Traffic Sign Recognition Model on Mobile Device

Wong Hwee Ling[1, 2], Woo Chaw Seng[2]
[1]Faculty of Engineering, Multimedia University,
Persiaran Multimedia, 63100 Cyberjaya, Malaysia.
Phone: 603-831205419, Email: hlwong@mmu.edu.my
[2]Faculty of Computer Science and Information Technology, University of Malaya,
50603 Kuala Lumpur, Malaysia.
Phone: 603-79676400, Email: cswoo@um.edu.my

Abstract- **Incorporation of a good machine vision into mobile device can create a powerful application. With such advancement, mobile users can use their existing smartphone as their *third-eye* to perceive the world. This paper introduces a framework using the smartphone to increase ambient intelligence and road safety of moving vehicles through traffic sign alert application. The driver is alerted of the incoming traffic signs in different modes depending on the user's preference. However, the competency of the application is still bounded with the capability of a normal human vision. Unlike conventional research work, this paper emphasizes on portability and expandability. Portability means there is no additional installation or placement of electronic hardware required in the vehicle in order to make the system work. Expandability focuses on the transparency of the traffic sign recognition output to other applications within same hardware device or external devices. This provides an option to other developers to use or expand their own applications by using the result of the proposed application. An interaction model of the traffic sign recognition system on mobile device is introduced.**

Keywords- **traffic sign recognition, mobile platform, interaction model**

I. INTRODUCTION

Road safety is important for drivers and the passengers inside the vehicle. The driver usually relies on his/her vision to comprehend the visual inputs while driving. The visual inputs can be the weather condition, road condition, road signs, billboards, landmarks, presence of other automobiles and pedestrians. It is difficult for the driver to pay attention in a monotonous journey. The driver's attention span can also be affected by poor health physically or emotionally. Hence, there is a motivation to use machine to increase road safety via driver alert system. Machines can utilize information from proximity sensors, global positioning system (GPS) or/and camera to interpret the condition outside the vehicle. The problem of the current in-vehicle driver assistant unit is the requirement to purchase an additional electronic unit or software license. This usually comes with certain cost for the purchase and maintenance.

The objective is to propose a vision system based traffic signs warning application that can be downloaded and used in a smartphone without installation of any additional electronics hardware. Subsequently, this offers a lower cost to use the application as the users can use an existing but compatible smartphone. The application needs to be downloaded into that smartphone and then be installed. In order to utilize the application, the phone needs to be placed on the windscreen where the camera faces the road. Eventually, the maintenance cost tends to be lower, as a user only needs to maintain one smartphone instead of extra electronic devices. The cost friendly solution facilitates the second motivation, which is to enable the technology to benefit more users. The third objective is to make the result of the traffic sign recognition transparent to other applications or devices. An example of a third party application is an autonomous vehicle that is required to navigate on the road. The third party applicant can use the result to enhance their own application or system, without having to develop their own traffic sign recognition algorithm. Fig. 1 portrays few main functions performed on the smartphone for traffic sign recognition and the access to the recognition result by one or more users.

