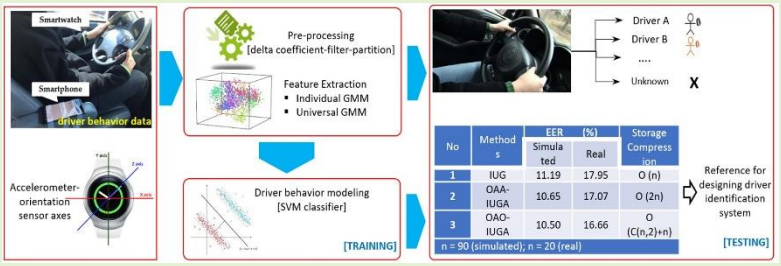
Smartwatch-based Open-set Driver Identification by Using GMM-based Behavior Modeling Approach

Rekyan Regasari Mardi Putri, Ching-Han Yang, Chin-Chun Chang and Deron Liang, Member, IEEE

***Abstract*— Driver identification must be studied because of the development of telematics and Internet of Things applications. Many application services require an accurate account of a driver's identity; for example, usage-based insurance may require a remote collection of data regarding driving. Recently, a Gaussian mixture model (GMM)-based behavioral modeling approach has been successfully developed for smartwatch-based driver authentication. This study extends the GMM-based behavioral modeling approach from driver authentication to open-set driver identification. Because the proposed approach can help for identifying illegal users, it is highly suitable for real-world conditions. According to a review of the relevant literature, this study proposed the first smartwatch-based driver identification system. This study proposed three open-set driver identification methods for different application domains. The result of this research provides a reference for designing driver identification systems. To demonstrate the feasibility of the proposed method, an experimental system that evaluates the performance of the driver identification method in simulated and real environments was proposed. The experimental results for the three proposed methods of driver identification illustrated an equal error rate (EER) of 11.19%, 10.65%, and 10.50% under a simulated environment and an EER of 17.95%, 17.07%, and 16.66% under a real environment.**



***Index Terms*— biometric identification, driver identification, Gaussian mixture model, smartwatch.**

1

I. Introduction

WITH the development of telematics and Internet of Things applications, many in-vehicle sensing devices, such as Global Positioning System (GPS), On-Board Diagnostics-II (OBD-II), Inertial Measurement Unit (IMU), and smart mobile devices are widely used in car networking. Because the Internet of Vehicles has numerous application services, further studies must be conducted for driver identification since many application services require proof of the driver's identity (Fig. 1); For example, usage-based insurance may remotely collect data regarding driving (driving time and driving habits). A backend platform may convert that driving data into a risk score to be used for adjusting the premium level or for offering various rewards. Driver identification can be further subdivided into two categories: closed-set driver and open-set driver identification. Closed-set driver identification is based on a set of known users; according to the behavior characteristics of the target, the most similar user from the set is selected. Open-set driver identification is not limited to a known-user set and must reject unknown users. Open-set driver identification is suitable for real-world situations, in which illegal users impersonate registrants to invade the personal application service.

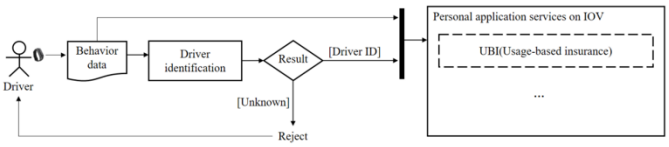


Fig. 1. Application services structure that require the driver's identity.

Biometric identification can be applied to driver identification. This identification can be categorized based on two factors: physiological and behavioral characteristics. The physiological characteristics are identified using features, such as fingerprint, palm shape, and iris and retina patterns, whereas behavioral features include the signature, pace, and keyboard

Manuscript received xxx; revised yyyy; accepted zzz. Date of publication zxzx;date of current version May, 2020. (Corresponding author: Rekyan Regasari Mardi Putri)

Rekyan Regasari Mardi Putri is with the Department of Computer Science and Information Engineering, National Central University, Taoyuan City, 32001, Taiwan (e-mail: rekyan.rmp@ub.ac.id)

Ching-Han Yang is with the Software Research Center, National Central University, Taoyuan City 32001, Taiwan (e-mail: yang.chinghan@gmail.com)

Chin-Chun Chang is with the Department of Computer Science and Engineering, National Taiwan Ocean University, Keelung City, 20224, Taiwan (e-mail: cvml@mail.ntou.edu.tw)

Deron Liang is with the Department of Computer Science and Information Engineering, National Central University, Taoyuan City, 32001, Taiwan (e-mail: drliang@csie.ncu.edu.tw)

tap rhythm. Both types of characteristics can be used for identifying individuals. Because physiological characteristics are distinct, they are suitable for developing highly reliable identification methods, such as one-off person identification. By contrast, behavioral characteristics may not be as accurate and stable as physiological characteristics because human behavior is inconsistent. However, behavioral traits have certain advantages over physiological traits for developing continuous and transparent recognition methods, such as personal re-identification [1, 2]. Personal re-identification is essential in driver identification, for example, under a theft scenario. It is the theft that happens when the car's owner is making a short stop and leaves the car without turning the engine off. In this case, without re-identification, the thief can easily drive the stolen car. Regarding this scenario, re-identification must be taken continuously because the theft can happen anytime. Therefore, this study used behavioral characteristics. Based on the types of behavioral characteristics of drivers, the current literature have examined throttle and pedal pressure signals [3-6], angle rotation of the steering wheel [6, 7], a force of the steering wheel [8], the weight of the driver's seat and back of the chair [9], operating habits of the driver's hand [10-12], vehicle' 's CAN bus through the On Board Diagnostics 2 (OBD-II) and CarbigsP (OBD-II scanner [13-15], inertial and exteroceptive sensor [16], and physiological feature [14] Most of the current research uses or embeds many measurement and control sensors.

Smartwatches are crucial examples of the development of light, small, versatile, mobile, wearable, and intelligent devices. According to statistics calculated by Gartner, a market research company, from 2017 to 2021, the number of smartwatches worldwide will increase from 41.5 million to 81 million. Furthermore, smartwatches are used by several car manufacturers as a car door lock remote controller, and in some cases, to start car engines; this is because smartwatches have multiple built-in sensors, which are suitable for these applications. Sensors on smartwatches can monitor the physiological status of the user and analyze the hand movement. Lee et al., [10] have used smartwatches to capture the driver's hand movements to detect fatigue, whereas Yang et al. [11] captured the exercise habits of the driver's hand and obtained driver certification.

