UNIVERSITI TEKNIKAL MALAYSIA MELAKA

FAKULTI KEJURUTERAAN ELEKTRIK

SEMESTER 2 SESSION 2015/2016

4 BEKM

BEKU 4792 FYP (FINAL YEAR PROJECT)

OBJECT RECOGNITION USING BOWS MODEL

MEMBERS	:	SOON WEI JUN	B011210124
SUPERVISOR	:	NURSABILLILAH BINTI MOHD ALI	
DATE SUBMITTED	:	20 MAY 2016	

OBJECT RECOGNITION USING BOWS MODEL

SOON WEI JUN

A report submitted in partial fulfilment of the requirements for the Degree of Mechatronics Engineering

Faculty of Electrical Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

YEAR 2016

“I hereby declare that I have read through this report entitle “Object Recognition Using BOWS Model” and found that it has comply the partial fulfilment for awarding the Degree of Bachelor of Mechatronics Engineering”

Signature	:	....
Supervisor’s Name	:	....
Date	:	....

I declare that this report entitle "Object Recognition Using BOWS Model" is the result of my own research except as cited in the references. The report has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : ....

Name : ....

Date : ....

ACKNOWLEDGEMENT

I would like to dedicate my deepest gratitude to my FYP supervisor, Mrs Nursabillilah Binti Mohd Ali for accepting me under her supervision. I had received her thoughtfulness, irreplaceable support, beneficial guidance and direction throughout my research. Her knowledge in solving problems using digital imaging processing had always been a great help for my study. Apart from that, her constructive comments and technical advices had also aided me in the completion of this research.

Besides my FYP supervisor, I would also like to give my appreciation to my family. My parents deserve a special mention for their unconditional support when I am conducting my research. I would like to thank them for being supportive not only during my FYP, but also during my entire educational career. I have no doubt that I could not have reached this far without their support and encouragement.

Lastly, I want to use this chance to thank my industrial training supervisor from ViTrox Technology Sdn Bhd, Mr Leong Kok Yeong, who had always provided me guidance and assistance even after I finished my internship. His support was also one of the elements that allowed me to complete this research and understand a lot of new things along my work.

ABSTRACT

This paper presents the findings on Object Recognition using Bag-of-Words model done by the author. For the past decades, object recognition had been used in image processing of many fields to carry out tasks such as image classifying, video searching, robot localization and optical character recognition (OCR). Among all these object recognition method, Bag-of-Words (BOWs) or Bag-of-Features (BOFs) is a very popular approaches because of its simplicity in object recognition. Bag of words model is used for documents classification where the occurrence of every word is treated as a feature to be used for classifier training. Due to the fact that BOWs model is a multi-step process, there are still many possible combination of method that has yet to be tried. Hence, in this project, researching and developing a better object recognition application in terms of feature recognition and processing speed using BOWs model will be emphasized. Several tests would be run to prove the functionality of the system and also to acquire the value for library size and threshold value that would generate the best performance for the system. The conclusion will be including the best combination of feature extractor descriptor – FAST SURF model along with recommended library size and threshold value. In the end, FAST SURF was proven to be faster than SIFT+SURF and also 1.373% more accurate than current existing system.

ABSTRAK

Kertas laporan kerja ini membentangkan penemuan Pengenalan Objek dengan menggunakan model beg-fakta yang dijalankan oleh penulis. Bagi beberapa dekad yang lalu, pengenalan objek telah digunakan dalam pemprosesan imej dalam pelbagai bidang untuk menjalankan tugas-tugas seperti klasifikasi imej, pencarian video, penyetempatan robot dan pengecaman aksara optik (OCR). Antara semua ini kaedah pengenalan objek, Beg-Perkataan (BOWs) atau Bag-Ciri (BOFs) adalah kaedah yang paling popular kerana keringkasannya dalam pengiktirafan objek. Asalnya, model Bag-Fakta digunakan dalam klasifikasi dokumen di mana setiap perkataan dianggap sebagai ciri-ciri yang akan digunakan untuk latihan pengelas. Disebabkan kenyataan bahawa model ini terdiri daripada process pelbagai langkah, masih terdapat banyak kombinasi mungkin kaedahnya yang belum dicuba dan dibicara. Oleh sebab itu, dalam projek ini, penyelidikan dan pembinaan aplikasi pengenalan objek yang lebih baik dari segi ketepan pengenalan ciri dan kelajuan pemprosesan menggunakan model Beg-Fakta akan diberi tumpuan dalam kertas laporan kerja ini. Kombinasi terbaik untuk pengekstrakan penghurai ciri - model FAST SURF bersama dengan saiz perpustakaan dan nilai ambang yang disyorkan akan dimasukkan dalam bahagian Kesimpulan keras laporan ini. Pada akhir eksperimen, FAST SURF telah dibukti bahawa lebih cepat daripada SIFT+SURF dan juga 1.373% lebih tepat daripada sistem yang sedia ada.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATION

