KIPS Trans. Softw. and Data Eng.
Vol.10, No.3 pp.99~108
ISSN: 2287-5905 (Print), ISSN: 2734-0503 (Online)

https://doi.org/10.3745/KTSDE.2021.10.3.99

Face Identification Using a Near-Infrared Camera in a Nonrestrictive In-Vehicle Environment

Min Song Ki† · Yeong Woo Choi††

ABSTRACT

There are unrestricted conditions on the driver's face inside the vehicle, such as changes in lighting, partial occlusion and various changes in the driver's condition. In this paper, we propose a face identification system in an unrestricted vehicle environment. The proposed method uses a near-infrared (NIR) camera to minimize the changes in facial images that occur according to the illumination changes inside and outside the vehicle. In order to process a face exposed to extreme light, the normal face image is changed to a simulated overexposed image using mean and variance for training. Thus, facial classifiers are simultaneously generated under both normal and extreme illumination conditions. Our method identifies a face by detecting facial landmarks and aggregating the confidence score of each landmark for the final decision. In particular, the performance improvement is the highest in the class where the driver wears glasses or sunglasses, owing to the robustness to partial occlusions by recognizing each landmark. We can recognize the driver by using the scores of remaining visible landmarks. We also propose a novel robust rejection and a new evaluation method, which considers the relations between registered and unregistered drivers. The experimental results on our dataset, PolyU and ORL datasets demonstrate the effectiveness of the proposed method.

Keywords : Face Identification, Near-infrared Image, Multi Support Vector Machine (Multi-SVM), Light Overexposure

적외선 카메라를 이용한 비제약적 환경에서의 얼굴 인증

기 민 송† · 최 영 우††

요 약

차량 내부에는 조명 변화, 부분적인 가림 및 운전자의 상태 변화와 같은 제한되지 않은 조건들이 존재한다. 본 논문에서는 비 제약적인 차량 환경에서의 운전자 얼굴 인증 시스템을 제안한다. 제안한 방법은 차량 내부 및 외부의 조명 변화에 따라 발생하는 얼굴 이미지의 변화를 최소화하기 위해서 근적외선(NIR) 카메라를 사용한다. 특히 정면에서의 강한 빛에 노출된 얼굴 이미지를 처리하기 위해서, 학습 이미지의 평균과 분산을 사용하여 정상적인 얼굴 이미지로부터 빛에 과다하게 노출된 이미지로 변환하여 사용한다. 따라서 정상적인 조명에서의 얼굴 분류기와 강한 정면광에서의 얼굴 분류기를 각각 동시에 만들어진다. 제안하는 얼굴 분류기는 얼굴 랜드마크를 추출하고 각 랜드마크의 신뢰도 점수를 합산하여 얼굴을 최종적으로 식별한다. 특히 각 랜드마크를 인식하여 부분적인 얼굴 가림에 강하기 때문에 안경이나 선글라스를 착용하는 상황에서도 높은 성능 향상이 가능하다. 즉 가려지지 않은 남은 랜드마크의 점수를 사용하여 운전자를 인식할 수 있다. 또한 등록 운전자와 미등록 운전자 간의 관계를 고려한 새로운 인식 거부 방법과 새로운 평가 방법을 논문에서 제안한다. 자체 취득한 데이터 셋, 공인된 PolyU 및 ORL 데이터 셋으로 실험한 결과 제안한 방법이 효과적임을 확인할 수 있었다.

키워드 : 얼굴 인증, 적외선 이미지, 멀티 서포트 벡터 머신, 정면광 노출

1. Introduction

Recent studies have focused on the combination of artificial intelligence-based autonomous vehicles with information technology. These studies focus on the automatic recognition of the external environment using a camera located inside the vehicle to provide safety and convenience to the driver without human intervention [1-3]. It also requires a monitoring system that detects the inside situations of the vehicle and provides personalized functionality through facial recognition of the driver.

※ This research was supported by a grant from the National Research Foundation of the Korean government (No. NRF-2017R1D1A1B0403 5633).
† 준 회 원 : 연세대학교 컴퓨터과학과 박사과정
†† 정 회 원 : 숙명여자대학교 소프트웨어학부 교수
Manuscript : October 22, 2020
Accepted : November 29, 2020
* Corresponding Author : Yeong Woo Choi(ywchoi@sookmyung.ac.kr)

Generally, face identification has to extract and train on facial features of the driver. It is also a crucial issue in the field of computer vision and has been actively studied in [4,5]. The existing methods perform well under constrained environments. However, in-vehicle systems can be exposed to several unconstrained variations, such as pose, occlusion, and exposure to illumination.

