Pedestrian Walking Direction Classification for Moroccan Road Safety

Safaâ DAFRALLAH
ENSA Kenitra, LGS Laboratory, Ibn Tofail University, Morocco
INSA Rouen, LITIS Laboratory, Rouen Normandie University, France
dafrallah.safaa@uit.ac.ma , safaa.dafrallah@insa-rouen.fr

Zakaria SABIR and Aouatif AMINE
ENSA Kenitra, LGS Laboratory
Ibn Tofail University
Kenitra, Morocco
zakaria.sabir@uit.ac.ma , aouatif.amine@uit.ac.ma

Stéphane MOUSSET and Abdelaziz BENSRHAIR
INSA Rouen, LITIS Laboratory
Rouen Normandie University
Rouen, France
stephane.mousset@insa-rouen.fr , abdelaziz.bensrhair@insa-rouen.fr

Abstract

This paper tackles one of the first causes of death all around the world. In fact, third world countries including Morocco suffers more from road accidents caused by undisciplined human behavior specially pedestrians. To address this issue, our work as part of a road safety project named SAFEROAD, aims to detect pedestrians and classify their walking direction in order to alert the driver of their presence using a new neural network algorithm named Capsule Network. As we work for the Moroccan case, we collected the first Moroccan pedestrian dataset from several Moroccan cities so as to consider the Moroccan pedestrian behavior in the proposed system. The collected dataset was used to train the proposed network giving an accuracy of 78.95% on the validation phase. The proposed approach shown better results compared to convolutional neural network algorithms.

Keywords
Pedestrian orientation, Capsule Networks, Road Safety, Moroccan Case, Moroccan Database.

1. Introduction

According to the World Health Organization (WHO) report, around 1.35 million deaths are reported every year worldwide due to road accidents (WHO, 2018). According to the same report, 54% of deaths are pedestrians, cyclists, and motorcyclists. Another report published by the Center for Accident Research and Road Safety-Queensland (CARRS-Q) concluded that 30\% of road fatalities are caused by the drowsiness of the driver (A.R.R.S, 2015). This rate can reach 50% in very specific cases such as fatal accidents involving a single vehicle.

This drowsiness is manifested by abnormal behaviors (slow reaction, inability to maintain the trajectory ...) as well as particular physiological characteristics (frequent yawning, difficulty keeping eyes open and head straight, etc.). There is an increasing need for knowledge of human behavior, this can be achieved gradually through the development of a variety of image and video processing methods.

In Morocco, road safety corresponds to a major public health and personal protection issue. According to statistics from the Ministry of Equipment, Transport, Logistics, and Water (METLE), the number of personal injury accidents reached 89,375 and every day an average of 10 people die and 361 are injured as a result of traffic accidents in 2017 the last statistics delivered by the ministry (METLE). Therefore road safety is placed at the heart of public policy in our country (METLE, 2017).

In this context, METLE launched calls for road safety research projects to remedy the problems of traffic accidents in Morocco. Our project named "SAFERAOD: a Meta platform for Road Safety", is one of the 6 accepted projects among 130 projects submitted throughout Morocco. It aims to provide the community with a new tool that is reliable and easy to use with new technologies. We are focusing on the realization of an intelligent system for road safety that solicits universal interest, caused by the increasing number of road accidents. We are interested in this research subject because it deals with a problem related to the safety of car drivers in road traffic. It will also highlight scientific and technical issues that constitute difficulties in the field of perception by artificial vision or computer vision.

In this paper, we are going to present this research project along with its main parts, and we will focus on the part that deals with humans, especially pedestrian's orientation. In this part we will exploit a new Moroccan pedestrian's dataset that we collected from two Moroccan cities in order to classify the pedestrian's walking direction using Capsule Networks.

2. Literature review

Significant studies have been done in pedestrian detection over the last decade using different methods of classification starting from hand crafted features to deep learning techniques. From the well-known methods classified as hand crafted features for human detection we can find the Haar-like features proposed by (Viola et al. 2003) that was firstly used for real time face detection, and improved later by researches like (Cui et al. 2007) and (Zhang et al. 2014) to perform pedestrian detection. The Histogram of Oriented Gradients (HOG) descriptor proposed by (Dalal and Triggs, 2015) was widely used for object detection and in particular for pedestrian detection by (Kobayashi et al. 2007) and (Keller et al. 2011). Belonging to hand crafted methods a new algorithm called the Aggregated Channel Features (ACF) for pedestrian detection was proposed by (Dollar et al. 2014), it contains 10 channels for each input image representing the normalized gradient feature, the Histogram of Oriented Gradients and LUV color. Those descriptors are generally followed by a classifier like the Support Vector Machine (SVM), random forests, AdaBoost and others.

