Department of Industrial and Systems Engineering

BSc(Hons) in Logistics Engineering and Management (Code: 02004)

ISE414 - Individual Project

Project

Monitoring and Assessment

Student Name:	**NG Kam Hung**	**PolyU ID No.**	**130 486 72D**
Supervisor:	**Dr K.M. Lee Carman**	**Project Code:**	
Project Title:	**Real Time Vehicle Routing Problem using Image Processing under Big Data Environment**		

Submit the completed workbook on 28 Nov 14 (Week 13) to the General Office (ISE) after approval by your Project Supervisor.

ISE414 - Individual Project

Working on your Project
The objective of this workbook is to guide you to work *effectively* through all the stages of your project in a systematic manner. It will help you to:
Analyse the nature and scope of your project and to formulate its objectives
Conduct a Literature Review
Suggest and select the methods of achieving the Project's objective(s)
Manage the Project's progress

The completed workbook will reflect your effort, skill and progress. It will be part of the Project's assessment and account for 10% of your final project grade.

Supervisor's comment: **Stage 1** **(Week 5 recommended)**

Stage 2 **(Week 12 recommended)**

Log Sheet

Note: To be updated regularly by the student during the Final Year.

Date of Meeting	Brief description of purpose and major conclusions	Duration of meeting
3 Jul	Project Identification Discussion – Big Data	1.5hrs
4 Aug	Discussion the possibility for using Image Processing in VRP	1 hr
3 Sep	FYP Progress – Review on Image Detection Algorithm	0.5 hr
8 Sep	FYP Progress – Review on Background Subtraction Algorithm	1 hr
29 Sep	FYP Progress – Proposed Image Processing Algorithm	1 hr
21 Nov	Workbook Presentation	0.5 hr

Continue on extra page if required.

I - Project Background

Write down:

1. The background of your project
2. The scope of the project
3. The issues/problems that need to be tackled

1. INTRODUCTION

1.1 Research Background

1.1.1 Complex Issue for Traffic Condition

The travel time is greatly influenced by distances and traffic condition. The result of routing path becomes a vital issue for the company on reducing total travel distance and delivery time, which allows the company to access a competitive marketing with a lower transportation cost and higher delivery efficiency. Considering the delivery stage from warehouse to a set of customers, a shorten path for delivery indicates the company performance and customer satisfaction.

Besides the issue of travel distance, traffic condition is another factor for the total travel time. To counterbalance the uncertainties of travel condition by the impacts of weather, accidents and traffic congestion, the company should take into account for those factors to achieve their delivery goals, like on-time delivery.

1.1.2 Congestion Problem in High-Density City

Nowadays, agile and on-time delivery linked with the company performance and customer satisfaction. In order to compete a competitive advantage towards other competitors, vehicle routing problem solution provide a shorten pathway for delivery activities. Since Hong Kong is a high-density city, the diversity of vehicle routing network solutions would be narrow. However, traffic congestion cites an important factor in delivery time or travelling time for the company. Time becomes a crucial factor to cost elimination and resources utilization, as traffic would not return any benefit for the company. The risk of late delivery will also occur, when inappropriate route has been selected for delivery.

It is important to evaluate and examine the significance level between delivery goals and traffic condition. Traffic Congestion is an example of information and road construction failure, which could be summarized in several characteristics, such as long vehicle queue, uncertain transportation time and slow-movement of vehicle, which frequently occur in high density city with low traffic flow. Substantial traffic congestion incurs great expenses towards all companies, which threatens the growth of economic of a city (Lee, 1998). A dynamic change of traffic condition may disturb the predetermine delivery plan and time management. And thus, this consequent the increased delivery cost.

1.1.3 Importance with Real Time Traffic Surveillance

A real-time surveillance for traffic risk mitigation in vehicle routing could avoid high traffic site, reduce the likelihood of long delivery journey and construct a smooth outbound logistics flow. With the reference of real time traffic information and periodic updated traffic information, a wise and constructive suggested vehicle path could reduce the variances of transportation cost and confirmative delivery time (Kim, Lewis & White III, 2005). Moreover, an up-to-date routing guidance and information providing to the road users could enhance the overall traffic condition (Arnott, De Palma & Lindsey, 1991). The traffic data collected from surveillance system allows us to access in "root causes" of congestion problems and other disciplines, like continuous speculation for civil engineering in road construction, driving behavior and planning for public transport routing.

1.2 Problem description

1.2.1 Current Approach for traffic information extraction

In this research, we identify the limitation of using public data to extract traffic information, investigate the possibility for processing traffic image with moderate resolution and frequent change of location on real time webcams. Based on current limitation, we adopt a modified image processing approach for Real-Time Vehicle Routing Problem (RTVRP).

1.2.2 Historical data and personal experiences

In general, drivers base on their past experience and traffic information broadcasting by radio station to evaluate the current traffic condition and plan the delivery route. The estimation of traffic condition will be subjective and the complexity of traffic estimation will increase with large customer size of delivery.

1.2.3 Intelligent design on image processing

According to the properties of image snapshot provided by HKTD, the traffic density estimation could be retrieved in real time basis by an appropriate image-processing algorithm. Due to the technology advance in traffic management system, traffic snapshot and data are available in real time. Traffic pattern could be retrieved as historical data for the traffic prediction. Considering the real time traffic factor for VRP, a reasonable computation time for retrieve traffic information could be estimated under a dynamic traffic management system.

Most of the image processing techniques has been well developed. However, the performance is subject to the resolution of webcams, illumination change, traffic density, weather, camera's orientation and other noises.

Information of traffic condition can be collected by different sensor, which involves a higher cost of hardware and software development. It is not feasible for private company to adopt a traffic surveillance system in Hong Kong. To extract the real time traffic information, Hong Kong transportation Department provides public data and traffic snapshot.

1.2.4 Limitation for current image processing algorithm

Image processing can be classified as two major groups, object detection and foreground subtraction, which are technological infeasible with RTVRP.

The current image processing approach may not be feasible for extracting traffic information. When considering the objective for traffic information, calculating the actual number and classification of vehicles can guarantee the optimal outcome of traffic extraction. However, the long process speed for one location may not be able to cope with RTVRP. The accuracy of object identification is subjected to the large quantities of machine learning process with high- image resolution, which is limited by time constraint.

The foreground subtraction approach relies on background update method to extract moving object on the image. The slight change of noise is defined as a background. However, the stopping units in video will be considered as "noise". When a long traffic occurs, the reliability of vehicle-units estimation would not be accurate, as temporary stopping units in video will be considered as “noise”.

1.3 Scope

This paper studies on real time intelligent routing system under civil area with frequently traffic congestion. The real-time vehicle routing solution becomes the core part for logistics sector, as pre-determine routing plan ignores the traffic condition, which mainly alerts the actual performance of delivery. This brings a significant cost towards company. The variance of vehicle routing solutions by heuristic approach would not be too diverse in a condense network. However, real-time traffic condition provides insight for time-optimal routing solutions with quick response and on time delivery approach. A modified image processing algorithm will be proposed in this paper with the consideration of agile computation speed and traffic information retrieval using video cam.

This paper reviewed the concept and methodology of vehicle routing problem. Our proposed vehicle routing algorithm included the real-time image processing technique. The traffic snapshot will be periodically downloaded from web sources, and proceed to foreground extraction algorithm. The retrieved information will be considered as a reference for constructing the Real Time Vehicle Routing Problem under Big Data Environment.

1.5 Organization of Report

The rest of this paper is organized as follows. After a brief introduction to vehicle routing problem, the concept of big data and a brief comparison of object detection and motion detection, we discussed the well-known object detection from image and motion detection by video in section 2. The proposed model formulation of Real Time Vehicle Routing Problem (NTVRP) with a Dynamic Contour Foreground Extraction Method will be discussed in the same section, which consists of several parts to retrieve the foreground in a quick method to solve the new NTVRP and examine the feasibility to implement that in current traffic snapshot hardware using Big Data Warehouse in section 3.1.

Population based VRP algorithm, which is Modified Artificial Bee Colony (ABC) Algorithm, will be described in section 3.2. In section 4 of the Performance Evaluation, we describe the computation result of the reference case comparing the performance of ABC and ABC with Map Reduce Approach and analyze the differentiation of the optimal solutions. Finally, the conclusion, limitation and future development are addressed in ending part.

Continue on extra page if required.

II - Problem Formulation

Write down:

1. The Objective(s) of the project
2. Its expected deliverables (e.g. workflow, process improvement, systems design, application software, product/process design, a business model, etc. etc.)
3. What needs to be done to ensure that the deliverables are met within the time, funding and other constraints.

1.4 Objectives

This paper mainly focuses on the following objectives:

- To design a traffic image processing technique to retrieve the traffic status
- To design a big data environment
- To minimize the overall traveling distance with the consideration of real time traffic condition
- To compare the selected population based algorithms for vehicle routing problem case

Expected deliverables

The real time vehicle routing model will be proposed in this research. The vehicle routing model is based on the shorten distance to develop the routing networks. The traffic condition will become a reference to guidance the VRP solutions. The model will be demonstrated the best three routes for the users.

VRP is a NP hard problem, if more factors add in the VRP model. The feasible solution will be increased significantly. Thus, this research also consider the Map-Reduce approach in Big Data to increase the calculation speed and increase the optimal searching rate for heuristic algorithm.

The system design and algorithm processing workflow will be shown in methodology part.

The frameworks including:

1. Problem identification
2. Analysis the image processing approach for Hong Kong Situation
3. Proposed Modified image processing approach
4. Implemented the RTVRP with traffic image processing
5. Parameter evaluation
6. Evaluating selected VRP algorithm and compare the performance between modified ABC algorithm and modified ABC algorithm using Map Reduce
7. Limitation for current studies and further studies

Limitation

The schedule has been arranged for this research. There are many image processing algorithms proposed by different scholar. It is important to understand the current algorithm and modified the algorithm to fix for current situation. The Literature review becomes the important part of the whole research. The program code is also a major part to implement the algorithm in VRP model. This is a tighten schedule for doing this research with large literature review and programming development.

Continue on extra page if required.

III - Literature Review

You will need to consult relevant background information and do a search for books, papers, studies, thesis, etc. from libraries, web-sites, and any other available sources that will be useful for you to successfully complete your project. Moreover, this is necessary to ensure that you make good use of work that has already been done and not try to “re-invent the wheel”.

