COMPUTER VISION

A MOBILE APPROACH TO MODERN AUTOMOTIVE

Dao Huu Hung | September 20, 2017

ABSTRACT

The automotive industry has been being revolutionized by three important trends (1) the shift to Electric Vehicles, (2) Internet of Vehicle for interconnections between vehicles and connections between vehicles and infrastructure, and (3) Autonomous Driving. The former is driven not only by the shortage of fossil energy but also by the full intervention in steering control of Electric Vehicles, facilitating Autonomous Driving function. The advances of camera technology and computer vision techniques have been emerging as indispensable keys to revolutionize the automotive industry. It enables various Advanced Driving Assistance System (ADAS) products and the recent rise of Autonomous Driving. Although there exist a variety of commercial computer vision-based ADAS products and Autonomous Driving systems, our quest for future generation of intelligent vehicles is still in the middle of nowhere. These products are implemented on embedded platforms which are still the most dominant in the automotive industry. However, the computation capability of personal mobile devices like iPad and Android tablets is increasingly powerful, making it very suitable to automotive applications. This white paper therefore addresses a mobile computer vision-based approach to modernize the automotive industry in years to come.

TABLE OF CONTENTS

I Introduction

Technology revolutions have been taking place in a fast face to improve every corner of our life. The proliferation of mobile devices and cloud technology make us connected everywhere. The advances in AI Vision and computing technologies add intelligences to the devices. Battery technology now expands the running distance per single charge and reduces significantly charging time of Electric Vehicles, etc. The breakthroughs of these supporting technologies have been driving a new revolution in the automotive industry.

Firstly, all car makers have been paying more attentions to Electric Vehicles (EV). According to a prediction in Nikkei paper, the production of EV will globally reach 15000000 by 2025. Chinese government recently imposes a rule of significantly restricting combustible engine vehicles. Similar sanctions against combustible engine vehicles will also be imposed by other governments in the near future. EV will be apparently our future in consideration of its eco-friendly factor and the shortage of fossil energy. Moreover, electric motor drives offer high fidelity of steering control to EV, enabling the second wave of attention in automotive industry – the recent rise of autonomous driving. Finally stemming from the concept of IoT, there is an increasing need to connect cars together and with supporting infrastructure, so-called Internet of Vehicle. The data collected from operating vehicles are valuable for makers in key applications such as predictive troubleshooting and maintenance. The infrastructure also provides critical information to vehicles to ameliorate driving experience which has never seen before as shown in Fig. 1.

Fig. 1: Electric Vehicles and Internet of Vehicles have been driving a new revolution in automotive industry.

It is inevitable that computer vision is at the heart of the new revolution in the automotive industry. Although various kinds of sensors like laser, radar, sonar, and LIDAR etc. are used in automotive, vision sensor is still considered as the most important. It is like when a person would like to go to somewhere, the eyes are the most important among 6 types of human sensory. A variety of computer vision-based commercial ADAS products such as lane departure warning systems, adaptive cruise control systems, and obstacle collision avoidance systems, etc. are widely used in reality. The main purposes of ADAS include provision of comfortable driving conditions and the increase in driving safety. ADAS can be deployed either in internal combustible engine vehicles or in EV, and is a crucial foundation of autonomous driving. Recent successful autonomous driving technology presented by Tesla and NVIDIA is also relied solely on camera images as input. However, the technology is not mature or intelligent enough to avoid fatal accidents which likely happen even under a small percent of chance. It is simply unacceptable in consideration of human safety first. Our quest for future generation of intelligent vehicles is still in the middle of nowhere.

In terms of computation devices, the automotive industry is relying solely on embedded platforms since it is impossible to deploy a computer in the vehicle. The main disadvantage of embedded devices is its relatively weak computation capability. It prevents us from realizing various interesting applications, for instance, ADAS. However, computation capability of personal mobile devices like iPad and Android Tablet, smart phone is increasingly powerful. Some are even equipped by GPU. It is feasible to replace some embedded devices on the vehicle by the personal mobile ones. For example, it is a trend to replace current navigators by personal mobile devices. Navigation functions can be purely carried by mobile software. Other available services on the mobile devices like location and route recommendation, public chat platform, chatbot, etc., definitely enrich our navigation experience.

Therefore in this white paper, we introduce a mobile vision-based approach to modernize the automotive industry. Mobile tablet and smartphone are necessary and indispensable stuffs in our daily life. Now it becomes even more useful when we are driving. We mainly focus on implementing various vision-based ADAS problems on the personal mobile devices such as surrounding-view image reconstruction, surrounding-view image-based auto parking, object detection, recognition, and tracking, etc. These ADAS functions are crucial foundations of increasingly automated driving. We strongly believe that our proposed mobile approach can be regarded as a roadmap of this field. Not only the concepts but technical features are also presented and discussed along with some preliminary results.