Whereas, automated feature extraction methods also known by the deep learning methods are the most used in recent studies due to their competitive results in computer vision. In his paper (Redmon et al. 2016) proposed You Only Look Once (YOLO) algorithm for real time object detection (like car, cat, dog, bicycle ...) that contains a deep network of 53 convolutional layers. In this paper, we used YOLO for pedestrian detection and bounding box extraction. Once the pedestrian is detected the bounding box is then used to predict its orientation. In this sense, a Hidden Markov Model (HMM) was proposed by (Gandhi et al. 2006) to model transition between pedestrian orientations over time using 8 orientation bins. (Hara et al. 2017) used Deep Convolutional Neural Networks (DCNN) to predict pedestrians' orientation, while (Sanchez et al. 2017) used different CNN architectures (AlexNet, ResNet and GoogleNet) for pedestrian movement direction recognition where ResNet achieved the best accuracy with 94%.

As a matter of fact, Convolutional Neural Networks performed admirably in computer vision and object detection; however, it has some limitations related to the use of the pooling layer as it makes CNN lose valuable information from the input image. As an alternative, Sara Sabour and Geoffrey Hinton proposed in (Sabour et al. 2017) a new neural network called Capsule Networks (CapsNet), it has been tested in several researches but none of them treat the pedestrian classification or walking direction prediction this is the aim of our paper.

The rest of the paper is arranged as follows: Section 3 describes the Saferoad project, Section 4 represents the proposed approach, in this section we describe briefly Capsule Networks, our collected dataset and the classification algorithm, this section presents also the experimental results and the comparison with other neural networks algorithms; Section 5 concludes the paper.

3. SafeRoad Project Description

SAFEROAD is part of current projects that affect the scientific and civil community, it is a project that is part of a well-thought-out scientific study in road safety, indeed it deals with several aspects of this problem; namely: driver monitoring and behavioral analysis of road users, measurement of road safety indicators, accident scenarios and forecasting.

This project aims to analyze the existing in terms of scanning studies of methods and applications deployed internationally, proposing modeling of a reliable road safety system, and carrying out prototyping of this system. This will allow us to set up the SAFEROAD application which brings together surveillance, sensory geolocation, accident control and planning precautions in civil engineering. Our project is divided into three main factors (Road Safety with Vehicle as a Factor, Road Safety with Environment as a Factor and Road Safety with Human as a Factor). Each one plays an important role in the Intelligent Transportation Systems (ITS) as can be seen in Fig.1