Although numerous verification and identification applications, such as speaker identification [17, 18], signature verification and recognition [19], handwriting recognition [20], human identity verification [2], biometric person authentication [1], and fingerprint verification [21], driver identification [12-16] have been developed, no reliable methods have been proposed for applying the operating habits of the driver's hand for identification using a smartwatch. Therefore, this research aimed to use the smartwatch to capture the behavioral traits of the driver's hand as a feature for identification.

Driver identification is performed in two stages, training and testing. A training stage is used to develop the model, whereas the testing stage is used to identify the user based on the provided inputs. A Gaussian mixture model (GMM) and the improvement is frequently applied to perform feature extraction, such as speaker and person identification [18, 22] and driver identification [3-5, 15]. GMM have been broadly selected because GMMs allow for mixed membership of points to clusters and are very flexible. Other approaches are used for driver identification, such as deep learning [13], K-Nearest Neighbor (KNN), Random Forests, Multiplayer Perceptron, Adaboost, Decision Tree [12, 14], improvement of GMM [15, 23]. Li, Zhengping et al. [12] stated that KNN, Random Forests, Multiplayer Perceptron, and Adaboost Algorithm have good accuracies when the data is limited and will decrease when the data is larger. Wang, Wenshuo et al. [15] concluded that the BGGMM-HMM would suffer a substantial computational cost due to its structural complexity.

The GMM approach uses the likelihood value of the GMM to determine if the input pattern is drawn from the data distribution modeled by the GMM. The GMM can be adopted to model for individual driver and this model is called an individual driver model (IDM). The IDM can appropriately identify individual drivers when the patterns of the drivers are different; however, identification is difficult for drivers having similar patterns. Yang et al. [11] proposed a GMM-based behavioral modeling approach, which combines the IDM with a universal driver model (UDM) modeling to overcome the problem encountered by a conventional GMM. The IDM is a model that captures the pattern of a single driver, and the UDM is a GMM established based on the patterns of many drivers. The IDM is specific and not general, whereas the UDM is general and not specific. Yang et al. used the IDM and UDM as base learners and combined them by stacking generalization which is called the IUG-based method. This IUG-based method is a baseline method in this study. However, the results of the IUG-based method were validated for driver authentication and not identification.

Driver identification requires multiclass classification because the number of drivers to be identified is usually more than two. Two common training approaches are available to train SVMs for multiclass problems: one-against-all (OAA) and one-against-one (OAO). In the OAA approach, a data point is classified to a class if its SVM accepts the point, and the SVMs of other classes reject it. This approach is accurate for tightly clustered classes; however this approach can leave regions of the feature space undecided, where more than one class accepts or all classes reject the data point [20]. The OAO approach involves $N(N-1)/2$ binary SVM classifiers. Each classifier is trained to separate each pair of classes. The OAO is often faster than OAA approach because the binary SVMs of the OAO approach are trained for two classes and fewer SV support vectors. The two approaches can provide different results on different cases, depending on the application domain, and the approach of classifier construction [19, 20, 23].

An open-set identification approach is required to identify an illegal user. Reynolds et al. [17] mentioned in their research on speaker identification that the problem of open-set identification can be solved using a closed set identification technology combined with identity authentication technology. On the basis of the Reynolds solution, this research extends and improves the modeling approach of Yang [11] for driver

identification. Three methods were proposed for smartwatch-based open-set driver identification: (1) IUG-based method, 'an individual and universal combining driver models for the GMM open-set identification method based on Yang's approach,; (2) OAO-IUGA, a combination of one-against-one training of the closed set identification model with the authentication model obtained through IUG modeling; and (3) OAA-IUGA, a one-against-all training of a closed set identification model with the authentication model obtained using IUG modeling.

To evaluate the methodology proposed in this study, the behavioral data of drivers were collected from driving simulation and real environments. The driving behavior data of each participant can be divided into straight, left-turn, and right-turn data, which will be used to construct a driver behavior model for each type of behavior. The experiment aimed to validate the superiority of the three proposed methods compared with the conventional approaches (GMM) and to provide guidelines of driver identification based on the three methods. The performances of the methods were evaluated by the respective equal error rate (EER) and space storage.

The experimental results showed that the proposed methods can considerably improve the accuracy of the GMM method on both simulated and real environments. The contributions of this study are two-fold: 1) the first smartwatch-based open-set driver identification, and 2) a reference guide of designing driver identification systems.

The remainder of this paper is organized as follows: Section II examines related studies. Section III describes the basic concepts of the proposed approach. Sections IV and V present the proposed methods and experimental results. Finally, the conclusions and future prospects are provided in Section VI.

II. DRIVER IDENTIFICATION

Researches on the analysis and identification the driver, can be classified based on the types of behaviors, such as the driver stepping on the throttle and brake pedal pressure signals, steering wheel angle and handgrip force, weight of the driver's seat and back of the chair, and the operating habits of the driver's hand. Yang et al. [11] proposed a new GMM-based method that can improve the GMM for driver authentication based on the motion sensor of the smartwatch. This method used a stacking approach to integrate two driver behavior models, namely the IDM and UDM, for driver authentication. The experimental results indicated that this approach had EERs of 4.62% and 7.86% for simulation and real environments.

In this study, the GMM-based behavioral modeling approach [11] was extended to the open-set driver identification problem, which is more difficult than the driver authentication problem and has rarely been studied in the relevant literature. Therefore, this study will be the first smartwatch-based driver identification. The proposed methods are designed to improve the performance of driver identification.

According to the previous research, no conclusive result has been obtained on the superiority of any approach in all domains. The performance of these approaches depends on applications and the construction of classifiers. For example, OAA is better than OAO for fingerprint-based identification [22]; however, OAO is superior to OAA for finger vein authentication. Özgündüz and other scholars [19] concluded that their OAA was superior to OAO for signature recognition, and thus, OAA was used. In handwriting recognition [20], no claim of an absolute superiority was made between the two types of support vector machine model training; OAA is considered superior for fewer numbers of classes, whereas OAA and OAO have similar results for moderate numbers of classes, and OAO is superior to OAA for large numbers of classes. As mentioned previously, no research has implemented these strategies for driver identification; thus, this study implemented the three proposed methods for driver identification and evaluated them in this domain. The conclusion is necessary to provide a reference for the future development of driver identification related to the field of biometric authentication and identification

III. GMM-BASED BEHAVIORAL MODELING APPROACH

The GMM-based behavioral modeling approach [11]. for driver identification is explained in the following three sections (Fig. 2). The first and second sections discuss preprocessing and feature extraction and the last section examines model construction, which is the decision of the driver model. The GMM-based behavioral combines two base models: the IDM and UDM. The two GMM-based driver models were developed to extract ten features from the preprocessed data. Then, the two types of features were separated to train two base SVMs. The output of the two-base SVMs were stacked to train another SVM for developing the driver behavior model.

