applied sciences

MDPI

Article

Detection of Emotion Using Multi-Block Deep Learning in a Self-Management Interview App

Dong Hoon Shin [1], Kyungyong Chung [2] and Roy C. Park [3,*]

[1] Department of Computer Science, Kyonggi University, Suwon 16227, Korea; dhshin8222@gmail.com
[2] Division of Computer Science and Engineering, Kyonggi University, Suwon 16227, Korea; dragonhci@gmail.com
[3] Department of Information Communication Engineering, Sangji University, Wonju 26339, Korea
* Correspondence: roypark@sangji.ac.kr

Received: 6 October 2019; Accepted: 7 November 2019; Published: 11 November 2019

Abstract: Recently, domestic universities have constructed and operated online mock interview systems for students' preparation for employment. Students can have a mock interview anywhere and at any time through the online mock interview system, and can improve any problems during the interviews via images stored in real time. For such practice, it is necessary to analyze the emotional state of the student based on the situation, and to provide coaching through accurate analysis of the interview. In this paper, we propose detection of user emotions using multi-block deep learning in a self-management interview application. Unlike the basic structure for learning about whole-face images, the multi-block deep learning method helps the user learn after sampling the core facial areas (eyes, nose, mouth, etc.), which are important factors for emotion analysis from face detection. Through the multi-block process, sampling is carried out using multiple AdaBoost learning. For optimal block image screening and verification, similarity measurement is also performed during this process. A performance evaluation of the proposed model compares the proposed system with AlexNet, which has mainly been used for facial recognition in the past. As comparison items, the recognition rate and extraction time of the specific area are compared. The extraction time of the specific area decreased by 2.61%, and the recognition rate increased by 3.75%, indicating that the proposed facial recognition method is excellent. It is expected to provide good-quality, customized interview education for job seekers by establishing a systematic interview system using the proposed deep learning method.

Keywords: self-management interview application; emotion analysis; facial recognition; image-mining; deep convolutional neural network

1. Introduction

Recently, Korea's youth unemployment rate has been high, to the extent that people aged 30 to 40 constitute more than half (56.7%) of the highly educated but economically inactive population (i.e., they cannot find good jobs). Accordingly, the severity of social waste through unemployment is increasing [1]. According to one analysis, many highly educated people who could have high-level careers have been produced through university education, but they are unable to find a good position right after graduation because they graduate without an appropriate interview clinic and without information and coaching on practical employment skills [2]. A situation in which they cannot work full time is problematic in job-seeking, and a common problem is that they have a lot of idle time as they go to graduate school, work as a freelancer, or work part-time due to parenting duties, despite their ability to hold down a good job. In addition, the hardest part when a job applicant seeks employment is preparing for the interviews [3]. In particular, since there are no set answers for

an interview, interviewees must exhibit their capabilities differently, depending on their individual living environment and values. Also, since there are big differences among individuals (e.g., posture, eye contact, unhelpful language habits), one-on-one consulting is required, and content is needed by which students can practice their interview techniques anywhere and at any time (e.g., the night before the interview, in the train when going to an interview, and in the waiting room before the interview). The demand for online interview content is increasing, which allows a last check briefly prior to the interview [4]. Facial recognition systems can be broadly divided into face area detection and facial recognition. Face area detection determines the position of the face, size, posture, etc., in the video, and helps create a certain image for facial recognition [5,6]. Types of face detection include (1) the knowledge-based method that uses information about the typical face, (2) the feature-based method that looks for easily detected characteristics, despite changes in posture or lighting, (3) the template-matching method, which stores the basic shape of a few faces and performs a comparison with the input images, and (4) the appearance-based method, learning the face model from training images representative of the diversity in faces [7–9]. As a study for facial recognition, algorithms such as Haar, scale invariant feature transform (SIFT), ferns, modified census transform (MCT), histogram of oriented gradients (HOG), etc., are used to extract the feature factors of an image, and face analysis is actively performed based on them [10,11]. Recently, deep learning-based facial recognition has also been widely used, and a method of automatically extracting feature factors using a convolutional filter based on a convolutional neural network (CNN) has been used [12–14]. When a face is recognized using a specific factor, it is difficult to extract and select an optimal specific factor, depending on the original image state and application, and it is also difficult to determine a feature factor through various experimental and empirical factors.

We developed a self-management interview system and conducted a study on deep learning-based face analysis for emotion extraction to provide accurate interview services. Unlike the basic structure for learning the whole-face image, in this paper, a deep convolutional neural network (DCNN) method [15] for image analysis through a multi-block process helps the user learn after sampling the core facial areas, which is important for emotion analysis during face detection. The system proposed in this paper is expected to contribute to the creation of job opportunities by providing customized interview education that enables efficient interview management that is not constrained by space and time, and that provides an appropriate level of interview coaching. The figure included in this image is the author, who agreed to provide the figure.

The study is organized as follows. Section 2 describes the research related to facial recognition-based application services, and technology using facial recognition. Section 3 describes detection of user emotion using multi-block deep learning in the self-management interview application. For that purpose, also described is the image multi-block process to extract the face's feature points, plus multi-block selection and extraction of main features, and a proposed experiment with the deep learning process in face detection. Section 4 describes the proposed mobile service for real-time interview management, and Section 5 provides a conclusion.

2. Related Research

2.1. Facial Recognition-Based Application Services

Recently, various services based on facial recognition have been provided. The face recognition process is as follows. A camera captures a face image. Then, the eyes, eyebrows, nose, and mouth, which are the main factors for emotion extraction, are analyzed to extract characteristics data, and they are compared with feature data in a database provided for face analysis in facial recognition. The facial recognition technology analyzes facial expressions to determine emotional states, such as happiness, surprise, sadness, disgust, fear, confusion, etc., and is used for advertising-effect measurement, marketing, and education [16]. Founded by the Massachusetts Institute of Technology Media Lab, Affectiva released Affdex, a solution for recognizing facial expressions and identifying emotional

states [17]. Figure 1 shows the Affectiva facial recognition platform. Affectiva modularizes emotion recognition-related artificial intelligence (AI) technology, distributes it through its website in the form of a software development kit (SDK), and opens it for use by various engineers and in business fields. It applies an emotion recognition solution to Tega, a robot that teaches foreign languages, and presents functions that provide appropriate content and gives rewards by understanding children's facial expressions. In addition, the facial recognition technology has been widely applied to various fields, such as locating crime suspects and lost children, and enabling mobile payments, in particular. It is used to arrest criminal suspects and locate lost children through artificial intelligence cameras attached on the chest, based on an agreement with US police [18].

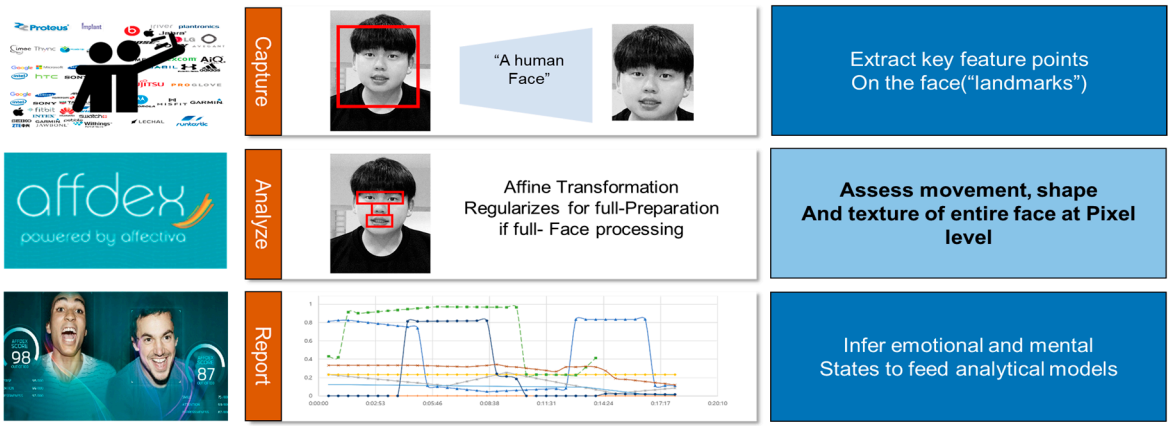


Figure 1. Facial recognition platform with respect to Affectiva [17]. * The figure included in this image is the author, who agreed to provide the figure.

