Continuous Real-Time Vehicle Driver Authentication Using Convolutional Neural Network Based Face Recognition

Ekberjan Derman[1] and Albert Ali Salah[2],[3]
[1] CuteSafe Technology, Gebze, Kocaeli, Turkey
ekberjan.derman@cutesafe.com
[2] Department of Computer Engineering, Bogazici University, Bebek, Istanbul, Turkey
[3] Future Value Creation Research Center, Nagoya University, Nagoya, Japan
salah@boun.edu.tr

***Abstract*— Continuous driver authentication is useful in the prevention of car thefts, fraudulent switching of designated drivers, and driving beyond a designated amount of time for a single driver. In this paper, we propose a deep neural network based approach for real time and continuous authentication of vehicle drivers. Features extracted from pre-trained neural network models are classified with support vector classifiers. In order to examine realistic conditions, we collect 130 in-car driving videos from 52 different subjects. We investigate the conditions under which current face recognition technology will allow commercialization of continuous driver authentication.** [1]

I. INTRODUCTION

Real-time biometrics for authenticating car drivers is a potentially important application. According to [1], there were an estimated 765,484 thefts of motor vehicles nationwide in the United States, just in 2016. Smarter cars that can authenticate their drivers could be a solution to mitigate this. But there are other applications. Car rental companies authorize designated drivers to driver their cars, and driver authentication can be used to enforce this. Similar concerns exist for long-distance drivers and taxi drivers, where the designated driver should use a vehicle, for specified periods. Systems that authenticate the driver's face can also be used in parallel for other face-based analyses, like drowsiness detection.

Current driver authentication approaches use several security features, including biometric and non-biometric methods. Biometric methods utilize driver's biometry, such as fingerprint or voice. For instance, BMW announces that its Z22 concept car would adopt fingerprint recognition to ensure that the car can only be used by authorized individuals [17]. The engine will only start when the driver places her index finger on the provided rotary switch. Honda FCX also uses fingerprint biometry for authentication, which requires the user to press and hold her thumb against a scanner for about three seconds to recognize its owner. Volvo SCC requires the user to hold the door handle for a few seconds to let the fingerprint recognition unit authorize the driver to use the car. Other car brands use voice recognition to ease in-vehicle operation, such as navigation or infotainment. Non-biometric approaches use driver's secret knowledge such as a password, or personal possession, like a physical token. Apart from the standard ignition key, in most cases, an authorized driver has a personal password to enter, a magnetic card, beacon, or an RFID tag to let the engine start.

The biometric approaches used by car manufacturers are typically restricted to their vehicle brands, and not always open to the more general public. Almost all those biometric and non-biometric approaches are performed either offline, or as a driving session starts, and thus are vulnerable to mid-session driving attacks, such as carjacking. A typical scenario could be that an authorized driver of a transportation firm can use some credentials to make the control center believe that the vehicle is used by a certain driver at the onset of the driving session, and then let another, unauthorized person to drive, violating the related regulations. A reliable, low-cost, and transportable real-time authorization system would help to improve detection of such frauds.

In this paper, we develop a system for real-time authentication of vehicle drivers. Our approach uses recent advances in deep neural network based face recognition, and deals with problems specific to the application scenario. We introduce an in-car face database collected from public videos, where we report experimental results in realistic acquisition conditions.

Facial verification systems can provide reliable, low-cost, and accurate online approaches for driver authentication tasks. Such a system usually consists of an image capturing device and computation unit that performs the verification. For vehicles, a camera that can capture the driver's image could be easily mounted inside, regardless of the car brand. Face verification based on images acquired from such a camera can be performed either by some kind of embedded computation device in car, or on a remote server. Here, we present a real-time driver facial verification framework that can be used to detect or avoid en route driving attacks. In our scenario, we first collect authorized driver's facial images, train a classifier based on their features, and perform real-time verification in some predefined time frequency during driving. For testing, we collected 130 in-car driving videos of 52 unique subjects.

For feature extraction, we applied transfer learning and

[1]This is the uncorrected author proof. Please cite this paper as E. Derman, A.A. Salah, Continuous Real-Time Vehicle Driver Authentication Using Convolutional Neural Network Based Face Recognition, Int. Workshop on Human Behavior Understanding, 2018.

fine-tuning on a pre-trained deep neural network model [18]. For classification, we examined Support Vector Machine (SVM) [3] and cosine similarity [6] based approaches. The advantage of our method is that it provides fast, low-cost, and reliable driver authentication compared to classical approaches.

The rest of this paper is organized as follows: In Section II, we present a brief overview of recent publications regrading driver authentication. In Section III, we describe our methodology. Our experimental results are provided in Section IV, while Section V concludes this paper.

