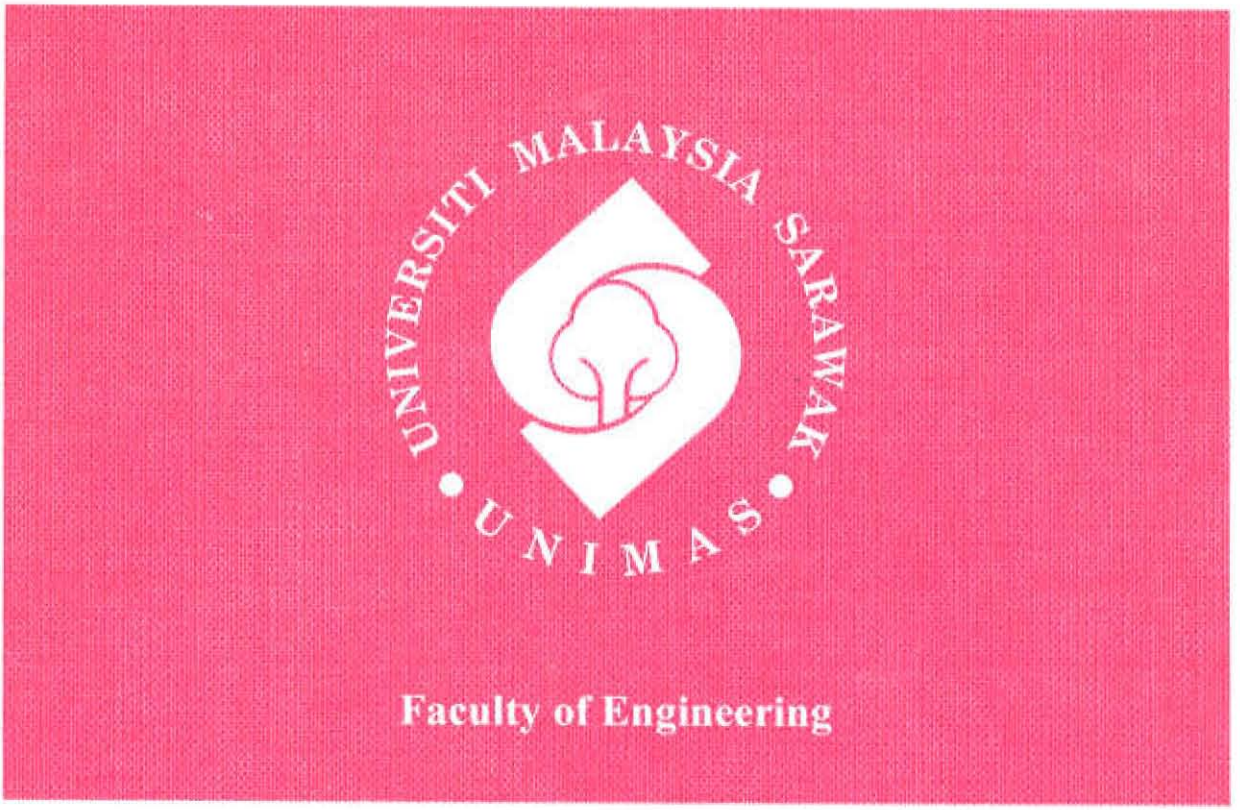

IMAGE SCENE RECONSTRUCTION FROM DASHCAM VIDEO

PUA ZE LONG

Bachelor of Engineering (Hons) in
Electronics (Computer)
2017/2018

UNIVERSITI MALAYSIA SARAWAK

Grade: _____

Please tick (√)

Final Year Project Report	☑
Masters	☐
PhD	☐

DECLARATION OF ORIGINAL WORK

This declaration is made on the ...11th.....day of...June.....2018.

Student's Declaration:

I, **PUA ZE LONG, 48724, FACULTY OF ENGINEERING** hereby declare that the work entitled **IMAGE SCENE RECONSTRUCTION FROM DASHCAM VIDEO** is my original work. I have not copied from any other students' work or from any other sources except where due reference or acknowledgement is made explicitly in the text, nor has any part been written for me by another person.