For face identification, a unique feature of a detected face can be obtained using global or local features [6,7]. But, it is not desirable to obtain global features for the entire face, especially for faces that are partially obscured. To identify a driver, it is first necessary to acquire the facial images and register the ID of the driver. It is difficult to acquire a large amount of data for registration as the facial images must be obtained within a short period for real-time training.

In this paper, we propose a multi support vector machine (multi-SVM) based face identification and rejection method. Our proposed face identification method is optimally designed to be used inside a vehicle, and only registered and identified drivers can drive. When registering a driver, we generally do not register a large number of people and may include family or friends. Therefore, the face recognition for in-vehicle can be regarded as a problem of recognizing a limited number of persons. We set the number of registrants up to five; however, it can increase the number of registrants. The number of unregistered drivers is not limited; therefore, the system can reject those who are not registered.

Fig. 1 shows the proposed method in two steps. In the first step, the driver is registered through training, and in the second step, the driver is identified and whether or not it is possible to drive. We develope an effective feature extraction method using local binary patterns (LBPs) for each landmark based on high-dimensional features [8]. Recognizing each landmark independently has the effect of recognizing the occluded faces using several landmarks that are not affected by occlusion.

We also propose a method of processing an over-exposed face image, which simulates an over-exposed face by adjusting the brightness and contrast of a normal face image. And the simulated face image is trained by an additional classifier, so normal and severely exposed face images are also trained and generated. There is no standard dataset of monitored faces of a driver inside the vehicle. The most common illumination variation in vehicles is sunlight; In addition, a set of driver data [11] containing images of faces exposed to excessive sunlight and various lighting conditions was collected. As we classify non-registrants first and recognize the registrants if they are not non-registrants, the recognition rates of registrants may vary depending on the rejection rate of non-registrants. Therefore, we also develope a performance evaluation method considering the relations between registered and unregistered drivers. The effectiveness of the proposed method was verified using the standard PolyU NIR [7] and Olivetti Research Lab (ORL) dataset [10], and compared with previous approaches.

2. Related Work

2.1 Our Previous Work [11]

In the previous study, a face identification method was proposed focusing on registered drivers. In the case of non-registrants, the rejection was performed using only the average of confidence scores; consequently, the rejection performance was not sufficient. When creating images overexposed to light, we transformed the images empirically by multiplying a predefined constant without considering the ambient illumination. Further, there was also a limitation in mounting the system on an actual vehicle because it did not solve the memory problem for storing facial features or various learned files.

To address these limitations, we improved our results and have suggested new ideas based on our previous research to optimize system for real-time. The main purpose of this study is to identify registered drivers and reject non-registrants by creating a real system that can be mounted in a vehicle in a non-restrictive manner. To evaluate the recognition performance fairly, we evaluated two cases, i.e., measuring performance per frame, and measuring the performance as a final result per driver after voting. To enhance the peformance, we formulated the image conversion function by using the mean and variance of the current facial images while considering the ambient brightness. The rejection of non-registrants is essential as much as the recognition of the registered drivers. We also developed novel rejection

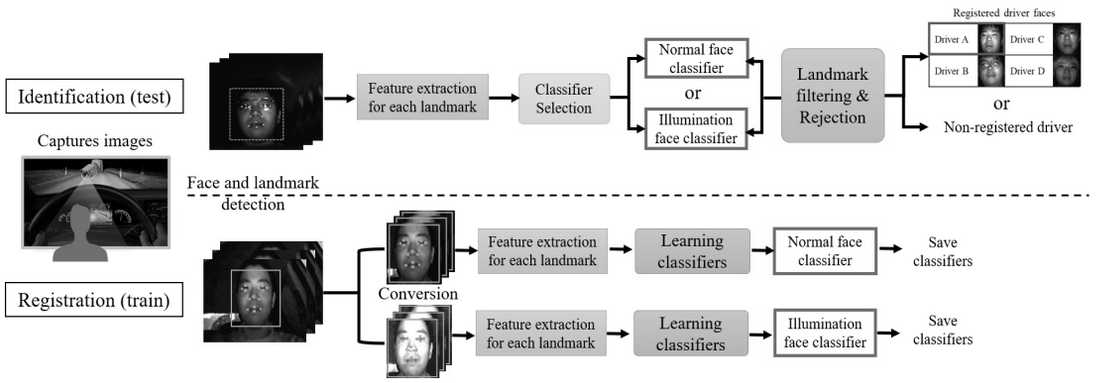


Fig. 1. The Overall Flow of Face Identification System Proposed for In-vehicle System

and performance evaluation methods that consider the relations between registered and unregistered drivers.