2.2. Technology Using Facial Recognition

The AdaBoost algorithm is a technology often used for face detection, and is a method for creating strong criteria for selection by combining weak criteria, which has advantages [19,20]. This reduces the probability of drawing wrong conclusions, and increases the probability of accurately assessing problems that are difficult to judge. For facial recognition, a face area image is required. In order to increase the success rate with face detection and facial recognition, the impacts of lighting and inclination should be minimized, and images should be normalized. So, as images are normalized, the probability of errors decreases [21]. Video-based emotion recognition analyzes the characteristics of the face in a video. At first, this study used classic machine learning and computer vision. For example, the characteristics of the face were extracted based on the gradients of the face extracted from video. The characteristics were analyzed, using algorithms like a support vector machine (SVM) or random forest, to figure out the facial expression. And yet, there is a singularity effect according to the surrounding background or the illumination intensity of the video. In addition, accuracy is greatly affected by the angle of the face. Figure 2 shows a facial recognition algorithm using a face database. The dataset used in the early stages was secured in a limited environment; however, videos that contain everyday situations are in the dataset [22].

Figure 3 shows how to recognize a face in a three-dimensional (3D) image [23,24]. The flow of the method can be separated from the training phase and the test phase. In the training phase, face data are collected from 3D images, and pretreatment is performed to obtain a clean 3D face without bending. Preprocessed data include facial features from a feature extraction system [25]. The features, such as extracted face data, are stored in a feature database [26,27]. Next, in the test phase, the entered target faces are the same as those used in the training phase and during the 3D face data collection, pretreatment, and feature-extraction steps. In the feature-matching phase, match scores are calculated by comparing the target face with the database saved in the training phase. When the match score is judged to be high enough, this algorithm determines that the target face has been recognized. Also,

facial recognition technology is used in the augmented reality field. It is a method to extract feature points, and draws the coordinate plane of the face to calculate the position, to make a 3D effect of the product, and overlap it onto the face [28,29].

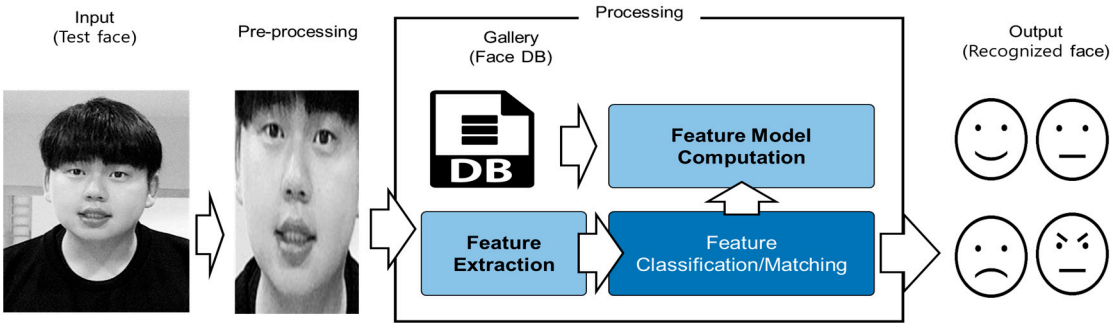


Figure 2. Facial recognition algorithm using a face database. * The figure included in this image is the author, who agreed to provide the figure.

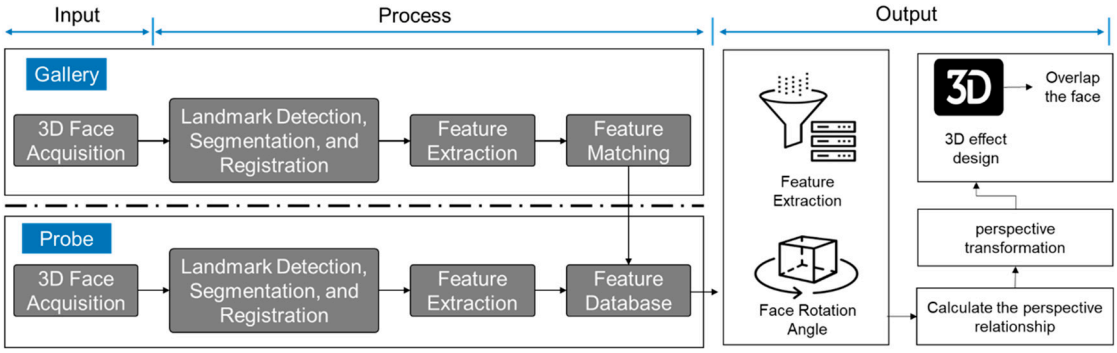


Figure 3. A 3D facial recognition of augmented reality system.

3. Detection of Emotions Using Multi-Block Deep Learning in Self-Management Interviews

Figure 4 shows the whole process of the system described in this paper. First, we proceed with facial recognition, where features are extracted. Multi-block sampling is performed by extracting feature points from the recognized faces. Sampled data are extracted through deep learning based on a DCNN. Analysis is conducted based on the extracted emotions, and the analyzed data are managed by the interview system proposed in this paper. Interview management is done through the application itself. The CAS-PEAL face database is used for facial recognition, and the Cohn-Kanade database is used for emotion extraction.

3.1. Image Multi-Block Process for Face Main Point Extraction

In this paper, we developed a self-management interview system and conducted a study on deep learning-based face analysis for emotion extraction to provide accurate interview services. Unlike the basic structure for learning the whole-face image, the deep learning method proposed in this paper is a model that helps the user learn from images of multi-block core areas, such as the eyes, nose, and mouth, which are important factors for emotion analysis during face detection. The proposed learning structure of the DCNN consists of a multi-block process of entered face images and multi-block deep learning. In the multi-block process, the input image is blocked based on multiple AdaBoost. The multi-block deep learning model is executed by considering the sizes of the original image and of the sampled image that is blocked for area extraction. When both processes are completed, the whole face image and the sampled multi-block image have been learned, making it possible to use them during the emotion detection stage afterwards. The recognition process of the multi-block deep learning algorithm consists of a multi-block process, multi-block selection, and a multi-block deep

performance process. In the existing deep learning model, facial recognition utilizes the whole face, which causes a problem in that areas such as the eyes, nose, and mouth (the key factors for analyzing emotions) are not recognized correctly. In this paper, therefore, the recognition rate was improved by extracting the specific parts of a face image required for emotion extraction by the multi-block method. In particular, if the multi-block is large or small in the blocking process, features of the main areas cannot be extracted accurately, which causes large errors in recognition and learning.

In this paper, multiple AdaBoost was used to carry out sampling by setting the optimal blocking. Figure 5 shows the process of detection and classification with multiple AdaBoost. Multiple AdaBoost creates a stronger classifier by combining weak classifiers, allows a weak classifier to determine whether the image is a face or not when there is a certain purpose. It is designed to select a feature of a rectangular shape with the fewest errors in order to let a weak classifier use the fewest improperly classified training videos, and, in turn, have the optimum threshold classification function.

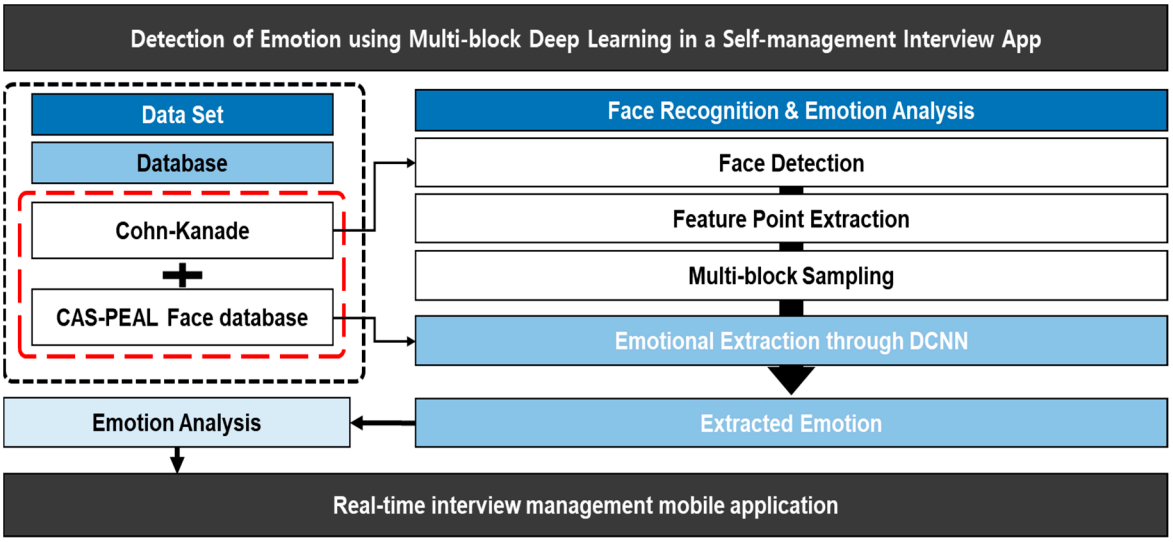


Figure 4. System-wide process for interview management.