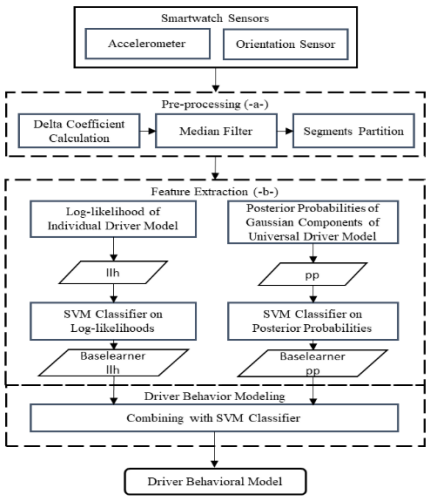


Fig. 2. GMM-based behavioral modeling approach

A. Preprocessing

In this research, the data required for the models were obtained from the smartwatch sensor, which covered 3-axis accelerometer (Acc) and 2-axis orientation (Ori) sensor signals. In the data preprocessing section, the signal data was collected, the noise was removed, and several features such as delta-coefficient values of the sensor signals were calculated. The complete sequence of data was then partitioned into segments to ensure that each segment was focused on a particular behavior of the driver: straight, left and right turning. The data were segmented in that manner because the behavior in the three segments is different.

The four types of signals obtained from the accelerometer and orientation sensors were used as features for the GMM-based behavioral modeling approach. The four types of features which derived into ten features are: (1) three-dimensional Acc, (2) two-dimensional Ori, (3) Delta Coefficients (DC) Acc and (4) Delta Coefficients Ori [11].

B. Feature Extraction

In feature extraction (Fig. 2), the IDM and UDM were developed for feature representations of the preprocessed data. Furthermore, two SVMs were trained and combined through stacked generalization to produce a driver behavioral model.

GMMs are frequently used to represent data distribution. The mixture density of a GMM can be provided as follows:

$$P(\omega|\theta) = \sum_{i=1}^{M} w_i \, G(\omega|\mu_i \sum i) \qquad (1)$$

where M denotes the number of Gaussian components; ω is a D-dimensional random vector, $\sum_{i=1}^{M} \omega = 1$ are the mixture weights and $G(\omega|\mu_i \sum i)$ represents the Gaussian component density function. $\mu_i$and $\sum i$ are the mean and mixture proportions, respectively. i = 1, M are the covariance matrices of the M Gaussian components.

GMM can used to represent the distribution of sensor data for a driver, which is referred as IDM for this driver, such as [3-5]. However, there are two deficiencies of the GMM approach, as explained in [11]. First, the log-likelihood value of the model is the total sum of each log-likelihood value of the GMMs based on each sensor. Since each of the four features differed in its effectiveness to authenticate the genuine drivers, the IDM log-likelihoods of four features were combined using SVMs in a weighted manner on GMM-based behavioral modeling. This approach will enhance the individual characteristics of a driver. Secondly, for a simpler behavioral pattern of a driver (driver B) that is a subset of the behavioral pattern of another driver (driver A), the GMM approach may misclassify driver A as driver B. To distinguish Gaussian components of the driver behavior, the UDM was estimated to develop a GMM for the collective behavior in a particular driving scenario. A segment of the smartwatch sensor data of all driver was mapped to vector f_t in a new d-dimensional space by using the formula:

$$f_t = \left[f_{1;1;t}, f_{2;1;t}, \ldots, f_{M\,1;1;t}, \ldots, f_{1;4;t}, f_{M\,4;4;t} \right]^T \qquad (2)$$

Where $f_{j;i,t}$is the posterior probability that $\omega_{i;t}$is generated by the $_j$th Gaussian component of the $_i$th UDM, and d-dimensional refer to the total number of Gaussian components of the four UMDs.

C. Driver Behavioral Modeling

In this study, two modalities (log-likelihood of IDM and posterior probability of UDM) based on linear SVMs were trained for different feature vectors based on [11]. Two GMM-based driver models (IDM, UDM) are applied to combine the specific features of IDM and a general feature of UDM to achieve greater predictive accuracy. The IDM captures each participant as the specific model. It appropriately detects individuals when the patterns are different; however, detection is difficult for similar or subset patterns. Meanwhile, UDM captures the patterns of driving behavior based on all drivers to represent the collective behaviors of all drivers that are more general. It will complete the distinctive feature of the Gaussian component using IDM on the feature extraction process.

Furthermore, SVMs on the log-likelihood of the Gaussian components of IDM and SVMs on the posterior probabilities of the Gaussian components of the UDM were built as bases learner. These two bases SVMs for each driver were combined through stacking (stacked generalization) to form each driving behavior model. Fig. 2 shows the combiner used as a meta-learner SVM to combine base-learner log-likelihood (llh) and base learner posterior probability (pp). In this study, the three driving behavior models were developed for a driver in three specific driving scenarios: straight, turning left, and right.

IV. Open-set Driver Identification Method

On the basis of the Reynolds solution [17], this study proposed an open-set identification method that combined the closed set identity by using OAA/OAO and authentication GMM-based behavioral modeling methods. In this section, the three proposed methods of open-set driver identification were described. The baseline was the GMM, whereas the proposed methods were IUG-based, OAO-IUGA, and OAA-IUGA methods. Each method included a training part for modeling and a testing part for the identification.