We conducted additional experiments using standard datasets and verified the superiority by analyzing the performance, execution time, and storage size. Consequently, we designed the system to enable near real-time learning and real-time testing as well as experiments in a PC environment.

2.2 Learning-based Face Recognition

Pan et al. [8] proposed a part-based recognition method. This approach decomposes the NIR facial image into several parts, such as eyes, nose, and mouth. Then, AdaBoost learning was applied to each region to build a classifier. Chen et al. [12] developed a high-dimensional feature by extracting multiscale patches. However, it requires a long time to compute because it obtains the multiscale features by changing the size of the input image.

The deep learning method has excellent features and performance in data learning. Convolutional neural networks (CNNs) are well-known architectures inspired by the natural visual perception mechanism. DeepFace [13] was the first to incorporate deep CNN into face recognition. Y. Taiman et al. [13] aligned faces after landmark extraction using pre-trained 3D face geometry model. FaceNet [14] can embed face features into the networks by learning 260 million images. The approaches are limited to RGB datasets and are also challenging to be used in the systems of the actual vehicles since real-time learning is not possible. Thus, our method only compares the performances of existing studies using infrared datasets.

Since the deep learning-based method requires a large amount of training data and a high-performance processing system, it is not yet suitable for real-time use in vehicle systems [13]. Even if a new driver is added to the vehicle, it should be able to learn in real time.

3. Proposed Method

The system we propose consists of face registration and identification. The overall system is shown in Fig. 1. For training and testing, 40 images per person were used. For the identification of one driver, each image generates a recognition result, and the final decision is made by voting over the 40 results.

3.1 Face and Facial Landmark Detection

For training, we randomly select 8 images for each predefined position. The driver looks at five predefined locations in counterclockwise direction: the dashboard, the start button, the navigation system, the front window and the left side mirror. The driver looks at each point for a few seconds, at which time images are acquired. Finally, we obtain 40 training images for each driver. The training data does not include images with occlusion(only normal category); however, the system is tested with considerably occluded or rotated images at the identification stage. First, a face is detected using a trained AdaBoost classifier with Haar-like features. The detected face is normalized to a size of 200 x 200 pixels, and then 15 facial landmarks in both eyes, nose, mouth and facial contours are placed on each face using the algorithm proposed in [15].

We re-trained a model using our new NIR face dataset. We annotated 15 landmarks for each image

to train. A total of 854 images are used for training. To obtain more details, the images consisted of three directions: front, left, and right. Thus, 257 frontal images, 250 images from the left, and 377 images from the right are used for learning.

3.2 Feature Extraction

Distinctive features are extracted from each location of the landmark. The previous method [12] extracts all features using an image pyramid to cope with the size variations. But since this method has a large number of features [12], the execution time increases accordingly.

Since the proposed method extracts multi-scale features for each landmark, processing time is reduced and it is resistant to size change and partial occlusion. The LBP features only need to be extracted once for the image without changing the input image size to generate a multi-scale image. Also, a multi-scale patch of 4 is defined around each landmark, and the size of each patch is 60 x 60 pixels. Then, this patch is scaled down by 20% to create a new patch, and enlarged twice by 20% to create two new patches. Thus, a total of 4 multi-scale patches are created. For each landmark, 4 multi-scale patches from the LBP functional map are cut from the LBP functional map.

Each patch is divided into 2 x 2 cells, with each cell generating a 256-dimensional histogram vector normalized by the number of pixels. Thus, each patch is a 1,024 dimensional normalized histogram vector. Then, four 1,024 dimensional vectors are combined to form a single feature vector. Finally, for each landmark, a 4,096-dimensional feature vector is created. Then, the extracted 4,096-dimensional feature vectors are reduced to 400 dimensions for each landmark by principal component analysis (PCA)[16].

3.3 Training the Facial Classifiers for Each Landmark

The facial classifiers (Fig. 2) are trained by using reduced feature vectors. These classifiers correspond to normal facial classifiers. The feature values change when the face is exposed to strong frontal sunlight. When the brightness of the face image increases to an extreme value, the LBP feature value approaches zero. Thus, we add a classifier trained with simulated data on the exposed image to handle

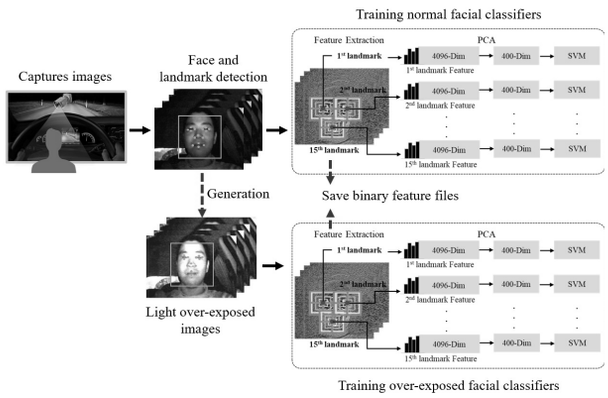


Fig. 2. Overall Architecture of the Proposed Face Registration

very strong frontal sunlight as shown in Fig. 2. The brightness and contrast of the face are adjusted to mimic the sunlight in front of the face, and the pixel conversion function for this is defined as Eq. 1.