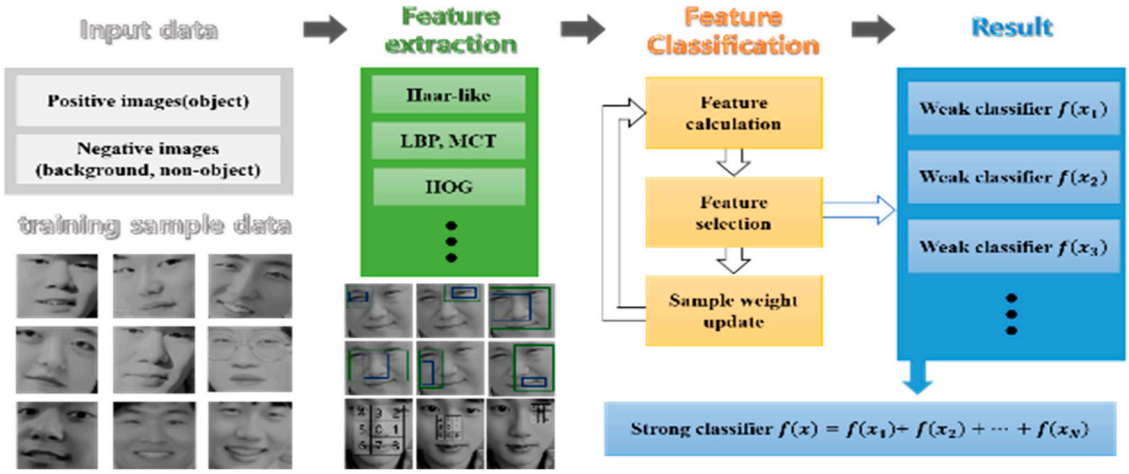


Figure 5. The process of detection and classification with multiple AdaBoost. * The figure included in this image is the author, who agreed to provide the figure.

For this process, training images and sample images were required, so by using the CAS-PEAL face database, our database included 99,594 images with a variety of poses, expressions, and lighting levels from 1040 individuals (595 male and 445 female). Domains of faces to be extracted were defined as positive (object) samples, while images other than a face were defined as negative (non-object, background) samples. Also, we use the Cohn-Kanade database to analyze perceived facial emotions

from data in this database that include 486 sequences from 97 poses [30,31]. At this time, positive images must have pixels of the same size, and detection should be made by aligning the positions of eyes, noses, and mouths so they are the same as much as possible. Learning data should include information on whether the image belongs in the positive or negative category. In addition, features for distinguishing a face from the background are also required. Such features could be presented as a classifier to distinguish/classify an object. Since these features are a base classifier and a candidate for a weak classifier, it was necessary to decide how many times the process of weak classifier selection should be repeated [32,33]. In other words, it was necessary to determine how many weak classifiers should be combined into one stronger classifier, and to select one feature having the best performance in classifying training samples by class and to calculate a weak classifier for the corresponding iteration [34]. Therefore, we used a weighted linear combination of T weak classifiers, as shown in Equation (1).

$$\mathrm{E}(x) = a_1 e_1(x) + \ldots + a_T e_T(x) = \sum_{t=1}^{T} a_t e_t(x) \quad (1)$$

E: final strong classifier,
e: weak classifier,
a: weighted of weak classifier,
t: iteration round (1,2, ... ,T).

3.2. Multi-Block Selection and Extraction of Main Area Features

The multi-block selection process selects blocks to be used for actual recognition among the multi-blocks previously delivered through the feature numerical analysis. For accurate emotion analysis, the user's eyes, nose, and mouth, which are the main feature points, should be clearly identified, and they can be classified according to the degree of rotation of the face. If there is no information on specific points in the whole image, the rotation information should be detected during the multi-block process. Figure 6 shows the whole facial recognition and emotion analysis process.

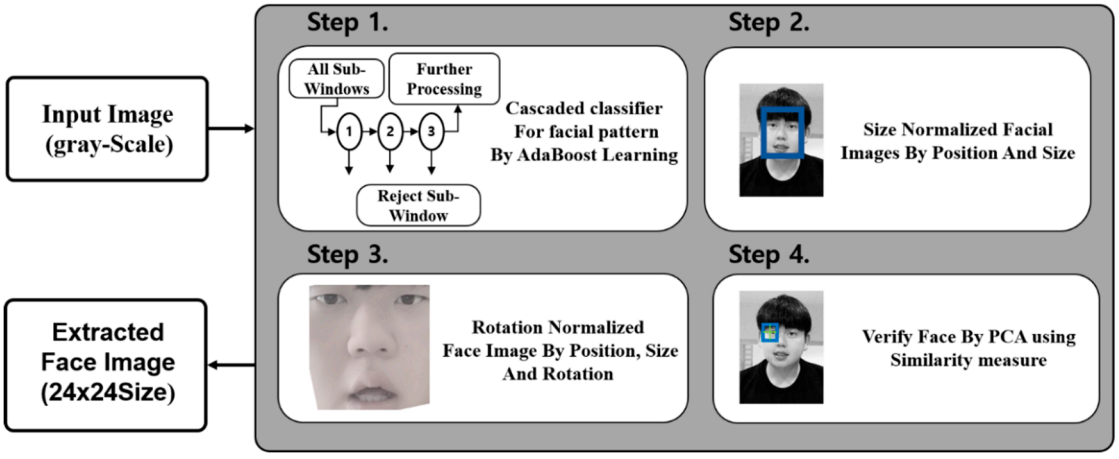


Figure 6. The whole facial recognition and emotion analysis process. * The figure included in this image is the author, who agreed to provide the figure.

Face detection was made by moving a 24 × 24 pixel block; for simple patterns in multiple AdaBoost learning, basic patterns were used. In addition, the number of simple detectors to be searched by the learning process was selected as 160, and the learned detectors became serialized, in turn enhancing the processing speed. The learned detectors were serialized into 10 stages in which 16 learned detectors belong in an arbitrary manner. Parameters for each stage were adjusted, and as for images in multiple AdaBoost learning, 24 × 24 resolution was used. For detection by size, the input images were classified

(based on the degree of down-sampling) into three types, and the face was detected from among the down-sampled images. In detection by rotation, facial images rotated at −5° to +5°, +15° to +25°, and −15° to −25° were learned by AdaBoost. Then, by using the serialized detector, they were each analyzed and, in order of detection, rotation of the face was classified. The detected faces were classified into nine types, and the information about the locations of the detected faces was provided as well. Figure 7 shows the face image–detection process. For face detection, 80 × 60 down-sampled images were used for detecting a large face, 108 × 81 down-sampled images were used for a medium-sized face, and 144 × 108 down-sampled images for a small face. The sequence of detection by size was selected to enhance the detection speed and was done as follows: Detection of 80 × 60 down-sampled images was first, followed by the 108 × 81 down-sampled images, and then, the 144 × 108 images. If a detected face overlapped the block detected in the face from the down-sampled image in the preceding step, that detection was not valid. The input image was searched for among the down-sampled images, and when it was detected, a block of the face from the detected image was cut and then normalized to the predesigned size and passed to the next process. At the time, principal component analysis (PCA) was used to measure similarity with the input image, verifying the face. This is a process of rotating an image by using the verified information on rotation of the face, until the rotation of the face in the image becomes almost zero. When the rotation of the face was verified to be between +15° and +25°, the face was rotated by −20°, and when the rotation of the face was between −15° and −25°, the face was rotated by +20°.