The rest of the paper is organized as follows. A mobile vision-based approach to modern automotive is presented in section II. We demonstrate various computer vision techniques solving a class of ADAS problems on the mobile devices in section III. The white paper is concluded in section IV.

II A mobile computer vision-based approach to modern automotive

Computer vision-based ADAS is not a new concept. Various commercial ADAS products are available on our doorstep. The number of academic publications dedicated to this field is also significantly increased. Most of prototypes reported in academic literature are implemented on strong computer since academic interests focus more on accuracy. Meanwhile, the industry relies solely on embedded capability which generally cannot realize state-of-the-art technology. Nowadays, computing capability of personal mobile devices are more and more powerful, that is, various modern computer vision methods can be realized in real-time on the mobile devices.

The main purpose of this section is to describe our mobile computer vision-based approach to modern automotive in general and ADAS in particular. Since the main application of computer vision in automotive is about ADAS, the rest of the paper treats ADAS as the main subject. We also strongly aim at defining a roadmap of this field in years to come because such mobile ADAS system is urgent by popular demand.

ADAS is generally broken into 2 categories: (1) in-car and (2) out-car. The former mainly concerns with the safety of both drivers and passengers even in case of no fatal accidents actually happening. The bus safety surveillance system in [Dao, ほか 2015] epitomizes the former. Other systems in the fomer include driver drowsiness detection [Hu , Zheng 2009] and driver monitoring system [Boguslaw , Slawomir 2014]. The latter focuses more on understanding the environment surrounding vehicles to assist drivers in making driving decision, including pedestrian detection [Zhang, ほか 2016], lane departure warning system [Son, ほか 2015], traffic sign detection system [Jin, Fu , Zhang 2014], sourrounding view reconstruction [Kalus, ほか 2017], automatic parking system [Kazukuni, ほか 2015], blind spot detection system [Lin, ほか 2012], etc.

Above systems assume that a strong computer can be deployed on the vehicle but it seems to be impractical. Some techniques are really heavy in computation and some require GPU computing resources. To deal with the heavy computational burden, ADAS chip makers often design several image processing techniques such as surrounding view reconstruction, smoothing, and lighting equalization, etc in embedded hardware. However, we believe that capability of software is tremendous which needs to be selected (simple but effective methods) and optimized to fit in mobile computational platforms.

Recent breakthroughs in mobile computing devices show promising feasibility in realizing the algorithms, even deep learning, in real time. Embedded platforms with GPU, like NVIDIA Jetson TX1 and TX2, are now available on our doorstep. Even our personal mobile devices like iPad and Android Tablet are also equipped by GPU to render high-quality images and video for entertainment and gaming purposes. Moreover, squeezed convolutional neural network [Forrest, ほか 2016] is a simplified model of AlexNet [Alex, Ilya , Geoffrey 2012] which achieves relatively competitive accuracy with original AlexNet model but its model size is less than 0.5 Mb. Its attractive attribute makes it potential for deep learning implementation in mobile devices.

Therefore in this white paper, we introduce our systematic computer vision-based approach to build a complete ADAS system on mobile platforms. Various computer vision techniques, especially the low-cost ones, are optimally implemented and integrated to solve a class of ADAS problems, including multiple lane detection, surrounding-view image, assisted parking, object detection, object-vehicle distance, and driver drowsiness detection. Our systematic approach is illustrated in Fig. 2. Details of each method will be presented along with preliminary results in section III.