Write down:

- A list of the references that you have found to be relevant to your project.
- State briefly their relevance and how you can make use of them.

2. LITERATURE REVIEW

2.1 Vehicle Routing Problem

The vehicle routing problem (VRP) is a NP-hard problem, which refers to a comprehensive optimization problem. The computation time for a NP-hard problem to calculate the shortened route using traditional method will be significantly increased when the number of nodes increases. It is not feasible to solve the huge routing network due to the inefficiency in solving a complex network by exact algorithm (Baade & Puranik, 2012).

2.1.1 VRP Modeling

The basic idea for vehicle routing problem is to minimize the total travel distance to delivery a certain amount of goods from warehouse to a set of customers with known position and definitely quantity of demand. The development of VRP has been amended due to different considerations and objective function. There are several VRP models considered different conditions and limitations to their applied situation. For instance, Capacitated Vehicle Routing Problem (CVRP) considers the vehicle capacity limit. It constructs the paths to serve all customers and make the return decision when the vehicles are fully loaded. Multiple Depots Vehicle Routing Problem (MDVRP) deals with the problem of routing paths from two or more warehouses to serve their customers. The solution for delivery to specific customers under a time window constraint has been proposed in Vehicle Routing Problem with Time Window (VRPTW) (Chen, Hsieh & Wu, 2012). Vehicle Routing Problem with Backhauls (VRPB), clarifies a more realistic situation, determines the Routing Path if the customers demand or return some goods during the delivery (Brandao, 2006). Environmental Vehicle Routing Problem (EVRP) maintains the balance between travel distances and Carbon footprint (Zhang, Lee, Choy, Ho & Ip, 2014).

2.1.2 VRP Solution

The complexity of VRP has been widely studied regarding to different objectives by proposed heuristics algorithm, which has been confirmed to be a NP –complete, like Local search method - Taboo Search (Wang & Lang, 2008) and population search based method - Genetic Algorithm (Baker & Ayechew, 2003), Ant Colony Optimization (ACO) (Chen, Hsieh & Wu, 2012), Artificial Bee Colony (ABC) Algorithm (Zhang, Lee, Choy, Ho & Ip, 2014).

2.1.3 Algorithm Selection

Comparing with the exact algorithm, approximate algorithms provide a near-optimal solution for the large-scale problem, like VRP, so as to satisfy the constraint and computation speed for operation. There are two types of approximate algorithms, which are classical heuristics and meta-heuristics. Meta-heuristics allows infeasible operation and re-combination as new solutions, which are possible to achieve a high quality solution than classical heuristics (Cordeau, et al., 2002).

Meta-heuristics divides into two categorized, which are Local search and Population search. Local search provide the basic iteration from the current solution to another better solution by neighbor operation. Population search is based on a pool of parent solution and produces offspring solutions for updating feasible solutions pool by iteration (Zhang, Lee, Choy, Ho & Ip, 2014). It is important to make a justification for the suitable algorithm evaluation.

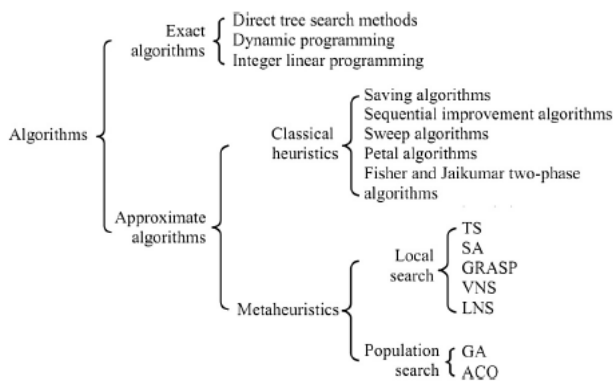


Figure 1.1: The algorithm for VRP and their relation

Adopted from Lin et al., 2014

2.1.4 The Benefit of including Real-Time Traffic Consideration

In real world, pre-determine routing without considering traffic information is not valid and ignores the different kinds of uncertainties that would alert the actual travel time, such as peak and non-peak hours, road maintenance and car clash. Examining the traffic information as a reference for VRP could reduce the variation of travel time and cost saving for the company (Fleischmann, Gnutzmann & Sandvoss, 2004). Traffic information could be retrieved from vehicle's GPS, driver's location services, traffic image and different sensors. However, those unstructured data is too large to handle and to be stored by a single node service. The size and variety of unstructured data input will dramatically increase and requiring to be transformed by a big data tool.

2.2 What is Big Data?

Due to the enormous data booming, a cohesive platform for processing structure and unstructured data becomes an essential element for any enterprise. The challenge of exploded data has been reviewed in term of speed and variation in comparison to the storage limitation and traditional data mining methods (Katel, Wazid & Goudar, 2013). Various industries notice the size of data which has been exponentially increasing and accelerating in data-flow. Notwithstanding the data scalability crisis, researchers discover the opportunities to analyze diverse Big Data type and Streaming Big Data. META Group initially established the concept of Big Data and identified three key features to define Big Data: Volumes, Velocity and Variety (Laney, 2001). Other authors also mention another V: Value that should be included in Big Data Analytics (Xin & Ling, 2013; Zheng, Chen & Li, 2013; Wang & Yuen, 2013).

2.2.1 Big Data Crisis or Opportunities

The major component of Big Data is unstructured data. International Data Corporation (IDC) addresses that the Compound Annual Growth Rate (CAGR) of structured data and unstructured data in "Worldwide File-based Storage 2010-2014 forecast" report were 21.8% and 60.7% correspondingly (Villars & Greyzdorf, 2010). Computer Sciences Corporation (CSC) further discusses the current domination of data types, which have been shifted to flexible, disorder and inter-related data. The rise of unstructured data reflects the need of new technology on data warehouse. (Shek & Gustafson, 2012) The solution for managing unstructured data has been introduced for industries, namely education, social media and Government sectors, which are Big Data Infrastructure. The transition from databases to big Databases is not only upgrading the storage capacity, but also requiring an infrastructure to process and handle structured and unstructured data (Madden, 2012).

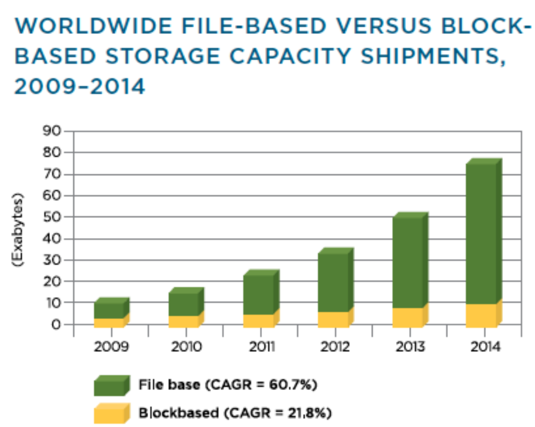


Figure 1.2 The CAGR of structured data vs. unstructured data from 2009 to 2014

(Adopted from IDC, Worldwide File-Based Storage 2010–2014 Forecast: Consolidation, Efficiency, and Objects Shape Market)

2.2.2 Automated Big Data Retrieval

Spatial data retrieved from various sensor input, mobile devices, social media, automatic sound or vision recording, eBusiness, satellite navigation device (GPS) and automatic identification devices (RFID, NFC) enables manufacturers to identify any data patterns and valuable messages in order to develop a precise strategy and respond to the market (Miller and Han, 2009). Big data storage infrastructure provides a base for spatial data mining in remote sensing (RS), Geospatial Information System (GIS) and business intelligence with a reliable, extensible and outstanding operation solution (Wang & Yuan, 2013). As the complexity of data type becomes the major challenge, the current database and data mining tools could not meet the demand of big data expansion. Zhang (2013) suggests a method to improve the quality of spatial data analysis in the view of data management. Moreover, the impact on inconsistent of unstructured big data and the classification of four types of problems were highlighted. The pre-processing of data input classification could carry out the easiest way to enhance the usability of big data mining.

2.2.3 Value Proposition for industries using Big Data Analysis

Current business intelligences are able to handle the structured data by specific algorithm (Capgemini, 2012). Yet, there are several papers reviewing that unstructured data had taken the first priority in decision-making and prediction that required big data analytics due to the rising diverse, real time and iterative data. Researchers suggest that investigating the integrated, transparent unstructured data sequence could benefit the entrepreneurs in accessing the huge potential for commercial organizations (Capgemini, 2012; Wielki, 2013; Muhtaroglu et al., 2013; Li et al., 2013). Despite the encouragement by researchers, entrepreneurs found difficulties in leveraging the effect of big data. However, Rajpurohit (2013) proposes a literature to deliver the key for the gap between big data mining and business value, which is beneficial to knowledge development, enhancing decision-making process, social improvement and so on. Furthermore, Amit & Zott (2001) identify four aspects of data that create business value and introduce value chain analysis to study the economic effects and business opportunities of Big Data Analytics. According to the nature of the Big Data, the growth, diversity and acceleration of data flow established the need for data warehouse management, but generated the deeper insight towards forecasting technique and greater probability to invest in customer behaviors (Balar et al., 2013; Zheng, Zhu & Lyu, 2013).

2.2.4 Benefit of Big Data Analytics for Commercial

Investigating the unstructured data could create the transparency, value and support decision-making among business with more vigorous and sophisticated data mining tools. Unstructured data becomes the essential element in strategies making (Kaisler et al., 2013). In addition, Big Data approach enhances the visibility and real time tracking in management through wireless sensor networks (WSNs). Deployed Automated data collection enables communication between servers and devices in Cloud Platform or Big Data Infrastructure (Tracey & Sreenan, 2013). There are two types of data being considered for analytics in business, which are structured data and unstructured data.

2.2.5 Structured Data for VRP

Structured data refers to the data model, which stores, processes and accesses data in a highly structuralized format. The predefined data model includes the data types (integer, string, currency, name, address) and constraint of data input (numeric only, number of characters, Person Identification Format). Since structures for data are limited, the processing speed would be quick, stored and visualized in graph or table.

Given a set of customers for VRP, the distances related to the vehicle networks could be formed as a matrix table. In order to optimize the transportation cost, a minimized total travel distance indicates a shorten length of the delivery path for vehicle path. In addition, the Number of Trucks and Truck's load capacity refer to the capability on handling a set of customers and corresponding demand for goods. The demand of each customer could be structured as a vector for computation.

