

데이터 분석 기반 운전자 프로파일링 연구 동향

곽 병 일*

요 약

운전자의 편의성 및 안정성 향상을 위해, 차량에 탑재되는 다양한 센서 및 전자제어장치들은 주행 중 많은 양의 데이터들을 생성한다. 이렇게 생성된 많은 양의 데이터들의 분석은 개인화 서비스, 자동차 보험, 사고 예측과 같은 곳에 활용되고 있다. 최근 주행 중의 다양한 데이터 종류와 머신러닝 및 딥러닝 기반의 방법론을 통해 차량의 운전자를 식별하는 연구들이 진행되고 있다. 본 고에서는 차량에서의 데이터 분석에 기반한 운전자 식별 연구 동향을 설명하도록 하겠다.

I. 서 론

빠르게 발전하고 있는 IT 기술과 함께 차량의 중요 기능을 담당하는 부분들은 기계적인 제어에서 전자적인 제어로 많은 부분 변화되고 있다. 또한, 차량에 탑재되는 수많은 센서는 주행 중의 차량 상태 파악을 위해 측정된 데이터들을 서로 송수신하여 실시간으로 차량 제어를 할 수 있도록 도움을 준다. 많은 센서를 탑재한 최신의 차량은 운전자와 탑승자에게 이동의 편리성 제공하지만, 그 센서들의 다양한 취약점 및 차량에 탑재된 전자 장치들의 취약점으로 인해서 금전적인 목적으로 차량 도난 범죄들의 표적이 되고 있다[1].

차량의 보안성 강화 및 도난 방지를 위해 차량 제조사들은 스마트 키를 사용하거나 최근 스마트폰을 차량의 키로써 사용하는 기술들도 적용되고 있다[2]. 하지만, 전자적으로 작동하는 스마트 키나 스마트폰 키를 이용한 기술들은 차량에 적용되어 편리성뿐만 아니라 차량에 탑재된 보안장치를 우회한 절도와 같이 악용될 우려가 있다. 이러한 부분에서, 주행하는 도중 운전자를 지속해서 식별하는 방법이 필요하다. 데이터 기반의 운전자 식별 방법은 주행 중, 차량에 탑재된 센서들로부터 생성되는 차량 상태정보들을 기반으로 운전자의 행위를 파악하고, 더 나아가 운전자를 식별할 수 있는 연구이다.

본 논문에서는 데이터 기반 운전자 식별 및 프로파일링에 대한 데이터, 모델 및 알고리즘, 행위 분석에 대한 이해를 제공하고, 그 동향에 대해서 설명한다.

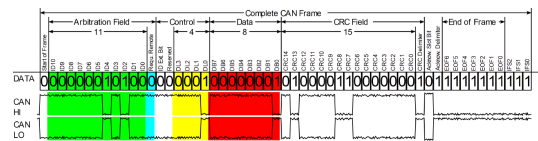
II. 차량 주행 데이터

본 장에서는 주행 행위 분석 및 운전자 식별을 위해 사용된 적용 가능한 데이터를 소개한다. 또한, 운전자 식별 연구에 현재 접근 및 사용 가능한 차량 주행 데이터셋을 소개한다.

2.1. 주행 데이터 분류

2.1.1. Controller Area Network bus traffic

차량에 탑재된 장치와 차량 자체를 제어하기 위한 컴퓨터로써 여러 개의 전자제어장치 (Electronic Control Unit, ECU)가 설치되어 있다. ECU들은 차량 내부에서 서로 통신하기 위해 Control Area Network (CAN) 프로토콜을 사용하며, 해당 CAN 통신 프로토콜의 Dataframe을 통해 차량의 상태정보들이 전송된다. [그림 1]은 CAN Dataframe 구조를 나타낸 것이다. CAN Dataframe에 포함되는 정보로 차량의 브레이크 페달, 가속 페달, 엔진 RPM, 연료 분사량, 차량



(그림 1) CAN Dataframe(3)

이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2022R1A4A1033600).