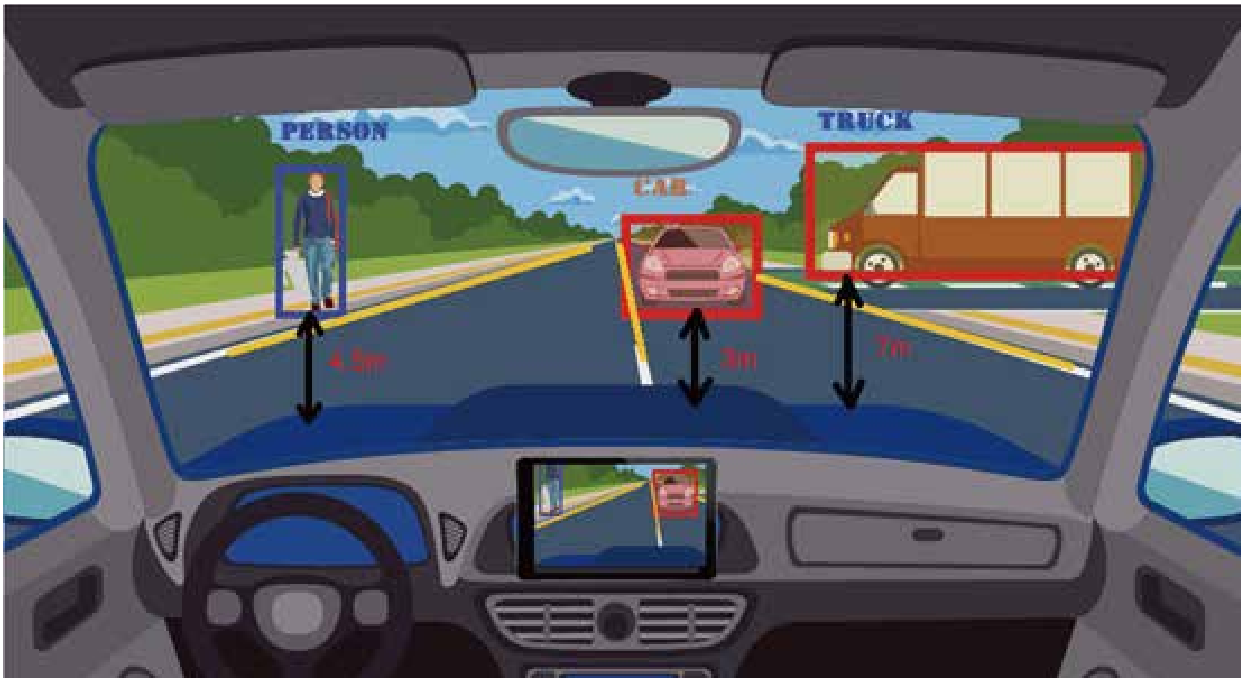

Figure 2: Our proposed mobile approach to ADAS system including multiple lane detection, object detection and tracking, object-vehicle distance measurement, surrounding view, and driver monitoring functions.

III Methods and Experiments

In order to realize a complete system of ADAS on mobile platforms, we focus mainly on simple but effective methods of computer vision. The tradeoff between computational cost and accuracy is taken into consideration. Besides low-cost computer vision methods, we also take advantages of the likes of squeezed CNN to enhance the accuracy in object recognition which is crucial in any ADAS system. The complete system presented in this section is able to detect multiple lanes, detect objects including vehicles, pedestrian, measure object-vehicle distance, and supervise driver consciousness.

1. Multiple Lane Detection

Lane information is critical in ADAS since it allows us to localize our vehicle on the road in relation with the others and also to show the moving forward direction of the vehicle. Lane facilitate various ADAS functions like Lane Departure Warning, Lane Change Assistance, and Forward Collision Warning, etc. Lane is also the first thing we need to consider when developing autonomous driving cars. In this section, we present a low cost solution to detect multiple lanes from images captured by an ego camera so that it can be deployed on mobile platforms.

Figure 3a shows the flowchart of our method, containing major modules such as lane segmentation, curve fitting and lane visualization. To reduce the computational cost, ROI (region of interest) must be specified in advance so that lane within 30 meters in front of the vehicle will be localized. Since lane is often painted by white and yellow colors, it had better to convert original images from RGB space to YUV space so that lane pixels are able to be well isolated. Once lane pixels are isolated, lane edges are segmented by canny algorithm and lane or line segments are extracted by Hough Line algorithm. In straight roads, one lane line is generally determined by one-line segment but in curved roads, multiple line segments form one lane line. The final results are visualized on the image as shown in Fig. 3b. The method of multiple lane detection is simple but effective, especially being suitable to mobile platforms.

In lane departure warning, lane change assistance applications, road curvature is indispensable information. Therefore, curve fitting algorithm is necessarily implemented. Lane segments are fitted into a curve model for computing road curvature. In other applications like forward collision warning, this module should be omitted for the sake of computation reduction.