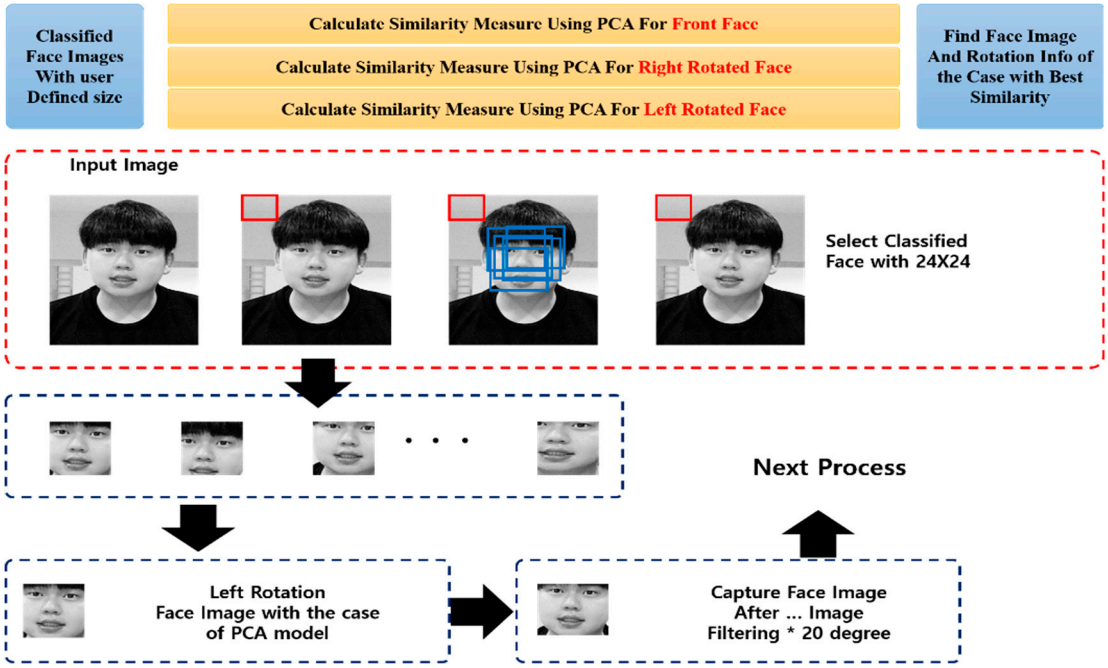


Figure 7. The face image detection process. * The figure included in this image is the author, who agreed to provide the figure.

When the user's face was extracted in the aforementioned process, the positions of the eyes and nose should be extracted. The patterns for the person's eyes could be extracted by using the facial image obtained through the face detection process. Eyes and nose extraction can be classified into three stages. The first stage was to designate a region to search for the eyes in the facial image obtained by the face detector. From this stage, it was possible to roughly estimate the position of the eyes, even if they were not precise, and such an estimated position could be defined as a certain domain. In the second stage, the region for the eyes must be clearly defined, as shown in Figure 8 After defining the eyes region, we used multiple AdaBoost to determine 12 × 12 pixel eye images and 12 × 12 pixel non-eye images to prepare the serialized eye detector. This was to classify these eyes from other eyes.

Then, AdaBoost went through a process of detecting block images that had the eyes in the designated region. The last stage was to use PCA, trained with eye images, to measure the similarity of each eye image and to select the image with the highest similarity. As shown in Figure 8, the position of the eyes could be defined as the center point of a verified eyes image.

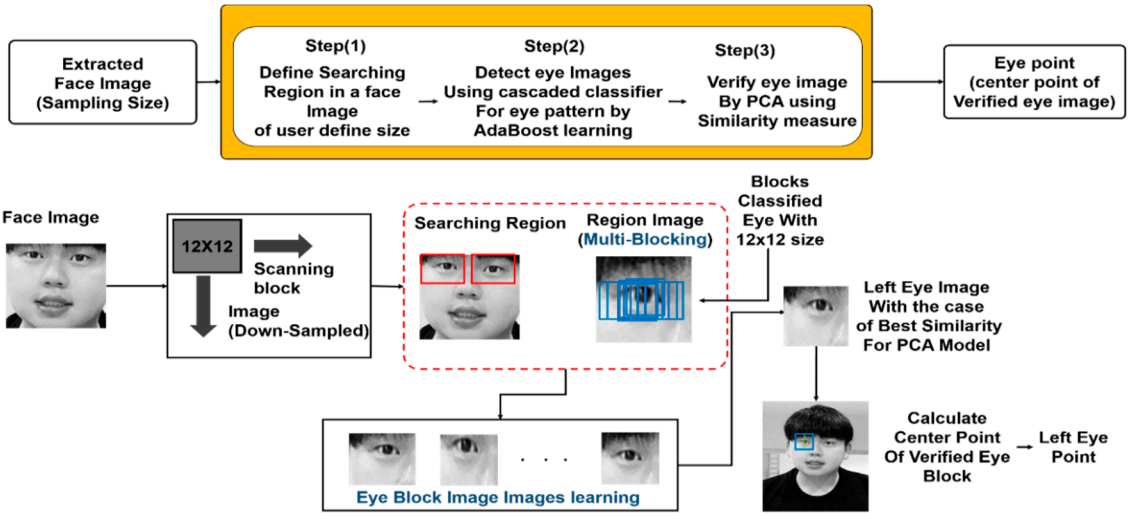


Figure 8. The eye detection process. * The figure included in this image is the author, who agreed to provide the figure.

In order to detect a nose's location, it was necessary to designate a nose search region on the face image acquired during the face search, which is the same process as required for the eye search. Although the exact location of the nose cannot be specified, a rough location can be estimated, and the predictive value of the location of the nose can be defined for certain regions. Figure 9 shows the nose detection process.

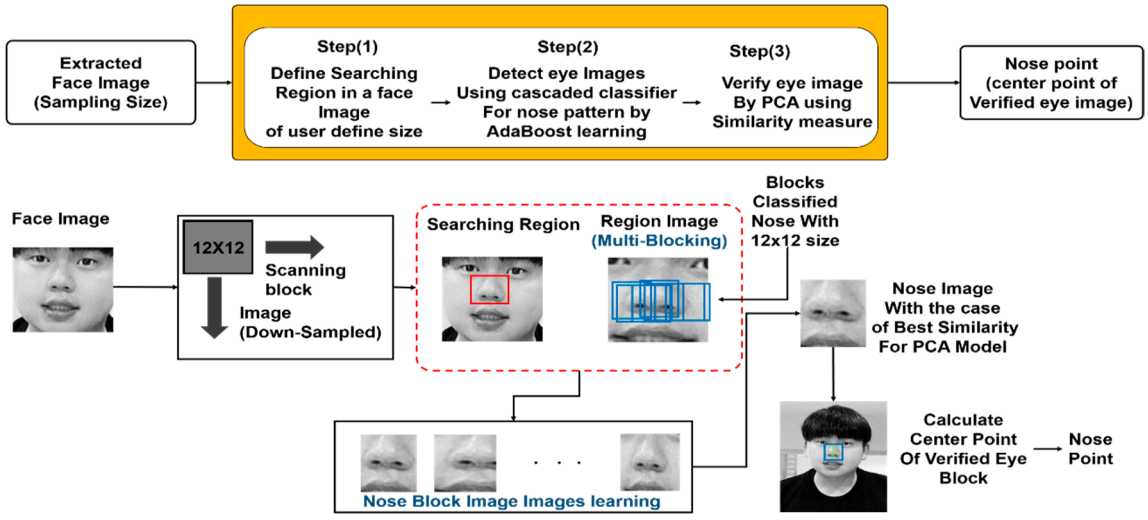


Figure 9. The nose detection process. * The figure included in this image is the author, who agreed to provide the figure.

In order to determine whether the image was actually a nose or not, multiple AdaBoost was used to learn 12 × 12 nose images and non-nose images in order to create a serialized nose detector and go through the process of finding the block images that were detected as noses within the defined region in the first step. The last step was to calculate the similarity of the nose images acquired during the second step, and compare nose images to find one with the best similarity. As with the eye image

search process, the nose location was the center point of a verified nose image. After the detection of eyes and nose locations, the face normalization process was followed. Face normalization is a process to calculate the accurate location, size, and rotation of the face using the locations of both the eyes and the nose. The face image was warped to ensure consistency among the different forms of a face. The actual emotion images can be created by finding the eyes and the nose in the image and by going through image warping based on that information. At this time, the size of the normalized images and the location of each part may vary depending on the design of the recognizer.

3.3. Face Detection Using the Multi-Block Deep Learning Process

For a deep learning model of the proposed user emotion extraction, this experiment extracted emotions using a DCNN based on the multi-blocked sample images of the major face areas, and the images with completed feature extractions, which was intended to minimize the performance time from entry, and the classification of the images. Figure 10 shows the multi-block deep learning structure proposed in this study. The emotion model was extracted by delivering a block target that included information about the features from the images learned by the DCNN in the multi-block and block selection stages. A convolution operation was conducted between the original images and the multi-block images extracted by sampling. This brought into relief the features of the major face areas for the extraction of emotions through the feature extraction filter. The filter coefficient for the feature extraction filter was set to a random value in the early stages, and was then set to the optimal filter coefficient with the least error rate through learning. Next, the process of reducing the images was executed, analyzing the features of the extracted images, and filtering the optimum features. At this time, the general DCNN launched a method for minimizing the cross-entropy loss function so it can be similar to the softmax result from image data entered from the multi-block and feature extraction stages.