11/6/2018
Date submitted

Pua Ze Long (48724)
Name of the student (Matric No.)

Supervisor's Declaration:

I, **IR DR DAVID BONG BOON LIANG** hereby certifies that the work entitled **IMAGE SCENE RECONSTRUCTION FROM DASHCAM VIDEO** was prepared by the above named student, and was submitted to the "FACULTY" as a * partial/full fulfillment for the conferment of **BACHELOR OF ENGINEERING (HONS) IN ELECTRONICS (COMPUTER),** and the aforementioned work, to the best of my knowledge, is the said student's work.

Received for examination by: **IR DR DAVID BONG BOON LIANG** Date: 11/6/2018
(Name of the supervisor)

I declare that Project/Thesis is classified as (Please tick (√)):

☐ **CONFIDENTIAL** (Contains confidential information under the Official Secret Act 1972)*
☐ **RESTRICTED** (Contains restricted information as specified by the organisation where research was done)*
☑ **OPEN ACCESS**

Validation of Project/Thesis

I therefore duly affirmed with free consent and willingness declare that this said Project/Thesis shall be placed officially in the Centre for Academic Information Services with the abiding interest and rights as follows:

- This Project/Thesis is the sole legal property of Universiti Malaysia Sarawak (UNIMAS).
- The Centre for Academic Information Services has the lawful right to make copies for the purpose of academic and research only and not for other purpose.
- The Centre for Academic Information Services has the lawful right to digitalise the content for the Local Content Database.
- The Centre for Academic Information Services has the lawful right to make copies of the Project/Thesis for academic exchange between Higher Learning Institute.
- No dispute or any claim shall arise from the student itself neither third party on this Project/Thesis once it becomes the sole property of UNIMAS.
- This Project/Thesis or any material, data and information related to it shall not be distributed, published or disclosed to any party by the student except with UNIMAS permission.

Student signature _____ (11/6/2018)

Supervisor signature: _____ (11/6/2018)

Current Address: 3, Lorong D, Jalan Masjid, 83000 Batu Pahat, Johor.

Notes: * If the Project/Thesis is **CONFIDENTIAL** or **RESTRICTED**, please attach together as annexure a letter from the organisation with the period and reasons of confidentiality and restriction.

[The instrument is duly prepared by The Centre for Academic Information Services]

IMAGE SCENE RECONSTRUCTION FROM DASHCAM VIDEO

PUA ZE LONG

A dissertation submitted in partial fulfilment
of the requirement for the degree of
Bachelor of Engineering (Hons) in Electronics (Computer)

Faculty of Engineering
Universiti Malaysia Sarawak

2018

To my beloved family and friends.

ACKNOWLEDGEMENTS

This work supported by Ministry of Higher Education Malaysia through the provision of FRGS research grant, grant number F02/FRGS/1492/2016.

First and foremost, I would like to express my deep gratitude to my supervisor, Ir Dr David Bong Boon Liang, for his patient guidance in finishing and completing this project. He has given me a lot of supports and helped me to finish this project as plan.

The appreciation also given to all of my friends, academic and non-academic staffs of Faculty of Engineering, University of Malaysia Sarawak, their helps and motivations given throughout my undergraduate study.

Lastly, I wish to thank my parents and family members for their encouragement in my study.

ABSTRACT

This project presents a method for street scene image reconstruction from recorded driving video. The aim of the project is to combine street scene image along the road. The image generated will be the left-side and right-side street scenes along the street for two types of road which are straight and bend roads. Main technique applied in this project is the SURF feature based panoramic image stitching. Reconstructed street scene image was validated with subjective Image Quality Assessment. This report discusses on the project background, literature review on related research, methodology, result, discussion and future development of this project. The entire process of the project was demonstrated using MATLAB. Without the need to install high-end devices, street scene image with satisfying visual summary along the way was reconstructed from captured driving video.