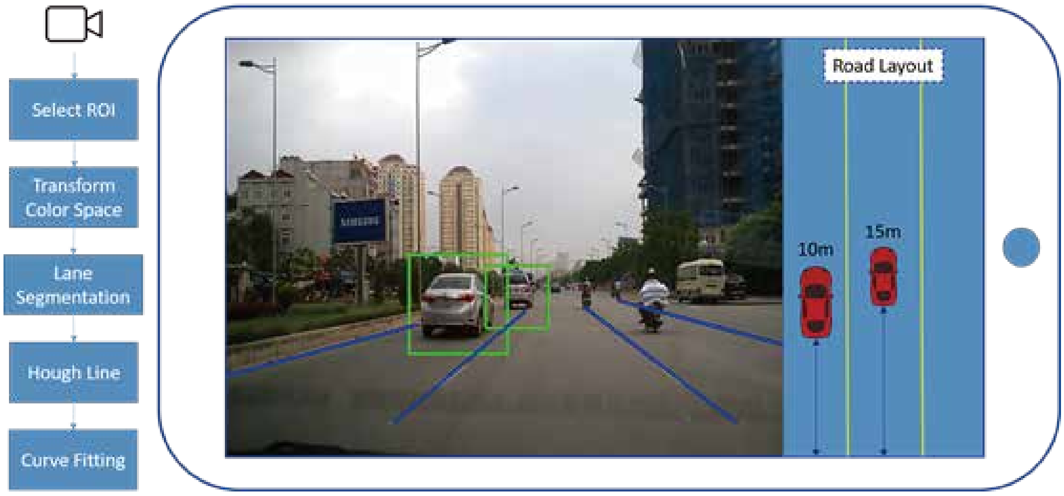

Figure 3: a (Left) Flowchart of our solution and b (Right) Results of multiple lane detection, vehicle detection and vehicle-to-vehicle distance estimation are implemented in mobile platforms.

2. Surrounding-View image and its applications

It is necessary to provide full view surrounding the vehicle to drivers for safety enhancement purpose. To this end, at least four cameras usually with lenses are positioned on the vehicle so that they can capture the full view surrounding the vehicle. A virtual image is synthesized and reconstructed from images captured simultaneously by those cameras, but from the top viewpoint, so-called surrounding-view image. An example of a surrounding-view image is shown in Fig. 4.

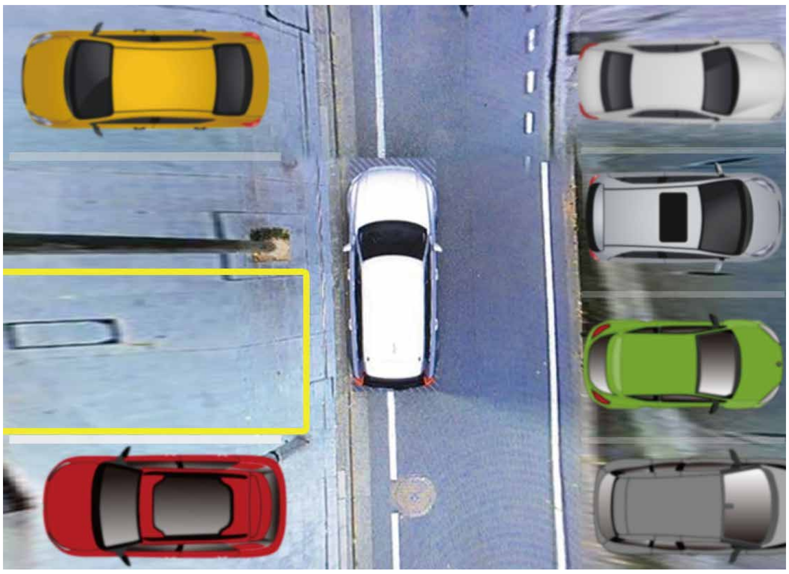

Figure 4: An example of surrounding view and its application in parking assistance. Parking slots are detected and marked by parallel black lines. Empty parking slots are detected and marked by yellow rectangles.

There are many applications of surrounding-view images, including blind spot detection and parking assistance, a foundation of an auto-parking system. In this section, we introduce our solution to reconstruct the surrounding-view images and its application to assisted parking as well as blind spot detection.

To reconstruct and synthesize surrounding-view images, original images captured by cameras must be undistorted since images captured by fish-eye cameras are highly distorted. We use one more camera, positioned 10 meters above the vehicle, to capture the top-view image for registration purpose. It is noted that the viewpoints of those five cameras are fixed for calibration and registration. The registration process is done in a semi-auto manner. The image region occupied by the vehicle in the top-view image is manually specified. Four special distinctive markers are placed in the views of the four side cameras for estimating homography matrices between these four side cameras and the top-view camera. Once the homography matrices are obtained, the top view camera can be removed. It is because the top-view image can be reconstructed by images captured from the four side cameras by using homography transformation.

One potential application of surrounding-view images is the assisted parking system in which empty parking spaces are automatically detected from surrounding-view images for guiding drivers to a perfect parking. Parking spaces are generally painted by white or yellow lines. The process of detecting parking line segments is similar to the one presented in section 1. A parking space is defined by two parallel lines. It is noted that the surrounding-view image is non-distorted then it is straightforward to detect two parallel lines. We also assume that parking spaces are located around 5 meters from either sides of the vehicle. Users must drive the vehicle to parking areas and turn on this assisted parking mode. It is therefore not necessary to detect parking spaces which are very far from the vehicle. The determination of ROI image is also straightforward to decrease the computation cost. The vacant parking spaces are distinguished from the occupied ones by the number of corners. Vacant parking spaces contain road surface which exhibit few corners in comparison with occupied ones. In the future, we will upgrade the system to fully automated parking by localizing relative position between the vehicle and the vacant parking space and driving the vehicle a perfect parking.