$$\begin{aligned} f(x) &= \theta x + \mu \\ \theta &= \|var/orig_{var}\| \\ \mu &= mean - \theta * orig_{mean} \end{aligned} \tag{1}$$

Here, x is the intensity of each pixel, θ is the appropriate value for the contrast adjustment, and μ is utilized for the brightness adjustment. As the normal images have various lighting conditions according to the place of the registration, we first predefined the mean and variance of lighting criterion. The mean and var correspond to the criteria of mean and variance, which are determined through experiments. For image conversion to a simulated exposure to light, we multiplied the intensity of each pixel with θ and added μ. Fig. 3 shows more examples of the converted images.

If a new driver needs to be registered, multiple SVM face classifiers are created for each landmark to handle both normal and extreme lighting cases. It can be used to train other classifiers by acquiring and transforming a regular image only once, without getting the actual data exposed to light. Each landmark is recognized individually, and an SVM with a low reliability score has the effect of lowering the importance of the results of several SVMs. In addition, when the face is partially covered by glasses or sunglasses, the face can be recognized using the score of the remaining landmarks.

If the registrant class is changed, we need to

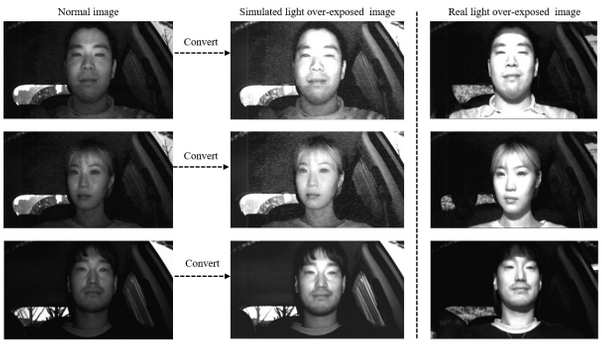


Fig. 3. Examples of the Conversion to Simulated Light Overexposed Cases when Compared to the Real Cases

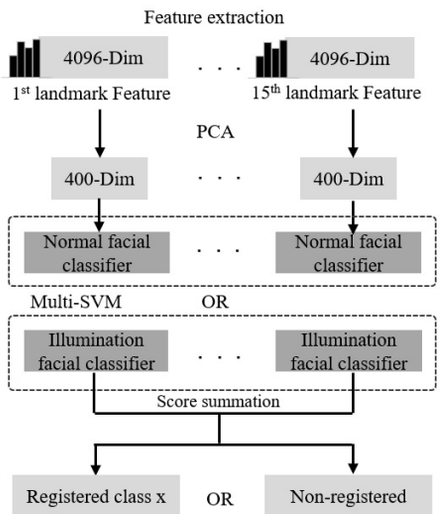


Fig. 4. Architecture of the Proposed Face Identification

re-train the model to reflect this change. Re-training can be conducted by loading the images used in the previous training. Therefore, our system stores the previously extracted features in binary feature files for each registrant; therefore, saving all the acquired facial images is not required.

The re-training time of the classifiers after adding or deleting a registrant is only 2s in our method. We require approximately 15s for training, which includes the time for acquiring 40 images from the camera to register a new driver. When the program is run for the first time, it requires 10s to load the learned PCA files and SVM classifiers from the repository for each landmark. It is not time consuming after the first loading.

3.4 Face Identification

Even if the appearance of the driver changes, we must recognize the driver in the same class as at the time of registration. The proposed identification method is shown in Fig. 4. The process from facial image acquisition to feature extraction is the same as that in registration. After extracting the features, we select whether to use a normal facial classifier or a facial illumination classifier depending on the exposure to light. The extracted LBP features are used to determine whether the face is exposed to light, and a classifier is selected accordingly.

If the sum of the values for each landmark is equal to or less than a predefined threshold, we filter the landmark that the surrounding area is exposed to light. If more than 5 landmarks are filtered, we use the classifier trained with simulated data. When the light exposure was not severe, we use the classifier trained using normal images for each landmark. Each classifier outputs a confidence score for all the registered classes per image.