2.2.6 Unstructured Data for VRP

Unstructured data refers to the data storage did not required any pre-defined data model and no be able to organize in a format for storage. Those data contains semantic meaning that would need a specified computation or data retrieval technique to transform into desired information.

Real time traffic snapshot becomes a consideration for VRP model, any traffic congestion with the trade-off between distance and time as an acceptable alternative could be implied.

2.2.7 Consideration on Big Data VRP

The reason for using Big Data Environment is the computation speed for image processing under a huge vehicle routing surveillance would significantly time-consume. Moreover, it is not feasible to use a computer for processing enormous data simultaneously. Multiple nodes for processing big data is required in this situation if we would like to consider a better iteration result from VRP.

2.3 Image Processing for Real Time Traffic Consideration

Manual analysis for traffic condition is not feasible and applicable. Developing an intelligent traffic surveillance system is crucial in traffic management. Image mining can be applied through feature detection or motion detection, which are proposed by different scholars recently.

2.3.1 Feature Detection from an image

There are several denoted algorithms for image processing by feature detection. Canny Edge Detection provides a basic idea for image processing by detecting a wide range of edge (Canny, 1986). In 2005, DALAL proposed a Histogram of Oriented Gradients for detecting pedestrians on street (Dalal & Triggs, 2005). HOG computes the local area of edge gradient by definite size of detection window, and compares the distribution of gradient direction by fed image and current image to determine the object. The performance of this algorithm is determined by the clearness of object's appearance and shape (Fei & Lin, 2013). Similar algorithms have been applied into other types of object detection, like vehicle (Chen, Chen & Chen, 2013; Wu et al., 2014). Recently, Deep Neural Networks shows an outstanding result for object detection. Google develops a "Google Brain" for object automatically learning and recognizing. However, this requires a large data size and long iteration for machine learning (Szegedy, Toshev & Erhan, 2013). Modified Approach, like Artificial Deep Neural Networks, is also applied in microscopy image for medical (Ciresan, Giusti, Gambardella & Schmidhuber, 2012).

2.3.2 Review on Image Processing by using Histogram of Oriented Gradients (HOG)

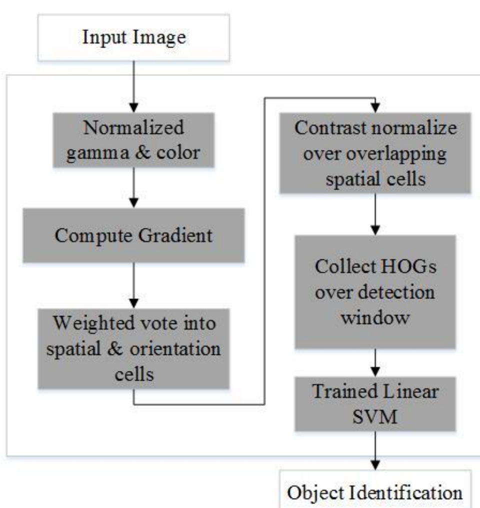


Fig. 2.1 processing step of Histogram of Oriented Gradients (HOG)

2.3.3 Motion Detection from a set of frame / video

Background subtraction for previous research can be divided in two groups – Semi-static and Feedback background subtraction (Intachak & Kaewapichai, 2011). The major difference of these two categories is the background updating method. As for the Semi-static approach, the background pixel is retrieved from fixed or adaptive frames in different time interval. The complicity of background model would be depends on the variation of illumination change, shadow, and weather (Horprasert, Harwood & Davis, 2000). As for feedback background subtraction, each pixel of the image is presented by density function or statistic model, such as K-Means and Gaussian Mixture Model. It is to identifies the foreground and background by the motion changes of each frame from video to eliminate the shadow factor with a various background changing from time to time, such as adaptive Gaussian Mixture Model for Background Subtraction (Zivkovic, 2004).

2.3.4 Review on Video Background Subtraction using Mixture Of Gaussians (MOG)

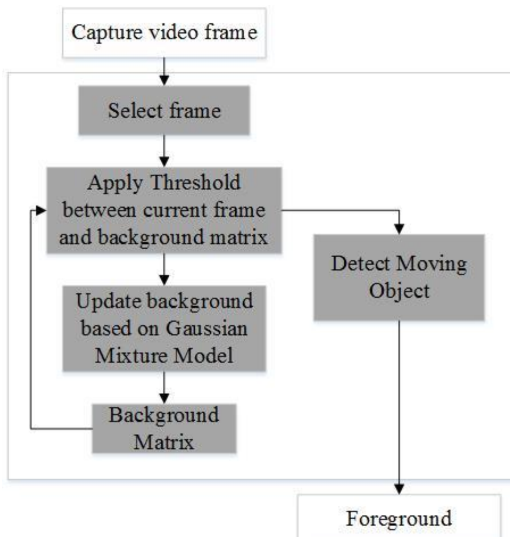


Fig. 2.2 processing step of Background Subtraction using Mixture of Gaussians (MOG)

2.3.5 Comparison between Feature Detection and Motion Detection

The dilemma of different methods for object recognition and traffic surveillance depends on the algorithm objective and software requirement. Feature detection could only apply to image with high resolution to retain the object recognition performance. In addition, manual selected “feeding” image for machine training would be quite time-consuming for initiating the learning vector. Most algorithms for feature detection could identify stopping vehicle. However, the accuracy will be alerted by the complexity of background environment and objects overlap.

Motion detection is required for processing a set of frames or videos with continuous time intervals to update the background model. For example, Mixture of Gaussian Model for background subtraction using Static location of surveillance cam provides a moderate calculated speed for real time traffic condition mining.

Traffic Detection	Feature Detection	Motion Detection
Medium	Image	A set of frame / Video
Resolution Requirement	High	Moderate
Continuous Time Interval	Not Required	Required
Training for Detector	Large Dataset	Not Required
Process Time	Relatively Low	Relatively High
Stopping Object Detection	✔	✖
Object Overlap Detection	✖	✔
Moving Surveillance Cam	Depends	✖

Figure 2.3 Comparison between feature detection and motion detection

Feature detection and video surveillance have been widely used in traffic inspection. In vehicle detection technique for several literatures, distinctive algorithm and robust solution have been developed in terms of traffic conditions, weather, and illumination changes to estimate vehicle density estimation under the limitation and hardware specification of traffic surveillance system.

There are 179 surveillance webcams in Hong Kong, which are located in the main road, highway and the road with traffic congestion. The traffic image is free to access by road users as reference.

Traffic surveillance snapshot and video are public resources provided by Hong Kong Transportation Department. Traffic snapshot is updated periodically in two-minutes interval with moderate resolution, which is not sufficient for applying feature detection. Traffic video is updated periodically from different sites in 7-seconds interval. This alerts the performance of motion detection for detecting background and foreground. The modified traffic surveillance system is required for object detection under the limitation of surveillance system.

2.4 Research gap

In extent literatures on image processing and VRP, wide ranges of researches or studies have been proposed in variety of journal. Most of the literatures focus on object detection and foreground subtraction to retrieve the useful information from selected frame, in while VRP aims to minimize the total travel distances or cost. However, there is less journal articles review on the real time traffic information in VRP using traffic density estimation, which applies image mining to traffic snapshot. There are a room of spaces to develop and improve the RTVRP from varies traffic information in Big Data, which is the major motive for this research.

2.5 Summary of Literature Review

This paper considers the real time traffic situation by using image processing and applies Ant Colony Optimization and Artificial Bee Colony Algorithm to evaluate the performance for Real Time Vehicle Routing Problem (RTVRP) in Hong Kong under Big Data Environment.

References

Amit, R., & Zott, C. (2001). Value creation in e-business. *Strategic Management Journal*, *22*(6-7), 493--520.

Arnott, R., De Palma, A., & Lindsey, R. (1991). Does providing information to drivers reduce traffic congestion?. *Transportation Research Part A: General*, *25*(5), 309--318.

Baker, B., & Ayechew, M. (2003). A genetic algorithm for the vehicle routing problem. *Computers & Operations Research*, *30*(5), 787--800.

Balar, A., Malviya, N., Prasad, S., & Gangurde, A. (2013). Forecasting consumer behavior with innovative value proposition for organizations using big data analytics (pp. 1-4).

Bhagade, A., & Puranik, P. (2012). Artificial bee colony (ABC) algorithm for vehicle routing optimization problem. *International Journal Of Soft Computing And Engineering (IJSCE ISSN: 2231-2307, Volume-2, Issue-2*.

Brandao, J. (2006). A new tabu search algorithm for the vehicle routing problem with backhauls. *European Journal Of Operational Research*, *173*(2), 540--555.

Canny, J. (1986). A computational approach to edge detection. *Pattern Analysis And Machine Intelligence, IEEE Transactions On*, (6), 679--698.

Capgemini,. (2012). *Resources search | Capgemini Worldwide*. Retrieved 16 October 2014, from http://www.capgemini.com/insights-and-resources/by-publication/ the-deciding-factor-big-data-decision-making/?

Cecere, L. (2012). *Big Data: Go Big or Go Home?*. Philadelphia: Supply Chain Insights LLC.

Chen, R., Hsieh, F., & Wu, D. (2012). Heuristics based ant colony optimization for vehicle routing problem, 1039--1043.

Chen, Z., Chen, K., & Chen, J. (2013). Vehicle and Pedestrian Detection Using Support Vector Machine and Histogram of Oriented Gradients Features, 365--368.

Ciresan, D., Giusti, A., Gambardella, L., & Schmidhuber, J. (2012). Deep neural networks segment neuronal membranes in electron microscopy images, 2843--2851.

Cordeau, J., Gendreau, M., Laporte, G., Potvin, J., & Semet, F. (2002). A guide to vehicle routing heuristics. *Journal Of The Operational Research Society*, 512--522.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection, *1*, 886--893.

Fei, L., & Lin, L. (2013). Pedestrian Detection Based on Histogram of Oriented Gradient in Intelligent Transportation System, 78--81.

Fleischmann, B., Gnutzmann, S., & Sandvoss, E. (2004). Dynamic vehicle routing based on online traffic information. *Transportation Science*, *38*(4), 420--433.

Horprasert, T., Harwood, D., & Davis, L. (2000). A robust background subtraction and shadow detection, 983--988.

Intachak, T., & Kaewapichai, W. (2011). Real-time illumination feedback system for adaptive background subtraction working in traffic video monitoring, 1--5.