Another application of surrounding-view images is the blind spot detection system. We believe that there is no blind spot existing in surrounding-view images. In practice, the blind spots generally happen when the vehicles, especially heavy trucks, take a turn. Similarly, the assisted parking system is merely used in parking. In other words, both these systems are solely used in very short durations. The deployment of the two systems on mobile platforms is highly suitable.

3. Object Detection

Detecting objects surrounding the vehicle such as pedestrians, traffic signs, and other vehicles is extremely important in ADAS. Pedestrian protection system, forward collision warning system, intelligent speed advice system and wrong-way driving warning system, etc., to name a few, are among ADAS applications which take advantage of object detection algorithms.

Even though a variety of object detection algorithms have been proposed in the literature [Piotr, ほか 2011] , the state-of-the-art, the likes of Deformable Part Model [Pedro, Ross , David 2010], Regional Convolutional Neural Networks [Zhang, ほか 2016], Sharp Mask [Pedro, ほか 2016] often focus on accuracy. They are generally computational expensive and are not feasible to be deployed on mobile platforms. To balance the tradeoff between accuracy and computational cost, we opt for traditional HOG-based cascaded classifier method but with intelligently optimized strategy to alleviate the processing time as demonstrated in Fig. 5.

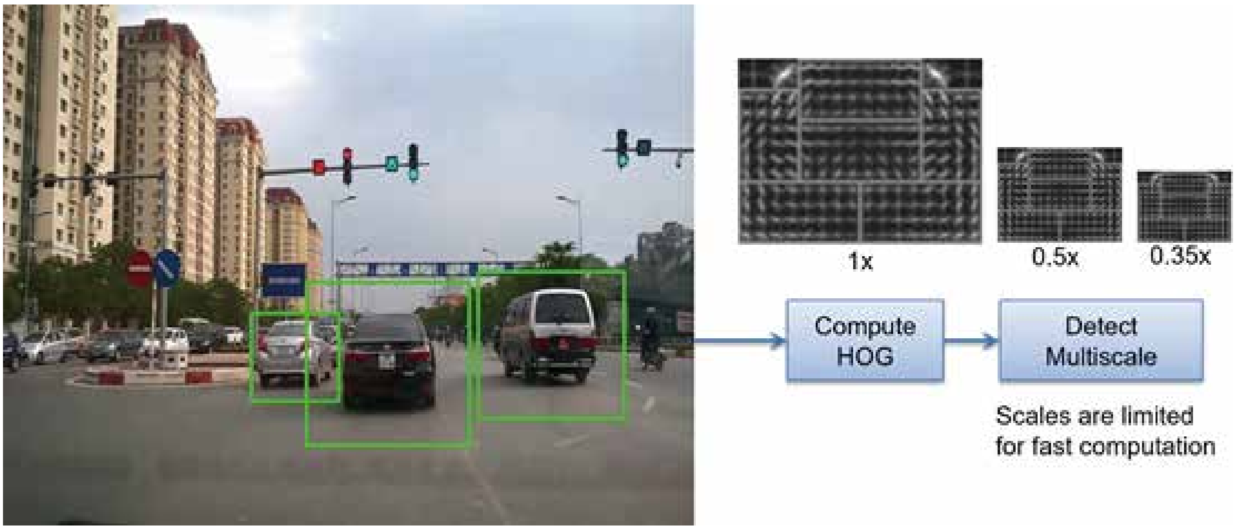

Figure 5: Objects are detected by using HOG features and selective-scale mechanism.

First we select a ROI by taking horizontal perspective information into account, generally excluding regions associated with sky, our own vehicle, and irrelevant parts. HOG features can be computed by integral operation whose computation is efficient [Piotr, ほか 2011]. To localize objects on images, multiscale sliding window method is often utilized. Since the camera resolution and viewpoint are fixed, scales of objects on images like people and vehicles are both upper and lower bounded. We therefore limit the searching scale within the upper and lower bounds, in turn, leading to tremendous deduction in computational cost but attaining a competitive detection accuracy. This intelligent strategy was adopted in [Rodrigo, ほか 2012], achieving a performance of 135 fps on street scenes. Figure 3b and 5 demonstrate our experimental results of detecting vehicles.