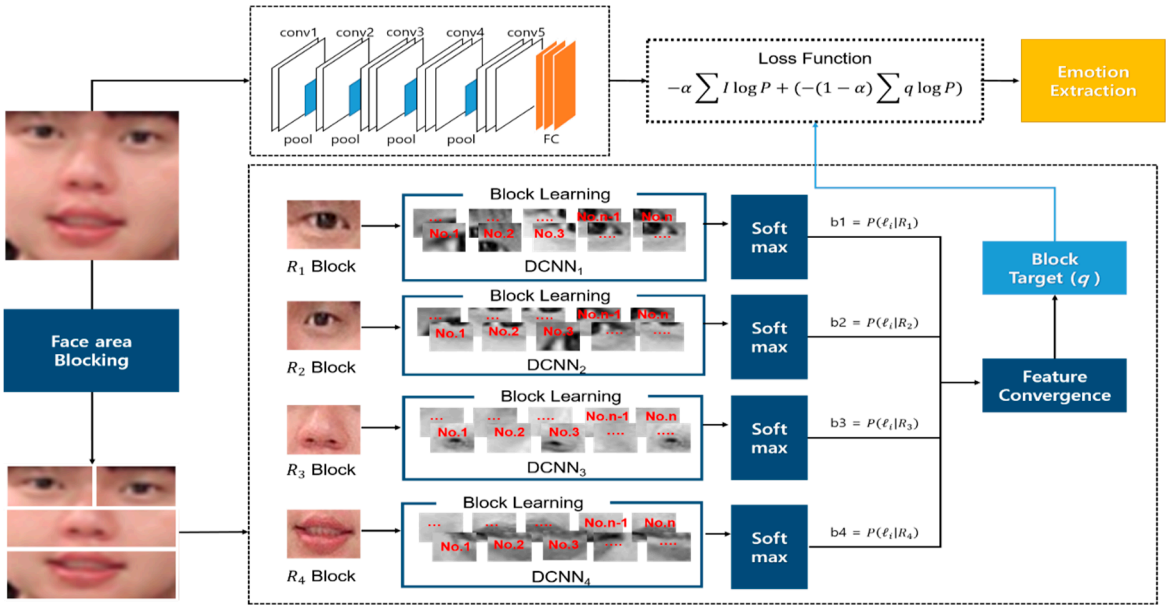


Figure 10. The proposed multi-block deep learning structure. * The figure included in this image is the author, who agreed to provide the figure.

This study defines two cross-entropy loss functions like those in Equations (2) and (3) to deliver knowledge: In Equation (2), loss function L_1 is a cross-entropy function based on a recognition result error for the label. In Equation (3), loss function L_2 is the cross-entropy function representing the error with the block target representing the predicted probability value of the DCNN.

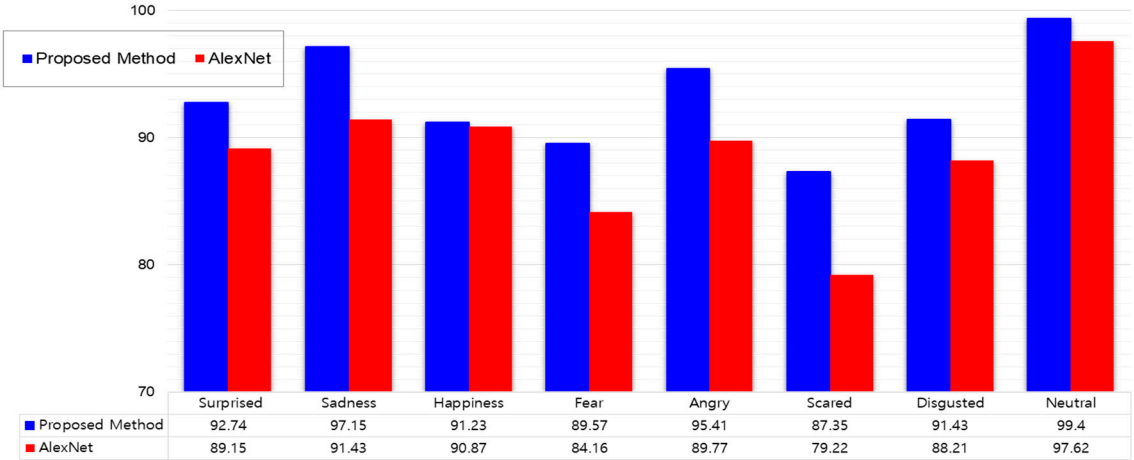
$$L_1 = -\sum_{n=1}^{|V|} H(y=n) \times logP(y=n|x;\ \theta) \tag{2}$$

$$L_2 = -\sum_{n=1}^{|V|} q(y=n|x;\ \theta_E) \times logP(y=n|x;\ \theta) \tag{3}$$

In the formula, q is the softmax probability value formed by learning the features of multi-blocked images, while $P(y=n|x;\theta)$ is the probability the DCNN learned by utilizing the features of the whole images, n is the index of the feature category, and $|v|$ is the total number of classes. This study used both the knowledge block target containing the feature information of the multi-block images delivered by the DCNN while learning, and the existing true class target value so as to allow learning everything. It extracted accurate emotions, utilizing feature extraction of the whole image area and the features of the blocked images of the key areas, giving a different weighted value to each of the two loss functions, L_1 and L_2, as seen in Equation (4):

$$L = \alpha \times L_1 + (1-\alpha) \times L_2,\ 0 < \alpha < 1 \tag{4}$$

In addition, for face area detection and estimation analysis, the CAS-PEAL face database was employed. The learning data in the database used consisted of classes of facial expression information for a total of 1040 persons, and consisted of a total of 1240 sheets of images for each class. The data for deep learning was composed of 10 sheets per emotion class, with noise added to the learning data. On the other hand, the AlexNet [35] structure, which is used a lot for facial recognition, was selected for comparison with the deep learning method proposed in this paper. Figure 11 shows a comparison of emotion recognition accuracy with the proposed method against the accuracy with AlexNet. The facial expressions were recognized through extraction of the entire face area and the main areas, and the accuracy of extracting emotions according to the expressions was compared, with the proposed method showing accuracy about 3.75% higher.



	Surprised	Sadness	Happiness	Fear	Angry	Scared	Disgusted	Neutral
Proposed Method	92.74	97.15	91.23	89.57	95.41	87.35	91.43	99.4
AlexNet	89.15	91.43	90.87	84.16	89.77	79.22	88.21	97.62

Figure 11. Comparison of emotion recognition accuracy of the proposed model and AlexNet.

Figure 12 shows the distribution of the times required to extract the face area, and the results of face area detection. As a result of one experiment, the proposed method had a faster processing time and a lower error rate than the basic method that did not go through smoothing. In addition, as the dispersion of the processing time was only a little, it turned out to be a normalization method suitable for real-time processing.

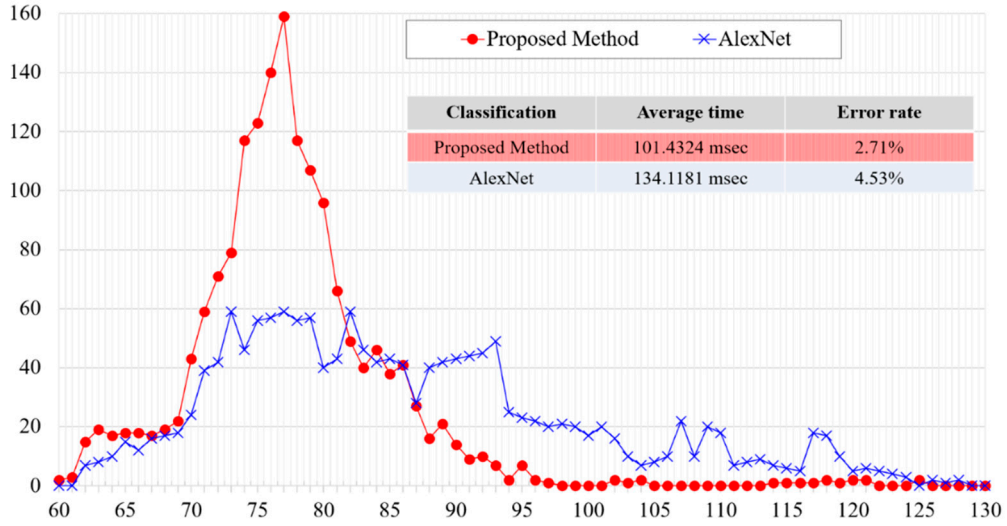


Figure 12. Face area detection time (in milliseconds).

