IEEE Access* 7 (2019), pp. 135621–135631.

[53] *Aira*. URL: https://aira.io/.

[54] *Be My Eyes*. URL: https://www.bemyeyes.com/.

[55] *Tap Tap See*. URL: https://taptapseeapp.com/.

[56] *Viz Wiz*. URL: https://vizwiz.org/.

[57] Li Q. Lin C. Zhao Y. Guo-A. Stangl A. Gurari D. and J.P Bigham. “VizWiz-Priv: A Dataset for Recognizing the Presence and Purpose of Private Visual Information in Images Taken by Blind People”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019, pp. 939–948.

[58] Little G. Black J. Krishna S. and S. Panchanathan. “A Wearable Face Recognition System for Individuals with Visual Impairments”. In: 2005.

[59] Sato D. Oh U. Ishihara T. Kitani-K. Ahmetovic D. and C. Asakawa. “ReCog: Supporting Blind People in Recognizing Personal Objects”. In: 2020.

[60] J. Ishikiriyama and K. Suzuki. “An interactive virtual mirror to support makeup for visually impaired persons”. In: *International Conference on Systems, Man, and Cybernetics (SMC)*. 2017, pp. 1393–1398.

[61] Seo J. M. Park M. J. and K. S. Park. “PC-based asymmetry analyzer for facial palsy study in uncontrolled environment: A preliminary study”. In: *Computer Methods and Programs in Biomedicine* 99.1 (2010), pp. 57 –65.

[62] Feng G. D. Wu H. Y. Zhuang-Y. Zhao Y. and Z. Q. Gao. “Measuring facial paralysis using the three-dimensional dynamic quantitative analysis system of facial motion: correlation with subjective grading systems”. In: *Chinese journal of otorhinolaryngology head and neck surgery* 51.7 (2016). in Chinese, 511 – 516.

[63] Lee K. Lee S. Lodhi B. Cho-J. G. Seo W. K. Barbosa J. and J. Kang. “Efficient quantitative assessment of facial paralysis using iris segmentation and active contour-based key points detection with hybrid classifier”. In: *BMC medical imaging* 16.23 (2016), 511 – 516.

[64] Huang W. Hsu G. J. and J. Kang. “Hierarchical Network for Facial Palsy Detection”. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2018, pp. 693–6936.

[65] Bevacqua E. Urbain J., Moinet A. Niewiadomski R. Pelachaud C. Picart-B. Tilmanne J. Dutoit T., and J. Wagner. “The AVLaughterCycle Database”. In: Sept. 2010.

[66] *OptiTrack Expression*. URL: https://optitrack.com/.

[67] Yu-Hsin C. Jing L. and F. Xiaolan. “How Fast are the Leaked Facial Expressions: The Duration of Micro-Expressions”. In: *Journal of Nonverbal Behavior* 37.4 (2013), pp. 217 –230.

[68] Richmond-S. Zhurov A. I. Rosin P. L. Popat H. and D. Marshall. “A geometric morphometric approach to the analysis of lip shape during speech: development of a clinical outcome measure”. In: ().

[69] Bhama P. K. Park J. Hadlock C. R. Banks C. A. and T. A. Hadlock. “Clinician-graded electronic facial paralysis assessment the eFACE”. In: *Plastic and reconstructive surgery* 136.2 (2015), 223e–230e.

[70] Kim S. Y. Kim Y. H. Kim H. S. and K. S. Park. “A smartphone-based automatic diagnosis system for facial nerve palsy”. In: *Sensors* 15.10 (2015), pp. 26756 –26768.

[71] Apple Computer Vision Machine Learning Team. “An On-device Deep Neural Network for Face Detection”. In: *Apple Machine Learning Journal* 1.7 (2017).

[72] Chong E. Ruiz N. and M. R. James. “Fine-Grained Head Pose Estimation Without Keypoints”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. 2018, pp. 2074–2083.

[73] Lei Z. Liu X. Shi H. Zhu X. and S. Li. “Face alignment across large poses: A 3d solution”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 146–155.

[74] Sharlet D. Geiss R. Adams A. Barron J.T. Kainz F. Chen J. Hasinoff S. and M. Levoy. “Burst photography for high dynamic range and low-light imaging on mobile cameras”. In: *ACM Transactions on Graphics (TOG)* 35.6:192 (2016).

[75] A. Bulat and G. Tzimiropoulos. “How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks”. In: *The IEEE International Conference on Computer Vision (ICCV)*. 2017, pp. 1021–1030.