II. RELATED WORK

While there are many academic and industrial approaches concentrating on face recognition and verification, only few focus on scenarios related to driver authentication [25]. For in-car scenarios, analysis mainly focuses on driver fatigue monitoring, cognitive load estimation, driver behaviour estimation, and driver emotion recognition [9], [10], [14], [24].

One of the earliest work related to driver face detection is described in [21], where a technique is proposed to detect stable face and eye regions both during day and night time. In their work, the authors used Infrared Rays Light Emitting Diodes (IR-LED) to illuminate the face of the driver. An IR-filter is also attached to the camera, which is placed at the inside of the inner mirror. The approach takes sequential images by blinking the IR-LED, turning it on and off alternately. By subtracting the on- and off-IR images, factors other than the light are eliminated, and a stable image of the face region is obtained. This system was used for driver drowsiness detection.

Several driver vigilance detection approaches are proposed for in-vehicle vision systems that monitor driver facial features [20], [23]. In these systems, a video camera is used without supplementary light. The processing range is sometimes reduced by motion detection on the captured video sequence. The face region is detected using a Viola-Jones-like face detector. After that, eye and mouth positions, as well as the face direction are determined using a predefined face model.

Illumination is a major problem for in-car face analysis. In [22], an image compensation method was proposed that includes illumination and color compensation to improve facial feature extraction of the driver. In this approach, a region of interest (ROI) is pre-defined as the region in which driver's face most frequently appears. After successfully detecting a face, the ROI is updated. Each image is compensated using a histogram equalization approach that matches the desired histogram specified from a reference image.

Other than using IR illumination and histogram equalization, an obvious way of dealing with challenging illumination conditions is by using 3D face sensors, but there are few applications that work with 3D [14].

The processing of the driver's face is often considered within a larger framework, where computing resources are not on board. In such scenarios, the acquired images are sent to a cloud server for processing, reducing the computation burden for in-car systems [15]. On the other hand, computational resources in cars are projected to be steadily developing, as autonomous or semi-autonomous driving (or parking) becomes available.

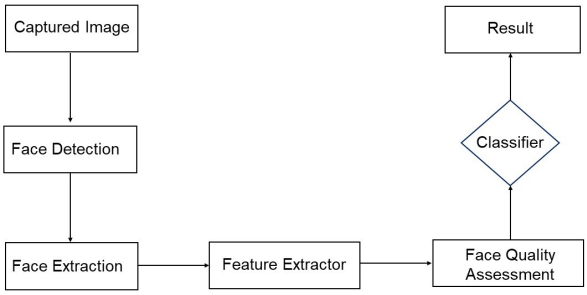

Fig. 1. The overall system flowchart of our proposed framework.

In [13], a framework is presented for performing driver's facial recognition for enhancing the security of vehicle parking spaces. In this work, a security camera installed on the side of the parking entrance is used to spot the vehicle. Once the vehicle enters the area, the driver has to pull down the side mirror to display her face. The facial region of the driver is detected using Haar-like features, followed by a simple recognition approach based on Eigenfeatures and Euclidean distance. In such a scenario, convenience is far more important than security, and a simple algorithm could be sufficient to serve the purpose. The evaluation pays more attention to false rejections than false accepts in such cases.

In [8], a novel facial landmark detection algorithm was proposed for real driving situations, based on an ensemble of local weighted random forest regressors with random sampling consensus (RANSAC) and explicit global shape models. This approach takes into account the irregular and dynamic characteristics of driving. In order to detect driver's facial landmarks, first the face of the driver is captured using a near-infrared camera and illuminator. The nose is localized and used as a reference point. The facial landmarks can serve for evaluation of driver fatigue, and similar states, as well as for proper geometric normalization of the face for authentication. Since sunglasses are quite frequently used during driving, the algorithm should be robust to partial occlusions such as caused by hair or sunglasses.

III. METHODOLOGY

The main flowchart of our proposed framework is described in Figure 1. We first acquire the driver's image using an in-vehicle camera, and send it to further processing, including face detection, feature extraction and classification. In this section, we describe the main components individually, starting from the database we have constructed.

A. Dataset Construction

There are no public resources for continuous driver authentication, but there are some related databases collected for driver face and head pose analysis. The PANDORA dataset was recently collected from car-driving scenarios, using a

Microsoft Kinect One RBG-D camera, and contains 110 annotated sequences using 10 male and 12 female actors, each of whom has been recorded five times [2]. However, while this database contains realistic driver poses, it has a different purpose and it is not recorded in a car. Subsequently, it does not possess the realistic and challenging illumination conditions we require, and the number of subjects is too small for authentication experiments.

The RS-DMV dataset was collected from outdoor driving scenarios with a dashboard camera, and contains different illumination conditions [16]. However, the purpose is to evaluate secure driving, and only seven outdoor driving sequences are available. Similarly, the CVC11 Driver Face dataset contains 606 image sequences from only four drivers, with gaze direction labels [5]. Obviously, the number of subjects is too small in either case for a realistic authentication scenario.