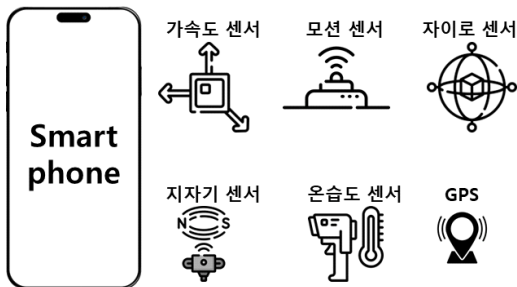
* 한림대학교 정보과학대학 소프트웨어학부 (조교수, kwacka12@hallym.ac.kr)

속도 등이 있다.

2.1.2. Inertial Measurement Unit

차량에 적용 가능한 데이터로써 차량의 상태정보들을 간접적으로 파악할 수 있는 관성 측정 장치(Inertial Measurement Unit, IMU)가 있다. IMU는 물체의 움직임 변화 정도를 측정할 수 있는 가속도 센서, 물체의 회전에 대한 변화 정도를 측정할 수 있는 각속도 센서, 물체의 방향을 파악할 수 있는 지자기 센서 등이 포함된다. 그 외에도, 스마트폰에 탑재되어 파악이 가능한 온습도 센서, 모션 센서, GPS, 근접/조도 센서 등이 존재한다. [그림 2]는 스마트폰에 탑재된 다양한 센서들을 나타낸 것이다.

주행 중 차량에 거치하여 스마트폰에 탑재된 다양한 센서들로부터 특정 정보들을 추출 및 이용할 수 있는데, 대표적으로, GPS 데이터가 있다. GPS의 경우, 해당 정보를 이용하여 차량의 위치 파악이 가능하며, 이 데이터를 통해 차량의 속도를 간접적으로 파악할 수 있다. 이러한 스마트폰 센서 데이터들은 데이터 분석 기반 운전자 식별 연구에 있어 차량 상태를 간접적으로 파악할 수 있는 정보로써 사용되고 있다.



[그림 2] 스마트폰에서의 IMU 센서를 포함한 추출 가능한 센서 데이터

2.1.3. Multimedia Data

[그림 3]은 차량에서 추출 가능한 데이터를 나타낸 것으로 멀티미디어 데이터인 이미지나 카메라 영상, 음성 데이터는 차량에 설치 가능한 블랙박스 또는 스마트폰의 카메라를 통해서 수집할 수 있다. 특히, 운전자의 차량 주행 상황을 빠르게 이해할 수 있는 장점이 있다. 하지만, 영상 데이터의 경우, 설치하는 위치에



[그림 3] 차량 블랙박스의 영상 및 소리 데이터

따라 수집 가능한 정보들의 차이가 있으며, 소리 데이터의 경우 데이터 수집 시 도로의 노면 소음으로 인한 노이즈가 함께 포함될 수 있어, 적절한 후처리 작업이 필요한 부분이 있다.

2.1.4. Simulator Data

주행 시뮬레이터를 이용한 데이터는 실제 도로와 유사하게 도로 환경을 가상으로 구성하여 주행 중 발생 가능한 시나리오를 통해 추출되는 데이터이다. 주행 시뮬레이션 상황을 가정하기 때문에, 운전자들의 주행 상황을 반복적으로 유사하게 구성할 수 있는 장점이 있다. 또한, 운전자들의 주행 특성 및 패턴을 비교 분석하는데 용이한 부분 역시 장점으로 볼 수 있다. [그림 4]는 Carsim[4]을 이용한 주행 시뮬레이터를 나타낸 것이다.