4. Object-Vehicle Distances

Sections 1 and 3 delineate where our vehicle and surrounding objects are on the images. Now we need to discover relative relationship between our vehicle and surrounding objects by computing so-called object-vehicle distances. By measuring the distance, it enables various ADAS applications such as forward collision warning, intelligent speed adaptation, and adaptive cruise control, etc.

Our proposed system is composed of two phases: offline calibration and registration and online measurement. In the former, we applied a similar method presented in section 2. Beside the ego camera, an auxiliary camera is positioned about 10 meters above the vehicle but look obliquely forward. We perform calibration and registration for the two cameras. The registration process is to estimate homography of the ground between the two cameras. The homography establishes the transformation of images captured by the ego camera to the images captured by the auxiliary camera. The calibration process which is carried on the auxiliary camera, is to create a unit transformation from image coordinates to real-world coordinate.

In the latter phase, we detect objects appearing in images captured by the ego camera and marked them by bounding boxes. The images are transformed to the view of the auxiliary camera by using homography transformation. The distances between our vehicle and other vehicles are defined from the bottom border of transformed images to the bottom line of bounding boxes. Thus, we perform the homography transformation of the bottom line of bounding boxes only to reduce computational cost. By using calibration information, the measurement in pixel can be converted into meters. The measurement errors of our proposed approach are associated with the error in object localization and homography transformation. The error tolerance is about 2 meters in our in-house experiments.

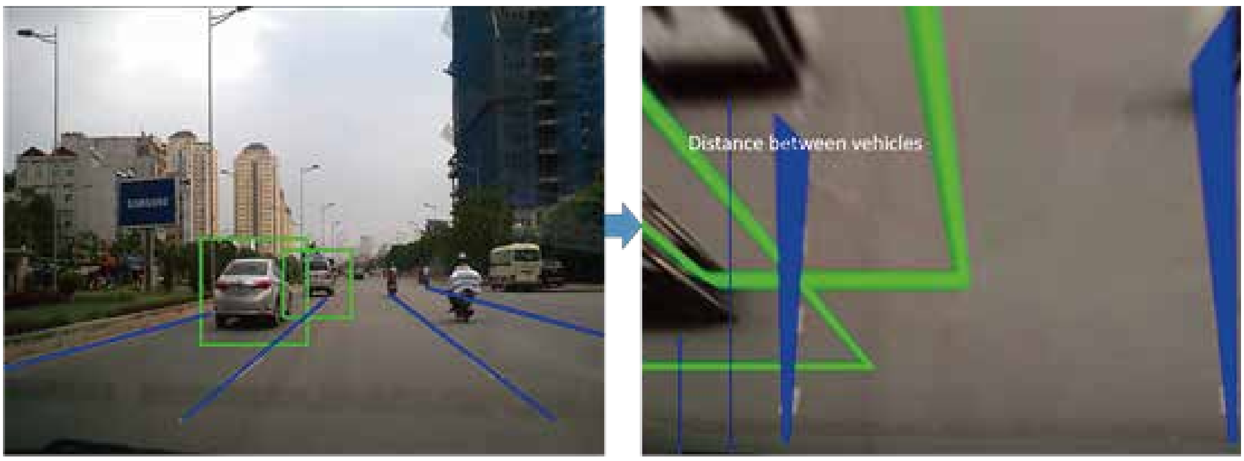

Figure 6: vehicle to vehicle distance measurement

5. Driver Drowsiness Detection

Driving in fatigue conditions is extremely dangerous not only for drivers but also for other surrounding people. It had better to have a system keeping an eye on the drivers non-intrusively to detect whether or not they are in fatigue conditions. A warning will be issued or an appropriate action will be applied in order to guide the drivers to safe conditions. A variety of reports in the literature [Arun, Kenneth , Murugappan 2012], [Dong, ほか 2011] attempted to determine consciousness of drivers by various kinds of sensors such as electrocardiogram, electromyogram, electrooculogram, camera, and other in-vehicle sensors. However in this section, we focus solely on a low-cost vision-based approach which can be deployed on mobile platforms.

In order to detect drowsiness symptoms, it had better to brief stages of sleep which are generally divided into (1) transition from awake to asleep (drowsy), (2) light sleep, and (3) deep sleep. The two latter stages are extremely dangerous to driving, certainly leading to fatal accidents. The former is also unacceptable although it may not lead to fatal accidents immediately. All drowsiness detection methods focus on the former stage in an attempt to prevent from fatal accidents. Drowsiness – a transition from awake to asleep, is characterized by behaviors of drivers such as rapidly repeated yawning, eye closure, eye blinking, and disrupted head pose, etc. These drowsy symptoms can be effectively discerned by vision-based approaches.