To deal with the shortcomings of available datasets, we have collected 1172 videos of 1042 different drivers from YouTube searched by using the keyword "CarVLog", and use it as our evaluation dataset[2]. Vlog, i.e. video blog, is a popular form of multimedia recording, and there are many users who record videos from dashboard cameras or mobile phones placed on holders in the dashboard area while driving. These typically include speech. Continuous driver authentication should be performed with segments including speech, as well as segments that do not contain speech. We acknowledge here that using video blogs biases the database towards more segments with speech, but since these present more challenging conditions than still faces, we argue that this is acceptable.

All the collected videos are taken in real-world driving scenarios, and thus present challenges such as extreme lighting conditions, head pose variance, etc. Among these collected videos, only 52 subjects have more than one video available, which gives us a total of 130 videos. Among them, 29 are female and 23 are male, and 26 videos contain drivers wearing sun-glasses on most of the driving session. The videos were manually checked, and non-driving sequences were eliminated. Also, vlogs typically start with non-driving segments, where the vlogger enters the car, or sets up the system. These frames are eliminated from the analysis.

For each subject, we use images from one video as positive samples to train our classifier, and the rest of the videos from the same person are used in the test database. We selected 990 videos from subjects with only one video to serve as negative samples during the training step. An alternative usage would be to use them for increasing impostor accesses in the testing protocol. However, reducing false rejections is more important than reducing false accepts for the proposed scenario, and this suggests that additional videos are more useful for improving training.

[2]Youtube has typically two licenses for its videos. Creative Commons license allows the download and use of the video, whereas Standard Youtube License requires permission from the owner for any re-publication. We do not make any videos with Standard Youtube License available.

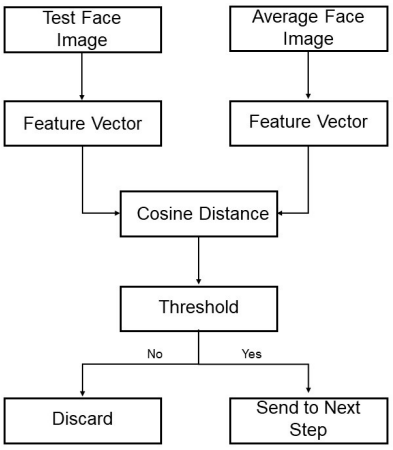

Fig. 2. Face quality assessment flowchart.

B. Face Detection and Extraction

For detecting the face region, we use the Dlib frontal face detector [12]. This detector uses a Histogram of Oriented Gradients (HOG) feature, combined with a linear classifier, an image pyramid, and a sliding window detection scheme. Comparing to the popular OpenCV face detector, which uses the Viola-Jones algorithm, Dlib's face detector gives more robust detection results, even under some challenging situations [11]. Once the face is detected, we crop the facial region from the background, and send it to our feature extractor. On our test dataset, the typical face region ranges between 80x80 to 180x180 pixels.

C. Feature Extraction

We use pre-trained convolutional neural networks as feature extractors. The pre-trained VGG-Face CNN descriptors described in [18] were used with fine-tuning paradigm. The original structure of this model is shown in Table I. We remove the last layer, which specializes in the classification task of the original training, and kept all the previous layers (i.e. up to fc7), and used this structure to extract our feature vector. We updated pre-trained model parameters with our training dataset. The output of this layer gives us a vector of size 4096 for each face image, which is used with the classifier.

D. Face Quality Assessment

Due to the camera view angle and driver's head pose, not each and every extracted face image is good enough to be sent to classification. Besides, even though the Dlib's frontal face detector is robust, there are still some false detections in the initial step. Therefore, we need to first assess the quality of the face image before further processing. By this step, we can filter out low-quality images to reduce unnecessary computation on the classification side. Also, these poor detections will invariably fail to pass the authentication test, causing false rejections.

To implement the quality assessment subsystem, we selected one thousand different frontal face images from the Labeled Face in the Wild (LFW) dataset [7], obtained their

TABLE I

DETAILS OF THE VGG FACE CNN ARCHITECTURE. FULLY CONNECTED (FC) LAYERS ARE LISTED AS "CONVOLUTION" AS THEY ARE CONSIDERED A SPECIAL CASE OF CONVOLUTION BY AUTHORS OF [18].