The final result is obtained by summation of the 15 confidence scores for each class using Eq. 2 and calculating the maximum summation value.

$$R = max(\sum_{l=1}^{n} S_{dl}) \tag{2}$$

Here, d represents each registered class (driver ID), and l represents a sequence of facial landmarks (unfiltered landmarks). dl represents the confidence score of the lth landmark of the registered class d. The final output of R is the recognized class with the highest total score. The corresponding class is assigned a relatively low score when a landmark is incorrectly detected or severe occlusion. Therefore, landmarks with scores of less than 0.5 were filtered out without being used in Eq. 2.

We describe the effect of landmark filtering in cases with occlusion and rotation. The filtered landmarks are visualized in the image as shown in Fig. 5 (red dot with a score less than 0.5). The reasons for filtering are the eye position is located incorrectly owing to severe facial rotation or sunglasses that are similar to the background color.

For rejection, we set three conditions to distinguish between registrants and non-registrants: i) Non-registrants have more images with lower confidence scores than registrants; ii) The average number of recognized frames of each class is less than that of the registrant (the number of recognized images per

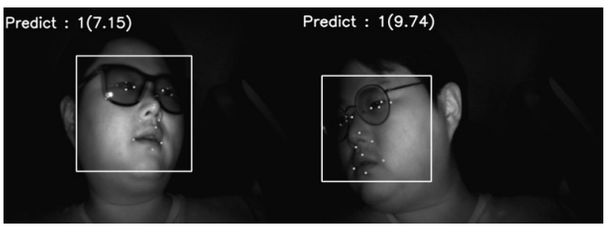


Fig. 5. The Accurately Recognized Results after Filtering Landmarks having Low Confidence in Cases with Occlusion and Rotation

class is scattered); iii) Confidence scores for non-registrants are distributed near the minimum score.

As a total of 40 images are captured per driver, we output the final score per frame and obtain the overall average score. If any one of these conditions is satisfied, we reject the driver as a non-registrant. The thresholds for the three conditions are obtained empirically. The recognition time is 0.03 s for each image.

4. Experiments

The performance evaluation of the proposed system uses a newly created dataset and two benchmark databases: the NIR in-vehicle facial dataset [11], the PolyU NIR [7] and the ORL dataset [10].

4.1 Datasets

The NIR camera is mounted in front of the vehicle dashboard. We use two types of vehicles: sports utility vehicle and passenger car. The captured images are of 720 x 480 pixels. In our dataset [11], as shown in Fig. 6, 12 drivers are included in about 5 categories: normal, wearing glasses, wearing sunglasses, wearing hats, and overexposure to light. The system recognizes the facial image after the first registration. Currently, the number of registered drivers is 5, but the number of registered drivers can be increased.

The PolyU NIR dataset [7] has 350 subjects, and each subject consists of 100 images, including a total of 35,000 images. Each image has a resolution of 768 x 576 pixels. The image set contains only the frontal faces, while the probe set contains images with resize, pose deformation, and time lapse. When evaluating performance using PolyU NIR, B. Zhang et al. [7] designed three types of experiments, and we experimented in the same way. For performance comparison with other methods, we evaluate the

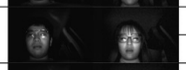
No.	Category	Sample images	Number of images acquired per person	Total number of images
1	Normal		Train: 500 images Test: 630 images	14,000 images
2	Glasses		Train: 500 images Test: 620 images	13,500 images
3	Sunglasses		Train: 500 images Test: 600 images	13,500 images
4	Hat		Train: 500 images Test: 600 images	13,500 images
5	Extreme illumination		Train: 300 images Test: 300 images	7,200 images

Fig. 6. Summary of the NIR In-vehicle Facial Dataset[11]

mean accuracy, which is the average of the accuracy of the three experiments.

The ORL dataset [10] contains 10 different images for each of the 40 individual subjects. Some images are captured at different times by changing the lighting, facial expressions and facial details (with or without glasses). Each image is 92 x 112 pixels in size and 256 gray levels per pixel.

4.2 Results from the NIR Face Dataset [11]

In the experiment, recognition rate was evaluated and a new evaluation method was developed and evaluated. As we reject the non-registrants and recognize the registrants simultaneously, the recognition rate of registrants could be varied according to the rejection rate of the driver. We first measured the recognition rate while only considering the registrants for each frame. The recognition rate (driver unit) is also measuring by combining the results of several frames. Next, we measure the recognition rate of the registrants reflecting the rejection rate of the drivers because registrants can also be rejected if they have a low confidence score.

Each test set consists of 5 registrants randomly selected from the dataset. We got 40 images per driver and the final result for each image. Fig. 7. shows an experimental scenario; this configuration measures the average recognition rate after performing 300 iterations for each category. Further, each test set is not duplicated.