Among various features characterizing drowsy symptoms which can be extracted from image like blinking frequency, average eye closure speed (AECS), and eye saccade, etc., PERCLOS (percentage of eyelid closure) is taken into consideration. PERCLOS is defined as the proportion of time the eyelids are at least 80% closed over the pupil. Since the eyes of drowsy person are not necessary to be closed, PERCLOS is the most reliable and valid determination of drivers drowsiness [David , Richard 1998].

We setup a camera, frontally oriented to the faces of drivers so that faces can be localized reliably. To efficiently localize the face, face tracking is more preferable than face detection since the face in the current frame should be in the neighborhood of its position in the previous frame. The best strategy is to utilize face detector periodically and to use face tracker between the intervals. To compute PERCLOS, we use DLIB to localize facial landmarks very efficiently, taking less than 1 millisecond. The determination of eyelid opened or 80 percent closed based on the eye landmarks is straightforward. The timing for each state is recorded to calculate PERCLOS.

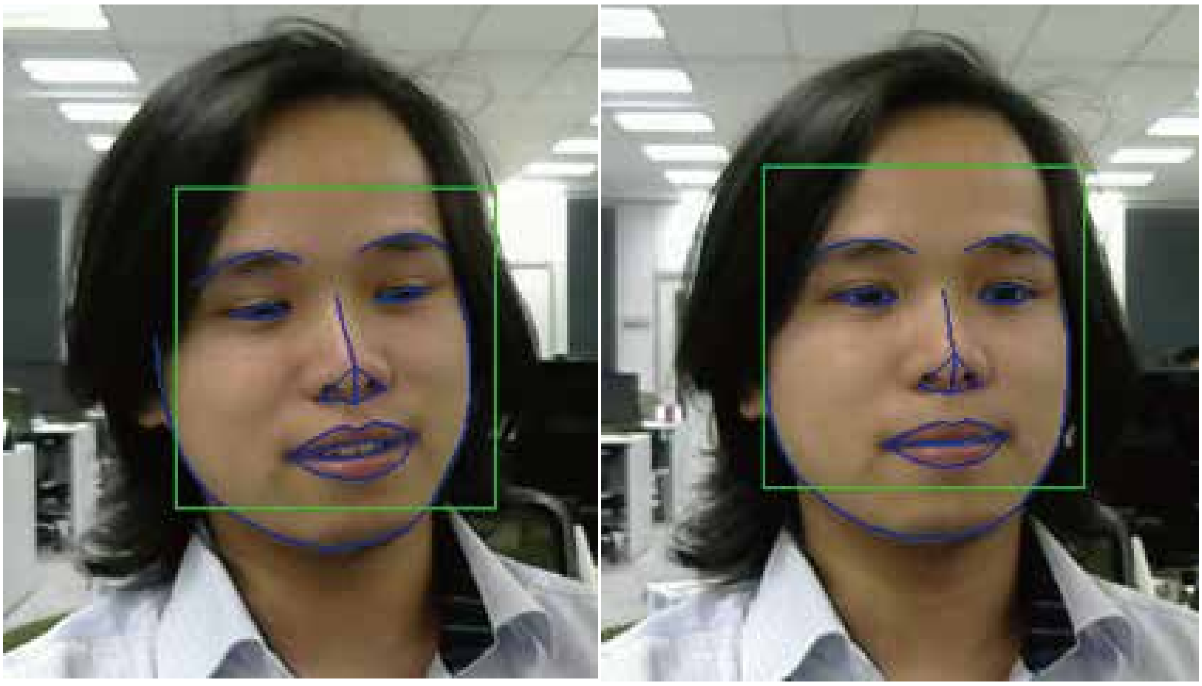

Figure 7: Facial landmarks for computing PERCLOS features. Fatigue and conscious people are depicted in the left and in the right, respectively

IV. Conclusions

We have presented a methodology of implementing various vision-based ADAS functions in mobile platforms. Vehicles, pedestrians and other objects are detected and mapped on the road layout with detected lanes. Moreover, distances between our vehicles and surrounding objects can be roughly estimated so that our vehicle can be self-aware its positions and surrounding objects. The heath conditions of drivers are also monitored to prevent them from driving in drowsy conditions which may lead to fatal accidents. Surrounding-view images can be constructed to assist drivers in parking and in case of inspecting blind spots. A variety of low-cost computer vision techniques are selected and optimized for the sake of deploying them into mobile platforms. The personal stuffs like iPad, Tablet and smartphone will become more useful when we are driving. It is flexible for users to easily customize hardware options for ADAS functions.