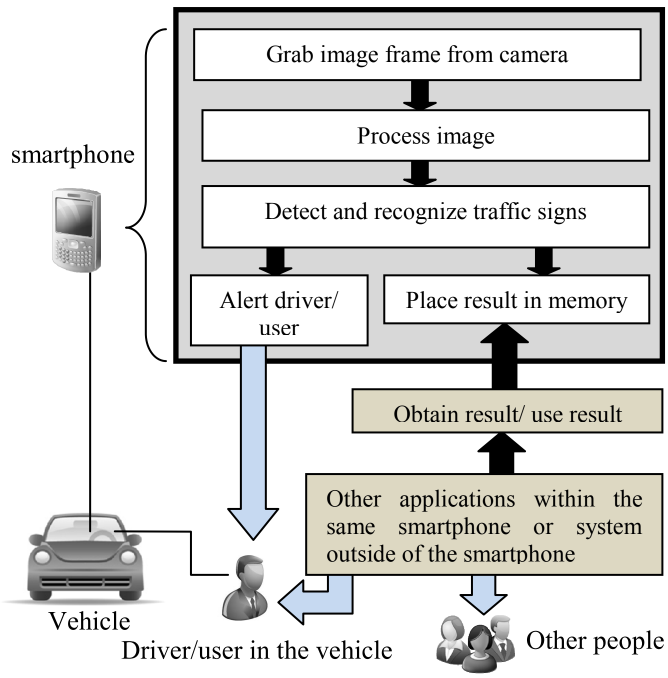

Fig. 1. Traffic sign recognition on smartphone

The difficulty faced in vision-based mobile application lies within both hardware and software. Smartphone application developers process image either fully on the phone itself or partially on the server side, where the phone is the client. For the later, a wireless connection is required. Most smartphones are equipped with camera with at least two mega-pixels. The processing power of smartphones is advancing. Memory capacity can be extended and the internal memory size is growing too. Hence, smartphone is a suitable candidate to be explored for the vision-based traffic signs warning application without requiring additional electronics hardware. The challenge is machine vision interpretation tends to consume high processing power and large memory space. Those features usually lack in smartphones, even today. Improvement in vision algorithm will be used to overcome the problem.

In this paper, background study is presented in section II to recapitulate the topics involved and to emphasize the importance of vision based traffic sign recognition framework. Later, framework of the application is described in section III together with a preliminary result of traffic sign recognition based on image processing. Last but not least, conclusion for the paper and the future work is shared in section IV.

II. BACKGROUND STUDY

The background study comprises of four main areas, which are traffic sign recognition, overview of mobile computing, vision application on mobile device and a quick review on in-vehicle infotainment trend. Traffic signs can be recognized through vision or through wireless identification. Traffic sign is erected by human and subjected to changes. Thus, it is difficult to keep track the exact location in order to enable its visibility on a GPS later. Traffic sign cannot be recognized by a very good resolution satellite image because the traffic signs are erected facing the horizon but not towards the sky. A wireless identification of traffic was proposed by sensing a transmitter placed on the traffic signs [1]. The disadvantage of the method is the need to have an access point facing the incoming vehicle for each traffic sign. The authors reported it is difficult to identify traffic signs which are relevant to the driver. Additional hardware required for that method may not be practical in near future. Vision based recognition is more practical than wireless identification method.

A. Traffic signs Recognition

Traffic sign recognition usually starts with detection, rectification and then recognition. Tracking method can also be employed to speed up recognition as a smaller area of detected object can be targeted for recognition. Research on traffic sign recognition has started since 1984 in Japan [2]. It is an important and widely researched topic in machine vision for road safety and autonomous ground vehicle. Several challenges of vision based traffic signs recognition are listed below:

1. Different road signs (e.g. shape, color, standardization)
2. Lighting variation (e.g. day, night)
3. Weather (e.g. rain, snow, dust, mist, reflection)
4. Location to deploy application (e.g. urban, rural area)
5. Background (e.g. presence of building, static/moving vehicle, road, signboard, trees)
6. Affine transformation of the road sign (e.g. scaling, rotation, shearing)
7. Occlusion of the road sign
8. Deformation of the road sign due to vandalism or weather
9. Difference in the hardware devices
10. Different placement or positioning of the camera
11. Different speed of the moving vehicle

Nguwi *et al.* [3] performed a study on road sign recognition methods. In other to perform road sign recognition, the road sign needs to be detected first. Color-based and shape-based methods are used due to distinction of the road signs based on colors and shapes. Examples of color based methods are static threshold, dynamic threshold and color indexing in various color spaces. Tsai *et al.* [4] proposed a novel Eigen color model for road sign detection based on the reflectance property that exist on the signage compared to the natural background. Comparison of separation ability between proposed method and various color spaces such as RGB, YIQ, LUV and L*a*b were performed. Fisher criterion was used to measure the separation ability and the result indicates that their proposed color model has higher separation ability. In contrast, shape based approach does not require color information. Thus, grayscale images are often used in this approach. Shape based method can be more robust during illumination invariance compared to color based method. However, shape based method tends to be difficult in situation where there is occlusion or affine transformation of the road signs viewed. Examples of shape based approaches are Hough transform, template matching and etc. Combination of both color and shape features are often utilized in detection process. Other more sophisticated methods that were used along with color and shape information for more robustness are support vector machines (SVM) [5], neural network and genetic algorithm.

