TRAFFIC SIGN AND LICENSE PLATE DETECTION BASED ON SALIENCY, MEAN-SHIFT, AND MATHEMATICAL MORPHOLOGY

CHEONG WEI SHEIK

UNIVERSITI SAINS MALAYSIA

2015

TRAFFIC SIGN AND LICENSE PLATE DETECTION BASED ON SALIENCY, MEAN-SHIFT, AND MATHEMATICAL MORPHOLOGY

by

CHEONG WEI SHEIK

Thesis submitted in fulfillment of the requirements
for the degree of
Master of Science

September 2015

ACKNOWLEDGEMENTS

To write this thesis is a great endeavor, and I could not have reached the finish line without the encouragement, advice, and support of my advisor and family. My thanks are due first and foremost to my advisor, Professor Lim Chee Peng, for being willing to take me on as a student, and for being unfailingly generous with his time, and his support throughout my research career. Professor Lim provided consistently good advice, helped me to situate my work in a right context, and gave me tremendous freedom in pursuing my ideas, while always ensuring that I was making progress towards some effectual outcome. I will be forever grateful for his help.

I owe special thanks to Dr. Khoo Bee Ee, who is willing to advice and support me in the period of Professor Lim's absence, and continue set me on a path which has resulted in this thesis.

Finally, I offer my deepest thanks to my wife, Sook Choo, whose love and support has sustained me; to my parents, Beng Hoong and Kiam Hing, who have instilled in me the love of learning that has fueled all my efforts; and to my lovely daughter, Ling Xi, who takes over my heart in the middle of research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ALPR	Automatic License Plate Recognition
ASIC	Application-Specific Integrated Circuit
ATSR	Automatic Traffic Sign Recognition
AUC	Area Under Curve
CCA	Connected Component Analysis
CIE	International Commission on Illumination (English)
COSFIRE	Combination of Shifted Filter Responses
CPU	Central Processing Unit
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DLPFC	Dorsal-Lateral Pre-Frontal Cortex
DTFT	Discrete-Time Fourier Transform
FCO	Frontal Cortical Oculomotor
FDN	Frequency Divisive Normalization
FDR	False Discovery Rate
FEF	Frontal Eye Fields
FFT	Fast Fourier Transform
FN	False negative
FP	False Positive
FPGA	Field-Programmable Gate Array
GABA	Gamma-Aminobutyric Acid
GPU	Graphics Processing Unit
HFT	Hypercomplex Fourier Transform
HOG	Histogram of Oriented Gradient
HSI	Hue, Saturation, Intensity
HSV	Hue, Saturation, Value
IDTFT	Inverse Discrete-Time Fourier Transform
IP	Inferior Parietal
IT	Inferior Temporal
ITS	Intelligent Transportation System
KDE	Kernel Density Estimation
LGN	Lateral Geniculate Nucleus
LIP	Lateral Intra-Parietal
LP	License Plate

MSER	Maximally Stable Extremal Regions
MST	Medial Superior Temporal
MT	Middle Temporal
PDF	Probability Density Function
PFT	Phase-spectrum of Fourier Transform
PPV	Positive Predictive Value
PQFT	Phase-spectrum of Quaternion Fourier Transform
RANSAC	Random Simple Consensus
RGB	Red, Green, Blue
SC	Superior Colliculus
SCi	intermediate region of Superior Colliculus
SCs	superficial region of Superior Colliculus
SEF	Supplementary Eye Fields
SIFT	Scale-Invariant Feature Transform
SLIC	Simple Linear Iterative Clustering
SNpr	Substantia Nigra par reticulata
SR	Spectral Residual
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
TPR	True Positive Rate
TS	Traffic Sign
V1	Primary Visual Cortex
V2	Secondary Visual Cortex
V3	Third Visual Complex
V4	Visual Area V4