ABSTRAK

Projek ini manyampaikan cara bagi pembinaan semula gambar pemandangan jalan raya dari video yang dirakam dengan kamera perakam kereta. Tujuan projek adalah untuk menggabungkan gambar pemandangan sepanjang jalan raya. Gambar pemandangan yang dibina semula adalah pemandangan sebelah kiri serta sebelah kanan sepanjang rakaman jalan raya. Rakaman video dijalankan dalam dua jenis jalan raya, iaitu jalan lurus dan jalan bengkok. Teknik utama yang digunakan dalam projek ini adalah penjahitan gambar panorama melalui ciri-ciri SURF. Gambar pemandangan jalan yang dibina semula telah disahkan melalui penilaian kualiti gambar secara subjektif. Laporan ini merangkumi kebelakangan projek, kajian literatur terhadap penyelidikan yang berkaitan, metodologi, hasil projek, perbincangan projek serta penambahbaikan projek ini pada masa depan. Proses keseluruhan telah dilengkapi melalui aplikasi MATLAB. Tanpa sebarang pemasangan alat yang berfungsi mewah, gambar pemandangan jalan raya yang telah dibina semula dari rakaman kamera perakam kereta masih mempunyai ringkasan visual yang memuaskan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Overview

This is a design based project to combine street scene images along the road from the driving videos captured by the vehicle camera.

Vehicle camera (also known as dash cam) is an on-board camera that record the driving scene continually which takes the view through the windscreen of the vehicle. A vehicle camera is normally attached by adhesive tape or suction cup to the dashboard of interior windshield of a vehicle. Vehicle camera is considered as a real time driving video recorder which mainly used to record the live video of the outside scene or environment of the vehicle. In recent years, with the rapid growth and cost-effectiveness of digital video recording devices, vehicle camera becomes a very popular device in the market[1]. All the incident or accident events that happen along the driving journey can be recorded in the driving video. The device has resulted in the rapid use of the video surveillance for safety purpose. For example, police may collect the videos recorded by the vehicle camera for investigation purpose during an accident. The captured driving video becomes the key element for traffic analysis as well as driving behaviour. A lot of information can be obtained from the captured driving videos through mining a large collection of driving video [2]. Hence, vehicle camera market is expected to grow due to its high demand.

The development in the technology of vehicle camera leads to the development of street scene image (also known as street view) generation. With vehicle camera, it provided the visual summaries which obtained along the way. The users are allowed to retrieve a video clip corresponding to a specific road section [1]. The street view techniques provide the real panoramic view of the street scene. Other applications of the

street scene image include virtual touring, building introductory, street navigation and the city modelling and profiling. In year 2001, Google began the inception on the Google Street View, and in year 2007, the system covered five major cities and their suburbs in United States of America. The collection of the Google street view images and geospatial information is done by the deployment of high-end devices and advanced computer vision. There are total of 9 digital cameras, 3 laser scanners and GPS in the system deployed by Google. The development of Google Street View considered as high budget and time-consuming work as it needs many cars and drivers to complete the street view image collection from region to region [3]. The fast development of the street view technologies makes the techniques to become more popular and common among the users and researchers.

Reconstruction of street scene images technologies involves the application of digital image processing. Image stitching and video stitching are the techniques to generate a panoramic image. Image stitching is a technique where images of overlapping domain of view are blended together. Video stitching perform stitching frame by frame independently [4]. To stitch image together, feature based techniques is applied to extract unique features from the target images or frames and correlate them. Feature points extracted is compared with other feature in other pair of image through the local feature descriptors. Other stages in image stitching and video stitching includes image calibration, image registration and image blending.

This project targets to reconstruct the panoramic street scene images along the street from driving videos. This can be achieved by applying the different techniques of image processing algorithms. The most effective algorithms will be selected to develop the image scene reconstruction system.

1.2 Problem Statement

The main objective of image scene reconstruction is to generate a panoramic street scene images from the driving video captured by the vehicle camera mounted in the car. The two common ways in street scene generation are professional generated videos, pre-processed such as Google Street View, and user generated videos such as hand-held camera and vehicle camera. However, the professional method needs a very high budget

and is considered as a time consuming approach to maintain high quality images. It causes the approach only applicable in popular cities and the information is updated at a very long intervals of time.

Enjoying the street scene can be described as one of the most enjoyable moment during travelling. However, drivers often face the problem that they are not able to enjoy the street scene as they have to focus on their driving. It is also inefficient to watch the substantial video collections in a sequential way for searching relevant driving videos. Therefore, image scene reconstruction proposed to reconstruct the side scenery images along the street from the driving video captured by the vehicle camera. With the panoramic street scene images constructed, drivers will be able to enjoy street view when they are not driving.