References

Alex, Krizhevsky, Sutskever Ilya, and E. Hinton Geoffrey. 2012. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems 1097-1105.

Boguslaw, Cyganek, and Gruszczynski Slawomir. 2014. "Hybrid computer vision system for drivers' eye recognition and fatigue monitoring." Neurocomputing 126 (27): 78-94.

Dao, Huu Hung, Saito Hideo , Yamamoto Keiichi, and Sato Hiromitsu. 2015. "An omnidirectional vision system for bus safety surveillance." 12th International Conference on Advanced Video and Signal Based Surveillance. Karlsruhe, Germany: IEEE. 1-6.

Forrest, N. Iandola, Han Song, W. Moskewicz Matthew, Ashraf Khalid, J. Dally William, and Keutzer Kurt. 2016. "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size." arXiv:1602.07360 1-13.

Hu, Shuyan, and Gangtie Zheng. 2009. "Driver drowsiness detection with eyelid related parameters by Support Vector Machine." Expert Systems with Applications 36 (4): 7651-7658.

Jin, Jun Qi, Kun Fu, and Chang Shui Zhang. 2014. "Traffic Sign Recognition With Hinge Loss Trained Convolutional Neural Networks." IEEE Transactions on Intelligent Transportation Systems 15 (5): 1991-2000.

Kalus, Huebner, Natroshvili Koba, Quast Johannes, and Kay Ulrich Scholl. 2017. Vehicle surround view system . US Patent 9679359 B2.

Kazukuni, Hamada, Zhen Cheng Hu, Meng Yang Fan, and Hui Chen. 2015. "Surround view based parking lot detection and tracking." Intelligent Vehicles Symposium (IV) . IEEE. 1106-1111.

Lin, Bin Feng, Yi Ming Chan, Li Chen Fu, Pei Yung Hsiao, Li An Chuang, Shin Shinh Huang, and Min Fang Lo. 2012. "Integrating Appearance and Edge Features for Sedan Vehicle Detection in the Blind-Spot Area." IEEE Transactions on Intelligent Transportation Systems 13 (2): 737-747.

Pedro, F. Felzenszwalb, B. Girshick Ross, and McAllester David. 2010. "Cascade object detection with deformable part models." IEEE Conference on Computer Vision and Pattern Recognition. 2241-2248.

Pedro, O. Pinheiro, Tsung-Yi Lin, Collobert Ronan, and Dollar Piotr. 2016. "Learning to Refine Object Segments." European Conference on Computer Vision. 75-91.

Piotr, Dollar, Wojek Christian, Schiele Bernt, and Perona Pietro. 2011. "Pedestrian Detection: An Evaluation of the State of the Art." IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (4): 743-761.

Rodrigo, Benenson, Mathias Markus, Timofte Radu, and Van Gool Luc. 2012. "Pedestrian detection at 100 frames per second." IEEE Conference on Computer Vision and Pattern Recognition. 2903-2910.

Son, Jongin, Hunjae Yoo, Sanghoon Kim, and Kwanghoon Sohn. 2015. "Real-time illumination invariant lane detection for lane departure warning system." Expert Systems wiht Applications 42 (4): 1816-1824.

Zhang, Liliang, Liang Lin, Xiaodan Liang, and Kaiming He. 2016. "Is Faster R-CNN doing well for pedestrian detection?" European Conference on Computer Vision. 443-457.

About FPT

FPT Corporation (HOSE: FPT) is the global leading technology, outsourcing and IT services group headquartered in Vietnam with over US$1.8 billion revenue and over 32,000 employees. Qualified with CMMI Level 5 & ISO 27001:2013, FPT delivers world-class services in Analytics, IoT, Mobility, Cloud, Product Engineering Services, Platform Modernization, Application Development & Management Service, and BPO services globally from delivery centers across the United States, Japan, Europe, Australia, Vietnam and the Asia Pacific. With focus on R&D activities to drive innovation and improve service quality, the company has been serving over 450 customers worldwide, of which over 50 are Fortune 500s in the industries of manufacturing, semiconductors, healthcare, utility, financial services, satellite TV and more. For more information, please visit https://www.fpt-software.com.

About the Author

Dao Huu Hung received B.Sc. and M.Sc., both in Electrical Engineering from Hanoi University of Science and Technology, Vietnam, and National Taiwan University of Science and Technology, Taiwan, in 2007 and 2010, respectively. He also received PhD. in Engineering from Keio University, Japan in 2014, where he continued to pursue a postdoctoral research position prior. Since 2015 he has been with FPT Japan. Co. Ltd., as a leading researcher in computer vision, machine learning, deep learning, and time-series data analysis. He published dozens of international proceedings and journal papers. He is a member of IEEE.