Figure 13 shows the distribution of the processing time to detect the eye area, and the results of eye area detection. In eye area detection, the distribution of the processing time was not affected by normalization; however, there was a difference in the error rate. As a result of the experiment, the proposed method was deemed excellent.

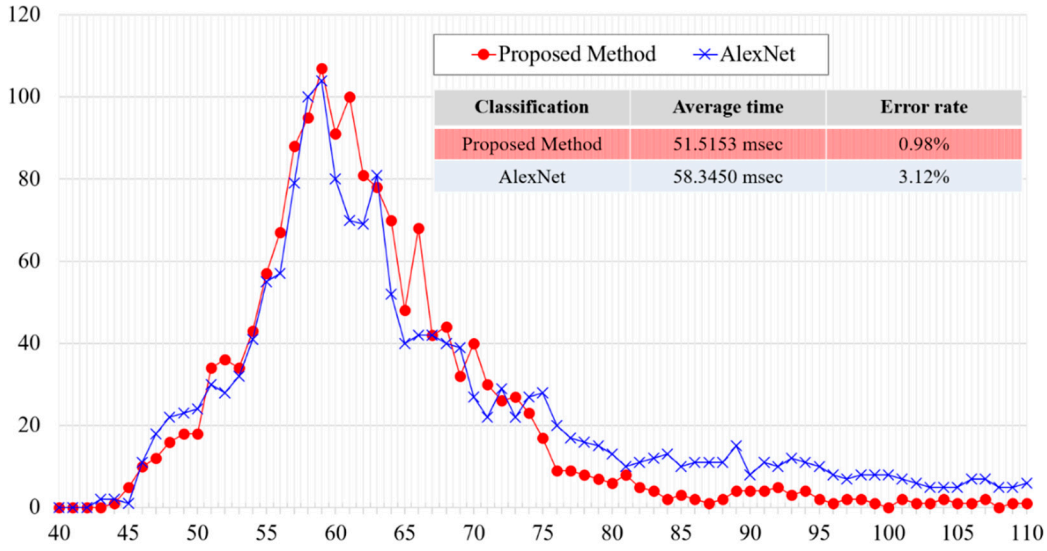


Figure 13. Eye area detection time (in milliseconds).

Figure 14 shows the distribution of the processing time to detect the nose area, and the results of nose area detection. As with eye area detection, the proposed method showed excellent performance in terms of average processing time and error rate.

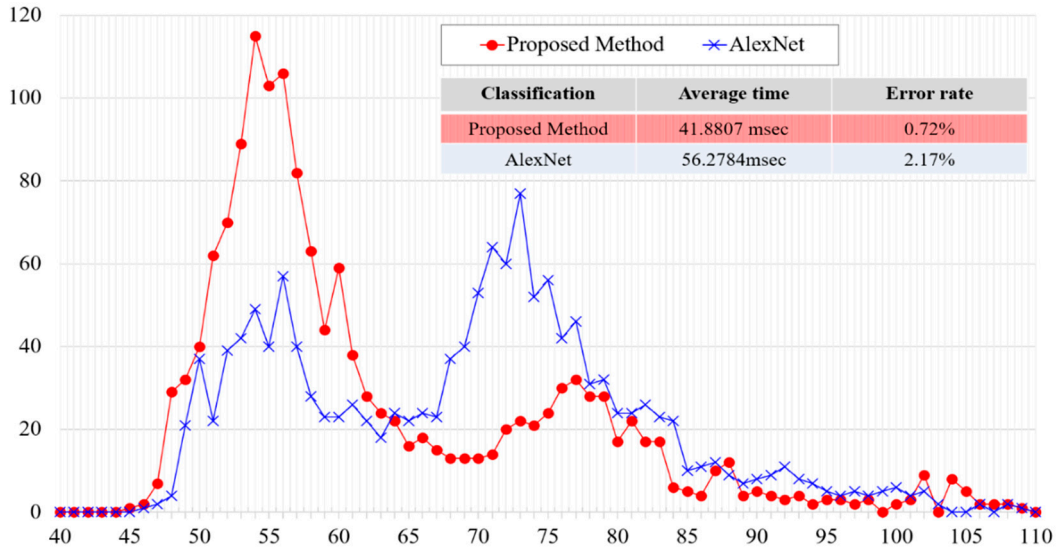


Figure 14. Nose area detection time (in millisecconds).

4. Mobile Service for Real-Time Interview Management

The self-management interview system was developed as a mobile application for smooth interview coaching. When the interview app is used, a real-time video is taken and transmitted to the server. At this time, the person's emotional state is presented through voice and facial recognition in the video, and real-time coaching is provided accordingly. In addition, including various types of interview coaching content and self-diagnosis programs, it is an effective system for speech practice as well as interviews. Figure 15 shows the image-analysis algorithm-based emotion matching. As for the image-analysis algorithm, the faces and eyes were detected, using an Extensible Markup Language (XML) classifier, and based on the detected images, emotions were extracted from a comparative analysis by the CAS-PEAL face database and Sort image. The figure included in this image is the author, who agreed to provide the figure.

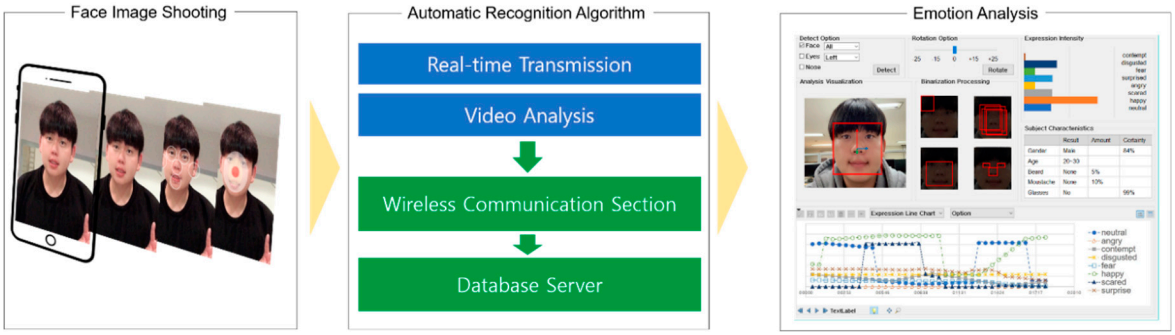


Figure 15. The structure of the interview management system. * The figure included in this image is the author, who agreed to provide the figure.

Figure 16 shows a system that analyzes images by capturing one frame after dividing a video into frame units. System functions include video playback, analysis visualization, recognition options, rotation options, binary processing, curve graph representation of emotions, object feature analysis, etc. After capturing the video, the user selects the part to be recognized with the recognition option and then recognizes that part through a binarization process. The binarization function finds the feature points of the image. There may be a rotated face in the captured image, so there is also an option that rotates the face to the correct position. This function offers a selection range of −25 to +25 degrees. There are eight emotions for analyzing a person's feelings through the recognition function: Neutral (usual expression), contempt, disgust, anger, happiness, surprise, sadness, and fear. There is also

a function that graphs the emotions in each image captured from the video. Analysis of the image in Figure 16 confirms the person is happy. Object characterization analysis shows the gender and age, and features like a mustache, beard, and eyeglasses. According to the analysis, the image in the current frame is male, 20–30 years old, without a mustache or beard, and no glasses. A screen shot from the facial recognition and emotion analysis results of the interview management system is shown in Figure 17.