BOG	-	Bag of Visual Graph
BoLSg	-	Bag of Local Sub-graph
BOW	-	Bag-of-Words
BoWSPM	-	Bag-of-Words using Spatial Pyramid Matching
CAD	-	Computer Aided Diagnosis
CPM	-	Cell Pyramid Matching
DoG	-	Difference of Gaussians
gSpan	-	Graph-based Substructure Pattern
HEp-2	-	Human epithelial type 2
HIK	-	Histogram Intersection Kernel
LLC	-	Locality-constrained Linear Coding
MKL	-	Multi-Kernel Learning
ORB	-	Oriented FAST and Rotated BRIEF
PCA	-	Principal Component Analysis
RBF	-	Radial Basis Function
SIFT	-	Scale-Invariant Feature Transform
SPM	-	Spatial Pyramid Matching
SPR	-	Spatial Pyramid Representation
SURF	-	Speeded Up Robust Features
SVM	-	Support Vector Machine

CHAPTER 1

INTRODUCTION

1.1 Motivation

It is to say that vision is perhaps the best senses that a human can possesses. It is often said that "Image is worth a thousand words". It is true that by simply taking a look at an object, a lot of information can be extracted from it. Over the decades, human are trying to develop machine eyes or machine vision to not just mimic our vision, but also overcome the limitation of human eyes. As proven by scientists, human eyes are limited to only the visual band of Electromagnetic (EM) spectrum. According to common scientific faces, a normal human eye will be able respond to EM radiation with wavelengths from about 390 to 700 nm [1]. In terms of frequency, this is equal to a band in the vicinity of 430 to 770 THz [1]. Machine vision on the other hand, is able to cover almost the entire spectrum. Hence their application such as X-ray imaging used for medical purpose, pest inspection on fruits using infrared imaging and night vision tool for the military [2][3][4].

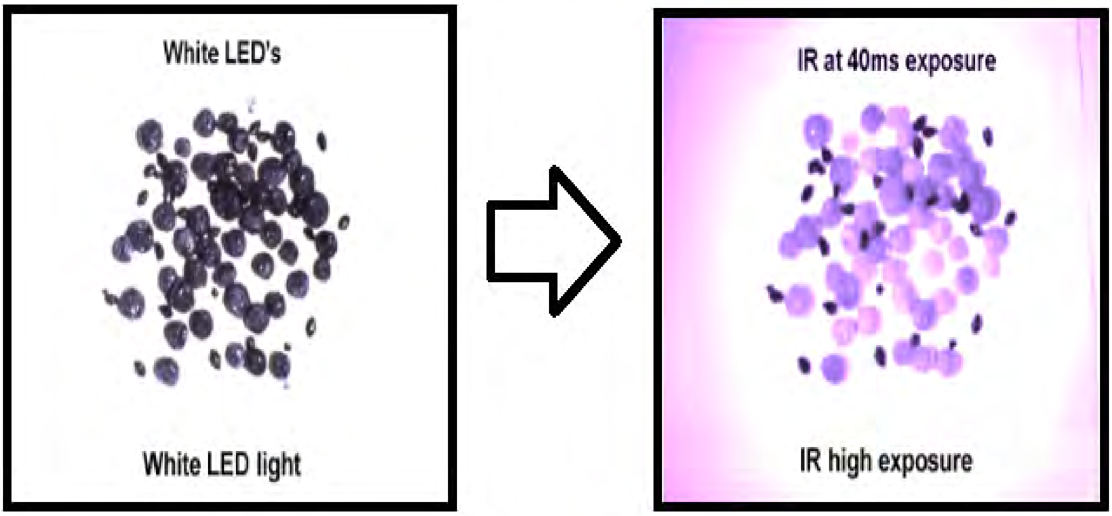

Figure 1.1: Infrared Vision for Bugs Detection

The BOWs model or BOFs model had always been the most popular representation methods for object recognition. It had been used in various computer vision tasks such as video searching, texture recognition, image classification and robot localization [5]. The main reason of this method is so widely used is because of its simplicity and its reliability [6]. The basic concepts of the way the model function is actually comparing the features of what the model "saw" with what it "knows", then it classify it accordingly to what it thinks the test object look alike from its "dictionary". It is actually like how human categorize things. First the common features of the training object were learnt and jotted down, then when a new test object was presented, it will be classify according to which common characteristic it share the most with the class that is presents.