Once the road sign is detected, the road sign needs to be recognized. Feature extraction needs to be performed to differentiate various road signs. Then, the feature will be compared with a standard representation of the specific road sign. The comparative value will determine whether the road sign is classified into the correct category. Examples of methods used for classification are template matching [6], Gaussian-kernel SVMs, character recognition for speed limit sign [7] and other hybrid classification techniques [8]. Saturnino *et al.* performed road sign detection and recognition based on SVM. The mean processing speed is 1.77s per frame and the best recognition success probability reported was 93.24%. Soigkas and Dermatas [6] reported an improvement of detection and recognition rate from below 45% to 81.2%, by employing tracking method. Fatmehsan *et al.* [8] obtained 93.1% accuracy using a hybrid one-vs.-rest SVM and naïve bayes classifier for detection and recognition of red color based road signs. The algorithms reported were mostly tested on generic computer and written in C/C++. It is also difficult to compare the actual performance as there is neither a standardized dataset nor a standardize procedure for evaluating the performance in terms of speed and accuracy.

B. Overview of Mobile Computing

A computer or controller that is portable is termed as a mobile hardware platform. Notebook, netbook, mobile phones or any other portable embedded systems are considered as a mobile platform. Hence, there could be millions of existing mobile platforms today. A mobile hardware platform has the capability to acquire data, process data and provide outputs. It has its own processor, memory, input-output devices and power supply to support portability.

Mobility of such platforms supports ubiquitous computing. Ubiquitous computing emerged as a post-desktop revolution with the advancement and miniaturization of technology. Mobile hardware interconnects with the sub-devices within the platform and connects with other devices using wired or wireless connectivity. The inter-connections and outer-connections are invisible to at the eyes of end user. Thus, mobile hardware platform has the ability to satisfy "*at anywhere, anytime*" utilization and "*disappearing computer*" concept of ubiquitous computing. The ambient intelligence paradigm builds upon ubiquitous computing. Ambient intelligence is accomplished through an electronic device that has reactive behavior to give the appropriate response to the human user in a specific environment. Mobile hardware platform is able assist in delivering ambient intelligence into dynamic environment, such as moving vehicles.

C. Vision Applications on Mobile Devices

Several vision-based mobile phone applications have been developed. The applications are a part of Augmented Reality (AR), where the real-world images or video are interpolated with computer generated graphics. Examples of vision based applications are landmark detection via Content Based Image Retrieval (CBIR) [9, 10], pose tracking of natural feature [11], face and eye detection [12]. In contrast with computer based vision approach, many of the AR approaches today rely mostly on the global positioning system (GPS) and sensors such as the accelerometer. According to Gammeter *et al.* [9], AR relaying on visual recognition tends to have more advantage on both stationary and non-stationary objects. Image processing on smartphone application can be conducted on the client and/or the server. The client is the phone, while the server is a super computer connected to the phone wirelessly. Most applications perform more complex image processing on the server, as the processing can be faster and huge database can be stored at the server side. Those applications require a persistent wireless connectivity, which may pose as a problem in some rural areas. On a single device processing, the challenge lies when the image processing algorithm gets complex that it takes up larger memory space and reduces the speed of recognition. The hardware limitation issue can be tackled either through software and/or hardware [13].

D. In-Vehicle Infotainment System

Few car manufacturers have turned to mobile application developers to built in-vehicle infotainment system. Ford introduces MyFord Touch which comprises of features such as media hub, phone, voice recognition, navigation, climate control, audio system, driver's personalization and other services [14]. MyFord Touch is a dedicated piece of electronics hardware embedded in certain models of Ford vehicles but with no capability for traffic sign recognition. Daimler has introduced Smart Drive Kit that works together with an iPhone to provide several in-vehicle infotainment functions for Smart car. In future, the designer intends to connect a camera to the Smart Drive Kit via wireless local area network (WLAN) to enable future applications such as traffic speed limit recognition [15]. Their proposed method requires three electronics devices which are the camera, a Smart Drive Kit and an iPhone. BMW was also spotted to seek Daimler for the vehicle infotainment solution as well. Other in-vehicle infotainment systems are Nokia's Terminal Mode [16] and Intel Atom chipset based in-vehicle infotainment system [17].

Several cars manufacturers have adopted Advance Driver Assisting System (ADAS), which includes traffic signs recognition. For instance in year 2008, Mobileye partnered with Continental AG launched three features in BMW 7 series, namely the lane departure warning, speed limit information based on traffic sign detection and intelligent headlight control [18]. Image processing is performed on a dedicated multi-core processor, namely the Mobileye EyeQ [19]. Mobileye uses both image and GPS mapping information to detect speed limit signs [20]. In contrast, Opel Eye on several models of Opel cars relies solely on image processing for traffic sign recognition. Opel Eye comprises of a set of dedicated hardware integrate into the car. The recognition capability is limited by speed and poor weather conditions [21].