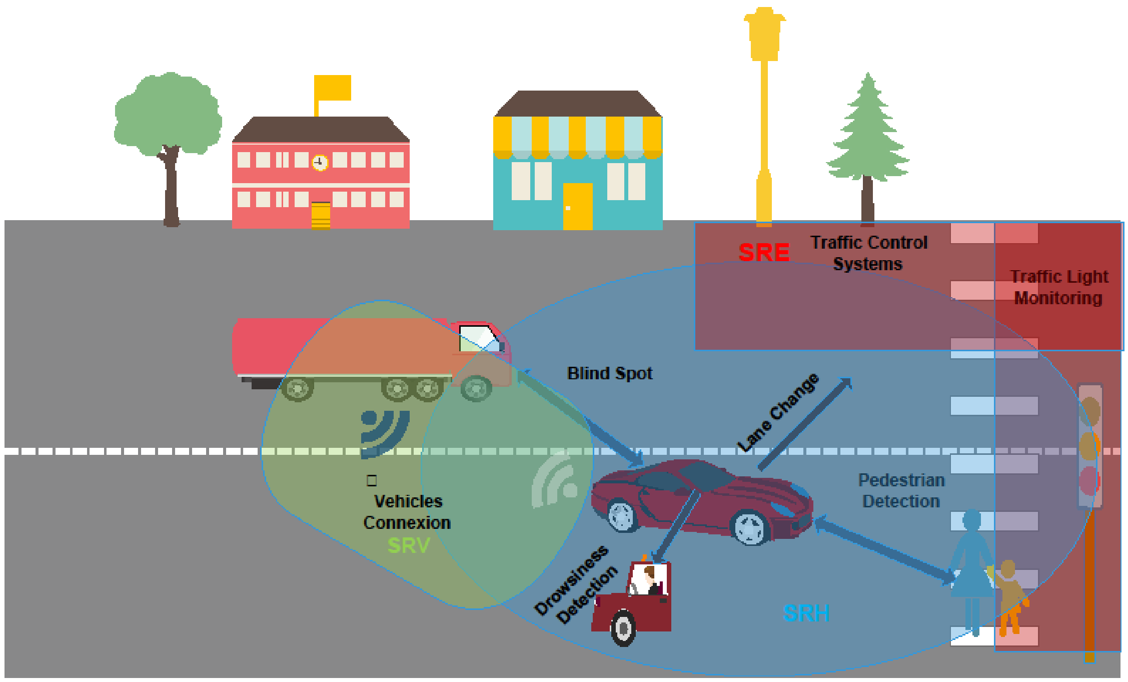

Figure1: SAFEROAD Project combine three main parts which are: Environment as a Factor (SRE), Vehicle as a Factor (SRV) and Human as a Factor (SRH)

3.1 Road Safety with Vehicle as a Factor (SRV)

The objective of the project in this part is to develop a system that acts to alert the driver of the risk of collisions, using the detection of entities around the vehicle, communication between vehicles (V2V): no infrastructure is used, vehicles communicate directly with each other) (Sabir and Amine, 2018) and between vehicles and infrastructure (V2I: use of stations installed along the road which also offer Internet access) (Sabir et al. 2019), and geolocation data (GPS) coupled with geographic information systems.

3.2 Road Safety with Environment as a Factor (SRE)

This part mainly focuses on three essential elements:

- Infrastructure conception: Priority in infrastructure conception must be given first to the following points: horizontal and vertical alignments, the width of the roadway, road markings, intersection design, dedicated space for pedestrians, parking areas, visibility distance, etc.
- Detection of traffic signs: The purpose here is to detect and recognize, from image analysis, certain characteristic shapes of road signs (square, triangle, round, octagon). The methods used in this part are based on the segmentation and classification in image processing.
- Road traffic management: We propose to install smart cameras with a controlled environment, in specific locations highlighting one or more accident-causing factors. In addition to the security that these cameras can provide, they will help to manage traffic via vehicle counting, for example.

3.3 Road Safety with Human as a Factor (SRH)

As the majority of accidents are human related, this part represents one of the main aspects of the project. We studied the human factor from two sides, the driver and the pedestrian side, where this last is what it concerns this paper.

Pedestrians in Morocco represents 27.70% of total victims in traffic accidents in 2017 (METLE, 2017) therefore it becomes a requirement to figure out this issue. By the aim of this work, we propose a system to detect pedestrian orientations that could be integrated in Advanced Driver Assistance Systems (ADAS) in order to alert the driver of the presence of a pedestrian. This paper tackle the pedestrian orientation classification using a new deep learning algorithm called Capsule Networks, the results of this work outmatch those proposed in (Dafrallah et al. 2019) by the use of different learning algorithms which increase the accuracy of the classification.

4. Proposed Approach

The aim of this work is to contribute in decreasing fatal injuries for vulnerable road users in Morocco, by detecting the pedestrian's orientation and alerting the driver if the pedestrian's direction is towards the vehicle, taking in consideration other constraints as the distance and the vehicle's speed.

In this paper we focused on the pedestrian direction detection using Capsule Networks by using different optimization algorithms which improves the positive detections accuracy in this work than the one published in (Dafrallah et al. 2019).

To evaluate the network, we collected a new pedestrian orientation dataset from several Moroccan cities so as to make this approach suitable for the Moroccan case, and as to the best of our knowledge, there is no dataset to date treating the Moroccan pedestrian.

4.1 Capsule Networks

A capsule is a group of neurons that takes as input the features map of an entity and outputs a vector which activity represents the instantiation parameters of the entity and its length represents the existence probability of the entity.