1.3 Objectives

The objectives of the project are as follow:

- To develop an image scene reconstruction algorithm for the generation of street scene images from captured driving video.
- To combine digital image processing techniques by using MATLAB as the assessment tool.
- To predict the quality of the stitched image by performing Image Quality Assessment (IQA).

1.4 Project Scope

In general, this project aims to develop an algorithm that is able to reconstruct and generate the panoramic street scene image. The image generated will be the left-side and right-side street scenes along the street. In this project, a vehicle camera will be used to record the driving video along straight and bend roads as to compare the performance of the algorithm under different types of road. Video recording is conducted in the morning, afternoon and night time to compare the quality of reconstructed scene under different

lighting conditions. Once a video is recorded by the vehicle camera, the algorithm will then generate the left and right street scene along the road. The driving condition is set to be under a constant speed of 20km/hour along the road. The vehicle camera used is a 120° wide angle with resolution setting of 1920x x1080 pixels. Research is done in order to identify the most suitable method in generating the algorithm for the image scene reconstruction.

1.5 Expected Outcomes

The expected outcomes of the project are:

- An algorithm with the capability of generating left and right-side scenes from driving video.
- Generation of side scenes from driving video along straight and bend roads.
- Evaluation of the quality of the stitched street scene image by using Image Quality Assessment.

1.6 Project Outline

This report is organized into 5 chapters. The chapters include Introduction, Literature Review, Methodology, Result and Discussion, and Conclusion.

Chapter 1 presents an overview of the project followed by the problem statement and the objectives to be achieved in this project. Besides, it provides the project scope and expected outcome of the project.

Chapter 2 summarizes the relevant works and research. The reviews provides a clear understanding of the project. Besides, it also reviews different methods and techniques in image processing in order to develop the algorithm for image scene reconstruction.

Chapter 3 discusses on the methods of the investigations, and methodology of the project. Different methods and approaches are clearly outlined and justified in this chapter.

Chapter 4 evaluates on the performance and results of the system. It also demonstrates the results obtained from the approach of image scene reconstruction algorithm.

Chapter 5 presents the summary of the overall process and performance of the project. Future works and suggestions for improvement is also being discussed in this chapter.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

This chapter includes the related work that has been done and different techniques which can be applied in the process of image scene reconstruction.

Different approach in video framing has been discussed in this chapter. Techniques in image processing, such as, feature descriptor, image registration, image matching, image mosaicking, image reprojection and image stitching are also been reviewed as well.

2.2 Digital Image Processing

Image is defined as two-dimensional function of *x* and *y*, f(x,y), in which *x* and *y* are spatial coordinates[5]. Intensity or gray level of an image refers to the amplitude of the function *f* at any pair of coordinates at any particular point. Digital image defined as the finite and discrete quantities of *x* and *y* and the intensity value of function *f*. By the means of digital computer to process a digital image, it refers to the field or study of digital image processing.

According to [6], the understanding of image by a computer can be considered as an attempt to characterize the relation between input image(s) and the previously established models of the observed world. It helps to reduce the information contained in the image to specific information for the application domain. The representation of an image falls into four levels according to the data organization, as shown in Figure 1.1. In

a simpler way of study, the hierarchy of digital image processing can be categorized into low-level image processing and high-level image processing. Low-level image processing is implemented by high-level algorithm for the image content understanding by computer. The methods include image compression, pre-processing method for edge detection, noise canceling and image sharpening. In high-level image processing, artificial intelligence methods are generally suitable according to the knowledge, goals and achievement planning.

2.3 Digital Video Processing

Videos can be defined as the visual information which includes still images and time-varying images [7]. Spatial distribution of intensity which is constant with respect to time is considered as a still image. Meanwhile, time-varying image is the spatial intensity pattern changes with time. Digital video contains the horizontal and vertical resolution, which is related to the number of pixels per line and the number of lines per frame respectively. It correlates to spatio-temporally sampled version of time-varying image. Figure 2.1 represents the block diagram for digital video formation, in which, "3-D Scene Modeling" considers as the modeling of motion and structure of objects; "Image Formation" refers to the geometric and photometric image formation; "Spatio-Temporal Sampling" corresponds to the sampling of time-varying images $s_c(x_1, x_2, t)$ in both spatial and temporal variables.