From the commercialized or ongoing activities surrounding in-vehicle infotainment system, the companies emphasize on both software and hardware approach to enable their system. Most solutions are dedicated towards certain car models. We viewed a mobile application development problem at a different angle, as a software solution will be emphasized while the existing hardware resource is used. In addition, traffic sign recognition solely based on image processing method processed on the phone itself is unavailable.

III. FRAMEWORK

The traffic sign alert application is proposed to operate under different kinds of behavior in order to adapt to the needs of the user. The system is dependent upon machine vision as the sole engine for recognition. The behaviors are generalized below:

1. Warning mode: The driver is alerted for warning signs. Warning signs at a further distance can be recognized.
2. Long distance driving mode in rural area: The driver is alerted for incoming traffic signs.
3. Tracking mode: The traffic sign is continuously tracked.

The Maddix Human-Computer Interaction (HCI) model is used to initiate the framework for the mobile traffic sign alert system. Maddix HCI model consists of physical, syntactic and semantic level [22]. A HCI modeling intends to describe the interaction between human and computer which includes the software and hardware levels of the computer itself. In Maddix HCI model, the physical level involves the physical contact between the human and the computer to exchange information.

The human commands the computer of a task at the semantic level. The structural information is translated into meaningful information by the computer at the syntactic level. Maddix HCI model has been employed in Human-Robot Interaction (HRI) as well [23]. A model is proposed in this paper to elucidate the components involved and to exhibit the affiliation between the user and the smartphone. Fig. 2 summarizes the model for traffic sign alert system. The following subsections provide explanation in more details of each component within the model and their interaction.

A. Physical Level

The physical level consists of the physical component that resides in the mobile device itself such as the input, output and memory. The application requires at least a touch screen or a keypad input, in order for the user to initiate or change command. The screen acts as the main output interface. It is capable of displaying a graphical user interface (GUI) for behavior selection. Besides that, the result of the traffic sign recognition can be displayed as AR on the screen. The user should also have an option to run the application in background mode. This means that the application can run using sound alert instead of graphical AR. The memory is where the database of a standardize traffic sign are stored and can be updated from time to time. The database does not consist of images but rather a feature set that have reduced dimensionality to represent each traffic sign. Other important components are the processor and the camera. Most smartphones today should have sufficient physical components to perform simple image processing.

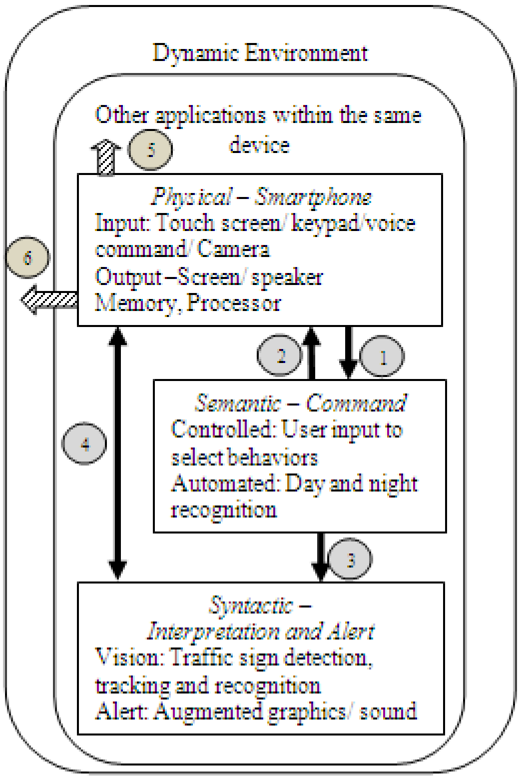

Fig. 2. A model for the traffic sign alert framework

B. Semantic Level

Semantic level consists of the input commands by the user. Upon starting the application, a simple GUI appears to prompt user on some information and options available. It is important to take note not to burden the user with too much information or options. Complexity of a GUI tends to make a user bored or impatient. Upon a mode selection, the application immediately detects day and night via the clock of the mobile device. Additional information such as the geographical location of the mobile device might be helpful to estimating the brightness of the environment. Identification of day and night is helpful for the image processing algorithm to adjust the recognition in dark and bright environment. Image processing methods can also be used to verify the brightness of the environment.