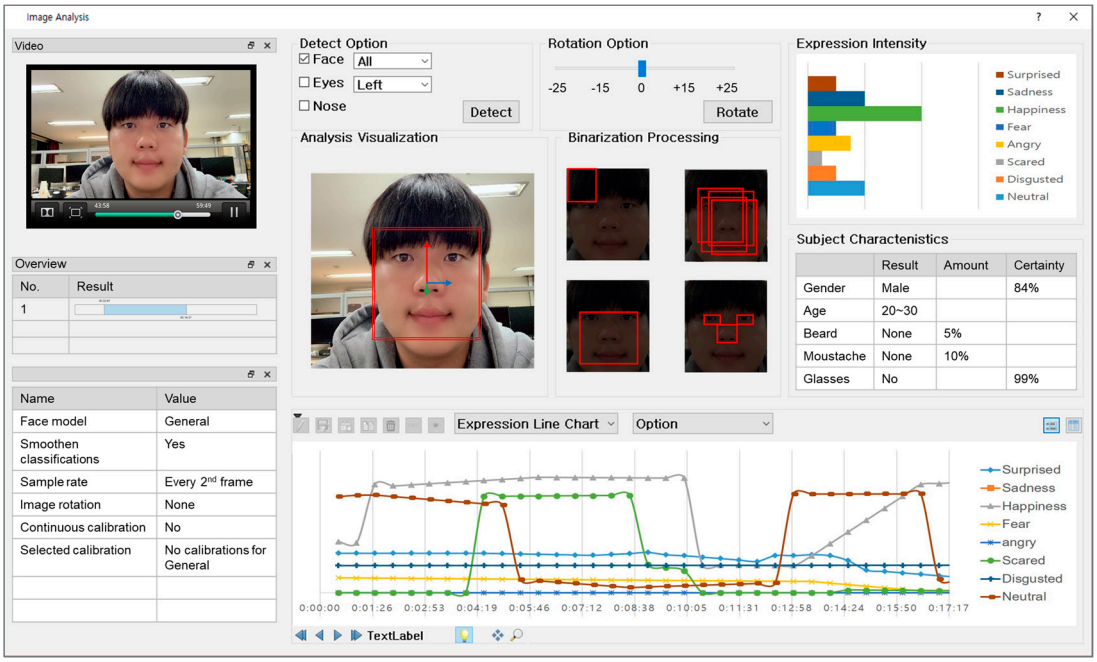


Figure 16. The real-time interview management system. * The figure included in this image is the author, who agreed to provide the figure.

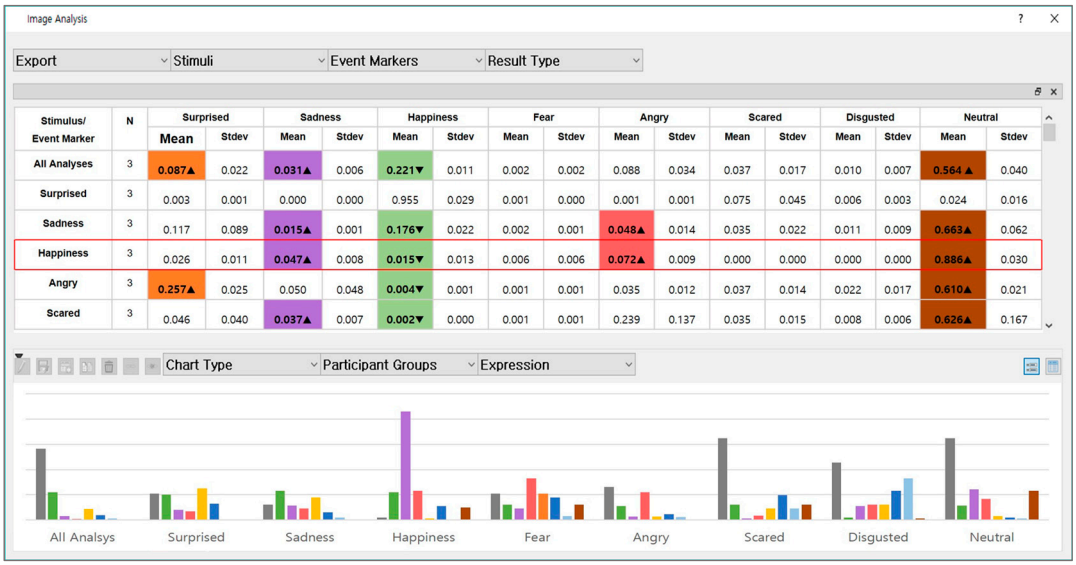


Figure 17. Facial recognition and emotion analysis results from the interview management system. * The figure included in this image is the author, who agreed to provide the figure.

For the mobile service configured in this study, an application was developed utilizing Android Studio 9 (Pie) on an Intel Core i7-4770 CPU at 3.40 GHz, with 16 GB of RAM running the Windows 10 Enterprise 64-bit environment. The figure included in this image is the author, who agreed to provide the figure. For the real-time interview and automatic coaching service, an app was configured that has

a server for interview management, a module for automatic coaching based on the interview when a user uses the service, and a user interface for the relevant services. When the user touches each button in the real-time interview management mobile application, including voice evaluation, interview evaluation, and comprehensive interview from the main screen, that input is passed to the service use information page, providing values for pronunciation, interview, and coaching, for the function interviewCode. On the service use information page, the value of interviewCode is forwarded as an intent that is distinguished as a value for each variable and is displayed, applying a message image for the corresponding voice evaluation, interview evaluation, comprehensive interview, and start button.

Splash screens for the facial recognition and emotion analysis results of self-managed interviews are shown in Figure 18. Once the interview evaluation begins, for evaluation questions, the application calls up the interview question API(Application Programming Interface) in the server, brings up the index of the relevant questions, the content of the questions, and information about the company that set the questions, and displays them in the application view. This was designed so that, once recording begins, the application calls the Android internal camera and conducts image recording and voice recording for encoding, so that both image and voice are included in the video. When the recording ends, the file-upload API in the server is immediately run to upload the user information, question index, and video file to the server, and once uploading is completed, the analysis procedure is launched through a module. On the module analysis information page, at regular intervals, the application continuously calls up the module analysis results API in the server. When the module analysis is completed, the user moves to the interview evaluation results page. Then, with the values coming from the module, the result is displayed in percentages of the emotions (including neutral, contempt, disgusted, angry, happy, surprised, scared, and fear) in the criteria for analysis.

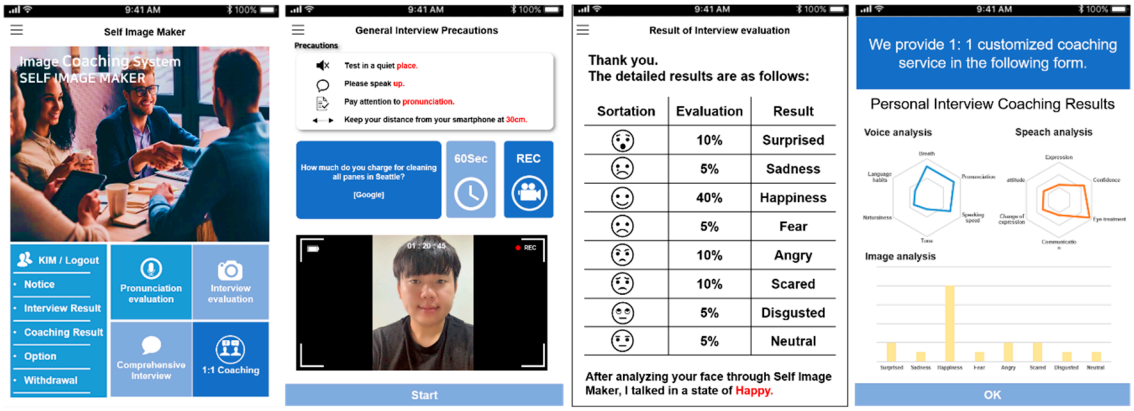


Figure 18. The real-time interview management mobile application. * The figure included in this image is the author, who agreed to provide the figure.

5. Conclusions

In this paper, we developed a self-management interview system and conducted a study on deep learning-based face analysis for emotion extraction to provide an accurate interview evaluation service. A self-management interview system was developed as a mobile application for smooth interview coaching. When the interview service is used, a real-time video is recorded and transmitted to the server. At this time, the person's emotional state is presented through voice and facial recognition from the video, and real-time coaching is provided accordingly. In addition, including a variety of interview coaching content and self-diagnosis programs, the proposed system is effective for speech practice as well as interview practice. Unlike the basic structure for recognizing a whole-face image, the deep learning method for image analysis in this system helps the user learn after sampling the core areas that are important for sentiment analysis during face detection through a multi-block process. In the multi-block process, multiple AdaBoost is used to perform sampling. After sampling, an XML classifier is used to detect the main features, which are set at threshold values to remove elements

that interfere with facial recognition. In addition, the extracted images are detected by using the CAS-PEAL face database to classify eight emotions (e.g., neutral, contempt, disgusted, angry, happy, surprised, scared, and fear), and services are provided through the application. In the experiment results, facial expressions were recognized through extraction of the entire face area and the main areas. The accuracy from extracting emotions based on the recorded expressions was compared, and the extraction time of the specific areas was decreased by 2.61%, and the recognition rate was increased by 3.75%, indicating that the proposed facial recognition method is excellent. The extracted emotions are provided through an interview management app, and users can efficiently access the interview management system based on them. We believe the interview coaching application will be utilized to provide an interview education that matches students with employment coaches, and it will provide quality job interview–education content for students in the future. The system proposed in this paper is expected to contribute to the creation of job opportunities by providing customized interview education that enables efficient interview management, is not constrained by space and time, and provides an appropriate level of interview coaching.