[그림 4] Carsim 주행 시뮬레이터(4)

2.2. 주행 데이터셋

운전자 프로파일링을 위한 다양한 종류의 데이터가 있지만, 그중 운전자 식별 및 프로파일링 연구에 있어 접근 가능한 데이터셋을 본 절에서 소개한다. [표 1]은 현재 학술적 용도로써 인터넷에서 접근 및 사용할 수 있는 데이터셋을 나타낸 것이다. Vehicular data trace[5]는 Intelligent Transportation Systems (ITS) 연구 및 Vehicular Ad-hoc Networks (VANET) 관련하여 수집된 데이터셋으로, 스마트폰에서 추출 가능한 IMU 센

[표 1] 활용 및 접근 가능한 데이터셋

데이터셋	데이터 종류	주행시간	운전자 수	도로 타입
Vehicular data trace [5]	CAN bus traffic, GPS	약 28시간 / 약 3시간	10 / 4	도심도로
HCILAB Dataset [6]	GPS, Multimedia, wearable sensors	약 300분	10	도심도로, 고속도로
OCSLAB Dataset [7]	CAN bus traffic	약 23시간	10	도심도로, 고속도로, 주차장
UAH-Driveset [8]	IMU, Multimedia, GPS	약 500분	6	고속도로, 보조도로

서 데이터를 수집하였다. 데이터 수집 시 사용한 차량은 Hyundai HB20 와 Renault Sandero 이며, 총 28시간, 36번의 주행, 10명의 운전자를 대상(약 25세~61세, 남성 6명, 여성 4명)으로 데이터를 수집하였다.

HCILAB Dataset [6]는 운전자의 체온, ECG, 심박수 데이터, multimedia data, GPS, 스마트폰에 탑재된 IMU 센서 데이터들을 수집하였다. 해당 연구진들은 도심도로, 고속도로에서 데이터를 추출하였으며, 10명의 운전자를 대상(약 23~57세, 남성 7명, 여성 4명)으로 데이터 수집하였다.

OCSLAB Dataset[7] 은 운전자 식별을 목적으로 추출한 데이터 셋으로, 차량의 상태정보를 보다 정확하게 확인할 수 있는 CAN bus traffic을 추출하였다. 데이터 수집에 사용한 차량은 KIA Soul이며, 총 10명의 운전자를 대상(약 26~32세, 남성 10명)으로 추출하였으며, 각 운전자 별 10번의 주행 데이터를 수집하였다.

UAH-Driveset[8]은 스마트폰의 센서들을 이용하여 IMU, multimedia data, GPS 데이터를 수집하였다. 도로의 유형은 고속도로 및 보조도로 2가지 유형이며, 총 6명의 운전자를 대상(약 21세~48세, 남성 5명, 여성 1명)으로 데이터를 수집하였다.

III. 데이터 기반 운전자 식별 및 인증 연구 동향

데이터 기반 운전자 식별은 주행 중인 운전자를 주행 중 차량에 탑재된 센서 및 다양한 데이터를 분석하여 차량의 소유주가 맞는지 아닌지를 인증 및 식별하는 연구이다. 또한, 해당 연구를 통해 차량 도난 탐지 연구에 적용이 가능하며, 최근 자동차 보험 회사에서의 Pay-as-you-drive 는 운전자의 성향 및 주행 습관에 따라 보험금 지급하는 부분들도 생겨나면서, 사기 탐지에도 적용될 수 있다.[9] [표 2]은 운전자 식별 및 프로파일링에 관한 주요 연구들과 그 특징들을 나타낸 것이다.

Enev et. al.[10]은 차량에서 발생 가능한 데이터를 통해 실제 도로상에서의 운전자 식별을 수행하였다. 해당 연구에서 사용한 데이터는 차량의 운전 조작에 직접적으로 연관되어 있는 브레이크 페달, 조향 핸들 각도, 차량의 속도 및 가속도, 기어 상태, 연료 소모량, 쓰로틀 포지셔닝 센서들을 사용하였다. 운전자 식별에 사용한 15개 종류의 센서 데이터 중 운전자 식별 알고리즘 적용 시 운전자의 특성을 잘 나타내는 피쳐 Top 3, Top 5로 구성하여 주차장과 도로 구간에서 운전자 식별 정확도를 측정하였다. 15명에 대한 운전자 식별 정확도가 Top 5 피쳐로 구성했을 경우 약 99% 이상