PENGESANAN TANDA ISYARAT LALU-LINTAS DAN PLAT LESEN BERDASARKAN KETONJOLAN, ANJAKAN PURATA DAN MORFOLOGI MATEMATIK

ABSTRAK

Sebuah model pengesanan objek yang terdiri daripada septrum ketonjolan, peruasan anjakan purata, dan anggaran bentuk morfologi telah dicadangkan dalam kajian ini. Kaedah pengiraan ketonjolan yang lebih baik berdasarkan perhatian penglihatan manusia telah diperkenalkan. Septrum ketonjolan adalah berdasarkan prinsip operasi de-konvolusi dalam domain log-spektrum, dan pengiraannya cepat dengan hanya parameter tunggal untuk ditala. Selain itu, septrum ketonjolan mempamerkan ciri keteguhan warna di bawah pelbagai pencahayaan, dengan skim warna biasa RGB boleh digunakan untuk imej warna. Bagi meningkatkan lagi prestasi model yang dicadangkan, anjakan purata tanpa parametrik dan kaedah Otsu telah digunakan untuk meruaskan objek daripada sekitarnya. Selain itu, kaedah faktor bentuk yang mudah berdasarkan morfologi matematik diperkenalkan untuk mengenal pasti objek yang diruas dengan mengukur bentuknya. Untuk menilai keberkesanan dan kesesuaian kaedah yang dicadangkan, dua masalah di sektor pengangkutan, iaitu, pengesanan tanda isyarat lalu-lintas dan plat lesen kenderaan dikaji secara terperinci. Berdasarkan dua set data daripada umum dan dikumpul secara tempatan, kaedah yang dicadangkan menunjukkan keseimbangan yang baik antara ketepatan dan kelajuan. Keputusan simulasi menunjukkan bahawa ia adalah tujuh kali lebih cepat daripada teknik perihalan bentuk dalam pengesanan tanda isyarat lalu-lintas, dan mengambil masa kurang daripada 0.6 saat dalam pengesanan plat lesen kenderaan berbanding degan kaedah pemadanan pencontoh dan pembelajaran mesin. Kajian ini menunjukkan kegunaan kaedah pengesanan objek dalam satu kerangka kerja bersepadu untuk kedua-dua masalah pegesanan tanda isyarat lalu-lintas dan plat lesen kenderaan, oleh itu, menyumbang ke arah kemajuan dalam sistem pengangkutan pintar.

TRAFFIC SIGN AND LICENSE PLATE DETECTION BASED ON SALIENCY, MEAN-SHIFT, AND MATHEMATICAL MORPHOLOGY

ABSTRACT

An object detection model that consists of cepstrum saliency, mean-shift segmentation, and morphological shape estimation is proposed in this research. An improved computational saliency method based on human visual attention is introduced. Cepstrum saliency is based on the principles of de-convolution in the log-spectrum domain, and is computationally fast with only single parameter to tune. Moreover, cepstrum saliency exhibits color consistency under various illuminations, where the normalized RGB color scheme can be used for color images. To further enhance the proposed object detection model, non-parametric mean-shift and Otsu's method are utilized for figure-ground segmentation. Besides that, simple shape factors based on mathematical morphology are introduced to identify the segmented objects by measuring shapes. To evaluate the effectiveness and applicability of the proposed method, two problems in the transportation section, i.e., traffic sign and license plate detection, were studied in detail. Based on two publicly available and locally collected data sets, the proposed detection method demonstrates a good equipoise between accuracy and speed. The simulation results indicate that it is seven times faster than shape descriptors in traffic sign detection, and has an average of less than 0.6 s in license plate detection as compared with template matching and machine learning methods. The findings indicate the usefulness of the proposed object detection method in providing a unified framework for both traffic sign and license plate detection problems; therefore contributing towards advancement in intelligent transportation systems.

CHAPTER 1: INTRODUCTION

1.1 Background

Traffic sign and license plate have been designed to be primarily distinguishable from their background, so they can be detected easily. While automated traffic sign and license plate detection have been widely studied over last decades (Lalonde and Li, 1995; Mogelmose et al., 2012; Watson and Walsh, 2008; Du et al., 2013), this area of research continues to attract immense interests in both academic and industrial domains to date. Numerous detection methods have been developed and deployed in many applications related to intelligent transportation systems (ITS) (Mogelmose et al., 2012), such as traffic surveillance, fraud alert, toll payment, and driver assistance. Despite many methods have been reported to produce good results, designing a robust detection model in an uncontrolled environment with large categories of traffic signs or license plates, and to execute detection in real time remain a challenging task.

Since traffic sign and license plate are simpler icons as compared with general objects, their detection can be conceptually seen as a sub-problem of general object detection. Specifically, it is known as the figure-ground problem (Wagemans et al., 2012), i.e., how to separate objects from the background? Many existing solutions use specific models from edges, shapes, and colors to discriminate objects from clutters. Other solutions formulate a binary classification problem (figure vs. background), and use machine learning methods to locate traffic signs or license plates from trained samples. As pointed out by Mogelmose et al. (2012) and Anagnostopoulos et al. (2008) in their survey papers, good results can be obtained either by using a large set of training samples, or making very specific assumptions in the model. This raises serious questions, viz., "How many training samples are adequate?", or "How many assumptions are satisfactory?". As we have no control over the environment, therefore, both approaches compromise practicality of the solutions in real-world applications.