Kaisler, S., Armour, F., Espinosa, J., & Money, W. (2013). Big data: Issues and challenges moving forward, 995--1004.

Katal, A., Wazid, M., & Goudar, R. (2013). Big data: Issues, challenges, tools and Good practices, 404--409.

Kim, S., Lewis, M., & White III, C. (2005). Optimal vehicle routing with real-time traffic information. *Intelligent Transportation Systems, IEEE Transactions On*, *6*(2), 178--188.

Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety. Retrieved from http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Lee, T. (1998). A study of the traffic congestion problem in Hong Kong: a case study of Wanchai district. *The HKU Scholars Hub*.

Li, L., Bagheri, S., Goote, H., Hasan, A., & Hazard, G. (2013). Risk adjustment of patient expenditures: A big data analytics approach, 12--14.

Lin, C., Choy, K., Ho, G., Chung, S., & Lam, H. (2014). Survey of green vehicle routing problem: Past and future trends. *Expert Systems With Applications*, *41*(4), 1118--1138.

Madden, S. (2012). From databases to big data. *IEEE Internet Computing*, *16*(3), 0004--6.

Miller, H., & Han, J. (2009). *Geographic data mining and knowledge discovery*.

Muhtaroglu, F., Demir, S., Obali, M., & Girgin, C. (2013). Business model canvas perspective on big data applications, 32--37.

Rajpurohit, A. (2013). Big data for business managersâ€”Bridging the gap between potential and value, 29--31.

Shek, S., & Gustafson, P. (2012). *Data Revolution*. Falls Church: CSC.

Shuliang, W., Gangyi, D., & Ming, Z. (2013). Big spatial data mining, 13--21.

Szegedy, C., Toshev, A., & Erhan, D. (2013). Deep neural networks for object detection, 2553--2561.

Tracey, D., & Sreenan, C. (2013). A holistic architecture for the internet of things, sensing services and big data, 546--553.

Villars, R., & Greyzdorf, N. (2010). Worldwide file-based storage 2010--2014 forecast update. *IDC, December*.

Wang, S., & Yuan, H. (2013). Spatial Data Mining in the Context of Big Data, 486--491.

Wang, Y., & Lang, M. (2008). Study on the model and tabu search algorithm for delivery and pickup vehicle routing problem with time windows, *1*, 1464--1469.

Wielki, J. (2013). Implementation of the Big Data concept in organizations-possibilities, impediments and challenges, 985--

989.

Wu, B., Kao, C., Jen, C., Li, Y., Chen, Y., & Juang, J. (2014). A Relative Discriminative Histogram of Oriented Gradients-based Particle Filter Approach to Vehicle Occlusion Handling and Tracking. *IEEE*.

Wu, Y. (2010). Model in Logistics Network Design for 3PL Companies, 1--4.

Xin, N., & Ling, L. (2013). How we could realize big data value (pp. 425-427).

Zhang, D. (2013). Inconsistencies in big data, 61--67.

Zhang, J., Chen, Y., & Li, T. (2013). Opportunities of innovation under challenges of big data, 669--673.

Zhang, S., Lee, C., Choy, K., Ho, W., & Ip, W. (2014). Design and development of a hybrid artificial bee colony algorithm for the environmental vehicle routing problem. *Transportation Research Part D: Transport And Environment*, *31*, 85--99.

Zheng, Z., Zhu, J., & Lyu, M. (2013). Service-generated big data and big data-as-a-service: An overview, 403--410.

Zivkovic, Z. (2004). Improved adaptive Gaussian mixture model for background subtraction, *2*, 28--31.

Continue on extra page if required.

IV - Methodology

1. You will need to formulate a framework for the methodology that you think most effective for achieving the objectives of the project.
2. Briefly describe the approach and method of investigation, design of any experiments and surveys, data collection and analysis methods, etc. to be used in the project. Think about the subjects you have studied over the last two years and any topics that could be made use of.
3. Think about any constraints that you may have to work within, such as time, equipment, budget and technical capability. How you will overcome them?

3. METHODOLOGY

3.1 System Architecture

In this chapter, we present the proposed system architecture of Real Time Vehicle Routing Problem (RTVRP). The framework architecture consists of sensor tier, data tier, logic tier and presentation tier.

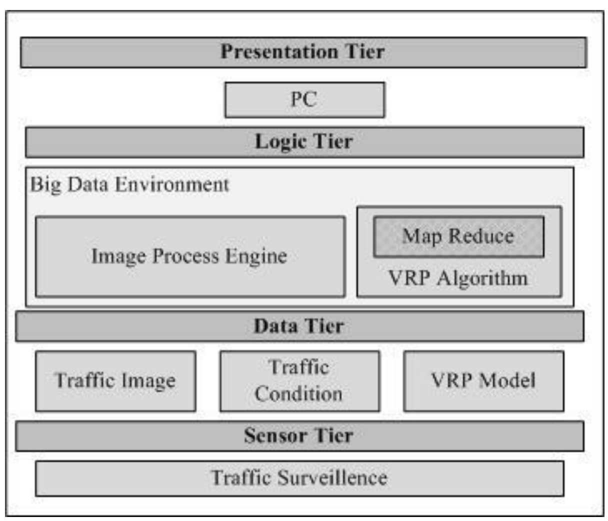


Fig 3.1 System Architecture for RTVRP using Image Processing Under Big Data Environment

3.2 Proposed Contour Extraction Method for Image Processing

By the reason of limitation of image processing and background subtraction, a modified approach for traffic density estimation has been proposed with RTVRP. The algorithm follows the below steps shown in Figure 3.1, which are region filtering, illumination detection, grey scale operation, threshold operation, morphological operation and traffic density estimation.

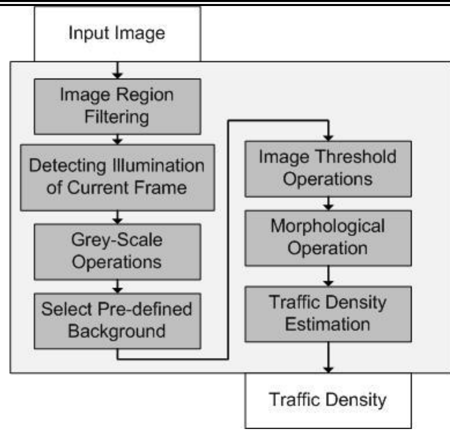


Fig. 3.2 processing steps of proposed dynamic contour extraction method for image processing

	Mathematical Model	Meanings
Image processing parameters	x, y	The coordinate of the image
	x^{ϕ}, y^{ϕ}	The maximum coordinate of the image
	$r_{x,y}$	The red colour in RGB model
	$g_{x,y}$	The green colour in RGB model
	$b_{x,y}$	The blue colour in RGB model
	$BG(r_{x,y}, g_{x,y}, b_{x,y})$	The subtracted background
	AIL	Average brightness level for current frame
	$ori(r_{x,y}, g_{x,y}, b_{x,y})$	The RGB model of input frame
	$F\left(r_{x,y}, g_{x,y}, b_{x,y}\right)$	The RGB model of filter region image
	$fil\left(r_{x,y}, g_{x,y}, b_{x,y}\right)$	The RGB model of filtered traffic image
	$gs_{x,y}$	The grey-scale model of a image
	$dst(x, y)$	The grey-scale model after applying threshold
	$threshold$	Threshold value
	$erode$	OpenCV library – erosion operator
	$dilate$	OpenCV library – Dilation operator
	$RD_{x,y}^{\zeta}$	Interested region of current input frame
	$RD_{x,y}^{\beta}$	Interested region of background frame
	IR^{β}	Accumulated value of targeted region of background frame
	IR^{ζ}	Accumulated value of targeted region of current frame
	TDE	Traffic density estimation
	t_{ij}	Traffic condition from position i and position j

3.2.1 Background Extraction

To enhance the computation speed for robust and real time traffic density estimation, a pre-defined background is required for Image Contour Extraction Method.

3.2.1.1 Background Extraction using Moving Average

Selecting the similar illumination level of image and low traffic as a feeding image for background extraction, the background will be generated by using moving average of each pixel with the formula (1) in RGB color model.

$$BG(r_{x,y}, g_{x,y}, b_{x,y}) = \frac{1}{n}\sum_{fr=1}^{n}\sum_{x=1}^{x\phi}\sum_{y=1}^{y\phi}F(r_{x,y}, g_{x,y}, b_{x,y}) \qquad (1)$$

Fig. 3.3 background subtraction using moving average

3.2.1.2 Detecting Background Illumination

Each pixels of RGB color model contains three values to display the color, which are red, green, and blue. The RGB coding of pure red, pure green and pure blue are (255,0,0), (0,255,0) and (0,0, 255) respectively, each value is under the range from 0 to 255.

The weighting on converting RGB model of each color to gray-scale model is not the same. The weighting for pure red, pure green and pure blue is described in formula (2), since brightness of RGB is different.

$$x = 0.299r + 0.587g + 0.114b \qquad (2)$$

In our model, we apply an average brightness level as a reference for identifying the background illumination and store the generated background with the brightness value. In formula (3), we apply the conversion formula to calculate the accumulated illumination value. The brightness value is a percentage conversion of accumulated illumination value divided by the number of pixels in formula (4).

$$AIL = \sum_{x=1}^{x\emptyset}\sum_{y=1}^{y\emptyset}\left(0.299r_{x,y} + 0.587g_{x,y} + 0.114b_{x,y}\right) \qquad (3)$$

$$Brightness = \frac{AIL}{x^{\emptyset} * y^{\emptyset}} \qquad (4)$$

3.2.2 Contour Extraction Method

The constructed background in previous part will be used as a foreground subtraction. Threshold operation and morphological operation will be applied to reduce the noise region and modify the interest region of the image.

3.2.2.1 Image Region Filtering

The road traffic is our major concern with this image mining approach. Other regions may alert the accuracy of the computation, like the pedestrians, swing trees and other moving objects. Image region filter will be applied in this step.