C. Syntactic Level

The syntactic level encompasses translation of the visual information into recognition of traffic signs for alert. The camera views the environment as video form. Hence, frame by frame of image must be grabbed to perform image processing. The image is processed to detect then recognize signs. Color and shape based features will be incorporated for the detection. Once detected, the region of interest will be segmented for recognition. The prominent feature of the region of interest will be computed. Then, a comparison will be made with the database of a standard feature set residing in the memory. Tracking method may be employed to increase the speed of detection and recognition at later stage. Upon recognition of the traffic sign, the user will be alerted through display or sound.

Fig. 3 portrays a preliminary methodology of a "no turning left" traffic sign recognition based on various low level image processing steps using color thresholding, shape indexing, edge detection and binary image template matching. Color and shape information are used to discard irrelevant information. Template matching is based on binary edge information. Similarity value between test image and template image is computed based on the equations (1) and (2):

$$Recognition = \begin{cases} True, & \text{if } Matching\ index > 0.3 \\ False, & \text{if } Matching\ index \le 0.3 \end{cases} \quad (1)$$

where,

$$Matching\ index = \frac{No.\ of\ overlapping\ pixels}{No.\ of\ white\ pixels\ in\ edge\ object} \quad (2)$$

The size of edge object must be normalized to match the size of template object before matching proceeds. The *Matching index* value for test image in Fig. 3 compared with a standard template for "no turning left" sign is 0.37. In contrast, the *Matching index* value for the same test image compared with a standard template for "no turning right" sign is 0.16. Via the *Matching Index* value, the recognition of the actual "no turning left" sign from the test image can be discriminated. However, the image processing algorithm needs to be refined further before implementation on the smartphone to improve its robustness against changes in a dynamic environment.