1.2 Motivation

The main challenge of traffic sign and license plate detection arises from the so-called object invariance (Riesenhuber and Poggio, 2000) issue, i.e., the same object alters its form in almost indefinite ways under changing of illumination, occlusion, viewpoint, and other factors. This problem has been tackled in many research domains for decades, ranging from neuroscience, psychophysics, to computer vision, and it still remains an open problem to be solved (DiCarlo et al., 2012; Pinto et al., 2008; Riesenhuber and Poggio, 2000).

Over the past 25 years of research in visual attention (Carrasco, 2011), findings from saliency detection, neuronal responses, to natural image statistics (Frintrop et al., 2010; Simoncelli and Olshausen, 2001; Olshausen and Field, 1996) have laid some fundamental understanding pertaining to what salient, or “pop-up” / “surprise”, features in an image are, in order to direct our gaze in the first 25–50 ms after the onset of stimulus, i.e., our bottom-up attention (Carrasco, 2011). Moreover, these features from natural images are correlated statistically in a way known as sparse coding (Olshausen and Field, 1996), and it leads to the study of natural image statistics.

Since the first computational saliency study introduced by Itti et al. (1998), saliency detection in computer vision becomes a popular research topic. More than 60 different models in computational saliency have been presented in a recent survey (Borji and Itti, 2013). The classic model from Itti et al. (1998) that involves many customized parameters and computational intensive procedures is hard to satisfy real-time and real-world requirements of traffic sign and license plate detection applications (Guo and Zhang, 2010). Hou et al. (2007) proposed spectral residual that computes saliency based on Fourier transform, which can be executed rapidly, and it relies only on very few parameters. The proposed method has attracted a lot of attention since then, and has been enhanced and deployed in many applications. Today, spectral saliency has become a standalone research topic under the umbrella of computational saliency (Borji and Itti, 2013).

Pertinent to advances of visual attention in the areas of neuroscience, computational neuroscience and computer vision, it is believed that an important research question is still under-explored, i.e., "Is there a more efficient and fundamental mechanism to detect objects, instead of increasing the number of training samples or making more specific assumptions?". This question forms the main motivation of this research.

1.3 Problem Statement

Conventionally, detection of traffic sign and license plate is treated as separated problems. Based on the literature review (as detailed in Section 2.4), traffic sign detection commonly uses color to locate the interested regions, which include Haar-like detectors to find relevant features, Hough transform to identify geometrical shapes, or combination of both (Mogelmose et al., 2012). On the other hand, license plate detection uses a myriad of techniques, from edge statistics based on repeating contrast changes of character properties in a plate, mathematical morphology of shapes filtering, to connected component analysis of spatial measurements, which include area, aspect ratio, and orientation (Anagnostopoulos et al., 2008). Machine learning methods such as classifiers (e.g., AdaBoost and support vector machine), artificial neural networks, and genetic algorithms are also commonly used in both traffic sign and license plate detection (Mogelmose et al., 2012; Du et al., 2013).

While many methods are available to tackle different issues in traffic sign and license plate detection (as in Section 2.4), a number of issues still exist. In general, there are four problems that continue to recur in reports published in the literature, as follows:

i) Illumination changes due to day-night, weather conditions, capturing devices, or wear-and-tear on the signs (or plates). Normalization to correct the contrast levels or color values is a very common technique against varying lighting conditions. However, when color segmentation is used in traffic sign detection, it requires a threshold that is either determined prior to using the data set (Soheilian et al., 2013), or specifically defined for different weather conditions (Gao et al., 2006). This compromises the robustness of the technique in real-world applications.

ii) Viewpoint changes with respect to the signs or plates caused by distance and/or angle. To tackle this problem, different sizes and affine transformations have been introduced. Many of the solutions use edge statistics or Haar-like detectors that are more viewpoint invariance (Mogelmose et al., 2012). Nevertheless, edge-based methods can hardly be applied to complex scenes, since they are too sensitive to undesired edges like the radiator in front of the vehicle (Anagnostopoulos et al., 2008). Haar-like detectors require a sliding window with different scales, transformations, and template patterns, which are not practical in situations when many signs are available, high-resolution images, and in real-time applications.

iii) Discrimination of objects from clutters (or background). In complex scenes like urban areas, the background can possess patterns that are similar to plates, such as advertisements, stickers, or building textures. Local features like edges are not able to discriminate them satisfactorily (Du et al., 2013). Therefore, advanced techniques like Hough transform, random sample consensus (RANSAC), support vector machine, are employed to find shapes or corresponding features for their similarity (or dissimilarity) (Du et al., 2013). However, these techniques are known to be computational intensive, and yet are subject to discrepancy in certain viewpoint deformation and occlusion conditions.

iv) Detection algorithms should run fast enough to meet real-time requirements, and no single object of interest should be missed when moving through the scene. Although more advanced technologies such as field-programmable gate array (FPGA) or application-specific integrated circuit (ASIC) can be used to markedly optimize and pipeline the processes, however, most investigations focus on comparing the performance based on general CPU (Gomez-Moreno et al., 2010).