Input image $ori(r_{x,y}, g_{x,y}, b_{x,y})$ combine with the filter region image $F\left(r_{x,y}, g_{x,y}, b_{x,y}\right) = (0,0,0)$, which is the image only with black color with RMG Code (0,0,0) and white color with RMG code (255,255,255) Black region represents the uninterested region of the image and the white region represents the interested region of the image.

$$fil\left(r_{x,y}, g_{x,y}, b_{x,y}\right) = \sum_{x=1}^{x\phi} \sum_{y=1}^{y\phi} \begin{cases} if\ F\left(r_{x,y}, g_{x,y}, b_{x,y}\right) = (0,0,0), fil\left(r_{x,y}, g_{x,y}, b_{x,y}\right) = fil(0,0,0) \\ \quad otherwise, fil\left(r_{x,y}, g_{x,y}, b_{x,y}\right) = ori(r_{x,y}, g_{x,y}, b_{x,y}) \end{cases} \quad (5)$$

Formula (5) is to exclude all uninterested regions of traffic snapshot that remain the road area as later part calculation and generate the filtered traffic image $fil\left(r_{x,y}, g_{x,y}, b_{x,y}\right)$.

3.2.2.2 Detecting Illumination of Current Frame

Formula (3) will be applied to calculate the brightness of current frame. The brightness value becomes the selection criteria for appropriate backgrounds in our database.

3.2.2.3 Grey-scale Operations

To increase the computation speed, the image is transferred into grey-scale. Formula (6) is based on formula (2) to convert the whole image from RMB model into Grey-scale model.

$$gs_{x,y} = \sum_{x=1}^{x\phi} \sum_{y=1}^{y\phi} (0.299 r_{x,y} + 0.587 g_{x,y} + 0.114 b_{x,y}) \quad (6)$$

Fig 3.4 Traffic Image after filtering and converting into grey-scale model

3.2.2.4 Image Threshold Operations

Threshold operation technique is common and widely used in foreground subtraction. Each pixel is computed by the formula (7). dst (0) represents a white color and dst (255) represents a black color.

$$dst(x, y) = \begin{cases} If\ fil(x, y) > threshold), dst(x, y) = 0 \\ \quad otherwise, dst(x, y) = 255 \end{cases} \quad (7)$$

The concept of foreground subtraction is based on the threshold line between current frame and background. This formula can retrieve and show the foreground mask. The occupied regions of vehicles are shown in white color in foreground mask and the black color represents a background or noise region.

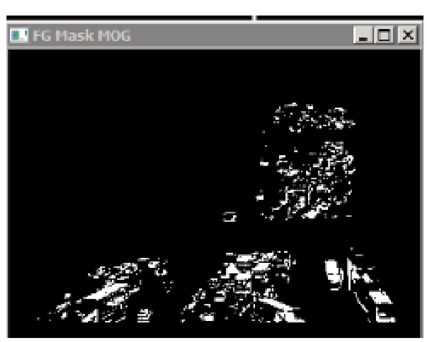


Fig 3.5 Traffic Image after threshold operations

3.2.2.5 Morphological Operation

The traffic image after applying threshold operations may not be able to show the vehicle sharp as proportions of vehicle sharp were filtered by threshold operations. Morphological operations have been considered to connect the central pixel by their compass coordinating with surrounding pixels. Morphological operations include two operations – which are dilation and erosion under OpenCV vision processing library.

Erosion Operations of foreground mask can concentrate the size of white region. The small objects that have not been filtered will be condensed. Dilation is the opposite operations of erosion. The size of whit region will be enlarged with neighbor white pixels. This operation allows connecting the broken parts of an object.

In our model, we apply formula (8) with OpenCV library to construct a mixture process of dilation and erosion operations.

$$RD^{\zeta}_{x,y} = \sum_{x=1}^{x^{\phi}} \sum_{y=1}^{y^{\phi}} dilate\left(\sum_{x=1}^{x^{\phi}} \sum_{y=1}^{y^{\phi}} erode\left(\sum_{x=1}^{x^{\phi}} \sum_{y=1}^{y^{\phi}} dilate(dst(x,y))\right)\right) \qquad (8)$$

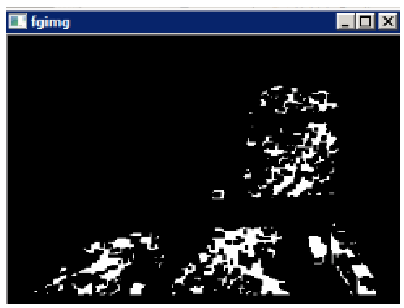


Fig 3.6 Traffic Image after Morphological Operation

3.2.2.6 Traffic Density Estimation

Formula (9) and (10) are to measure the size of interest regions of the background $RD^{\beta}_{x,y}$, while formula (11) and (12) are to calculate the occupied region of vehicles in the interest regions $RD^{\zeta}_{x,y}$. The percentage conversion of traffic density estimation is the occupied region of vehicles IR^{ζ} divided by the size of interest regions IR^{β}.

$$RD^{\beta}_{x,y} = \sum_{x=1}^{x^{\phi}} \sum_{y=1}^{y^{\phi}} \begin{cases} if\ BG(r_{x,y}, g_{x,y}, b_{x,y}) = (255,255,255), RD^{\beta}_{x,y} = 1 \\ otherwise, RD^{\beta}_{x,y} = 0 \end{cases} \qquad (9)$$

$$IR^{\beta} = \sum_{x=1}^{x^{\phi}} \sum_{y=1}^{y^{\phi}} RD^{\beta}_{x,y} \qquad (10)$$

$$RD^{\zeta}_{x,y} = \sum_{x=1}^{x^{\phi}} \sum_{y=1}^{y^{\phi}} \begin{cases} if\ BG(r_{x,y}, g_{x,y}, b_{x,y}) = (255,255,255), RD^{\zeta}_{x,y} = 1 \\ otherwise, RD^{\zeta}_{x,y} = 0 \end{cases} \qquad (11)$$

$$IR^{\zeta} = \sum_{x=1}^{x^{\phi}} \sum_{y=1}^{y^{\phi}} RD^{\zeta}_{x,y} \qquad (12)$$

The traffic density estimation value will be assigned in to the scale from 1 to 5 in formula (13). The estimation traffic condition (TDE) is under the scale from 1 to 5, where 1 represents low traffic and 5 represents extreme high traffic congestion from node i to j in formula (14).

$$TDE \begin{cases} if\left(0.0 \le \frac{IR^{\zeta}}{IR^{\beta}}\right) and \left(\frac{IR^{\zeta}}{IR^{\beta}} < 0.2\right), TDE = 1 \\ if\left(0.2 \le \frac{IR^{\zeta}}{IR^{\beta}}\right) and \left(\frac{IR^{\zeta}}{IR^{\beta}} < 0.4\right), TDE = 2 \\ if\left(0.4 \le \frac{IR^{\zeta}}{IR^{\beta}}\right) and \left(\frac{IR^{\zeta}}{IR^{\beta}} < 0.6\right), TDE = 3 \\ if\left(0.6 \le \frac{IR^{\zeta}}{IR^{\beta}}\right) and \left(\frac{IR^{\zeta}}{IR^{\beta}} < 0.8\right), TDE = 4 \\ if\left(0.8 \le \frac{IR^{\zeta}}{IR^{\beta}}\right) and \left(\frac{IR^{\zeta}}{IR^{\beta}} \le 1.0\right), TDE = 5 \end{cases} \tag{13}$$

$$t_{ij} = t_{ji} = TDE \tag{14}$$

3.3 Real-Time Vehicle Routing Problem (RTVRP)

There are two criteria for VRP algorithm selection in our evaluation of Real Time Vehicle Rouging Problem (RTVRP). First, alternative algorithms have to be well known and widely used in VRP. Second, similar parameter and algorithms from the same category are implied in this study for ease of comparable in nature. Modified Artificial Bee Colony (ABC) Algorithm and Modified Artificial Bee Colony (ABC) Algorithm with Map Reduce Approach will be applied in this paper, which are population-based heuristic approach.

3.3.1 Real-Time Vehicle Routing Model (RTVRP) Setting

VRP is the graphical based model to combine with s set of nodes and networks. The set of node $V = \{0,1,2,\dots,n\}$ represents the depot and customers location, and the set of edges $A = \{(i,j)|i,j \in V \ \ , i \neq j|\}$ represents the connective path between all nodes. The setting and notation of RTVRP are described in Table 2.

In this research, the single depot is denoted as node 0 and the number of customer I, to be served, is the remaining node from 1 to n. distance $d_{ij} = d_{ij}(i,j,i \neq j)$ describes the actual distance from each customers, which is associated with traffic condition $t_{ij} = t_{ij}(i,j,i \neq j)$ to indicate the current traffic situation under the scale from 1 to 5. Each customer requests a nonnegative demand $r_i \ (i = 1,2,\dots,n)$ which requests a delivery from depot to all customers. There are spare vehicles for emergency usage with the maximum number of available vehicles m. Each vehicle denoted with $k = \{0,1,2,\dots,m\}$ and each vehicle will be used one time in the vehicle routing plan under the maximum vehicle's capacity Q.

Table 2 - The setting and notations of RTVRP

	Mathematical Model	Meanings
Graph related parameters	$G = (V\ \ , A\ \ )$	The Graph
	$V\ \ = \{0,1,2,\dots,n\}$	The set of nodes in the graph
	$A\ \ = \{(i,j)\|i,j \in V\ \ , i \neq j\|\}$	The set of edges in the graph
Customer-related parameter	$d_{ij} = d_{ij}(i,j,i \neq j)$	The distance between i and j
	$t_{ij} = t_{ij}(i,j,i \neq j)$	The traffic condition between i and j
	$r_i\ \ (i = 1,2,\dots,n)$	The customer i demand
Vehicle-related parameter	m	The maximum number of vehicle
	$k\ \ = \{0,1,2,\dots,m\ \ \}$	Number of vehicle
	Q	The vehicle capacity
	V_0	The depot
Vehicle load-related parameter	q	The current vehicle load
	q_{ijk}	The load of vehicle k on edge (i,j)
	q_{0jk}	The initial load of vehicle k
	q_{i0k}	The final load of vehicle k

Objective Function for RPVTP

$$\min f_1 = \sum_i \sum_j \sum_k d_{ij} * x_{ijk} \tag{1}$$

$$\min f_2 = \sum_i \sum_j \sum_k t_{ij} * x_{ijk} \tag{2}$$

Constraint

s.t.