〈Table 1〉 shows the average recognition rates for four categories measured for each frame. The entire face recognition method extracts features from the entire face and is recognized using one classifier. The proposed method (1) extracted features that are recognized by training each of the 4 patches (both eyes, nose, and mouth) using 4 SVM classifiers. As a

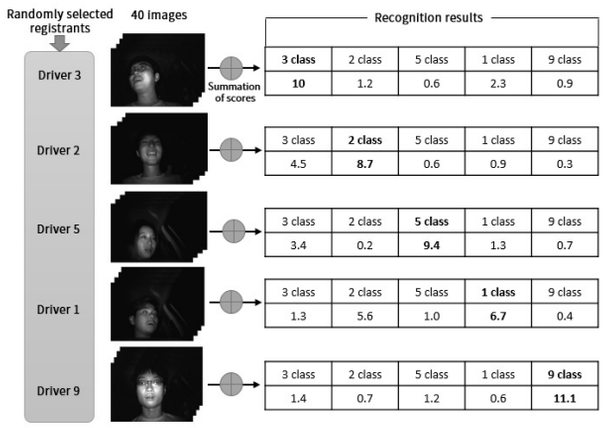


Fig. 7. Experimental Scenario for Evaluation of Recognition Rate

Table 1. Results for Frame Units using Our NIR Dataset

	Normal	Glasses	Sunglasses	Hat
Entire face recognition	91.0%	83.6%	81,8%	91.6%
Chen et al. [12]	95.1%	75.8%	81.8%	94.5%
Proposed (1)	97.3%	92.6%	90.0%	97.6%
Proposed (2)	**98.8%**	**96.4%**	**93.8%**	**98.7%**

result, we get 4 scores for each face image.

The proposed method (2) is a multi-SVM that recognizes each landmark using 15 SVM classifiers. The multi-SVM method achieves an average recognition rate of 96.9% in all four categories. Recognizing features from the entire face is significantly weak in partial occlusions. In particular, the performance improvement is the highest in the classes of glasses and sunglasses owing to the robustness to partial occlusions. Our method can recognize facial features by using the scores of remaining visible landmarks.

Table 2 shows the average recognition rate when the face is exposed to extreme light. The recognition rate improves from 60.5% to 91.4%. In contrast, when we used the normal face classifier directly, the recognition rate significantly dropped because the extracted features from the overexposed images captured during registration is different from the extracted features in identification.

In Tables 1 and Table 2, the performance are measured for each frame. However, the acquired image results must be aggregated so that the actual system (registrant or non-registrant) outputs a single result. Therefore, we measured the rate by voting 40 results for each driver. Table 3 lists the average recognition rates measured by the driver unit. Even if

Table 2. Recognition Rate when Overexposed to Light

	Method	Rate(%)
Training	With only normal images	60.5
Testing	With real light overexposed images	
Training	With normal and simulated light overexposed images	**91.4**
Testing	With real light overexposed images	

Table 3. Results for Driver units of the Proposed Method with Our Dataset

Mean Acc(%)	Normal	Glasses	Sunglasses	Hat	Light
Proposed (2)	**100**	**100**	**100**	**100**	**92.6**

there were some misidentified frames, the recognition rate of the driver unit increased because of the voting to determine the most frequently recognized class.

We propose a novel evaluation method that considers the rejection rates of non-registrants and the recognition rates of registrants simultaneously. The non-registrants are determined based on the final acceptance or rejection by aggregating the results of all images. Thus, the final rejection and recognition rates are determined after a rejection is measured by the driver unit rather than the frame unit.

In each test set, five registrants and one non-registrant are randomly selected. We also conducted repeatable experiments 300 times for each category. We first measured the rejection rates of non-registrants based on the aforementioned three conditions. Half of the test images of five registrants (similar to a validation set) are used to select three thresholds for three conditions. The other half of the test image is used to measure the registrant's recognition rate after rejection. The registrant's results are used to determine the threshold for evaluating the three conditions of the proposed rejection method. The three thresholds are defined as α, β, and γ. Here, α represents the average top-1 confidence scores of the five registrants corresponding to the correctly recognized images for each registrant; β represents the average of the number of correctly recognized images out of the 40 images for each registrant; γ is derived from the average of the minimum confidence scores of the five registrants corresponding to the correctly recognized images for each registrant. The three thresholds, as mentioned above, should be

Table 4. Average Accuracies of the Registrants that Reflect the Rejection of Non-registrants

Category	Rejection rates(%)	Recognition rates(%)
Normal	99.7	100.0
Hat	99.0	100.0
Sunglasses	99.0	92.4
Glasses	99.0	92.3
Light exposure	94.7	86.3

lower for non-registrants than registrants. Three values for three conditions can be calculated for each class. We compared the three values with three thresholds for each registered class (the validation set). If even one of these three values is lower than the threshold values, the corresponding class is rejected. Similarly, if the registered classes are all rejected, then the driver is regarded as a non-registrant.