Author Contributions: K.C. and R.C.P. conceived and designed the framework. D.H.S. implemented Multi-Block Deep Learning for a Self-Management Interview App. R.C.P. and D.H.S. performed experiments and analyzed the results. All authors have contributed in writing and proofreading the paper.

Funding: This research was funded by a National Research Foundation of Korea (NRF) grant funded by the Korea government (2019R1F1A1060328).

Acknowledgments: We appreciate very much the author and researchers who agreed to provide the images used in this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jang, H. The effectiveness of labor market policies on youth employment. *Korean Public Adm. Rev.* **2017**, *51*, 325–358. [CrossRef]
2. Na, E.M. The Identity of the Employment-oriented Genre and the Design of Self-introduction writing and interview. *Korean J. Lit. Res.* **2018**, *9*, 131–159.
3. Park, J.H.; Lee, K.J.; Lee, S.W. The Effects of Local Industrial Structures on the Probability of Youth Employment. *J. Korean Reg. Dev. Assoc.* **2018**, *30*, 133–159.
4. Choi, C.S. Effects of Position Interdependency and Competitiveness on Communication Strategy Selection in Groupdiscussion Employment Interviews. *J. Commun. Res.* **2019**, *56*, 330–371.
5. Ran, H.; Xiang, W.; Zhenan, S.; Tieniu, T. Wasserstein CNN: Learning Invariant Features for NIR-VIS Face Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 1761–1773.
6. Zhou, L.; Li, W.; Du, Y.; Lei, B.; Liang, S. Adaptive illumination-invariant face recognition via local nonlinear multi-layer contrast feature. *J. Vis. Commun. Image Represent.* **2019**, *64*, 102641. [CrossRef]
7. Cai, Y.; Lei, Y.; Yang, M.; You, Z.; Shan, S. A fast and robust 3D face recognition approach based on deeply learned face representation. *Neurocomputing* **2019**, *363*, 375–397. [CrossRef]
8. Lia, C.; Huang, Y.; Yu, X. Dependence Structure of Gabor Wavelets based on Copula for Face Recognition. *Expert Syst. Appl.* **2019**, *137*, 453–470. [CrossRef]
9. Shakeel, M.S.; Lam, K.-M.; Lai, S.-C. Learning sparse discriminant low-rank features for low-resolution face recognition. *J. Vis. Commun. Image Represent.* **2019**, *63*, 102590. [CrossRef]
10. Wu, Y.; Ji, Q. Facial Landmark Detection: A Literature Survey. *Int. J. Comput. Vis.* **2019**, *127*, 115–142. [CrossRef]
11. Corneanu, C.A.; Simon, M.O.; Cohn, J.F.; Guerrero, S.E.; Oliu, M.; Escalera, S. Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect-Related Applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 1548–1568. [CrossRef] [PubMed]
12. Iqbal, M.; Sameem, M.S.; Naqvi, N.; Kanwal, S.; Ye, Z. A Deep Learning Approach for Face Recognition based on Angularly Discriminative Features. *Pattern Recognit. Lett.* **2019**, *128*, 414–419. [CrossRef]
13. Guo, G.; Zhang, N. A Survey on Deep Learning based Face Recognition. *Comput. Vis. Image Underst.* **2019**, *189*, 102805. [CrossRef]

14. Elmahmudi, A.; Hassan, U. Deep Face Recognition using Imperfect Facial Data. *Future Gener. Comput. Syst.* **2019**, *99*, 213–225. [CrossRef]
15. Mayya, V.; Pai, R.M.; Pai, M.M. Automatic Facial Expression Recognition Using DCNN. *Procedia Comput. Sci.* **2016**, *93*, 453–461. [CrossRef]
16. Chung, K.; Park, R.C. Cloud based U-healthcare Network with QoS Guarantee for Mobile Health Service. *Clust. Comput.* **2019**, *22*, 2001–2015. [CrossRef]
17. Affectiva. Available online: https://www.affectiva.com/ (accessed on 9 October 2019).
18. Chung, K.; Park, R.C. Chatbot-based heathcare service with a knowledge base for cloud computing. *Clust. Comput.* **2019**, *22*, 1925–1937. [CrossRef]
19. Xu, Y.; Li, Z.; Yang, J.; Zhang, D. A Survey of Dictionary Learning Algorithms for Face Recognition. *IEEE Access* **2017**, *5*, 8502–8514. [CrossRef]
20. Fu, S.; He, H.; Hou, Z.G. Learning Race from Face: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 2483–2509. [CrossRef] [PubMed]
21. Chung, K.; Park, R.C. PHR Open Platform based Smart Health Service using Distributed Object Group Framework. *Clust. Comput.* **2016**, *19*, 505–517. [CrossRef]
22. Gong, D.; Li, Z.; Huang, W.; Li, X.; Tao, D. Heterogeneous Face Recognition: A Common Encoding Feature Discriminant Approach. *IEEE Trans. Image Process.* **2017**, *26*, 2079–2089. [CrossRef] [PubMed]
23. Violante, M.G.; Marcolin, F.; Vezzetti, E.; Ulrich, L.; Billia, G.; Di Grazia, L. 3D Facial Expression Recognition for Defining Users' Inner Requirements-An Emotional Design Case Study. *Appl. Sci.* **2019**, *9*, 2218. [CrossRef]
24. Wang, J.; Yin, L.; Wei, X.; Sun, Y. 3D facial expression recognition based on primitive surface feature distribution. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 17–22 June 2006; Volume 2, pp. 1399–1406.
25. Zhou, S.; Xiao, S. 3D face recognition: A survey. *Hum. Cent. Comput. Inf. Sci.* **2018**, *8*, 35. [CrossRef]
26. Yin, L.; Wei, X.; Sun, Y.; Wang, J.; Rosato, M.J. A 3D facial expression database for facial behavior research. In Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, Southampton, UK, 10–12 April 2006; pp. 211–216.
27. Kim, J.C.; Chung, K. Prediction model of user physical activity using data characteristics-based long short-term memory recurrent neural networks. *Ksii Trans. Internet Inf. Syst.* **2019**, *13*, 2060–2077.
28. Kim, J.C.; Chung, K. Associative feature information extraction using text mining from health big data. *Wirel. Pers. Commun.* **2019**, *105*, 691–707. [CrossRef]
29. Yoo, H.; Chung, K. Mining-based lifecare recommendation using peer-to-peer dataset and adaptive decision feedback. *Peer-to-Peer Netw. Appl.* **2018**, *11*, 1309–1320. [CrossRef]
30. Gao, W.; Cao, B.; Shan, S.; Chen, X.; Zhou, D.; Zhang, X.; Zhao, D. The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations. *IEEE Trans. Syst. Man Cybern.* **2008**, *38*, 149–161.
31. Chung, K.; Yoo, H.; Choe, D.; Jung, H. Blockchain network based topic mining process for cognitive manufacturing. *Wirel. Pers. Commun.* **2019**, *105*, 583–597. [CrossRef]
32. Yoo, H.; Chung, K. Heart rate variability based stress index service model using bio-sensor. *Clust. Comput.* **2018**, *21*, 1139–1149. [CrossRef]
33. Song, C.W.; Jung, H.; Chung, K. Development of a medical big-data mining process using topic modeling. *Clust. Comput.* **2019**, *22*, 1949–1958. [CrossRef]
34. Baek, J.W.; Kim, J.C.; Chun, J.; Chung, K. Hybrid clustering based health decision-making for improving dietary habits. *Technol. Health Care* **2019**, *27*, 459–472. [CrossRef] [PubMed]
35. Lin, L.; Yingzi, T. Detection of Cabinet in Equipment Floor based on AlexNet and SSD Model. *J. Eng.* **2019**, *2019*, 605–608.