<u>For CVRP constraint</u>

Binary constraint for decision variables:

$$x_{ijk} = \begin{cases} if\ node\ i\ is\ followed\ the\ sequence\ by\ \ vehicle\ k, x_{ijk} = 1 \\ otherwise\ , x_{ijk} = 0 \end{cases} \tag{3}$$

$$y_{ik} = \begin{cases} if\ node\ i\ is\ visied\ by\ vehicle\ k, y_{ik} = 1 \\ otherwise\ , y_{ik} = 0 \end{cases} \tag{4}$$

Vehicle Load: $\sum_{i=1}^{n} R_i\ \ z_{ik} < Q^w, \forall k$ (5)

Cumulative Vehicle Load: $q_{0jk} = \sum_{i=1}^{n} r_i\ \ y_{ik}, \forall k, j$ (6)

$q_{ijk} = q_{uik} - r_i\ \ , (u \rightarrow i \rightarrow j), \forall k,$ (7)

$X = (x_{ijk}) \in S, (S \subset V\{0\})$ (8)

Maximum Number of vehicle: $\sum_{k\ \ =1}^{m} x_{ijk} < m\ \ , \forall j$ (9)

Vehicle usage: $\sum_{i=0}^{n} y_{ik} = 1, \forall i$ (10)

Customer's assignment: $\sum_{i=0}^{n} x_{ijk} = y_{jk}, \forall i \neq j, j, k$ (11)

$$\sum_{j=0}^{n} x_{ijk} = y_{ik}, \forall i \neq j, \forall i, k \tag{12}$$

Objective (1) is to minimize the total travel distance and explore the optimal route. Objective (2) is designed to minimize the accumulate traffic factor to all vehicle. The two objectives are under different priority levels. Objective (1) is the primary objective and objective (2) is the secondary objective, as the travel distance is the major target for VRP and traffic factor is a reference guide for designing the vehicle paths.

Constraints (3) and (4) indicate there is no excess vehicle arrival to the same customers. Constraint (5) requires the vehicle load under the maximum loading weight for one trip delivery. Constraint (6) and (8) describe the initial and final vehicle routing load. The cumulative vehicle load from customers i to j is shown in constraint (7). There are spare vehicles for delivery plan and they would not excess the maximum number of vehicle under the constraint (9).

Each customer would be visited once for all vehicles with the constraint of (10). Constraint (11) and (12) define that the network of each sub-route is limited by only one for all visiting vehicle, departing vehicle and serving vehicle.

3.3.2 Description of Artificial Bee Colony Algorithm (ABC)

Artificial bee colony algorithm is a foraging heuristic approach based on the behavior of natural honeybee swarms presented. In ABC algorithm, three groups of honeybees have been imitated, which are employed bees, onlooker bees and scout bees.

The whole process of food sourcing by honeybees is explained as follow. Onlooker bees and scout bees are the unemployed bees, which carry a function of food sources selection and searching. Employed bees carry nectar extraction and return nectar to the hive.

The scout bees will start the initial search for surrounding food sources randomly. Each employed bee will be sent to extract nectar from the found food sources and return the food back to the hive. Employed bees perform waggle dance to recruit more employed bees if the selected food source is abundance of nectar. Otherwise, the bee may abandon the food source.

The onlooker bees will select the most attractive food sources according to the performance of employed bees' waggle dance. The employed bees will be assigned to the corresponding food source and extract nectar until the food source is exhausted. Then, the employed bees will become scout bees and start to discover the new surrounding food sources. The scout bees identify the position of next food source. Once they find the next food sources, they will become the employed bees to perform a waggle dance and provide signal to onlooker bees to recruit more employed bees until the nectar is exhausted.

The step of the ABC algorithm is shown in Figure 3.7.

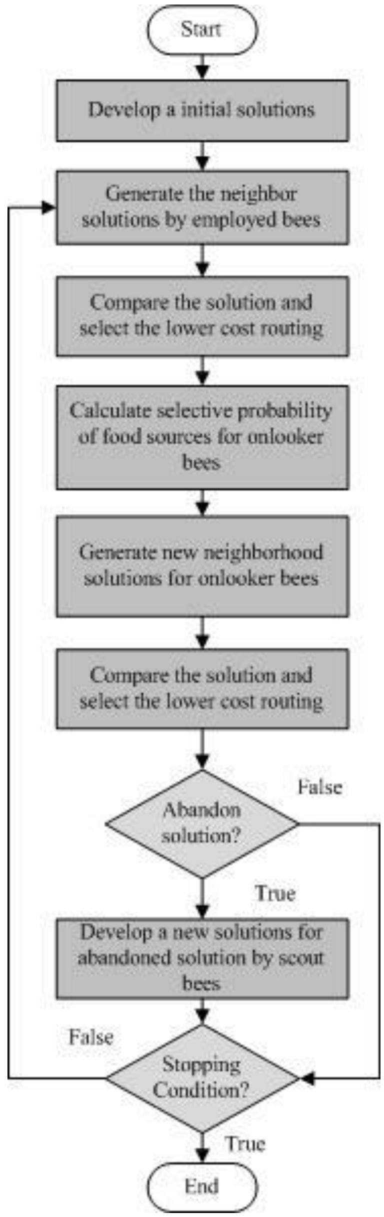


Fig 3.7 Process of Artificial Bee Colony (ABC) Algorithm

3.3.3 Solution Representation

3.3.3.1 Search Space

The searching heuristic of ABC algorithm is to consider the whole routing path of different customers for sequential delivery schedule. The dimension variables of each routing path are inter-correlated and identical, which are not able to adopt original ABC algorithm that aims to resolve continuous numerical problems.

The dimension of the routing plan is $N + K - 1$. For instance, $X = \{0,3,5,0,1,2,4,0,0,0\}$ describes that the vehicle start from depot 0 and ends at depot 0. The first route starts from depot 0 to serve customer 3 and 5, and then returns to the depot. This sequence will stop until last customer is served. The searching heuristic releases the restrictive consideration for only feasible solution to conduct a wide range searching of all possible solution by swarm algorithm.

3.3.3.2 Initial Solution

In original ABC algorithm, the initial paths are constructed from continuous variables within a range. However, the dimensional variables are discrete and identical in value to represent the specific customer. In order to facilitating the arrangement of feasible solution for VRP, each set of solutions generates randomly as initial food sources. Each customer will randomly be assigned to the k vehicle route. The capacity of vehicles is considered at initial stage to provide a feasible solution for each food source.

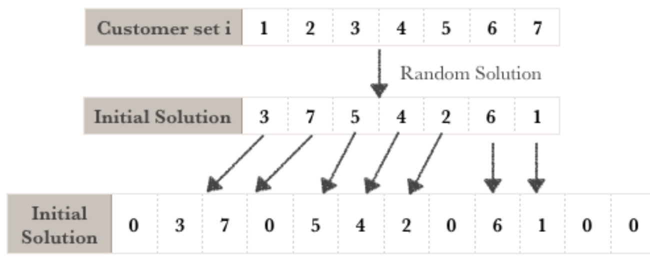


Fig 3.8 The construction of Initial solutions

The employed bees are assigned to each food source to generate the next neighbor solution.

3.3.3.3 Employed Bee Phase - Neighborhood Operators

Neighborhood operators of ABC heuristic, which are performed by employed bees, are used to create a new solution. Several neighbor operations of ABC algorithm have been proposed, which aim at increasing the robustness of searching of neighbor solution by iteration. The set of pre-defined neighborhood operators is applied once for each food sources solution.

(A) Neighbor Search - Swap Operator

This operator selects a randomly position in each food source to apply the swapping for two customers positions i, j.

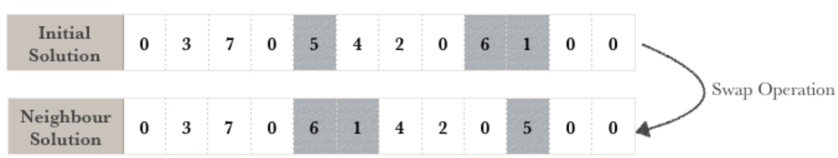


Fig 3.9a Neighbor Search - Swap Operator

(B) Neighbor Search – Insert Operator

This operator chooses positions i, j. Customer i is extracted from position i and inserted in position j.

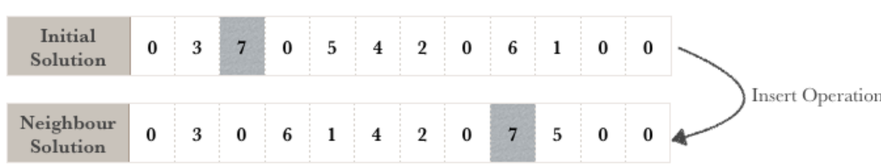


Fig 3.9b Neighbor Search - Insert Operator

(C) Neighbor Search – Reverse Operator

This operator selects a range at random, which is smaller than the size the dimension, and applies reverse order to the selected region.

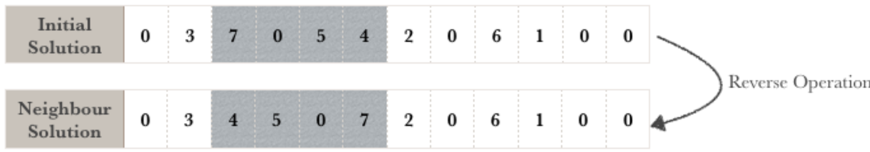


Fig 3.9c Neighbor Search - Reverse Operator

(D) Neighbor Search – Local Search

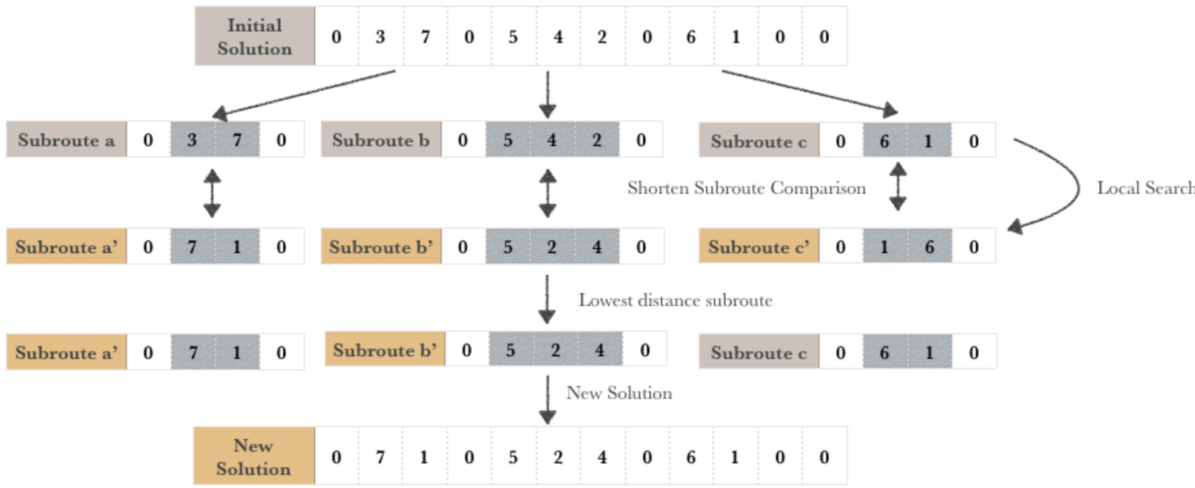


Fig 3.9d Neighbor Search - Local Search

In each neighbor search, the algorithm will compare the cost between the routes before each neighbor search. If the lower distance has been found after neighbor operation, the new route will replace the previous one. Otherwise, the route will not be changed.