The rejection rate is the number of rejections for a non-registrant divided by the total number of identification attempts. After rejecting a non-registrant, we utilize the test images that were not used to set the three thresholds for measuring the recognition rates of registrants. In the same way, we calculate α, β, and γ for the registrants and compare the values to the threshold values. We determine that the driver is a registrant if all calculated values for the three conditions are higher than the three threshold values.

In this study, we evaluate the recognition rates and the rejection rates, as shown in Table 4.

When the average rate of successful rejection of non-registrants is 99% (false acceptance rate was equal to 1%), we measure the recognition rate for each of the four different categories. The recognition rates of the final registrants, which reflected the rejection rates, are multiplied by the recognition rates before rejection, as shown in Table 3, and the rejection rates of the registrants.

〈Table 5〉 lists the storage size of the existing deep network backbone models. The two VGG and ResNet models are the backbone networks most commonly used in recent research. This is just the backbone network used by deep learning based methods; therefore, it requires more memory than the values notated in Table 5.

Our proposed system requires a total of 230 MB of memory. In detail, the binary feature files for five registered drivers utilize 10 MB, PCA files require 210 MB, the classifiers for landmarks require 7.5 MB, and finally, the landmark regressors require 2 MB. Therefore, it is suitable for mounting on a vehicle.

Table 5. Storage Size of Traditional Deep Learning Backbone Models

Model	Input size	Parameters	Features	Total
VGG-16	224 x 224	528 MB	58 MB	586 MB
VGG-19	224 x 224	548 MB	63 MB	611 MB
ResNet-101	224 x 224	170 MB	155 MB	325 MB

4.3 Results form PolyU NIR Face Dataset[7]

〈Table 6〉 shows the performance comparison with the previously proposed methods. Accuracy was measured in the same way as the evaluation used for the PolyU dataset [7]. Recognition time is the total time taken to process each image. Gabor-DBC [7] proposed a method called directional binary code (DBC) in which direction information was added and Gabor filter was applied to the binary pattern. However, there is a disadvantage that the feature extraction speed per image is 85 times slower than the proposed method. Zernike moment Zernike moment undecimated discrete wavelet transform (ZMUDWT) [6] uses a new algorithm based on the combination of local features extracted using undecimated discrete wavelet transform and global features extracted using Zernike moment. Zernike moments and Hermite kernels (ZMHK) [18] is a method of extracting features using PCA and linear discriminant analysis (LDA) and using both local and global information. LBPL [17] presented an active near-infrared imaging system capable of generating facial images in good condition regardless of the visible light of the environment. The performance of the proposed method is better than that of deep learning-based CNN[19], and the recognition time is 0.03 seconds.

4.4 Results from ORL Face Dataset[10]

We compared the average recognition rates with the four existing methods. Five face images of each person were used for training, and the remaining face images were randomly selected and used for testing [20].

DLSPC [21] proposed an explicit learning method of class-specific dictionaries for each category that captures the most distinctive features of this category. LC-KSVD2 [23] proposes a new dictionary

Table 6. Comparison to Previous Method using PolyU Dataset [7]

Method	Mean accuracy (%)	Recognition time (s)
LDA	93.25	0.18
Gabor-DBC [7]	75.62	0.24
ZMUDWT [6]	85.01	0.52
LBPL [17]	81.57	0.05
ZMHK [18]	87.22	0.19
CNN-based [19]	94.50	0.03
Proposed method	**95.30**	**0.03**

Table 7. Comparison to Previous Method using ORL Face Dataset [10]

Method	Mean accuracy (%)	Recognition time (s)
DLSPC [21]	84.60	0.013
SRC [22]	90.01	0.073
LC-KSVD2 [23]	88.85	**2.4e-5**
Dictionary learning [20]	92.15	0.020
Proposed method	**92.21**	0.030

learning approach for sparse coding, which has the advantage of fast execution time. Dictionary learning [20] proposed a new multi-resolution dictionary learning method for face recognition, and the proposed method has higher performance than other methods. The execution time is the lowest in LC-KSVD2. However, in the actual in-vehicle system, the proposed method offers the best trade-off between accuracy and time complexity.

5. Conclusion

This paper proposed a face identification system to be used inside the vehicle with NIR camera. We also proposed an evaluation method that considers the relationship between registrants and non-registrants. In addition, the experimental results showed that the proposed method recognizes very accurately even when the face is partially covered or the face is overexposed to the light. Through this study, it was possible to collect NIR facial image data of the driver inside the vehicle.