Capsule Network contains three layers, the first one is a convolutional layer fed with the input image and outputs the features maps. This output becomes the input of the second layer called the Primary Capsule layer (L), it predicts the output of the next layer called the class capsule layer (L+1), in other words it predicts which class relating to the entity represented by the capsule should be activate. This process is done using a routing algorithm called "routing by agreement". The predicted output $\hat{u}_{j/i}$ is first calculated using a product between the output vector u_i of each capsule *i* in layer L and a weight matrix W_{ij}:

$$\hat{u}_{j/i} = W_{ij} u_i \quad (1)$$

The first step of the routing process is to calculate the routing weight b_{ij} which is initiated by zero. This routing weight is fed afterward to a softmax function to give as output a coupling coefficient c_{ij}, this last is used to computes the weighted sum s_j of each primary capsule predictions using eq.2.

$$s_j = \sum_i c_{ij} \, \hat{u}_{j/i} \quad (2)$$

A squashing function v_j is then applied to the weighting sum giving the actual output of capsules in layer L+1:

$$v_j = \frac{||s_j||^2}{1+||s_j||^2} \frac{s_j}{||s_j||} \quad (3)$$

This process represents the first iteration of the routing algorithm. In the next iteration the routing weight bij, is recalculated using the scalar product between the prediction output and the actual output of the layer L+1 as represented in eq.4.

$$b_{ij} = b_{ij} + \hat{u}_{j/i} . v_j \quad (4)$$

This step is used to study the agreement between the prediction and the actual output, if the two outputs match together the routing weight becomes large, else it will be small. Thereafter, the routing process is calculated again and the capsule representing the class of the input entity will have the highest output vector.

4.2 Optimization Algorithms

Optimization Algorithms have a significant effect on the performance of the neural network, their objective is to find an optimal solution to minimize the loss of the network during the training process. Among existing algorithms Gradient Descent is the most popular one, the gradient is calculated in each iteration of the training thereafter the weights and biases are updates which helps to minimize the cost function and reach the global minimum value. The RMSProp is another optimizer that helps to find the global minimum by adjusting the learning rate for each iteration step and restricts the oscillation into one direction so as to converge faster. The Adagrad Optimizer adapts also the learning rate based on the parameters by referring to the past gradients using the decay mean of all past squared gradients, while its extension Adadelta limits the past gradients to a fixed size. The Adaptive Moment Estimation known by AdamOptimizer computes the adaptative learning rate for each parameter and stores an exponential decay average of past gradients. The optimization algorithms aforesaid are the ones used in this work to improve the performance of the network.

4.3 Data Collection

In order to explore the behavior of Moroccan pedestrians we recorded the Moroccan road in several main cities, using For the purpose of recognizing the pedestrian's orientation in a Moroccan environment, we collected the first Moroccan pedestrian dataset from different Moroccan cities, using an industrial camera with a resolution of 2.3 megapixels and 60 frames per second, on board a moving vehicle.

We extracted the pedestrians using YOLO algorithm and classified the pedestrians manually to four walking directions (Left, right, front and back) as shown in Fig.2. Actually the dataset contains 5160 pedestrians that we aim to increase for further work before making it publicly available.

Figure2: Samples of the four orientations in the Moroccan collected dataset

4.4 Pedestrian walking direction classification using Capsule Networks

The aim of this work is to build a pedestrian orientation classification system using Capsule Networks, for this purpose we fed to the system raw images captured from real road traffic containing pedestrians crossing in different directions. To extract the pedestrian bounding box we used the YOLO algorithm due to its high accuracy in real time detection. Thereafter, the pedestrian crop samples are resized to 48x48 grayscale images to be used after that as inputs to the capsule networks as illustrated in Fig.3.

The network contains in the encoder part two convolutional layers with 64 and 128 filters respectively and a kernel size of 5x5, a primary capsule layer of 16 channels and a class capsule layer of 4 channels as our system contains four classes. This last layer classifies the pedestrian image into one of the four walking directions aforementioned. While in the decoder part it contains three fully connected layers with 512, 1024 and 2304 filters.