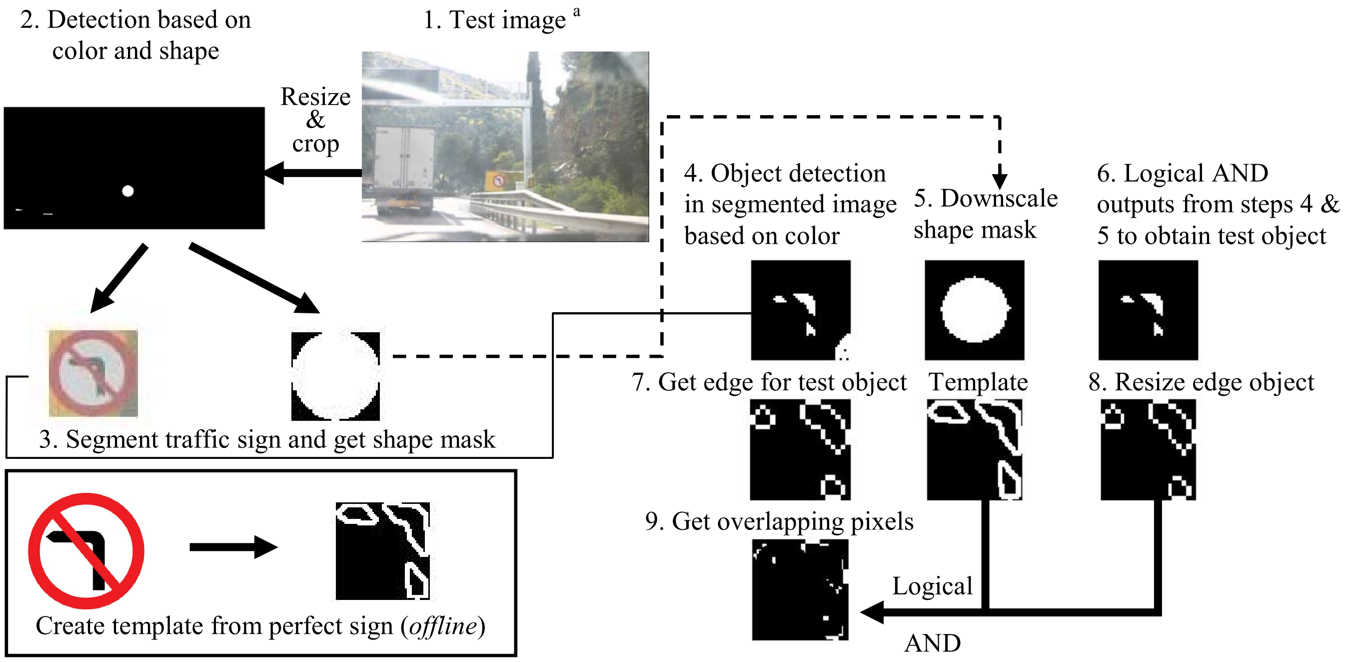

Fig. 3. Low level image processing for traffic sign detection and recognition

D. *Interaction*

Inter-connectivity of each component of the traffic sign alert model is illustrated in Fig. 2. Outer-connectivity of the application to the other applications on the mobile device and the environment are shown as well. The inter-connectivity is shown as arrows 1 to 4. Arrow 1 indicates that the input from the physical level is translated into commands. Arrow 2 indicates that a change in command can be displayed in the output device of the physical level.

Behavior selected by user and early information of day/night in semantic level is passed to the syntactic level via Arrow 3 in Fig. 2. Thus, the vision system could produce the suitable output for the corresponding behavior. The syntactic level also interacts with physical level directly and continuously for image processing. The processor and memory are used extensively during processing of the images. As recognition is performed, the user is alerted via the output devices in the physical level. The bidirectional data highway is represented as Arrow 4. Accuracy and speed of the application must meet the system's real-time and reliability constraints.

The traffic sign recognition result is continuously place in a location of the memory where it is made public to other applications. Other applications need to have their own means of method to retrieve the public data. The outer-connectivity is represented by arrows 5 and 6. Other applications refer to any other applications that reside within the mobile device. For instance, the traffic sign recognition result can be used along with a driver's attention detection application via Arrow 5. If the driver is detected to be sleepy, the volume of the speaker can be tuned up to inform the driver of a recognized traffic sign by the third-party application.

[a] Test image was obtained from http://agamenon.tsc.uah.es/Investigacion/gram/

Dynamic environment represents any element outside the mobile device. The environment is termed as dynamic as the vehicle is moving. Hence, the scene viewed by the mobile device is always changing. Optical flow of the scene may be inconsistent as a human can drive a vehicle at various speeds. The environment may also contain other electronic devices inside or outside the vehicle. Other electronic devices inside the vehicle can communicate with the smartphone wirelessly or through wired connection. Electronic devices outside the vehicle need to communicate with the smartphone wirelessly. These are represented by Arrow 6.

IV. Conclusions & Future Work

A conceptual framework of a traffic sign recognition system on mobile device is presented. The traffic sign alert framework proposed in this paper is an infant step to developing a fully functional application on a mobile device. The interaction model is designed to produce clarity on the complexity involved in realizing a machine vision on a mobile device that operates in dynamic environment. The model is divided into physical, semantic and syntactic levels to depict human and smartphone interaction during the mobile application utilization. Result of the application is proposed to be made available for applications that reside within the same mobile device or other electronic devices to assist in further expansion of other applications' capability.

The future work emphasizes on realizing the traffic sign recognition through vision techniques on a targeted smartphone. The aim is to enable the smartphone to recognize traffic signs as how a driver would perceive a sign. Through research on software based solution to overcome inadequacy of hardware computation power for image processing, traffic sign recognition algorithm is intended to run on the smartphone alone. The phone's hardware resources will be used to capture image, process image and interact with user. Hence, the purchase and maintenance cost by an end user aimed to be lower than having to plug in additional piece of electronic hardware to run the image processing algorithms or other processes for interaction. A preliminary traffic sign recognition result based on low level image processing technique was presented. The authors hope to develop a novel image processing algorithm in the area of traffic sign recognition that can help to compensate the hardware limitation of the smartphone, operates in real-time and adaptable to some variations in the environment.

REFERENCES

[1] A. Katajasalo and J. Ikonen, "Wireless identification of traffic signs using a mobile device", in *Proc IEEE 3rd International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, Sliema, Malta, pp.130 - 134, Oct. 2009.

[2] P. Paclik, "Road Sign Recognition Survey", [Online: 08 Sep 2010]. Available: http://euler.fd.cvut.cz/research/ rs2/files/skoda-rs-survey.html

[3] Y.Y. Nguwi and A. Z. Kouzani, "A study on automatic recognition of road signs", in *Proc IEEE Conference of Cybernetics and Intelligent Systems*, Bangkok, Thailand, pp.1 – 6, Jun. 2006.

[4] L. W. Tsai, *et al.*, "Road sign detection using eigen color", *Journal of IET Computer Vision*, Vol. 2, No. 3, pp.164–177, 2008.