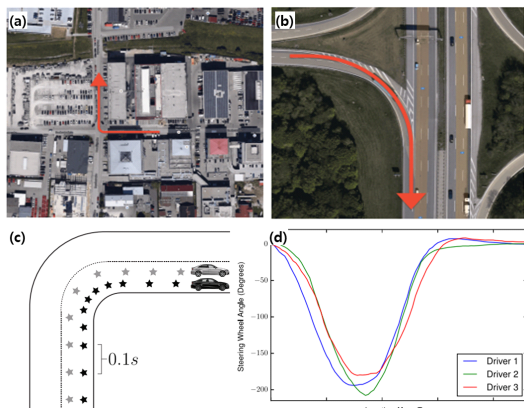
[표 2] 데이터 기반 운전자 식별 주요 연구

연구	특징
Enev et. al.[10]	- CAN bus traffic 사용 - 8분 미만의 데이터를 사용하여 높은 운전자 식별 정확도를 확보
Hallac et al.[11]	- 짧은 시간의 GPS 데이터를 사용 - 1회의 회전 데이터를 통해 운전자 식별을 수행
Sánchez et al.[12]	- 스마트폰의 가속도 센서를 통해 운전자 식별 및 차량 도난 탐지를 수행함 - 시계열 데이터를 2차원 이미지 형태로 변환하여 입력값으로 적용
Xun et. al.[13][14]	- CAN bus traffic을 사용 - 차량 도난 탐지 및 운전자 식별을 함께 수행[17], 운전자 행위 분석 수행 [13]
Kwak et. al.[15]	- CAN bus traffic을 사용 - 2개 도로(도심도로, 고속도로)에서 운전자 식별 실험을 수행했으며, 짧은 주행 시간에서 운전자의 특성이 나타남을 확인함
Hongyu et al. [16]	- CAN bus traffic을 사용 - 운전자 주행 데이터가 제한된 상태에서 식별이 가능한 모델을 제안함

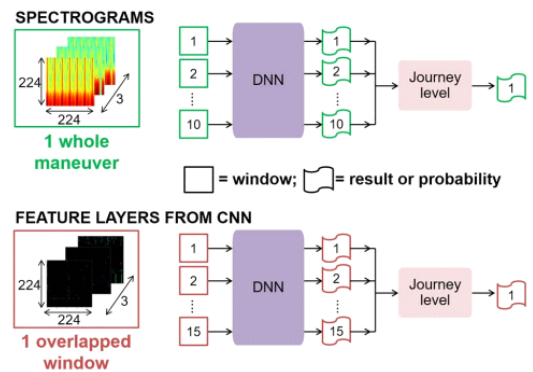
의 정확도를 나타냈으며, 전체 센서 데이터들을 사용하는 경우보다 낮은 시간 및 자원 소모를 가지면서 높은 성능을 확보할 수 있음을 확인하였다.

Hallac et al.[11]은 운전자 식별을 위해 차량의 CAN bus traffic과 함께 GPS 데이터를 사용하였다. [그림 5]은 차량 주행 중 우회전 구간에서 운전자별 조향 각도의 변화 정도를 나타낸 것이다. 회전 구간에서 운전자들은 운전자마다 주행 속도에 따라 조향 각도의 변화가 차이가 존재할 수 있으며, 차량의 주행 차선에 따라 그러한 패턴이 나타날 수 있다. 해당 연구진들은 이러한 회전 구간에서의 조향 각도를 기반으로 차량 속도, 브레이크 페달, 엔진 RPM 등 12개의 센서 신호를 통해 운전자를 식별하는 실험을 진행하였다. 또한, 1번의 우회전 또는 좌회전 데이터만을 학습하여 높은 정확도를 확보하였다.