Our future work will focus on the development of a way to learn online by continuously capturing data under various lighting conditions inside the vehicle. Also, we will study how to automatically obtain empirically calculated thresholds for image conversion and rejection of non-registrants.

References

[1] Q. Bi, M. Yang, C. Wang, and B. Wang, "An efficient hierarchical convolutional neural network for traffic object detection," in *Proceedings of IEEE Intelligent Vehicles Symposium (IV)*, 2018.

[2] K. Lim, Y. Hong, M. Ki, Y. Choi, and H. Byun, "Vision-based recognition of road regulation for intelligent vehicle," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, pp.1418-1425, 2018.

[3] Z. Yang, J. Li, and H. Li, "Real-time pedestrian and vehicle detection for autonomous driving," in *Proceedings of IEEE Intelligent Vehicles Symposium (IV)*, 2018.

[4] N. Crosswhite, J. Byrne, C. Stauffer, O. Parkhi, Q. Cao, and A. Zisserman, "Template adaptation for face verification and identification," *Image and Vision Computing*, 2018.

[5] E. M. Cherrat, R. Alaoui, and H. Bouzahir, "Convolutional neural networks approach for multimodal biometric identification system using the fusion of fingerprint, finger-vein and face images," *PeerJ Computer Science*, 2020.

[6] S. Farokhi, S. M. Shamsuddin, U. Sheikh, J. Flusser, M. Khansari, and K. Jafari-Khouzani, "Near infrared face recognition by combining zernike moments and undecimated discrete wavelet transform," *Digital Signal Processing*, 2014.

[7] B. Zhang, L. Zhang, D. Zhang, and L. Shen, "Directional binary code with application to polyu near-infrared face database," *Pattern Recognition Letters*, 2010.

[8] K. Pan, S. Liao, Z. Zhang, S. Z. Li, and P. Zhang, "Part-based face recognition using near infrared images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[9] T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, 1996.

[10] ATT Laboratories Cambridge, The orl database of faces, 1994.

[11] M. Ki, B. Cho, T. Jeon, Y. Choi, and H. Byun, "Face Identification for an in-vehicle Surveillance System Using Near Infrared Camera," in *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp.1-6, 2018.

[12] D. Chen, X. Cao, F. Wen, and J. Sun, "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.3025-3032, 2013.

[13] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.1701-1708, 2014.

[14] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.815-823, 2014.

[15] S. Ren, X. Cao, Y. Wei, and J. Sun, "Face alignment at 3000 fps via regressing local binary features," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.1685-1692, 2014.

[16] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, Vol.3, No.1, pp.71-86, 1991.

[17] S. Z. Li, R. Chu, S. Liao, and L. Zhang, "Illumination invariant face recognition using near-infrared images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.29, No.4, pp.627-639, 2007.

[18] S. Farokhi, U. Sheikh, J. Flusser, and B. Yang, "Near infrared face recognition using Zernike moments and Hermite kernels," *Information Sciences*, Vol.316, pp.234-245, 2015.

[19] X. Zhang, M. Peng, and T. Chen, "Face recognition from near-infrared images with convolutional neural network," in *2016 8th International Conference on Wireless Communications & Signal Processing (WCSP)*, 2016.

[20] X. Luo, Y. Xu and J. Yang, "Multi-resolution dictionary learning for face recognition," *Pattern Recognition*, Vol.93, pp.283-292, 2019.

[21] D. Wang and S. Kong "A classification-oriented dictionary learning model: Explicitly learning the particularity and commonality across categories," *Pattern Recognition*, 2014.

[22] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008.

[23] Z. Jiang, Z. Lin, and L. Davis, "Label consistent K-SVD: Learning a discriminative dictionary for recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013.

Min Song Ki

https://orcid.org/0000-0003-3458-5951
e-mail : kms2014@yonsei.ac.kr

She received the B.S degree in Computer Science from Duksung Women's University, Seoul, Korea in 2014. She is currently a Ph.D. student with the Department of computer science, Yonsei University, Seoul, Korea. Her research interests include intelligent vehicle, face recognition, object detection, object recognition, machine learning and deep learning.

Yeong Woo Choi

https://orcid.org/0000-0003-0364-236X
e-mail : ywchoi@sookmyung.ac.kr

He received B.S. degree in Electronic Engineering from Yonsei University, and M.S. and Ph.D. degrees in Computer Engineering from University of Southern California. He has been a professor in the Department of Computer Science at Sookmyung Women's University, Seoul, Korea. His research interests include image processing, pattern recognition and machine learning.