3.3.3.4 Onlooker Bees Phase - Selection of Food Sources

The information provided by employed bees will be share in hive with onlooker bees. The onlooker bees will select the food sources based on the abundance of the nectar. In ABC algorithm, the selection rule is according to the probability of each food source and is compute by roulette wheel mechanism.

3.3.3.5 Scout Bee phase – Abandoning the Exhausted Food Sources

Food source will be consumed from time to time until it exhausted. The employed bee will abandoned the exhausted food sources, and transfer to scout bees for food sources exploration. In ABC algorithm, each food source will be processed with neighbor search. If there is no updated from several iterations, the corresponding food source will be abandon and establish a new solution in scout bee phase.

3.3.4 Artificial Bee Colony Algorithm (ABC) using Map Reduce Approach

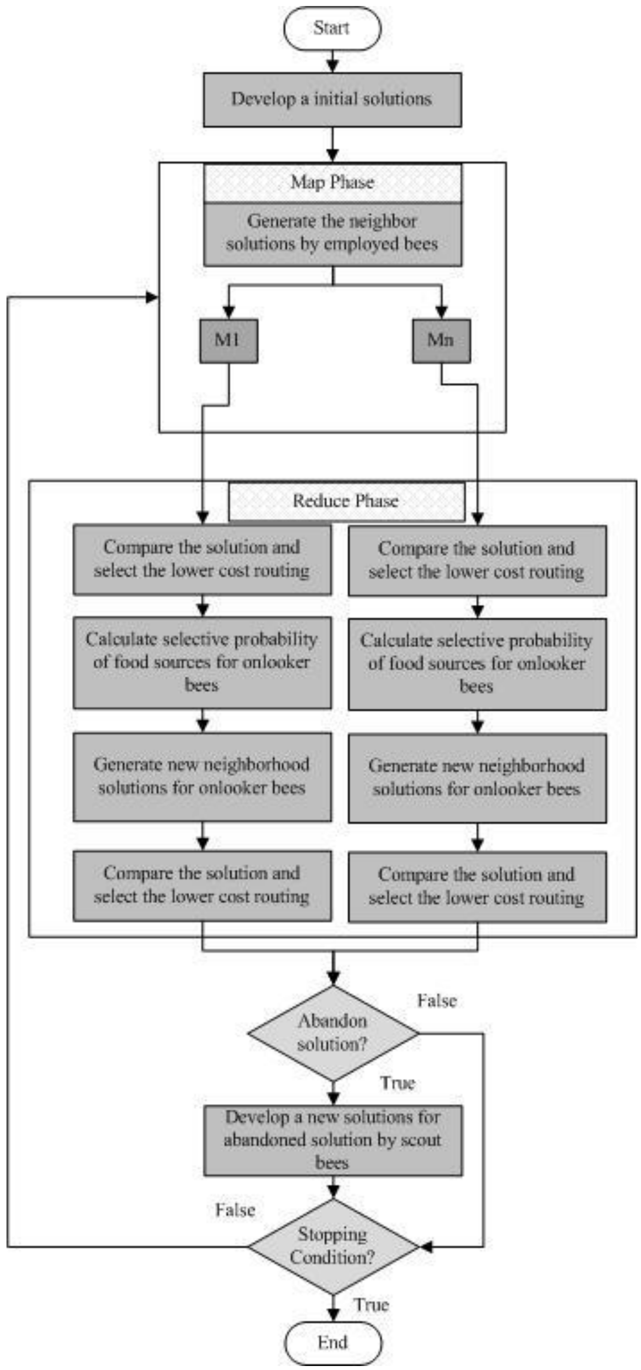


Fig 3.10 Process of Artificial Bee Colony (ABC) Algorithm using Map Reduce Approach

3.3.4.1 Map Reduce Approach

Map Reduce approach is a framework design for data mining with several clusters in order to accelerate the computation speed and the robustness of heuristic approach. The process of Map Reduce Approach under VRP is described in figure 3.10. According to the figure 3.11, the employed bees phases will become the Map Phase, which explore one solution into multiple solutions for neighbor operations to increase the searching size of each food sources. The figure 3.12 illustrates the process of Reduce Phase, which selected the optimal solution from the solutions of Map Phase.

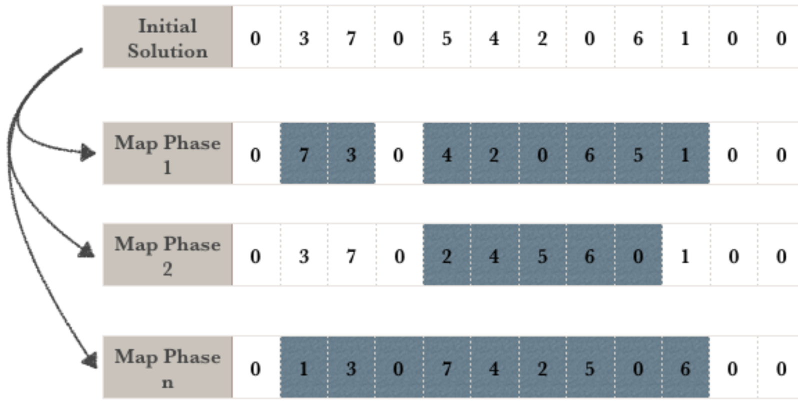


Fig 3.11 Employed Bees Phase with Map Approach

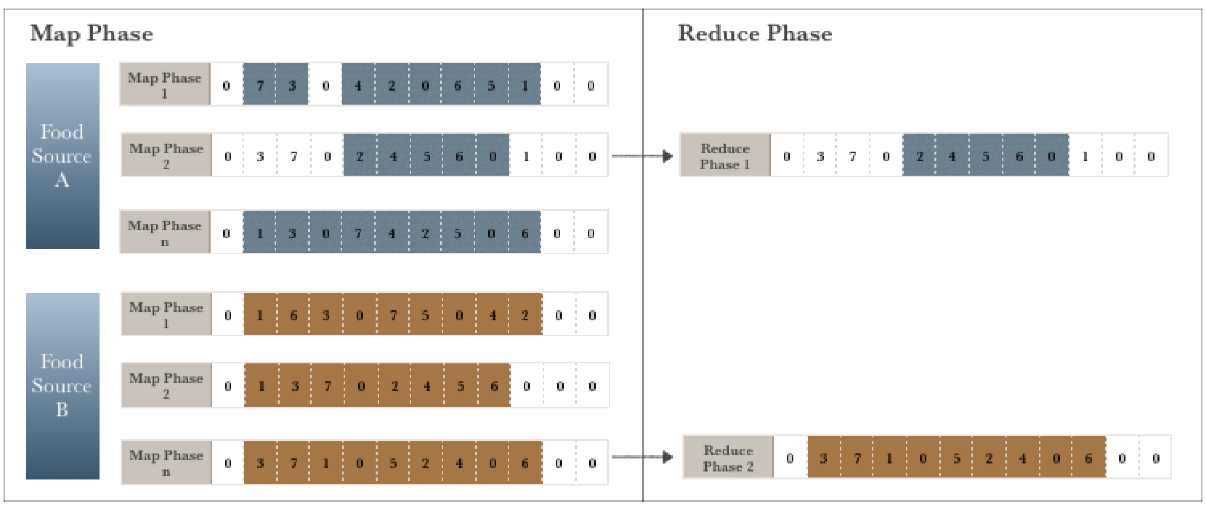


Fig. 3.12 Scout Bees Phase with Reduce Approach

Limitation

This research is to combine the image processing and VRP. The transferring speed and requirement should consider in both algorithms. In addition, the current object detection approach and background subtraction may not be feasible in Hong Kong situation. The modified version may not be appropriate in other research scope, as the objective function for traffic image processing modified as traffic density estimation rather than object identification. This approach may be limited for VRP environment.

Continue on extra page if required.

V – Project Supervisor's Comments (Stage 1)

The Project Supervisor is to make comments and provide constructive feedback on the work done by the student thus far with reference to:

- Communication with the Student
- Problem Formulation (II)
- Looking for References / Literature Review (III)
- Methodology (IV)

Kam has reported his progress regularly and has a very good commutation with supervisor through email or what's app. During summer, he has attachment in Perfect Green, a green service provider, to collect information for his FYP. Therefore, he can realize the problem during the summer attachment and generalize the problem by studying the related literature.

He has reviewed over 30 paper/books in his literature review. He conducted comprehensive literature review in big data and vehicle routing problem.

Regarding methodology, he has formulated a mathematical model for the real time vehicle routing problem and proposed to use artificial bee colony and contour extraction method to analyze the road traffic through image processing. He has shown his analytical skill and originality in his study.

Supervisor's Signature: **Date:**

Continue on extra page if required.

VI - Project Workflow and Completion Schedule

1. Identify the major tasks for the completion of the project and report.
2. Estimate the time required for each task and compile a schedule for your project. It may be useful to construct a Gantt chart to show these.
3. Use this schedule in order to track the progress of your project, although it is understood that it may need to be adjusted in the light of later events.

1. **Project Identification – Big Data**
   Topic identification is the basic requirement to identify the research scope. Big Data and Big Data Analytics become the major tactics to resolve large dataset. If the dataset follows three features – Big Volume, Velocity and Variety of dataset, which will be defined as a Big Data. The storage technique and data mining process would be different than traditional approach. At this stage, widely search for the latest Big Database, such as Hadoop, and Big Data Analytics, like Map Reduce Approach become a crucial part for later part development.