Sanchez et al.[12]은 스마트폰의 IMU 센서 데이터를 이용하여 운전자 식별을 수행하였다. [그림 6]은 IMU 센서를 스펙트로그램으로 변환하여 운전자 식별을 수행하는 전체 과정을 나타낸 것이다. 해당 연구에서는 IMU 센서 중에서도 가속도 센서의 X, Y, Z 3축에 대한 센서 데이터를 변환하여 종방향 및 횡방향 가속도와 각속도를 구하였다. 차량에 스마트폰을 탑재시키거나 운전자가 스마트폰을 차량에 보관하게 될 경우, 스마트폰의 위치나 방향에 따라 기존의 가속도 X, Y, Z 3축 센서 데이터의 값들이 차량의 방향 및 속도에 일치하지 않은 상태로 값이 기록된다. 스마트폰을 조작하거나 스마트폰의 위치와 관계없이 운전자 식별에



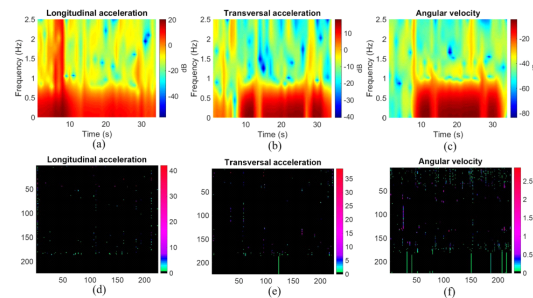
[그림 5] 주행 중 우회전 구간에 따른 운전자 별 주행 패턴 및 조향 각도 변화 (a: 도심 도로에서의 우회전 구간, b: 고속도로 상에서의 회전 구간, c: 회전구간의 주행 패턴, d: 운전자 별 조향 각도 변화 여부) [11]



[그림 6] 스펙트로그램 이미지 데이터 및 ResNet을 통한 운전자 식별 과정(12)

필요한 센서 데이터 값의 동일성을 위해 해당 연구에서는 일차적으로 데이터 전처리를 수행하여 종방향 및 횡방향 가속도와 각속도를 추출하였다. 또한, 데이터 전처리 이후, 스마트폰 센서의 1차원의 시계열 데이터를 2차원의 이미지 데이터로 변환시켰다. [그림 7]은 시계열 데이터의 스펙트로그램 변환한 결과를 나타낸 것이며, 2D 이미지로 변환시키기 위해 시계열 데이터에 Short-Time Fourier Transform (SFTP)를 이용했으며, 이미지로 변환된 데이터는 딥러닝 알고리즘인 ResNet에 입력값으로 적용된다. 결과적으로, 해당 연구에서는 스마트폰의 가속도 센서 1개만을 이용했음에도 불구하고 운전자 식별에 있어서 높은 정확도를 확보하였다.

Xun et al.[13]은 CAN bus traffic을 이용하여 운전자 식별과 함께 도난 탐지를 함께 수행하였다. CAN bus traffic 중 주행과 관련된 가속 페달, 브레이크 페달, 엔진 속도, 차량 속도, 조향각도 등에 대한 CAN Dataframe을 선택하여 입력값으로 적용하고, 해당 데이터를 CNN 및 Support Vector Domain Description (SVDD) 모델에 적용하여 운전자 식별과 도난 탐지에

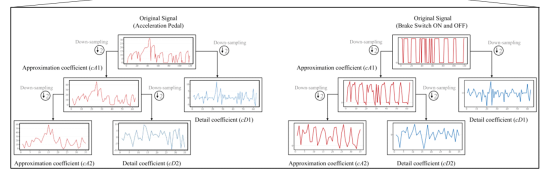
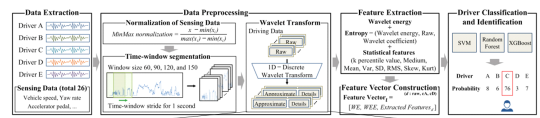


[그림 7] GPS 센서 데이터에서 2차원 스펙트로그램으로의 변환 [12]

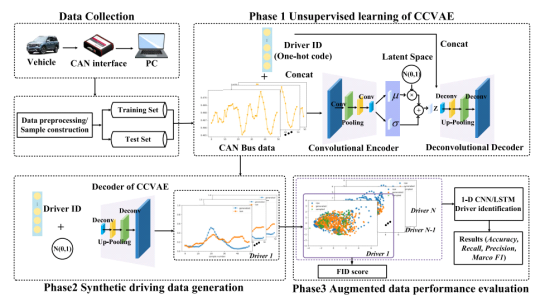
적용하였다. 해당 연구에서 운전자 식별 모델에 15명, 도난 탐지 연구에 5명의 인원을 적용하여 높은 성능을 확보하였다. 추가 연구에서, 추가로, Xun et al.[14]은 Multitask Learning (MTL)을 적용하여 동일한 전처리 및 학습 구조를 거친 후, 최종 출력을 도출하는 부분에서 3가지 기능을(운전자 식별, 주행 행위 평가, 불법 운전자 탐지-[그림 8] 참조) 수행하는 모델을 설계하였다.