This method could be referred as a wide combination of image processing techniques. Combining every different detection method, detection method and classification method, BOWs could produce similar results of object recognition but with different processing time and accuracy.

The BOWs model possesses the advantages as follow:

- By clustering the features of the local image into visual words, it helps to save a lot of computational time and memory if compared with storing and matching each feature of the local image individually [7].

- The process of gathering set of local features derived from an image into one global histogram of visual words allows the users to apply statistical learning algorithms, on them [7].

- The BOW representation could be classed (category) independent which will reduce the amount of supervision required in the classification assignment [7].

BOW is said to be the promising model of the future applications because with a robust and accurate object recognition system, a lot of things could be achieved without so much of lifting a finger. Law enforcers could find a stolen car within days, locate the last location of a known terrorist within hours or even send tickets to car owners that were speeding. Agriculture workers could identify the stage of the fruits without checking on them daily. Forest fire or house fire could be identified as soon as possible with the help of aerial vehicles or devices to minimise the damage.

Hence, the motivation to carry out research on this so that any improvement or suggestion could be made to bring the state-of-the-art to a better level that is possible to be implemented to most of real life applications.

1.2 Problem Statement

In order to make this project a success, the key step is to identify the problem to be addressed with the existing Bag-of-Word recognition model. The main problem of the BOW model is that the accuracy of the existing system is not accurate enough, which generate accuracy of averagely 94.7%. Ideally, a recognition system is expected to achieve 100% accuracy. Hence, in this project the BOW model will be put under research and development to solve the issue of such considerably low accuracy.

There is in fact several algorithms designed lately that achieved a recognition accuracy of nearly 95%. Yet it is not consider as the perfect solution as in order to achieve that kind of high accuracy performance, the researchers usually had to make the hard choice and decrease its processing speed. In other word, the problem of current Bag-of-Word recognition system possessing low processing speed will be addressed in the project as well.

Last but not least, a problem to be considered is that the feature detection ability. Algorithms with low feature ability cause detection drop-out, which means that the feature retrieval method was unable to collect the feature information of the images [1]. There are also cases where the feature detection is not functioning properly and detect the background of the image as part of the object's features. Hence, this project will also be addressing the problem of features drop-out of current Bag-of-Word recognition model.

All of the problems stated are those to be overcome within this project with proposed solution – FAST SURF Bag-of-Word recognition model.

1.3 Objectives

The main objectives of carrying out this project are:

1. To develop an algorithm for object recognition based on BOWs model
2. To identify the best feature extraction and detection based on the object recognition.
3. To evaluate the system performance in terms of speed and accuracy.

1.4 Scope

The scopes of the project are listed as follow:

1. The object categories are limited to detect and recognise cats, airplanes, tables, human and motorcycle.
2. The algorithm developed will not take into account on the environment of the object that might causes the object to be difficult to be detected and classified.
3. The vocabulary of the database development is limited to 100 images per class.
4. The developed system is only applicable to static images.
5. The system's performance analysed only included accuracy, error rate, false positive rate, true positive rate, precision and processing time of the system.

1.5 Report Outline

This report is divided into 5 chapters. Basic introduction on the background of this project will be covered in the Chapter 1 – Introduction of the report. Next, clear explanation on algorithms and techniques used will be presented in the Chapter 2 – Literature Review. The algorithm that would be discussed included the main Bag-of-Words process, several extractors descriptors such as SIFT, SURF, and ORB. Apart from that, techniques that were implemented by other authors such as the Principal Component Analysis (PCA) and Confusion Matrix will also be looked into in this section. At the end of Chapter 2, an analysis will be done and benchmarking will be carried out to propose a better solution for the recognition system. Explanation of the process flow and experiment to be done will be covered in the Chapter 3 – Methodology, followed by the results acquired from the tests and experiments together with analysis on the results that will be delivered in Chapter 4 – Results. Last but not least, at the end of all the tests and analysis, conclusions will be written to conclude the findings of this project in the final chapter, Chapter 5 – Conclusion.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

There are several researches done on how to improve the performances of image classification using Bag-of-Words models. It is wise for every researchers to study about them to know the state-of-the-art of the technique to develop the improvement required and also to avoid repeating research that is already been done by any other researchers.

Four journals of research were selected to be studied and the methods used were identified. Analysis was made to benchmark the ideas of improvement in order to combine the advantages of few methods and eliminating their drawbacks.