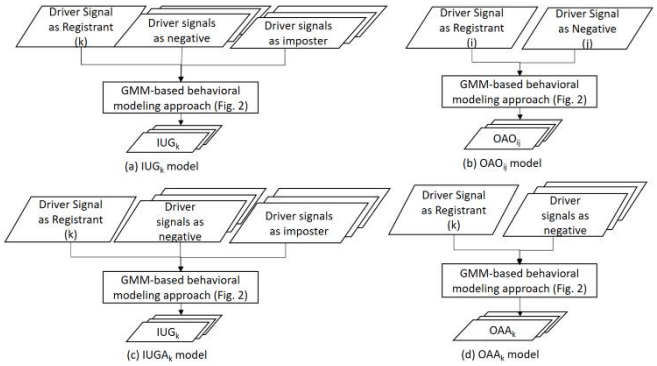


Fig. 3. IUG models: (a) IUG_k (b) OAO_{ij} (c) $IUGA_k$ (d) OAA_k

A. IUG Modeling

IUG modeling is a driver behavioral modeling approach that implements the GMM-based behavioral modeling approach proposed by Yang et al. [11] (Fig. 2). IUG modeling process uses different inputs, so different models were generated, as shown in Fig. 3. The models built are *IUG_k, OAO_{ij},* OAA_k and $IUGA_k$. The proposed methods of open-set driver identification incorporated IUG modeling differently. The $IUGA_k$ model was an IUG_k model in an authentication phase. It is called the $IUGA_k$ model, where k is the registrant.

B. Open-Set Driver Identification

The resulting model is used to identify the driver in the testing part. The baseline of this research is the GMM method that combined the closed set GMM identity and authentication because most of the current identification using GMM.

1) Individual-Universal Driver GMM Based Method (IUG-Based Method)

The IUG-based method combined a closed-set identity by using the IUG_k model and authentication with a threshold. During the training part, each registrant was trained using a driver behavior model.

Each behavior model used the registrant and non-registrant data by applying the GMM (Section III). In the testing part, data preprocessing and feature extraction were first performed for the input driving behavior signal, and the posterior probability of the registrant was then calculated with each driver's behavior model in the system. Finally, the posterior probability of the maximum value was selected to determine if it is higher than the threshold. If the value is higher than the threshold, the driver identity of maximum posterior probability was generated; otherwise, the driver was determined as unknown. Assuming the system has n registrants, this identification technique required the construction of n behavioral models. Fig. 4 shows the IUG-Based method.

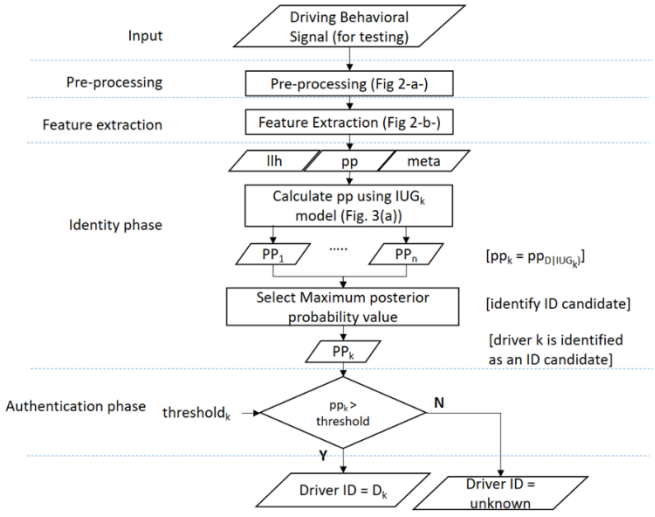


Fig. 4. IUG-Based Method

2) *One-Againts-One IDM-UDM GMM Authentication (OAO-IUGA)*

A one-against-one training method for the closed set identification model (OAO_{ij}) combined with the authentication model ($IUGA_k$). The process starts with data preprocessing and feature extraction for the input driving behavior signal. OAO_{ij} model used in the Identity phase to fit the testing data behavior with all registrant by calculating the posterior probability of the testing data conditional by each OAO_{ij} model that represents each registrant. In this phase, the driver ID candidate was obtained from the maximum value of the posterior probability. Finally, in the authentication phase, the posterior probability of driver ID candidate (Driver k) conditional by the $IUGA_k$ model calculated and then determined whether it is greater than the threshold or not. If the posterior probability is greater than the threshold, the output is driver k; otherwise, the driver is determined to be unknown. Assuming that the system had n registrants, this identification technology was used to construct $(C_2^n + n)$ behavioral models. Fig. 5 shows the OAO-IUGA Method.

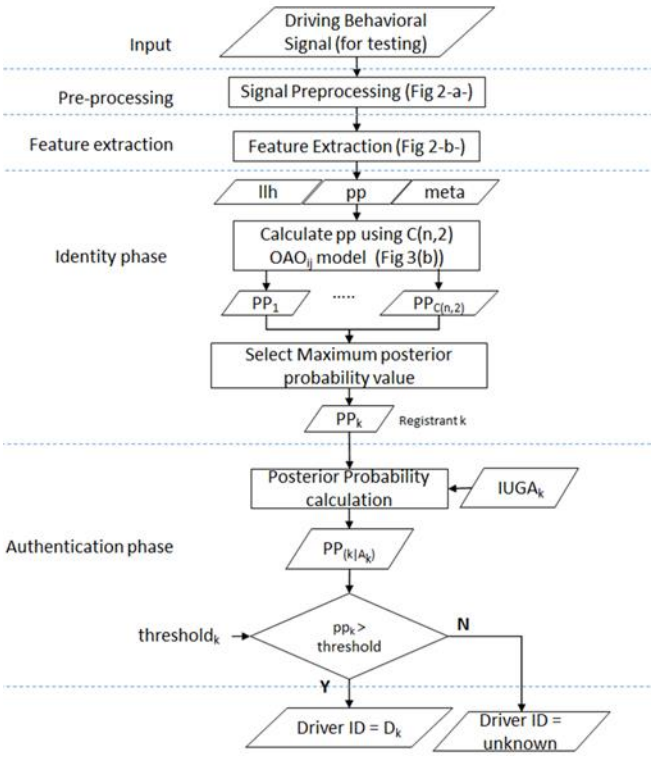


Fig. 5. OAO-IUGA Method

3) *One-Againts-All IDM-UDM GMM Authentication (OAA-IUGA)*

A one-against-all training method for the closed set identification model (OAA_k) combined with the authentication model ($IUGA_k$). OAA-IUGA method has the same algorithm as OAO-IUGA method (Fig. 5), but use the different model in the identity phase (OAA_k) as shown in Fig. 3.

Assuming the system has n registrants, this identification technique required the construction of 2n behavioral models.