Despite the aforementioned problems, it is observed both traffic sign and license plate are designed to be salient against their environment. Therefore, to enhance computational saliency and achieve a better understanding of human visual attention, the key research question undertaken in this thesis is: "Can the problem of traffic sign and license plate detection, which has been treated as separated problems conventionally, be solved together in a more effective (accuracy wise) and efficient (speed wise) way by exploiting biologically-inspired saliency detection methods?".

1.4 Objectives

The objectives of this research are:

i) to formulate a biologically-inspired saliency detection model to address the problems described in Section 1.3;

ii) to evaluate the proposed model using a publicly available traffic sign data set and compare the results with those from state-of-the-art methods;

iii) to apply the proposed model to license plate detection using a locally collected data set.

1.5 Scope and Methodology

This thesis focuses on enhancing the robustness of existing traffic sign and license plate detection methods towards object invariance. The proposed biologically-inspired saliency detection model should have fewer assumptions (parameters), and should be able to run faster in order to meet real-time requirements.

To understand the topic comprehensively, the mechanisms of visual attention in human visual system, including psychophysical experiments, neuronal responses and computational model, are first reviewed. Then, current saliency methods in computer vision are evaluated, with the focus placed on methods that can be executed rapidly and that have few parameters to set. State-of-the-art traffic sign and license plate detection methods are also evaluated, in order to understand their advantages and limitations. Based on the findings, the proposed model is formulated towards solving some of the existing problems in the literature.

For the scope of developing a prototype tool that implements the proposed model, the MATLAB software is used. Some C/MEX codes for rapid execution are developed. In addition, a number of third-party toolboxes like VLFeat from University of Oxford, Peter Kovesi's computer vision and image processing toolbox from University of Western Australia, Piotr Dollar's image and video toolbox from University of California San Diego, are deployed.

To evaluate the proposed model, one publicly available traffic sign data set and one real Malaysian license plate data set are used. The traffic sign data set (Grigorescu and Petkov, 2003; Azzopardi and Petkov, 2013) has 48 images of traffic signs in Netherlands. The Malaysian license plate data set (Zakaria and Suandi, 2010; Soon et al., 2012) comprises a total of 222 images collected by researchers of School of Electrical & Electronics Engineering, Universiti Sains Malaysia.

1.6 Thesis Outline

This thesis is organized in six chapters, as follows: Chapter 1: Introduction, Chapter 2: Literature review, Chapter 3: The proposed saliency detection model, Chapter 4: Evaluation of traffic sign detection, Chapter 5: Malaysian license plate detection, and Chapter 6: Conclusions and future work.

In Chapter 2, research related to object detection in three areas is reviewed. Firstly, visual attention in human visual system, including findings in psychophysical experiments, neuronal responses, and normalization model of visual attention, are examined. Secondly, saliency detection in computer vision and details of spectral saliency since the introduction by Hou et al. (2005) are studied. Thirdly, state-of-the-art methods in traffic sign and license plate detection are surveyed.

In Chapter 3, a new model is proposed for traffic sign and license plate detection. The theoretical aspects of spectral saliency, cepstrum analysis, and homomorphic systems are explained. The proposed model is also extended to color and bias saliency, in order to improve its robustness. Issues related to how clustering segmentation can be guided by sparse saliency to obtain the distinct boundary from objects are investigated. A fast and straightforward mathematical morphological to estimate shapes from binary images is deployed.

In Chapter 4, the proposed model is evaluated using a publicly available traffic sign data set from University of Groningen. The experimental setup and performance metrics are explained. The results are discussed and compared with those from state-of-the-art methods.

In Chapter 5, the proposed model is applied to detection of Malaysian license plates. The same methodology in Chapter 4 is followed. The experimental setup, performance metrics, as well as results and discussion are presented.

Finally, in Chapter 6, conclusions and contributions of the research are summarized. The existing limitations and possible areas for future work are discussed.