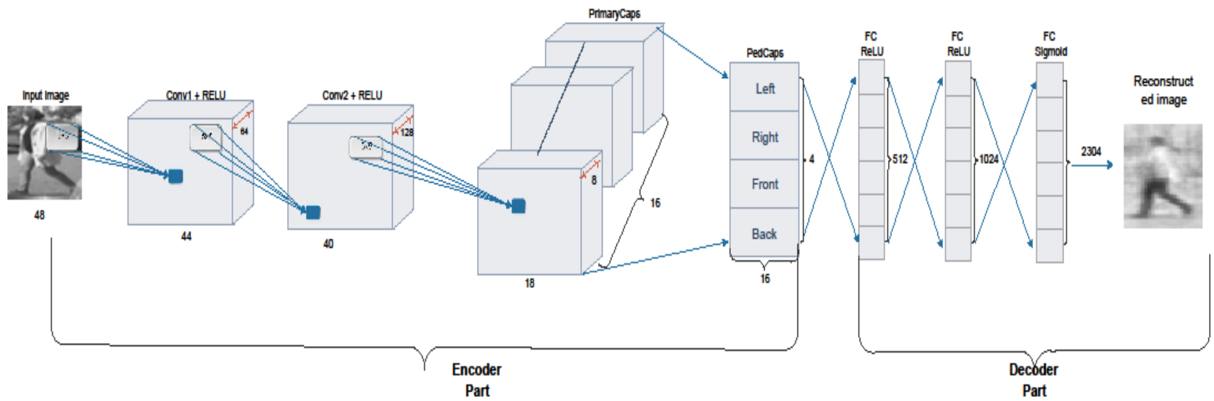

Figure 3: Capsule Network architecture for pedestrian walking direction detection.

4.5 Results and Discussion

The experimental results were validated using our collected dataset described above. We used in this paper multiple optimization algorithms such as the Gradient Descent, AdamOptimizer, Adadelta, Adagrad and RMSProp, in order to improve the work done in (Dafrallah et al. 2019). The above CapsNet architecture was compared by the use of CNN algorithms (AlexNet and ResNet) as it is described in Table 1.

Despite having the longest training time, Capsule Network, gives the best accuracy compared to CNN algorithms (AlexNet and ResNet) as it is described in Table 1. Practically for all the networks, RMSProp Optimizer gives the best result with 78.95%, 77.60% and 78.54% for CapsNet, AlexNet and ResNet respectively.

In almost all the optimization algorithms CapsNet gives the best detection accuracy, except for the Gradient Descent Optimizer where ResNet exceeds CapsNet with an accuracy of 78.12%. However, the best accuracy over all the evaluated cases was for CapsNet architecture using RMSProp Optimizer with an accuracy of 78.95% using a learning rate of 0.01, overpassing the previously published work with an accuracy of 73.64% using Adam Optimizer.

Table 1: Comparison of the Pedestrian Orientation Classification accuracy on different architectures and learning algorithms

Optimizers / Architectures	AdamOptimizer	Gradient Descent	Adadelta	Adagrad	RMSProp
CapsNet	73.64%	75.41%	66.87%	76.87%	**78.95%**
AlexNet	70.10%	77.60%	61.35%	74.37%	77.60%
ResNet	73.12%	78.12%	56.66%	75.41%	78.54%

5. Conclusion

In this paper, we presented a novel solution for road safety named SAFEROAD project that aims to reduce fatal accidents in Morocco using Machine Learning and Computer Vision Algorithms in three main factors: SRV that targets the Vehicle to Vehicle (V2V) and Vehicle to Infrastructure (V2I) communications, the SRE concerns infrastructure conception, detection of traffic signs and road traffic management, and the SRH which is the main part of this paper, it focuses on the pedestrian orientation detection in order to prevent pedestrian accidents by alerting the driver of the presence of a near pedestrian giving its walking direction. The classification was done using Capsule Networks giving an accuracy of 78.95% using the RMSProp Optimizer.

Another main contribution of this paper is the new Moroccan pedestrian orientation dataset gathered from several Moroccan cities so as to analyze the Moroccan pedestrian behavior.

For future works, we aim to detect the pedestrian orientation on a sequence of frames so that to include it in an anti-collision system to prevent intended accidents in poor structured areas.

Acknowledgements

This work is part of the project "SAFEROAD Meta-plateforme pour la Sécurité Routière (MSR)" which is supported by the METLE and the National Center of the Scientific and Technical Research (CNRST) under contract No: 24/2017 and by the LITIS laboratory at INSA Rouen as part of a joint supervision thesis.

References

A.R.R.S, *Monotony and hypovigilance screen*. Tech. rep, 2015.

Cui, X., Liu, Y., Shan, S., Chen, X., Gao, W., 3d haar-like features for pedestrian detection, *IEEE International Conference on Multimedia and Expo*, pp. 1263-1266, 2007.

Dafrallah, S., Amine, A., Mousset, S., Bensrhair, A., Will capsule networks overcome convolutional neural networks on pedestrian walking direction? *IEEE Intelligent Transportation Systems Conference (ITSC)*. pp. 3702–3707, 2019.

Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection, *Computer Vision and Pattern Recognition*, IEEE ComputerSociety Conference on. vol. 1, pp. 886–893, 2005.

METLE, *Compendium of Statistics of Road Traffic in Personal Accidents*. Tech. rep, 2017.