V. Experiments and Discussion

In this study, four experiments were conducted to evaluate the proposed driver identification method. These experiments aimed to 1) analyze the number of times required to perform the repeated sampling strategy, 2) Confirm that in identification field, IUG-Based give better result than GMM, as it does in verification, 3) evaluate the accuracy of the four identification methods in the simulated environment, and 4) evaluate the accuracy of the four identification methods in real environments. All analyses were performed on a personal computer with an Intel Core i7-7th Gen CPU, 32 gigabytes of RAM, and Windows 10.

As a preliminary experiment, the number of Gaussian components in GMM was determined. The number of Gaussian components required for the GMM was analyzed from 15 participants in the simulated environments concerning 2, 4, 8, 16, and 24 Gaussian components. A model's accuracy tends to increase with more GMM components at the cost of longer training time. This experiment examines the tradeoff of the accuracy gain and training time through different numbers of component settings and chooses the component number when

the accuracy saturates. The accuracy of IDM improves 6.17%, but the training time increases by 216% when the number of GMM components of IDM increases from 4 to 8. The results also show that the accuracy of UDM improves 14.15%, and training time raises 54.63% when the number of GMM components of UDM increases from 8 to 16. After evaluating the tradeoff between EER and training time, the number of GMM components was set to 4 for IDM and 16 for UDM.

A. Experimental Setups

1) Data Collection

The data simulated were collected from 90 participants. The driving behavior data of each participant can be divided as straight, left-turn, and right-turn data, which were then used to develop three types of driver behavior models.

A driving simulation system close to a real driving system was developed to analyze driving behaviors (Fig. 6). The simulation system included a desktop computer, liquid-crystal display monitor, simulator-grade wheel, and pedal unit. Driving simulation software City Car Driving was used to simulate realistic three-dimensional road scenes with dynamic traffic streams. Sony smartwatch 3 and Sony Xperia Z5 premium were the smartwatch and mobile phone adopted for collecting data. The exercise habits of the 'driver's hand were captured through the built-in accelerometer and orientation sensor of the smartwatch, while the mobile phone sent the data to the server.

Fig. 6. Simulated and real environment data collection

The driving behavior data of 20 participants driving a real vehicle (Honda CR-V) in the campus of National Central University were also collected.

Fig. 6 also shows the route and equipment of real environment. The route included five turns and was approximately 1.77 km long, clockwise and counterclockwise to ensure the collection of their driving behaviors when turning in both directions. The equipment is a smartphone which placed in car besides the driver, and the gyroscope readings of the smartphone were used to divide the driving session of each driver into separate segments for different driving maneuvers.

2) Evaluation and Performance Indices

To assess the effectiveness of each method, the repeated sampling strategy was used to generate the training and evaluate the set data required for the experiment. In this study, a car was assumed to be owned/shared by at most 4–5 people, and thus, 15 drivers were drawn from 90 people per sample, 4 of which were registrants (car owners), 10 of which were registered as illegal users for training, and the last of which was treated as illegal for testing purposes. For each experiment, each registrant provided 55 training materials and 10 test materials. Each illegal user provided 10 training and 40 test materials.

The following performance indicators were used to assess open-set driver identification, including False Acceptance Rate (FAR), False Rejection Rate (FRR), and MisLabeling Rate (MLR), Registrant Error Rate (RER), Equal Error Rate (EER), and Detection Error Trade-off (DET) curve. The FAR is the probability that an illegal user was judged as a registrant. The FRR is the probability that the correct registrant is judged as an illegal user. The MLR is the probability that the correct registrant is judged as another registrant. The FAR has a tradeoff relationship with the RER. With the increasing threshold, the RER increased and the FAR decreased. By contrast, for the reduced threshold, the RER decreased and the FAR increased. The EER is the value at which the FAR and RER are equal. DET curve shows all the corresponding FAR and RER when moving the set threshold.

The models obtained for each driving maneuver were annotated with S (driving straight), L (turning left), or R (turning right). The S + L + R referred to the approach that utilized the three segments, with each annotation representing one of the three maneuvers.

B. Experiments

1) Sensitivity Analysis of the Round Number of the Repeated Sampling Strategy

In this experiment, the number of executions required for the repeated sampling strategy by 90 participants was analyzed in the simulation environment. If all pairing combinations were executed, $C_{15}^{90} = 4579x10^{17} = O(10)^{17}$. This execution would have been excessive, so we used the cumulative average method to find out that when a certain number of execution rounds is run, the average value will not change significantly. Fig. 7 shows that 1500 rounds of GMM and IUG-Based modeling methods were performed, and the cumulative average of their EER per round was calculated. On the basis of the results, the EER average did not significantly change when the number of execution rounds reached 900, and thus, in the subsequent experiment, the number of executions of the repetitive sampling strategy (execution run) was set to 900. Furthermore, the Gaussian component numbers of the IDM and UDM used in all experiments in this study were set with reference to the parameters in Yang et al. [9]; that is, the IDM and UDM were 4 and 16, respectively.

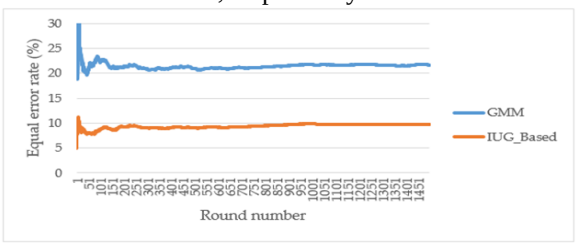
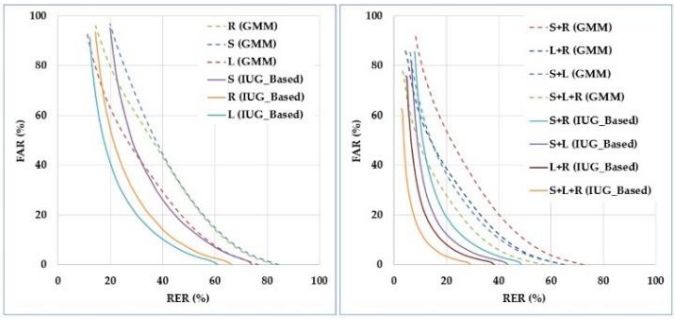


Fig. 7. Sensitivity Analysis of the Repeated Sampling Strategy

2) *Performance Comparison of the Proposed Driver Identification Methods and GMM in the Simulated Environment*

Fig. 8 shows the DET Curve for the GMM and IUG-Based in a simulated environment. Fig. 8(a) shows that the IUG-Based method is more accurate than the GMM in a single driving situation. Furthermore, Fig. 8 (b) shows that the IUG-Based method is more effective than the GMM in a multi-driving scenario and is more accurate than the GMM.