layer	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
type	input	conv	relu	conv	relu	mpool	conv	relu	conv	relu	mpool	conv	relu	conv	relu	conv	relu	mpool	conv
name	–	conv1_1	relu1_1	conv1_2	relu1_2	pool1	conv2_1	relu2_1	conv2_2	relu2_2	pool2	conv3_1	relu3_1	conv3_2	relu3_2	conv3_3	relu3_3	pool3	conv4_1
support	–	3	1	3	1	2	3	1	3	1	2	3	1	3	1	3	1	2	3
filt dim	–	3	–	64	–	–	64	–	128	–	–	128	–	256	–	256	–	–	256
num filts	–	64	–	64	–	–	128	–	128	–	–	256	–	256	–	256	–	–	512
stride	–	1	1	1	1	2	1	1	1	1	2	1	1	1	1	1	1	2	1
pad	–	1	0	1	0	0	1	0	1	0	0	1	0	1	0	1	0	0	1

layer	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
type	relu	conv	relu	conv	relu	mpool	conv	relu	conv	relu	conv	relu	mpool	conv	relu	conv	relu	conv	softmx
name	relu4_1	conv4_2	relu4_2	conv4_3	relu4_3	pool4	conv5_1	relu5_1	conv5_2	relu5_2	conv5_3	relu5_3	pool5	fc6	relu6	fc7	relu7	fc8	prob
support	1	3	1	3	1	2	3	1	3	1	3	1	2	7	1	1	1	1	1
filt dim	–	512	–	512	–	–	512	–	512	–	512	–	–	512	–	4096	–	4096	–
num filts	–	512	–	512	–	–	512	–	512	–	512	–	–	4096	–	4096	–	2622	–
stride	1	1	1	1	1	2	1	1	1	1	1	1	2	1	1	1	1	1	1
pad	0	1	0	1	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0

Fig. 3. The calculated average of the one thousand face images from the LWF dataset. The feature vector of this image is used to compare whether the input face image is of enough quality in our quality assessment step.

Fig. 4. Some examples of false or unqualified face detection result. We can filter these out with the proposed face quality assessment process.

average image, and extracted its feature vector according to method just described. The system calculates the cosine distance between this feature vector with each detected face during operation. Once this distance is higher than a pre-defined threshold value, we consider it as of enough quality and send it to the next step.

We employ Support Vector Machines for the classification task [3]. For the training of the classifiers, we first collect 50 face images of the driver. Taking into account the challenging situation of driver's head pose and illumination conditions in the real scenario, we enrich the set of the original facial images with rotated versions, as well as simulating dark and bright conditions. The data augmentation on original images uses the following steps:

1) Rotation: we rotate each image from -15^o to 15^o, with 5^o increments.

2) Brightness change: we obtain one dark and one bright version of each image after rotation by Equation 1:

$$I_{new} = \alpha I_{original} + \beta \tag{1}$$

where, $I_{original}$ refers to the original image intensity and I_{new} refers to new intensity after the operation. For obtaining dark image, we use $\alpha = 1.0$ and $\beta = -15$, while for bright image, we use $\alpha = 1.5$ and $\beta = 30$.

3) Flip: once the rotation and brightness change steps are performed, we flip the resulting images and add to the training set.

In this manner, we get 42 images from one single image, and a total of 2100 images using our original 50 face images from a single driver. This ensures that a very short sequence is usable for establishing the registration of the driver with the system.

The feature vectors of these 2100 images serve as our positive examples. We collect 5000 face images of different subjects, not included in the test dataset, extract their features and let them be serve as negative examples. We use five-fold cross validation to estimate the optimal values of SVM parameters γ and ν during training. For each subject, we train one classifier, save it locally and use it to predict whether the test image belongs to the registered driver or not. As our goal is to perform authentication, after having the test image, we use the expected driver's trained classifier to perform the prediction.

An obvious extension is using multiple genuine drivers in the training scenario. This is a typical case for taxi drivers, where the owner of the taxi leases the vehicle to two different designated drivers to maximize the usage. These drivers use fixed shifts, and it is not uncommon that for a given city, the shift times are fixed across all taxi companies to allow easy substitution of drivers. Our proposed method allows a straightforward extension in this case, where instead of a single driver, multiple drivers are used for extracting positive samples in the training stage, and the negative samples are left unchanged. Such a system has the advantage of not requiring the driver to use an additional step for entering ID to the system, as the classifier is not ID-specific, but vehicle-specific in this case.

E. Face Tracking

Once we get a positive feature classification result, we track the face during the upcoming frames so as to reduce the computation workload. The facial region's bounding box serves as input for our correlation tracker [4]. If the tracking fails, face detection and the consecutive steps are performed again. The correlation tracker returns the peak to side-lobe ratio, which is a value that measures the confidence of the tracker on whether the object is inside the target region or not. We use this ratio to decide whether our tracking is successful or not. In our test case, we choose ratio values of 8.0 and higher to accept tracker success.