[76] F. T. Chan and A. L. Vese. “Active Contours Without Edges”. In: *The IEEE Transactions on Image Processing* 10.2 (2001), pp. 266–277.

[77] S. H. Richard. In: *J. Opt. Soc. Am.* 12 (), pp. 985–995.

[78] *OpenCV*. URL: `https://opencv.org/`.

[79] *Python*. URL: `https://www.python.org/`.

[80] J. Brooke. *SUS: A “quick and dirty” usability scale*. sability Evaluation in Industry.

[81] J. Brooke. “SUS: A Retrospective”. In: *EURASIP Journal on Image and Video Processing* 8.2 (2013), pp. 29–40.

[82] T. Tarumoto. *Usability Survey Methodology*. in Japanese. Ohmsha, 2005.

[83] Z. Karel. “*Contrast Limited Adaptive Histograph Equalization.*” *Graphic Gems IV*. San Diego: Academic Press Professional, 1994, 474–485.

[84] Uijlings J. R. R. Gevers T. van de Sande K. E. A. and A. W. M. Smeulders. “Segmentation as selective search for object recognition”. In: *2011 International Conference on Computer Vision*. 2011, pp. 1879–1886.

[85] *Dlib C++ Library*. URL: `http://dlib.net/`.

[86] R. Brunelli. *Template Matching Techniques in Computer Vision: Theory and Practice*. Wiley Publishing.

[87] S. Baker and T. Kanade. “Limits on super-resolution and how to break them”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24.9 (2002), pp. 1167–1183.

[88] B. Simon and M. Iain. “Lucas-Kanade 20 Years On: A Unifying Framework”. In: *International Journal of Computer Vision* 56.3 (2004), pp. 221–255.

[89] Stefano M. Luigi D. and T. Federico. “ZNCC-based template matching using bounded partial correlation”. In: *Pattern Recognition Letters* 26.14 (2005), pp. 2129 –2134.

[90] *scikit-learn*. URL: `https://scikit-learn.org/`.

[91] T. Nishida. *Quantitatice Scoring of Facial Nerve Paralysis with Principal Component Analysis based on the Yanagihara Grading System*. Master’s Thesis, Nara Institute of Science and Technology. 2012.

[92] WernickRobinson M. Mehta R. P. and T. A. Hadlock. “Validation of the Synkinesis Assessment Questionnaire.” In: *Personality and Social Psychology Bulletin* 117.5 (2007), 923–926.

[93] Wang G. Chiu M. T. Tai Y. Gu Q. and C. Tang. “LADN: Local Adversarial Disentangling Network for Facial Makeup and De-Makeup”. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2019, pp. 10480–10489.

[94] Dantcheva A. Swearingen T. Chen C. and A. Ross. “Spoofing faces using makeup: An investigative study”. In: *IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*. Feb. 2017, pp. 1–8.

[95] L. S. Jeffrey F. C. Yanxi L. Karen and Sinjini M. “Facial asymmetry quantification for expression invariant human identification”. In: *Computer Vision and Image Understanding* 91.1 (2003). Special Issue on Face Recognition, pp. 138–159.

[96] Sakamoto D. Koyama Y. and T. Igarashi. “SelPh: Progressive Learning and Support of Manual Photo Color Enhancement”. In: 2016.

[97] Friedman J. H. Ungar L. H. Pirracchio R. Eaton E. Reichmann L. G. Interian Y. Luna J. M. Simone C. B. 2nd Auerbach A. Delgado E. van der Laan M. J. Solberg T. D. Gennatas E. D. and G. Valdes. “Expert-augmented machine learning”. In: *Proceedings of the National Academy of Sciences of the United States of America*. Vol. 117. 9. 2020, 4571–4577.

[98] Ruocco A. C. Daros A. R. and N. O. Rule. “Identifying mental disorder from the faces of women with borderline personality disorder”. In: *Journal of Nonverbal Behavior* 40.4 (2016), pp. 255–281.

[99] Roussos A. Chrysos G. Ververas E. Kotsia I. Shen J. Deng J. and S. Zafeiriou. “The menpo benchmark for multi-pose 2d and 3d facial landmark localisation and tracking”. In: *International Journal of Computer Vision* 127.6-7 (2019), pp. 599–624.

[100] Edler R. Haers P. Wertheim D. Ghoddousi H. and D. Greenhill. “Comparison of three methods of facial measurement”. In: *International Journal of Oral and Maxillofacial Surgery* 36.3 (2007), pp. 250–258.