(a) (b)

Fig. 8. Performance Comparison of the Proposed Driver Identification Methods and GMM in the Simulated Environment

Table 1 validates that the IUG-Based method is at least 5% superior to the GMM in terms of the EER. Furthermore, the experimental results show that using the three driving scenarios S + L + R at the same time had the optimal recognition result; therefore, the subsequent two experiments only used S + L + R for performance evaluation.

TABLE I
EER COMPARISON OF PROPOSED METHODS AND GMM RESULT

Classification Scenario	Simulated Environment	
	GMM	IUG_Based
S	41.54%	34.98%
L	35.68%	26.15%
R	41.28%	29.21%
S+L	26.23%	16.85%
S+R	31.94%	19.67%
L+R	27.47%	14.64%
S+L+R	23.18%	11.19%

3) *Performance Evaluation of Four Driver Identification Methods in the Simulated Environment*

In this experiment, four driver identification methods in a simulated environment were compared. Fig. 9 shows the DET curve for four driver identification methods. The effects of the IUG-Based, OAA-IUGA and OAO-IUGA methods were significantly superior to the GMM. EER values of each method in a simulated environment is as follows. GMM 23.14%, IUG-Based 11.19%, OAA-IUGA 10.65% and OAO-IUGA 10.50%. Insert of Fig. 9 shows the MLR and FRR results of the four methods. The IUG-Based, OAA IUGA, and OAO IUGA methods exhibited no difference between the FRRs; however, the IUG-Based method was slightly inferior to OAA IUGA and OAO IUGA.

4) *Performance Evaluation of Four Driver Identification Methods in the Real Environment*

In this study, 20 participants drove in a real environment. Fig. 10 show that the three proposed methods as the superior of the GMM method. Among them, the EER values for the GMM, IUG-Based, OAA-IUGA, and OAO-IUGA were 33.83%, 17.95%, 17.07%, and 16.66%, respectively.

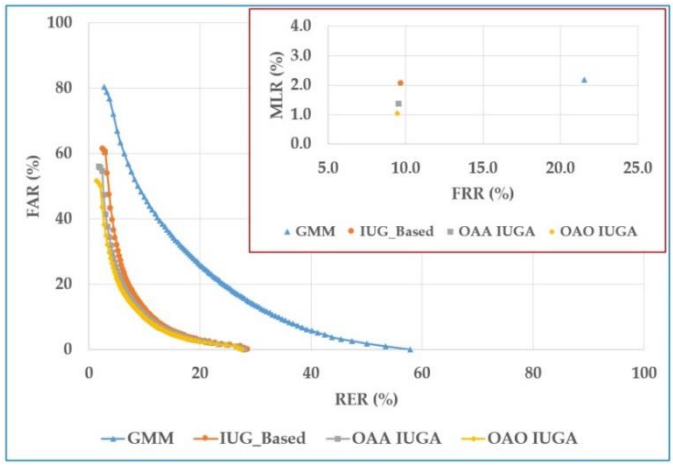


Fig. 9. Performance evaluation of the driver identification methods in the simulated environment (insert: MLR and FRR results)

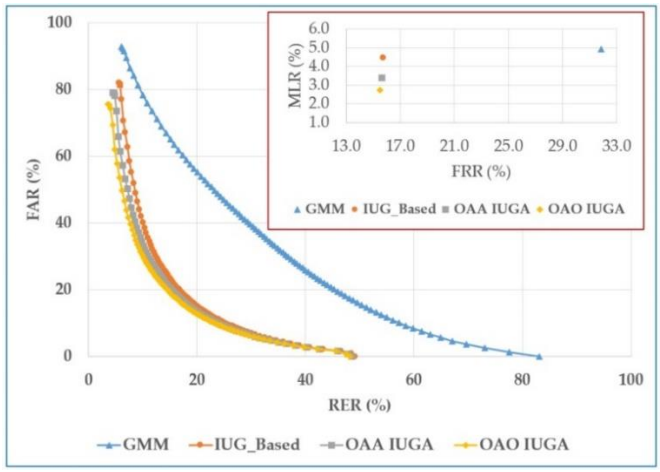


Fig. 10. Performance evaluation of the driver identification methods in the real environment (insert: MLR and FRR results)

C. Discussion

Some conclusions can be obtained based on the experimental results. Experiment 1 validated that in 1500 rounds when the number of execution rounds reached 900, the EER average did not significantly change. Therefore, 900 execution runs can be performed for Experiments 2, 3, and 4, and thus, sampling is the representative of the data and provides an almost constant result. Experiment 2 shows that the IUG-Based method is superior to the GMM in terms of driver identification, and thus, this finding can be extended to the OAA and OAO methods. Moreover, Experiment 2 shows that S + L + R classification must be used for performance evaluation because it provides the optimal identification results for the GMM and IUG-Based methods. The difference in the EER of S + L + R classification and other classification was 3.45%–23.79% for the IUG-Based based and 4.29%–18.36% for the GMM.

Performance evaluation stated that the three proposed methods: IUG, OAA-IUGA, and OAO-IUGA provided superior results to that of the GMM. A significant difference

was observed in the GMM and the three proposed methods. Experiments 3 and 4 showed that among the three proposed identification methods, IUG method was slightly worse than OAA-IUGA and OAO-IUGA methods; however, the OAA-IUGA and OAO-IUGA methods were not significantly different. The OAO-IUGA exhibited the highest accuracy, followed by the OAA-IUGA, IUG, and GMM methods.

TABLE II
COMPARISON OF THE EER AND SPACE STORAGE OF THE 3 THREE PROPOSED METHODS

No	Methods	EER (%) Simulated Environment	EER (%) Real Environment	Storage Compression (Big (O) notation)
1	IUG	11.19	17.95	O (n)
2	OAA-IUGA	10.65	17.07	O (2n)
3	OAO-IUGA	10.50	16.66	O (C(n,2)+n)
n = 90 (for simulated environment) n = 20 (for real environment)				

Although among the three proposed recognition methods IUG-Based method exhibited the lowest accuracy, it required the least number of classifiers only n. OAA-IUGA and OAO-IUG-Based methods required 2n and ($C_2^n + n$), respectively. The OAA-IUGA method required 2n classifiers because it applied 2 sets of models: the OAA_k model for the identity phase, and the IUGA model for the authentication phase. Likewise, OAO-IUGA methods required C_2^n models for the identity phase and n models for the authentication phase. The result indicated that the GMM and IUG-Based methods required less space, followed by the OAA-IUGA and OAO-IUGA methods.