F. Cloth Verification

Considering the challenging illumination and head pose conditions in our scenario, chances are high for falsely rejecting the correct driver. For that, we also take into account the cloth region of the driver to be verified and serve as a supplementary verification step in order to reduce false alarms. In our cloth verification step, once we successfully detected and tracked the authorized driver, we choose a rectangle region of size 50x50 below a certain pixel of the face, and consider it as our cloth region (see Figure 5). The system extracts foreground and background pixels of this selected region and decides whether the current frame's cloth region is the same as that of the previous frames or not. In a real-world scenario, since the possibility of a driver's changing their appearance during driving is relatively low, we use the cloth verification result to discard some false alarms raised by our face classifier.

To be specific, we use a Gaussian Mixture-Model (GMM) based foreground/background segmentation algorithm [26], [27]. This algorithm gives us a binary mask for the driver. We use five components per mixture in our tests, and obtain the cloth verification ratio using Equation 2:

$$R = N_{background}/N_{total} \tag{2}$$

where, $N_{background}$ and N_{total} refer to total background pixels, and the overall pixel count of the selected cloth region, respectively. R indicates the verification ratio. $R = 1.0$ indicates that all pixels are regarded as background, that is, the test cloth region is exactly the same as in the previous cloth region, while $R = 0.0$ means all the pixels are considered as foreground and the test cloth region is totally different from the cloth region of the previous frame. This ratio can be thresholded to decide whether the target cloth region belongs to the same user or not. We choose a threshold value of 0.8 to assume the cloth region is of the same person.

Occasional rejections by the face or cloth verifiers should not raise alarms. During our tests, after a positive face authorization, if there exist two consecutive frames that the face verifier gives negative results while the cloth verifier gives a positive, we discard the face verification result to still consider the person as a positive driver. If the face verifier gives a negative result in the next upcoming third frame, then we consider the driver as negative. In this manner, we are able to reduce some amount of false rejections. We only consider two consecutive frames for cloth verification, since in real-world scenarios, the cloth region selected below the driver's face sometimes points to regions other than the cloth area, such as the steering wheel (see Figure 6).

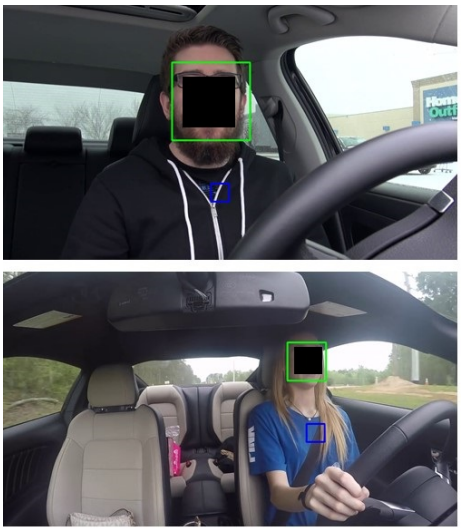

Fig. 5. Illustration of cloth region detection. The blue rectangle for the cloth area is best viewed in color. We intentionally added additional mask on face regions to avoid YouTube license conflict.

Fig. 6. Examples of false cloth region detection. Top: face (green) and cloth region (blue) detection results. Bottom: their corresponding cloth region binary mask extracted by GMM. As we can see, in the left image, the steer wheel is regarded as background (cloth) and this leads to the incorrect consideration of the true cloth region as foreground in the right image. We intentionally added additional mask on face regions to avoid YouTube license conflict.

IV. EXPERIMENTAL RESULTS

For continuously verifying the driver, we choose 30 frames out of each test video, which gives us a total of 52389 genuine scores (with some dropped frames) and 201600 impostor scores. We report the Receiver Operating Characteristic (ROC) curve [28] (see Figure 7), True Positive Rate (TPR) corresponding to 1% and 0.1% False Positive Rate (FPR) of our proposed framework on this sample (see Table II). For comparison purpose, we provided test results based on same architecture but the CNN-based feature extractor is the pre-trained VGG model without fine-tuning, as well as the popular FaceNet Inception based NN4 model [19].

As we can see from these test results, our proposed method gives us a relatively low FPR. Also, feature extraction based on VGG model with fine-tuning gives us better results than

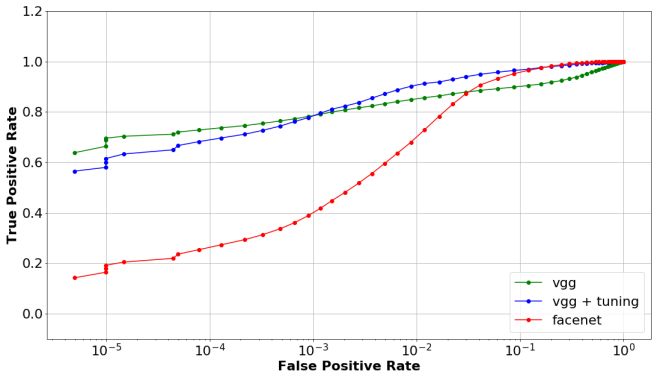

Fig. 7. Obtained ROC curve of our proposed framework based on our test dataset (using threshold values of 0.8 and 0.4 for cloth region verification and face quality assessment filter, respectively).