Dollàr, P., Appel, R., Belongie, S., Perona, P, Fast feature pyramids for objectdetection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(8),1532–1545, 2014.

Dominguez-Sanchez A., Orts-Escolano, S., Cazorla, M, Recognizing pedestrian direction using convolutional neural networks, *International Work-Conference on Artificial Neural Networks.* pp. 235–245, 2017.

Gandhi, T., Trivedi, M.M, Pedestrian collision avoidance systems: A survey of computer vision based recent studies, *IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 976–981, 2006.

Hara, K., Vemulapalli, R., Chellappa R, Designing deep convolutional neural networks for continuous object orientation estimation, arXiv, preprintarXiv: 1702.01499, 2017.

Keller, Christoph Gustav, Markus Enzweiler, and Dariu M. Gavrila. A new benchmark for stereo-based pedestrian detection, *IEEE Intelligent Vehicles Symposium (IV)*, p. 691-696, 2011.

Kobayashi, T., Hidaka, A., Kurita, T., Selection of histograms of oriented gradients features for pedestrian detection, *International conference on neural information processing*, p. 598-607, 2007.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A, You only look once: Unified real-time object detection, *Conference on Computer Vision and Pattern Recognition*, pp. 779–788, 2016.

Sabir, Z., Amine, A, Connected vehicles using NDN for intelligent transportation systems, *International Conference on Industrial Engineering and Operations Management,* vol. 2018, pp. 2433–2441, 2018.

Sabir, Z., Dafrallah, S., Amine, A.: A Novel Solution to Prevent Accidents using V2I in Moroccan Smart Cities, *International Conference on Computational Intelligence and Knowledge Economy*, p. 621-625, 2019.

Sabour, S., Frosst, N., Hinton, G.E., Dynamic routing between capsules, *Advances in Neural Information Processing Systems*, pp. 3856–3866, 2017.

Viola, P., Jones, M.J., Snow, D., Detecting pedestrians using patterns of motion and appearance, *International Journal of Computer Vision*, p. 734, 2003.

W.H.O, *Global status report on road safety,* 2018.

Zhang, S., Bauckhage, C., Cremers, A. B., Informed haar-like features improve pedestrian detection, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 947-954, 2014.

Biographies

Safaâ DAFRALLAH is currently a Joint PhD Student in computer science at the LGS Laboratory, at the National School of Applied Sciences ENSA, Kenitra, Morocco and at the LITIS Laboratory, INSA Rouen, France. She received her M.Sc. in computer science at ENSA, Ibn Tofail University, Kenitra, Morocco. Her research interests includes pedestrians detection and orientation recognition using deep learning based approaches in the field of Intelligent Transportation Systems.

Zakaria SABIR is a Ph.D. student of computer engineering at National School of Applied Sciences (ENSA), Kenitra, Ibn Tofail University. He received his Master's degree in Information Systems Security from the ENSA of Kenitra, and he is currently completing his researches on field of Connected Vehicles and Intelligent Transportation Systems with the BOSS team in the LGS Laboratory, ENSA of Kenitra, Morocco.

Aouatif AMINE is an associate professor at ENSA, Kenitra, Ibn Tofail University. She received in 2004 her M.Sc. and PhD degree from the Faculty of Sciences, Mohammed V Agdal University, Rabat, Morocco, in 2004 and 2009 respectively. Her research interests include but are not limited to dimensionality reduction, feature selection applied to face detection and recognition and driver hypo-vigilance. She has been a Vice-Chair of IEEE Signal Processing Society, Morocco Chapter and since 2015 she's the BOSS team header in the LSG Laboratory, ENSA Kenitra, Morocco.

Stéphane MOUSSET received in 1997 his Ph.D. degree in computer science from the University of Rouen, France. His Ph.D. dissertation focused on the estimation of axial motion using a stereo vision system. He is currently an Assistant Professor at the Information Processing, and Systems Laboratory (LITIS), at the National Institute of Applied Sciences (INSA), University of Rouen. His research interests include stereo vision analysis, road application, driving assistance systems, and sensors.

Abdelaziz BENSRHAIR received the M.S. degree in electrical engineering and the Ph.D. degree in computer science from the University of Rouen, France, in 1989 and 1992, respectively.
He was the head of Intelligent Transportation Systems (ITS) team from 2007 to 2012, and the co-director of the Information Processing, and Systems Laboratory (LITIS) from 2002 to 2016 at the National Institute of Applied Sciences. Currently he is a Professor in the Department of Information, Systems, and Architecture.