Tables 2 and 3 show that for a system with high recognition performance, then the OAO-IUGA method must be selected. If average recognition performance and high storage space are required, the IUG-Based method must be used. For a few drivers, no significant differences are observed in terms of storage space. However, only a slight difference in accuracy is observed between the OAA-IUGA and OAO-IUGA methods. The OAA-IUGA method has medium accuracy and high storage space. Thus, it can be an optimal option.

TABLE III
THE RANK OF ACCURACY AND PROVIDENT SPACE OF THE THREE PROPOSED METHODS

No	Simulation and Real Environment	
	The rank of an accurate method	The rank of provident space
1	OAO-IUGA	IUG
2	OAA-IUGA	OAA-IUGA
3	IUG	OAO-IUGA

D. Security Analysis

The security of Biometric Authentication Systems is an important issue that needs to be discussed. Newton, Elaine on National Institute of Standards and Technology (NIST) of the United States Government published an evaluation framework of biometric authentication, stated that the possible attacks are zero information attacks and targeted attacks [24]. The targeted attack is impersonation attacks. Zero information attacks is a typical attack because the attacker does not have to mimic even know the biometric pattern. Experiments 3 and 4 were conducted under the zero-information attacks scenario as explain on V.A.2 "Evaluation and Performance Indices." The experiment result shows that the EER average is 10.78 for the simulation environment and 17.22 for the real environment. The EER captures the legal drivers identified as illegal drivers (FRR) and illegal drivers identified as legal drivers (FAR). The DET curve (Fig. 9, 10) shows the tradeoff relationship between the FAR and RER. If the threshold increased, RER would increase, and FAR will decrease. Conversely, if the threshold is lowered, the RER will decrease, and the FAR will increase. Fig. 9 and 10 indicate that FAR is low, which means the methods able to secure the car from the illegal user.

Information theoretical analysis of impersonation attack [25] has been studied to demonstrate that the information taken by the proposed classifier has enough entropy against possible impersonation attacks. However, the metrics cannot directly be applied to the system because the method used regression (standard least-squares method) for trajectory based on the positions of one or more end-effectors, while the data set in this study consists of ten accelerometer and orientation features. Therefore, the metrics cannot easily evaluate our data set. Building metrics to evaluate that the proposed driver identification method is secure against the impersonation attack can further be researched on future work.

VI. CONCLUSIONS AND FUTURE WORK

In this study, three smartwatch-based open-set driver identification methods (IUG-Based, OAA-IUGA, and OAO-IUGA) were proposed as the first smartwatch driver identification methods. Moreover, it's were validated and compared with the GMM methods. The experimental results showed that the three proposed methods were more accurate than the GMM method. In the simulation environment, the EER values were 11.19% for IUG-Based, 10.65% for OAA-IUGA, and 10.50% for OAO-IUGA, whereas the EER of the baseline GMM method was 23.14%. In a real environment, the EER values for IUG-Based, OAA-IUGA, and OAO-IUGA were 17.95%, 17.07%, and 16.66%, respectively, whereas for the baseline, it was 33.83%.

This study can provide a reference for developers of driver identification systems with different requirements. If a system with high identification performance, the OAO-IUGA method is recommended whereas if it requires a low identification accuracy and can provide considerable storage space, the IUG-Based method is recommended otherwise if it moderately high identification performance and relatively economical storage use, the OAA-IUGA method should be used. According to the security analysis, the proposed method is secure from the zero-information attack. Further research concerned with the security of the proposed identification method against various attacks, including impersonation, will be conducted on the future work.

REFERENCES

[1] S. Bengio and J. Mariéthoz, *Biometric Person Authentication Is a Multiple Classifier Problem*. 2007.

[2] N. S. E. Hossain and G. Chetty, "Human Identity Verification by Using Physiological and Behavioural Biometric Traits," *International Journal of Bioscience, Biochemistry and Bioinformatics*, vol. Vol. 1, no. No. 3, September 2011.

[3] K. Igarashi, C. Miyajima, K. Itou, K. Takeda, F. Itakura, and H. Abut, "Biometric identification using driving behavioral signals," *IEEE International Conference on Multimedia and Expo, Taipei, Taiwan*, vol. 27–30 June 2004, no. 1, pp. pp. 65–68., 2004, doi: 10.1109/ICME.2004.1394126.

[4] C. Miyajima *et al.*, "Driver modeling based on driving behavior and its evaluation in driver identification," (in English), *Proc. IEEE*, vol. 95, no. 2, pp. 427-437, Feb 2007, doi: 10.1109/Jproc.2006.888405.

[5] A. Wahab, C. Quek, C. K. Tan, and K. Takeda, "Driving profile modeling and recognition based on soft computing approach," (in English), *IEEE Transactions Neural Networks*, vol. 20, no. 4, pp. 563-82, Apr 2009, doi: 10.1109/TNN.2008.2007906.

[6] H. Qian, Y. Ou, X. Wu, X. Meng, and Y. Xu, "Support Vector Machine for Behavior-Based Driver Identification System," *Journal of Robotics*, vol. 2010, pp. 1-11, 2010, doi: 10.1155/2010/397865.

[7] B. G. Lee, B. L. Lee, and W. Y. Chung, "Wristband-Type Driver Vigilance Monitoring System Using Smartwatch," (in English), *IEEE Sensors Journal*, vol. 15, no. 10, pp. 5624-5633, Oct 2015. [Online]. Available: <Go to ISI>://WOS:000360072500034.

[8] R. Chen, M. F. She, X. Sun, L. Kong, and Y. Wu, "Driver Recognition Based on Dynamic Handgrip Pattern on Steeling Wheel," presented at the 2011 12th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, 2011.

[9] A. Riener and A. Ferscha, "Supporting Implicit Human-to-Vehicle Interaction: Driver Identification from Sitting Postures," presented at the Proceedings of the First Annual International Symposium on Vehicular Computing Systems, 2008.

[10] B. L. Lee, B. G. Lee, and W. Y. Chung, "Standalone Wearable Driver Drowsiness Detection System in a Smartwatch," (in English), *IEEE Sensors Journal*, vol. 16, no. 13, pp. 5444-5451, Jul 1 2016. [Online]. Available: <Go to ISI>://WOS:000378509200042.