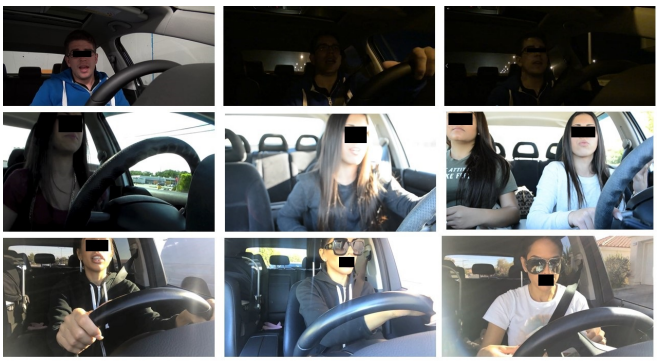

Fig. 8. Some examples of challenging videos. The images of the first column are from the videos used for registration, while the rest are from the verification step. The first row shows the extreme illumination change, the second row illustrates different head pose, camera location and multiple face appearances, and the third row demonstrates a driver wearing sunglasses. We intentionally added additional mask on face regions to avoid YouTube license conflict.

Fig. 9. One typical example of the same user's facial images captured during different time period, facial expression, and illumination conditions. The top left facial image is from the video used for the training step, and the rest are test images. We intentionally added additional mask on face regions to avoid YouTube license conflict.

TABLE II

TRUE POSITIVE RATE (TPR) CORRESPONDS TO 1% AND 0.1% FALSE POSITIVE RATE (FPR) OF OUR PROPOSED FRAMEWORK COMPARING WITH VGG WITHOUT FINETUNING AND FACENET

Methods	1% FPR	0.1% FPR
VGG	85.6%	75.16%
VGG + Finetuning	**90.8%**	**79.4%**
FaceNet	72.8%	41.6%

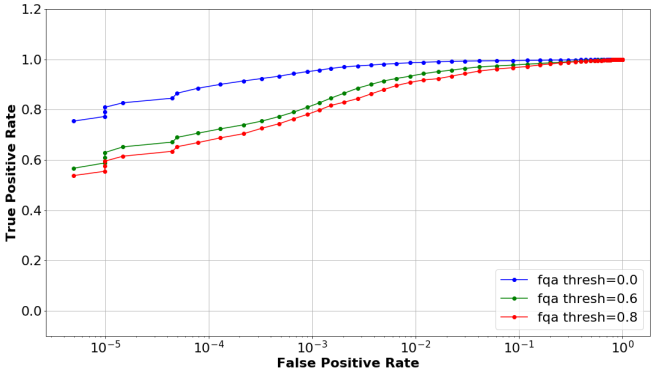

Fig. 10. ROC curve illustrating the impact of our proposed face quality assessment filter to the overall system performance.

VGG without fine-tuning and FaceNet models. The reason for our obtained TPR is that the test videos captured from actual scenarios present extremely challenging conditions, such as illumination change, head pose variation, sun-glasses, facial expressions and makeup, other user's appearance, etc. Also, the difference of time period for one user's multiple videos may vary. For instance, there are users that upload a new video just several days after their first one, while some upload videos after several months or even years (see Figure 8 and 9).

This large inter-session time span poses a great challenge, but corresponds to a possible scenario, where the user does not drive the car for a long time. Of course, it may be possible to mitigate some of the effects by setting a maximum amount of days before the stored template is considered to be old, and a new one is required by the system.

Meanwhile, we also investigated the impact of our proposed face quality assessment step and cloth region verification to our final system performance. The related ROC curves for these are illustrated in Figure 10 and Figure 11 respectively.

From these ROC curves we can deduce that the face quality assessment filter can help to reduce the FPR to some degree, which depends heavily on the false alarm of the face detector. Meanwhile, the FPR grows as the cloth verifier's threshold value increase, and vice versa. Thus, choosing a good threshold value for cloth verifier can help to reduce the overall false alarm rate.

We have investigated a scenario in which there are two authorized users for one vehicle, which is typically true for a taxi that has a day time and a night time driver, or a family car shared among a couple. As mentioned earlier, it is