Kwak et al.[15]은 CAN bus traffic을 이용하여 운전자 식별을 수행하였다. [그림 9]에서 볼 수 있듯이, 차량에서 추출되는 센서 데이터 중, 운전자 식별에 중요한 역할을 하는 센서 데이터 및 핵심 피쳐들을 구분하였으며, 운전자의 주행 행위 특성 파악 및 분석을 위해, 시계열 데이터 특성을 가지는 CAN bus traffic에 Wavelet transform을 적용하여 운전자의 주행 행위 특성을 추출하였다. 또한, 통계적 특성을 통해, 운전자들의 출발 시각 및 차량을 멈추는 부분에서 운전자들의 성향이 공격적 성향인지 또는 환경친화적이고 항상 유사한 패턴인지를 확인하였다.

Hongyu et al.[16]은 운전자 식별 과정에서 대부분의 기존 연구들이 많은 양의 데이터들을 항상 필요로 했는데, 적은 양의 학습 데이터로도 충분히 좋은 성능을 나타냈다. [그림 10]에서 볼 수 있듯이, Deep Generative model을 이용하여 필요로 하는 운전자의 학습 데이터를 증강하는 방법을 적용하였다. 1차로 수집된 데이터를 Convolution condition variant autoencoder



(그림 9) 운전자 식별 프레임워크 및 시계열 데이터의 wavelet transform 적용 과정 (15)

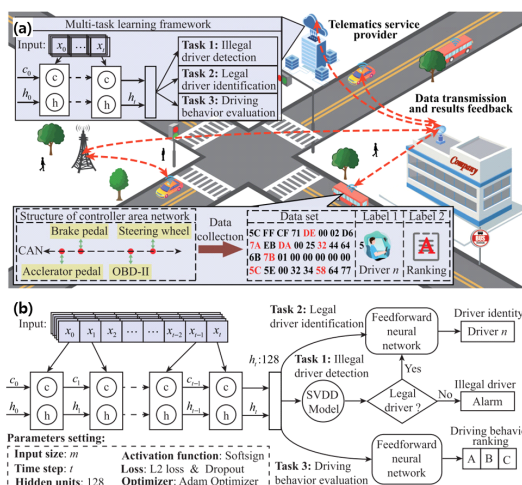


(그림 10) CCVAE 기반 data augmentation 과정 (16)

(CCVAE) 알고리즘에 학습한 후, CCVAE의 decoder에서 합성 데이터를 생성하여 기존의 주행 데이터와 결합 및 데이터를 증강 시킨다. 데이터를 증강할 경우, 증강하지 않은 경우보다 평균 10퍼센트 이상의 성능 향상이 있음을 확인하였다.

IV. 결론

본 논문에서는 운전자 식별 및 운전자 프로파일링에 대한 연구 동향에 대하여 알아보았다. 특히, 추출 가능한 데이터에 따른 운전자 식별 연구들을 분류 및 설명하였다. 주행 데이터의 개인정보보호 이슈로 인해, 학계에서의 연구 목적 외에 데이터를 공유하기 어려운 부분이 있지만, 이를 극복할 수 있는 새로운 방향의 연구들이 필요하다. 또한, 현재 차량에 적용 중인 새로운 형태의 센서 데이터를 적용한 연구들과 함께 다양한 주행데이터 분석 기반의 운전자 식별 및 행위 식별 연구들이 필요하다.