2.2 Bag-of-Words

Object recognition using Bag-of-Words is a strategy that was inspired by the text retrieval community and enables efficient indexing for local image features.

Methods for indexing and efficient retrieval with text documents are mature, and effective enough to operate with millions or billions of documents at once. Documents of text contain some distribution of words, and thus can be compactly summarized by their word counts (known as a bag-of-words). Since the occurrence of a given word tends to be sparse across different documents, an index that maps words to the files in which they occur can take a keyword query and immediately produce relevant content.

Researchers then implement this idea to carry out object recognition. The basic flow of the model is very straight forward and similar to document classification using Bag-of-Words. First, a training class has to be built by using several training images. Features will be detected by using different feature descriptor of choice and extracted. Extracted features are then put into a bag of features that can be used to represent that class of object. Few bags are formed together to create a category models or classifier.

Then, a test image which is not in those bags is passed through the systems. Features of the test image are extracted as well and to be used to match compare with the existing bag-of-features. Finally, the best category or class is chose to represent that image according to different matching algorithm. An illustration of the system explained can be seen clearly in Figure 2.1:

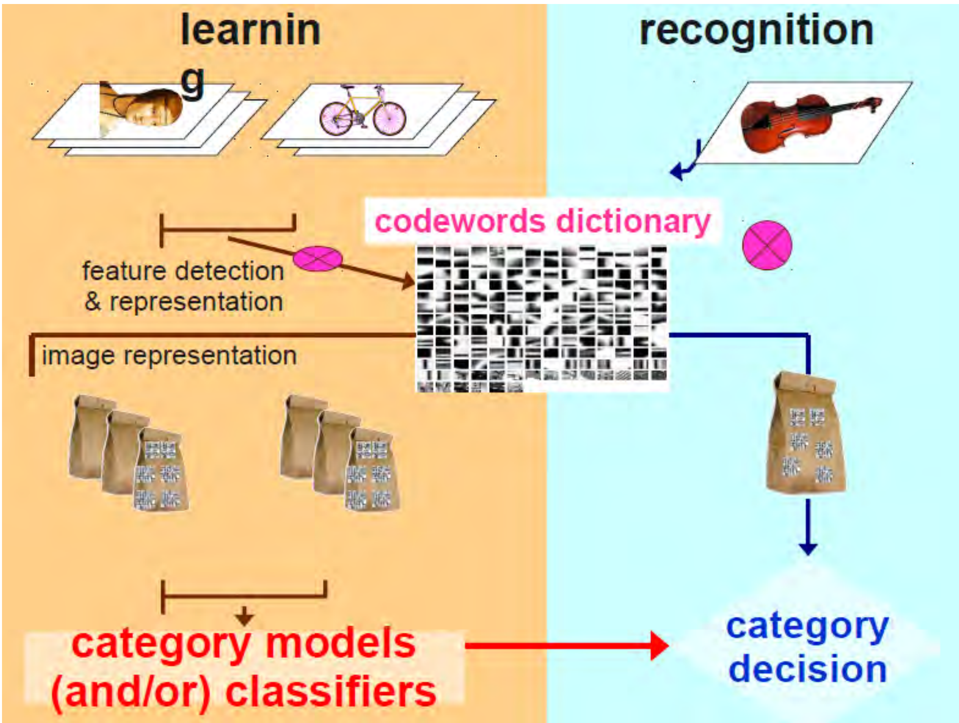

Figure 2.1: Basic Object Recognition Process using Bag-of-Words

2.3 BOW Model using Scale-Invariant-Feature-Transforms (SIFT)

In order for a BOW recognition system to work, keypoints or feature points of an object are required. For any object within an image, many “special points” could be extracted to provide a “feature description” of the object. This description can later be processed to trace the item in an image having various other objects.

SIFT is one of the detector or also known as extractor where its image features offer a set of features of an object that are not affected by many of the problems faced such as different object rotation and scaling between library and targeted images, making SIFT features very robust to the effects of noises in the image.

The SIFT method uses an image and transforms it into a large collection of local feature vectors. Each of these feature vectors is invariant to any scaling, rotation or translation of the original image. The features extraction is done through four different stages:

1. Scale-space extrema detection

   This is a filtering stage where a scale space is used to determine locations and scales that are identifiable from different views on the same object. The scale space is defined by the function:

$$\mathrm{L(x, y, \sigma)} = G(x, y, \sigma) * I(x, y) \quad (2.1)$$

Given:

* is the convolution operator.

$G(x, y, \sigma)$ = variable-scale Gaussian function.

$I(x, y)$ = the input image.