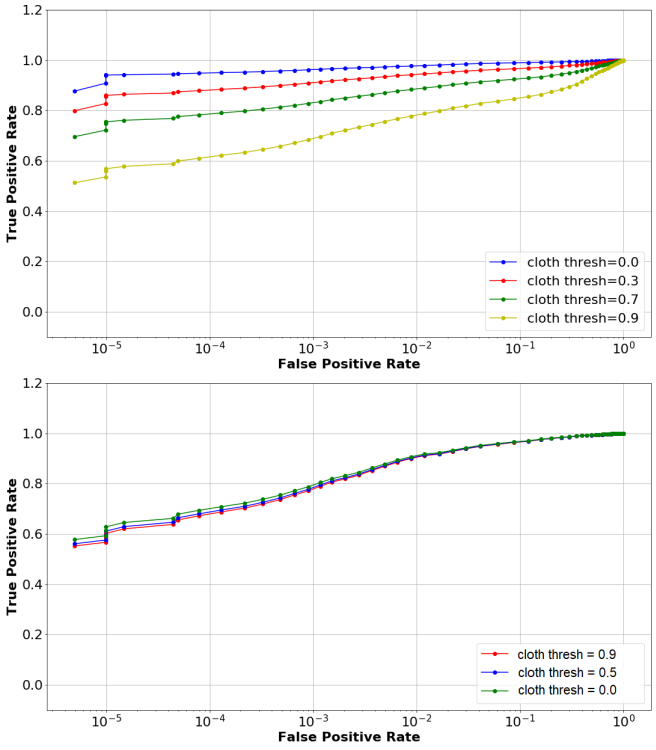

Fig. 11. ROC curve illustrating the impact of our proposed cloth verifier to the overall system performance. Top: system with pre-trained VGG model. Bottom: system with pre-trained VGG model plus fine-tuning operation.

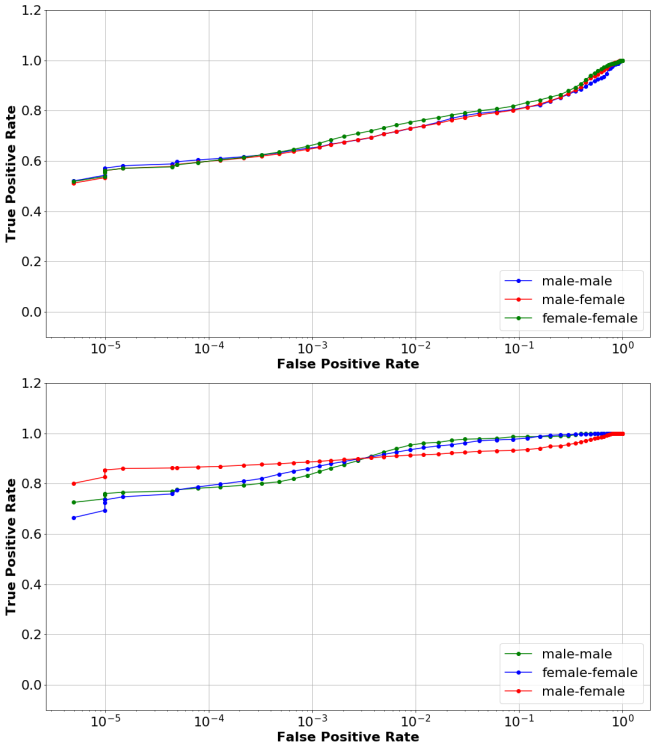

Fig. 12. ROC curve illustrating the system performance for multiple authorized drivers. Top: system with pre-trained VGG model. Bottom: system with pre-trained VGG model plus fine-tuning operation.

TABLE III

TRUE POSITIVE RATE (TPR) CORRESPONDS TO 1% AND 0.1% FALSE POSITIVE RATE (FPR) OF OUR PROPOSED FRAMEWORK FOR MULTIPLE DRIVERS TEST

User Group	1% FPR	0.1% FPR
Male + Male	94.4%	89.5%
Male + Female	90.3%	87.2%
Female + Female	92.7%	84.7%

possible to train one classifier that separates two drivers from the rest. For that, we organized three different test groups, in which the two drivers could have the same sex or not: i.e. male-male, male-female and female-female, respectively. Each group consists of 10 different subjects in total. We have used the same method for registration. For verification, the trained classifier of each group are used for prediction, and both drivers are tested for positives. For testing, we used one video from each driver and the videos used for registration of other drivers were treated as impostor accesses. The obtained ROC curve is shown in Figure 12, and related TPR and FPR are given in Table III. As expected, the performance for classifying two drivers as a single class is lower than classifying for a single driver. As far as we observe, to some degree, this performance is affected by the limited sample size and their quality chosen for our testing.

V. CONCLUSIONS

In this paper, we proposed a face and appearance based real-time driver authentication framework. We combine a CNN-based face classifier with a GMM-based appearance verifier to decide whether the operating driver in the vehicle is authorized or not. We collected real-world scenario based videos from the Internet, and constructed our own test database. Our test results show that we can provide relative acceptable true acceptance and very low false acceptance rate for our challenging scenario of real-time driver authentication. Our proposed method can deliver low-cost, continuous and real-time driver verification.

VI. ACKNOWLEDGMENTS

The authors would like to thank Kazim Esen and Esra Civik for their help on collecting and organizing test data.

REFERENCES

[1] FBI uniform crime reports: Motor vehicle theft data in 2016. 2016.