(그림 8) Multitask Learning에 대한 운전자 식별, 주행 행위 평가, 불법 운전자 탐지 관련 오버뷰 및 딥러닝 네트워크 구조 (a: 시스템 아키텍처, b: Multitask Learning Network 구조) (14)

참 고 문 헌

- [1] Sean McLain, "Kia, Hyundai Thefts Continue Three Months After Carmakers Deployed Fix", *The Wall Street Journal*, https://www.wsj.com/articles/kia-hyundai-thefts-continue-three-months-after-carmakers-deployed-fix-c4351e60?mod=tech_lead_pos4, Accessed: 2023-08-05
- [2] Digital key, <https://www.bmw.co.kr/ko/services/bmw-connected-drive/2022/digital-key.html>, Accessed: 2023-08-05
- [3] CAN BUS, https://en.wikipedia.org/wiki/CAN_bus, Accessed: 2023-08-05
- [4] CarSim, <https://www.carsim.com/>, Accessed: 2023-08-05
- [5] Vehicular data trace, <http://www.rettore.com.br/prof/vehicular-trace/>, Accessed: 2023-08-05
- [6] HCILAB Driving Dataset, <https://www.hcilab.org/research/hcilab-driving-dataset/>, Accessed: 2023-08-05
- [7] OCSLAB Driving Dataset, <https://ocslab.hksecurity.net/Datasets/driving-dataset>, Accessed: 2023-08-05
- [8] E. Romera, L.M. Bergasa and R. Arroyo, "Need Data for Driving Behavior Analysis? Presenting the Public UAH-DriveSet", *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, pp. 387-392, November 2016.
- [9] Desyllas, Panos, and Mari Sako. "Profiting from business model innovation: Evidence from Pay-As-You-Drive auto insurance." *Research Policy* 42.1: pp. 101-116. 2013.
- [10] Enev, M., Takakuwa, A., Koscher, K., & Kohno, T. "Automobile Driver Fingerprinting." *Proc. Priv. Enhancing Technol.*, 2016(1), pp. 34-50.
- [11] Hallac, David and Sharang, Abhijit and Stahlmann, Rainer and Lamprecht, Andreas and Huber, Markus and Roehder, Martin and Leskovec, Jure and others, "Driver identification using automobile sensor data from a single turn." *In 2016 IEEE 19th International Conference on Intelligent Transportation Systems*, pp. 953-958.
- [12] Sánchez, Sara Hernández, Rubén Fernández Pozo, and Luis Alfonso Hernández Gómez. "Driver identification and verification from smartphone accelerometers using deep neural networks." *IEEE Transactions on Intelligent Transportation Systems*, 23.1 (2020), pp. 97-109.
- [13] Xun, Yijie and Liu, Jiajia and Kato, Nei and Fang, Yongqiang and Zhang, Yanning, "Automobile driver fingerprinting: A new machine learning based authentication scheme." *IEEE Transactions on Industrial Informatics*, 16(2), pp. 1417-1426.
- [14] Xun, Yijie and Liu, Jiajia and Shi, Zhenjiang, "Multitask learning assisted driver identity authentication and driving behavior evaluation." *IEEE Transactions on Industrial Informatics*, 17(10), pp. 7093-7102.
- [15] Kwak, Byung Il and Han, Mee Lan and Kim, Huy Kang, "Driver identification based on wavelet transform using driving patterns." *IEEE Transactions on Industrial Informatics*, 17(4), pp. 2400-2410.
- [16] Hu, Hongyu and Liu, Jiarui and Chen, Guoying and Zhao, Yuting and Gao, Zhenhai and Zheng, Rencheng, "Driver Identification Using Deep Generative Model With Limited Data." *IEEE Transactions on Intelligent Transportation Systems*, 24(5), pp 5159-5171.

〈 저 자 소 개 〉

곽 병 일 (Byung Il Kwak)

증신회원

2013년 2월: 세종대학교 컴퓨터공학과 학사 졸업

2021년 2월: 고려대학교 정보보호대학원 정보보호학과 박사 졸업

2021년 8월: 고려대학교 정보보호대학원 연구교수



2021년 9월~현재: 한림대학교 정보과학대학 소프트웨어학부 조교수

<관심분야> 네트워크 보안, IoT 보안, 자동차 보안, 침입 탐지, 이상 탐지