[11] C. H. Yang, C. C. Chang, and D. R. Liang, "A Novel GMM-Based Behavioral Modeling Approach for Smartwatch-Based Driver Authentication," (in English), *Sensors-Basel*, vol. 18, no. 4, Apr 2018. [Online]. Available: <Go to ISI>://WOS:000435574800079.

[12] Z. P. Li, K. Zhang, B. K. Chen, Y. H. Dong, and L. Zhang, "Driver identification in intelligent vehicle systems using machine learning algorithms," (in English), *Iet Intell Transp Sy*, vol. 13, no. 1, pp. 40-47, Jan 2019. [Online]. Available: <Go to ISI>://WOS:000454695200006.

[13] J. Chen, Z. Wu, and J. Zhang, "Driver identification based on hidden feature extraction by using adaptive nonnegativity-constrained autoencoder," (in English), *Appl Soft Comput*, vol. 74, pp. 1-9, Jan 2019. [Online]. Available: <Go to ISI>://WOS:000454251200001.

[14] M. A. Rahim, J. M. Liu, Z. J. Zhang, L. H. Zhu, X. Li, and S. Khan, "Who is Driving? Event-Driven Driver Identification and Impostor Detection Through Support Vector Machine," (in English), *IEEE Sensors Journal*, vol. 20, no. 12, pp. 6552-6559, Jun 15 2020. [Online]. Available: <Go to ISI>://WOS:000536772100039.

[15] W. S. Wang, J. Q. Xi, and J. K. Hedrick, "A Learning-Based Personalized Driver Model Using Bounded Generalized Gaussian Mixture Models," (in English), *IEEE T Veh Technol*, vol. 68, no. 12, pp. 11679-11690, Dec 2019. [Online]. Available: <Go to ISI>://WOS:000518226900027.

[16] B. I. Ahmad, P. M. Langdon, J. Liang, S. J. Godsill, M. Delgado, and T. Popham, "Driver and Passenger Identification From Smartphone Data," (in English), *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 4, pp. 1278-1288, Apr 2019. [Online]. Available: <Go to ISI>://WOS:000463475900008.

[17] Reynolds and A. Douglas, "Automatic Speaker Recognition: Current Approaches and Future Trends," *Acoustics, Speech, and Signal Processing (ICASSP)*, 2001.

[18] D. A. Reynolds and R. C. Rose, "Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models," (in English), *IEEE T Speech Audi P*, vol. 3, no. 1, pp. 72-83, Jan 1995, doi: Doi 10.1109/89.365379.

[19] E. Ozgunduz, T. Senturk, and M. E. Karsligil, "Efficient off-line verification and identification of signatures by multiclass Support Vector Machines," (in English), *Lect Notes Comput Sc*, vol. 3691, pp. 799-805, 2005. [Online]. Available: <Go to ISI>://WOS:000232301200098.

[20] J. Milgram, M. Cheriet, and R. Sabourin, "One Against One or One Against All: Which One is Better for Handwritting Recognition with SVMs " *HAL archives-auvertertes*, pp. Hal id: inria-00103955, 2006.

[21] R. Cappelli, D. Maio, D. Maltoni, J. L. Wayman, and A. K. Jain, "Performance evaluation of fingerprint verification systems," (in English), *IEEE T Pattern Anal*, vol. 28, no. 1, pp. 3-18, Jan 2006. [Online]. Available: <Go to ISI>://WOS:000233172000001.

[22] J. L. Douglas, J. C. Junqua, C. Kotropoulos, R. Kuhn, F. Perronnin, and I. Pitas, "Recent Advantages in Biometric Person Authentication," *Acoustics, Speech, and Signal Processing (ICASSP)*, pp. pp.4060-4063, May 2002.

[23] J. H. Hong, J. K. Min, U. K. Cho, and S. B. Cho, "Fingerprint classification using one-vs-all support vector machines dynamically ordered with naive Bayes classifiers," (in English), *Pattern Recogn*, vol. 41, no. 2, pp. 662-671, Feb 2008. [Online]. Available: <Go to ISI>://WOS:000250695500020.

[24] E. Newton, "Strength of Authentication for Biometrics: An Evaluation Framework," *National Institute of Standards and Technology (NIST) gov*, 2020. [Online]. Available: https://www.nist.gov/system/files/documents/2020/07/30/08_newton_biometrics_presentation_final.pdf.

[25] M. Sherman *et al.*, "User-generated free-form gestures for authentication," presented at the Proceedings of the 12th annual international conference on Mobile systems, applications, and services - MobiSys '14, 2014.

Rekyan Regasari Mardi Putri received BS and MS degree in electrical engineering from Institute of Technology Sepuluh November, Surabaya Indonesia in 2002 and Brawijaya University in 2010 respectively.

She is currently pursuing the Ph.D. degree with the Department of Computer Science and Information Engineering, at National Central University, Taiwan. Her research interests include machine learning, decision support system, and driver recognition.

Ching-Han Yang received a BS degree in computer science and information engineering from National University of Tainan, Taiwan, in 2008, and an MS degree in software engineering from National Central University, Taiwan, in 2010. He received his PhD degree in computer science and information engineering from National Central University, Taiwan, in 2018. At present, he is an engineer at Digital Transformation Institute, Institute for Information Industry, Taiwan. His research interests include biometrics authentication, and driver behavior analysis

Chin-Chun Chang received the BS degree and the MS degree in computer science in 1989 and 1991, respectively, and the PhD degree in computer science in 2000, all from National Chiao Tung University, Hsinchu, Taiwan. From 2001 to 2002, he was a faculty of the Department of Computer Science and Engineering, Tatung University, Taipei, Taiwan. In 2002, he joined the Department of Computer Science and Engineering, National Taiwan Ocean University, Taiwan, where he is currently an associate professor. His research interests include computer vision, machine learning, and pattern recognition. Dr. Chang is a member of the IEEE.

Deron Liang received a BS degree in electrical engineering from National Taiwan University in 1983, and an MS and a PhD in computer science from the University of Maryland at College Park in 1991 and 1992 respectively. He is on the faculty of Department of Computer Science & Information Engineering, and serves as Director of Software Research Center, National Central University, Taiwan since 2008. His current research interests are in the areas of software fault-tolerance, system security, and system reliability analysis. Dr. Liang is a member of ACM and IEEE.