[5] M. B. Saturnino, L. A. Sergio, G. J. Pedro, G. M. Hilario and L. F. Francisco, "Road-sign detection and recognition based on support vector machines," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 8, No. 2, pp.264 - 278, Jun. 2007.

[6] G. K. Siogkas and E. S. Dermatas, "Detection, tracking and classification of road signs in adverse conditions," in *Proc. IEEE Mediterranean Electrotechnical Conference*, Benalmádena (Málaga), Spain, pp.537 - 540, May 2006.

[7] Y. P. Wong, M. P. Shi and T. Wu, "A method of fast and robust for traffic sign recognition," in *Proc. IEEE Fifth International Conference on Image and Graphics*, Shanxi, China, pp.891 - 895, Sep. 2009.

[8] Y. R. Fatmehsan, A. Ghahari, R. A. Zoroofi, R.A, "Gabor wavelet for road sign detection and recognition using a hybrid classifier", in *Proc. IEEE International Conference on Multimedia Computing and Information Technology*, Sharjah, United Arab Emirates, pp.25 - 28, Apr. 2010.

[9] S. Gammeter, A. Gassmann, L. Bossard, T. Quack and L. V. Gool, "Server-side object recognition and client-side object tracking for mobile augmented reality", in *Proc. IEEE Computer Vision and Pattern Recognition Workshops*, California, U.S.A, pp.1 - 8, Jun. 2010.

[10] Y. Zhou, X. Fan, X. Xie, Y. Gong and W. Y. Ma, "Inquiring of sights from the web via camera mobiles", in *Proc. IEEE International Conference on Multimedia and Expo*, Ontario, Canada, pp.661 - 664, Jul. 2006.

[11] D. Wagner, G. Reitmayr, A. Mulloni, T. Drummond and D. Schmalstieg, "Pose tracking from natural features on mobile phones," in *Proc. IEEE International Symposium on Mixed and Augmented Reality*, Cambridge, U. K., pp.125 - 134, Sep. 2008.

[12] A. Hadid, J. Y Heikkild, O. Silven and M. Pietikdinen, "Face and eye detection for person authentication in mobile phones," in *Proc. ACM/IEEE 1st International Conference on Distributed Smart Cameras*, Vienna, Austria, pp.101 - 108, Sep. 2007.

[13] S. E. Lee, Y. Zhang, Z. Fang, S. Srinivasan, R. Iyer and D. Newell, "Accelerating mobile augmented reality on a handheld platform", in *Proc. IEEE International Conference on Computer Design*, California, U.S.A, pp.419 - 426, Oct. 2009.

[14] "MyFord™ Defines a New Driver Experience," [Online: 29 Sep. 2010]. Available: http://www. thefordstory.com/smart-technology/myford%E2%84%A2-defines-a-new-driver-experience/

[15] "Smart Drive Kit for The iPhone," [Online: 29 Sep. 2010]. Available: http://500sec.com/smart-drive-kit-for-the-iphone/

[16] "Integration of Mobile Apps into the Vehicle HMI: Terminal Mode", [Online: 29 Sep 2010]. Available: http://www.nokia.com/NOKIA_COM_1/Microsites/terminalmode/pdf/terminal_mode.pdf

[17] "In-Vehicle Infotainment," [Online: 29 Sep. 2010]. Available: http://www.intel.com/design/embedded/ infotainment/ products.htm

[18] "BMW", [Online: 28 Jan 2011]. Available: http://www.mobileye.com/manufacturer-products/product-launches/BMW

[19] P. S. Gideon, R. Elchanan, H. Gaby and S. Amnon, "A computer vision system on a chip: a case study from the automotive domain" in *Proc IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA, USA , pp. 130 - 134, Jun. 2005.

[20] "Test: traffic sign recognition systems", [Online: 28 Jan 2011]. Available: http://www.inautonews.com/test-traffic-sign-recognition-systems

[21] "Reward 2010 – Opel Eye", [Online: 28 Jan 2011]. Available: http://www.euroncap.com/rewards/opel_eye.aspx

[22] F. Maddix, Human-Computer Interaction: Theory and Practice, Ellis Horwood Limited, England, 1990.

[23] H. L. Wong and S. H. M. Amin, "Skin color detection based on various chrominance pairs from multiple color spaces for face detection in human-robot interaction", Jurnal Elektrika, Vol. 7, No. 1, pp.35-44, Jun. 2005.