[2] Guido Borghi, Marco Venturelli, Roberto Vezzani, and Rita Cucchiara. Poseidon: Face-from-depth for driver pose estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[3] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, Sep 1995.

[4] Martin Danelljan, Gustav Häger, Fahad Khan, and Michael Felsberg. Accurate scale estimation for robust visual tracking. In *British Machine Vision Conference, Nottingham, September 1-5, 2014*. BMVA Press, 2014.

[5] Katerine Diaz-Chito, Aura Hernández-Sabaté, and Antonio M López. A reduced feature set for driver head pose estimation. *Applied Soft Computing*, 45:98–107, 2016.

[6] E. Garcia. Cosine similarity tutorial. In *Minerazzi Tutorial Section*, 04 October, 2015.

[7] Gary B Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, Technical Report 07-49, University of Massachusetts, Amherst, 2007.

[8] Mira Jeong, Byoung Chul Ko, Sooyeong Kwak, and Jae-Yeal Nam. Driver facial landmark detection in real driving situation. *IEEE Transactions on Circuits and Systems for Video Technology*, 2017.

[9] Hang-Bong Kang. Various approaches for driver and driving behavior monitoring: a review. In *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on*, pages 616–623. IEEE, 2013.

[10] Sinan Kaplan, Mehmet Amac Guvensan, Ali Gokhan Yavuz, and Yasin Karalurt. Driver behavior analysis for safe driving: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 16(6):3017–3032, 2015.

[11] Davis E King. Dlib vs opencv face detection. *comparison result is available at: https://www.youtube.com/watch?v=LsK0hzcEyHI*.

[12] Davis E King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10(Jul):1755–1758, 2009.

[13] Zahid Mahmood, Tauseef Ali, Shahid Khattak, Samee U Khan, and Laurence T Yang. Automatic vehicle detection and driver identification framework for secure vehicle parking. In *Frontiers of Information Technology (FIT), 2015 13th International Conference on*, pages 6–11. IEEE, 2015.

[14] Hyeonjoon Moon and Kisung Lee. Biometric driver authentication based on 3d face recognition for telematics applications. In *International Conference on Universal Access in Human-Computer Interaction*, pages 473–480. Springer, 2007.

[15] Syrus C Nemat-Nasser and Andrew Tombras Smith. Driver identification based on face data, June 3 2014. US Patent 8,744,642.

[16] Jesus Nuevo, Luis M. Bergasa, and Pedro Jiménez. Rsmat: Robust simultaneous modeling and tracking. *Pattern Recognition Letters*, 31:2455–2463, December 2010.

[17] Palmer Owyoung. Cars that use biometric locks, 2016.

[18] Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, et al. Deep face recognition. In *BMVC*, volume 1, page 6, 2015.

[19] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. *CoRR*, abs/1503.03832, 2015.

[20] Mohamad-Hoseyn Sigari, Mahmood Fathy, and Mohsen Soryani. A driver face monitoring system for fatigue and distraction detection. *International journal of vehicular technology*, 2013, 2013.

[21] Isamu Takai, Kazuhiko Yamamoto, Kunihito Kato, Keiichi Yamada, and Michinori Andoh. Detection of the face and eye region for drivers' support system. In *Sixth International Conference on Quality Control by Artificial Vision*, volume 5132, pages 375–381. International Society for Optics and Photonics, 2003.

[22] Jung-Ming Wang, Han-Ping Chou, Sei-Wang Chen, and Chiou-Shann Fuh. Image compensation for improving extraction of driver's facial features. In *Computer Vision Theory and Applications (VISAPP), 2014 International Conference on*, volume 1, pages 329–338. IEEE, 2014.

[23] Jung Ming Wang, Han-Ping Chou, Chih-Fan Hsu, Sei-Wang Chen, and Chiou-Shann Fuh. Extracting driver's facial features during driving. In *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*, pages 1972–1977. IEEE, 2011.

[24] Guosheng Yang, Yingzi Lin, and Prabir Bhattacharya. A driver fatigue recognition model based on information fusion and dynamic bayesian network. *Information Sciences*, 180(10):1942–1954, 2010.

[25] Hailing Zhou, Ajmal Mian, Lei Wei, Doug Creighton, Mo Hossny, and Saeid Nahavandi. Recent advances on singlemodal and multimodal face recognition: a survey. *IEEE Transactions on Human-Machine Systems*, 44(6):701–716, 2014.

[26] Zoran Zivkovic. Improved adaptive gaussian mixture model for background subtraction. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 2, pages 28–31. IEEE, 2004.

[27] Zoran Zivkovic and Ferdinand Van Der Heijden. Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern recognition letters*, 27(7):773–780, 2006.

[28] Mark H Zweig and Gregory Campbell. Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine. *Clinical chemistry*, 39(4):561–577, 1993.