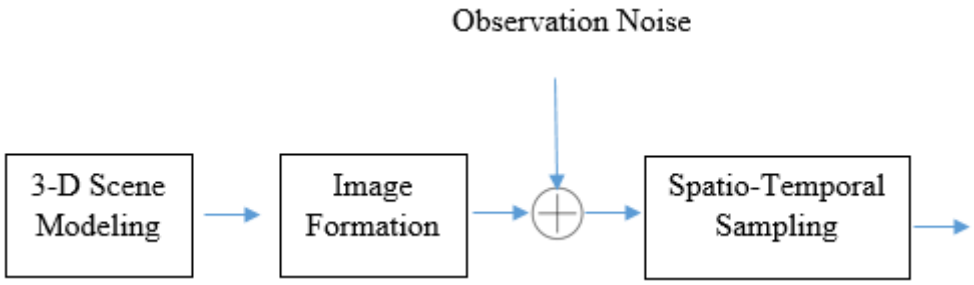

Figure 2.1: Digital Video Formation Block Diagram [7]

Digital video processing refers to the manipulation of the digital video bit stream. Digital video processing is also studied as the video imagery which contains a significant amount of temporal correlation between the frames.

2.4 Video Framing

A video is defined as the sequence of a number of frames running per unit of time. Colour information, textures and shapes in the original dimension are included in the video frames. Extraction or matching is then become a popular topic in digital video processing. The video sequence can be arrange in a more well-defined form. Sequence of video is divided into five levels which are *video shot*, *key frame*, *video scene*, *group of shots* and *video* [8]. Video shot is the basic block of the video which is a series of frames that took for a period without disruption; Key frame includes the principal content of the shot; Video scene compile the continuous shots and contents in a video; Group of shots is an intermediary singleton between video shots and video scenes; Video compile all the other components.

2.4.1 Key Frame Extraction

Basic unit for a video is known as *frame*. A content representation of a video is complex and unstructured, therefore, it is necessary to summarize the video [9]. Key framing extraction is a productive method to keep the overall video's contents. It is also the process to select the important frames in a video by sampling video frames randomly at certain period of time. There are several techniques applied in key frame extraction [10] -[11].

Ntalianis and Kollias [10] proposed the method of Singular Value Decomposition (SVD) to extract the key frame of a video. SVD summarizes with correct shot option according to the key frames. The process to prepare summarized video that provides abstract view of the original video sequence is known as the process of video summarization. This can be further applied for browsing and improvement system. However, the size of video frames affect the execution of the video summarization poorly.

Another method in key frame extraction is based on motion analysis [12], where the optical flow for each frame is evaluated of its differences along the sequence, and then, a simple motion metric is used as well. The key frames can be obtained at the places where the metric as a function of time have the local minima.

In [13], an approach to extract key frames from the amount of overlap is proposed to achieve the minimization of the computational cost.. The stitching only considers the frames for the mosaic image construction. Every frame is paired to its past frame. The frame is marked when it is found that the amount of overlapping is below a threshold.

2.5 Scene Construction

Several methods can be used for scene construction. It can be classified into four categories.

2.5.1 Depth Map Based

Methods of depth map based is by registering the input images to depth map [14]. This approach integrates the new viewpoint images and then it is further accelerated by quad tree decomposition and view-independent visibility priority. For image in between viewpoint, the image is formed by morphing the adjacent images. This process utilizes the correspondence maps which is pre-computed. The final new image formed does not dependent of the complexity of the scene. To complete the transformation of the scene coordination with the point source, morph map from the viewpoint is computed.

2.5.2 Monocular Image Based

Applying Markov Random Field (MRF) into infer plane parameters is able to identify the position and orientation of triangular facets of image [15]. The approach is to over-segment the image into several small homogeneous regions. The regions are then named as “super-pixels”. MRF is applied in order to deduce the 3D position and