2. **Project Modification – Big Data & Image Processing**
   After searching the latest technology of Big Data, review on image processing becomes the next step. In this research, traffic image or video will be used for processing. There are enormous algorithms could be applied in image and video processing. Study the appropriate algorithms is the key aspects for evaluating the feasible and suitable method to retrieve traffic information.

3. **Image Processing**
   Most of the image-processing focus on object detection. Analysis and combination of knowledge are required to adopt in our research, since our proposed algorithm would integrate different image processing skills to fit for Hong Kong Situation.

4. **Video Processing**
   The studies on Video processing most focus on foreground subtraction. The Hong Kong Transportation Department provides image and video as a traffic reference, and thus, studying the possibility in background subtraction by video processing is also a important part for proposed image processing algorithm.

5. **Proposed Image Processing Algorithm**
   To cope with the calculation speed of VRP, the current image-processing algorithm may not be feasible to adopt in RTVRP. In our methodology, the workflow of proposed image processing has been developed.

6. **Consultation**
   Consulting supervisor could let me to understand any mistake or error that would be found in my research or any misconception towards algorithm design and basic setting of VRP.

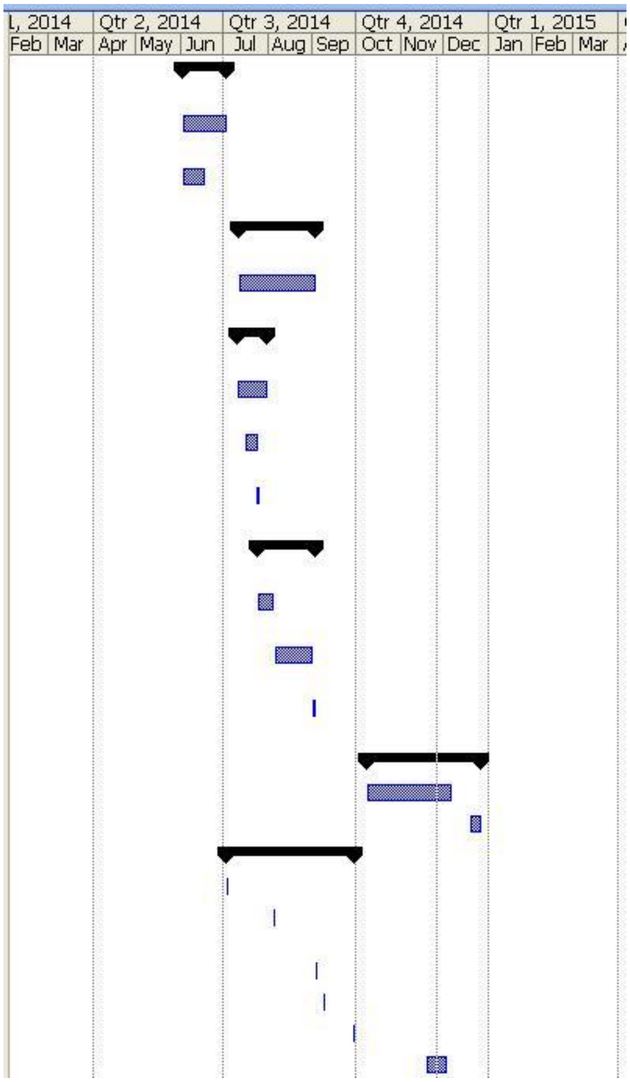
7. **Summer Internship – Case Company**
   As the RTVRP will be proposed to case company, it is important to study the current logistics problem in Hong Kong.

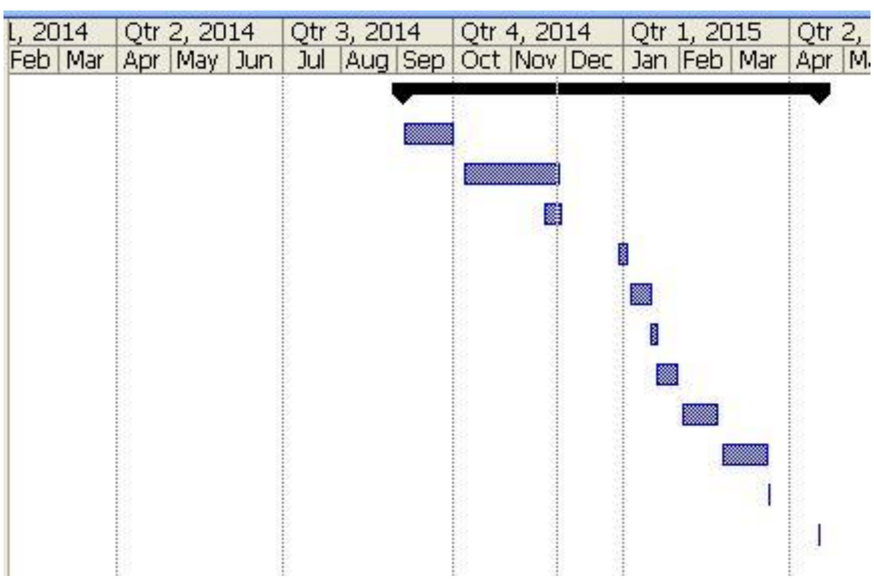
8. **Workbook**
   This is a simple draft for FYP to let my supervisor to have a brief understanding on my research study.

9. **Final Year Report and Presentation**

\

		Task Name	Duration	Start	Finish
1		**Project Identification - Big Data**	**23 days**	**Mon 6/2/14**	**Wed 7/2/14**
2		Study on Big Data Analytics Journal Articles	23 days	Mon 6/2/14	Wed 7/2/14
3		Problem Identification - Case Company	12 days	Mon 6/2/14	Tue 6/17/14
4		**Project Modification - Big Data & Image Processing**	**39 days**	**Fri 7/11/14**	**Tue 9/2/14**
5		Big Data Architecture	39 days	Fri 7/11/14	Tue 9/2/14
6		**Image Processing**	**15 days**	**Thu 7/10/14**	**Wed 7/30/14**
7		Study on Image Processing Technology Journal Articles	15 days	Thu 7/10/14	Wed 7/30/14
8		Image Processing Programming - Histogram of Oriented Gradients (HOG)	6 days	Wed 7/16/14	Wed 7/23/14
9		Investigating the current Limitation of Image Processing under VRP	3 days	Wed 7/23/14	Fri 7/25/14
10		**Video Processing**	**29 days**	**Fri 7/25/14**	**Tue 9/2/14**
11		Study on Video Processing Technology Journal Articles	6 days	Fri 7/25/14	Sun 8/3/14
12		Video Processing Programming - Background Subtraction	20 days	Tue 8/5/14	Sun 8/31/14
13		Investigating the current Limitation of Video Processing under VRP	3 days	Sun 8/31/14	Tue 9/2/14
14		**Proposed Algorithm**	**58 days**	**Wed 10/8/14**	**Fri 12/26/14**
15		Proposed Image Processing Algorithm - Phase 1	43 days	Wed 10/8/14	Fri 12/5/14
16		Proposed Image Processing Algorithm - Phase 2	6 days	Fri 12/19/14	Fri 12/26/14
17		**Consultation**	**64 days**	**Thu 7/3/14**	**Mon 9/29/14**
18		1st Consultation - Big Data	1 day	Thu 7/3/14	Thu 7/3/14
19		2nd Consultation - VRP + Traffic (using Traffic image processing)	1 day	Mon 8/4/14	Mon 8/4/14
20		3rd Consultation - Review on Image Detecion Algo	1 day	Wed 9/3/14	Wed 9/3/14
21		4th Consultation - Review on Background Subtract	1 day	Mon 9/8/14	Mon 9/8/14
22		5th Consultation - Proposed Image Processing Algo	1 day	Mon 9/29/14	Mon 9/29/14
23		Summer Internship - Case Company	10 days	Wed 11/19/14	Tue 12/2/14
24		**Final Year Project Report**	**162 days**	**Thu 9/4/14**	**Fri 4/17/15**
25		Background	19 days	Thu 9/4/14	Tue 9/30/14
26		Literature Review	38 days	Mon 10/6/14	Wed 11/26/14
27		System Architecture & Algorithm Description	8 days	Wed 11/19/14	Fri 11/28/14
28		Performance Evaluation	5 days	Mon 12/29/14	Fri 1/2/15
29		Result and Limitation	9 days	Mon 1/5/15	Thu 1/15/15
30		Further Development	3 days	Thu 1/15/15	Mon 1/19/15
31		Revision the whole report	10 days	Mon 1/19/15	Fri 1/30/15
32		prepare for presentation - powerpoint	15 days	Mon 2/2/15	Fri 2/20/15
33		Review on all presentation and report material	19 days	Mon 2/23/15	Thu 3/19/15
34		Final Draft for FYP	1 day	Fri 3/20/15	Fri 3/20/15
35		Presentation for FRP	1 day	Fri 4/17/15	Fri 4/17/15





Continue on extra page if required.

VII – The Content Page of your Project Report

Prepare a draft of the Contents page of your final report. This will serve to:

- Give a clear indication of the structure of the report.
- Help to ensure that the structure of your report is logical and that you have not missed important issues.

Your Contents page should clearly show the titles of the major Chapters or Sections and where possible, their sub-sections.

Also think about the numbering system you will use in your report. Page numbering itself will not be necessary at this stage.

VIII - Supervisor's Comments (Stage 2)

The Project Supervisor(s) is to make comments and provide constructive feedback on the work done by the student thus far with reference to:

- Project Workflow and Completion Schedule (VI)
- The Contents Page of the Project Report (VII)
- The overall progress of the Project from original commencement at the beginning of Semester 1.

The project workflow and completion schedule is written in detail. The Gantt chart clearly shows the task of his FYP and duration of each task. Kam shows a good progress of his final year project. He can manage well to learn the meta-heuristics method and image processing algorithm by himself. I believe that he can complete the task on schedule.

The content page has enlisted the main chapter of the report. The Contour Extraction Method for Image Processing and artificial bee colony will be carried and it is required to state the role of big data for this FYP.

In general, I am very satisfied with the progress of the project from original commencement at the beginning of Semester 1.

Supervisor's Signature: **Date:**

Continue on extra page if required.