Discovering Semantic Attributes from Visual Descriptors

Lim Phei Chin

A thesis submitted
in fulfillment of the requirement for the degree of Doctor of Philosophy
(Text and Data Mining)

Faculty of Computer Science and Information Technology
UNIVERSITI MALAYSIA SARAWAK
2016

UNIVERSITI MALAYSIA SARAWAK

Grade: ____________

Please tick (√)

Final Year Project Report ☐

Masters ☐

PhD ☑

DECLARATION OF ORIGINAL WORK

This declaration is made on the ......12th......day of......April......2016.

Student's Declaration:

I Lim Phei Chin (09011440) Faculty of Computer Science & Information Technology (PLEASE INDICATE STUDENT'S NAME, MATRIC NO. AND FACULTY) hereby declare that the work entitled, "Discovering Semantic Attributes from Visual Descriptors" 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.

24/6/2016

Date submitted

Lim Phei Chin (09011440)

Name of the student (Matric No.)

Supervisor's Declaration:

I Narayanan Kulathuramaiyer (SUPERVISOR'S NAME) hereby certifies that the work entitled, Discovering Semantic Attributes from Visual Descriptors (TITLE) was prepared by the above named student, and was submitted to the "FACULTY" as a * partial/full fulfillment for the conferment of Doctor of Philosophy (PLEASE INDICATE THE DEGREE), and the aforementioned work, to the best of my knowledge, is the said student's work

Prof. Dr. Narayanan Kulathuramaiyer
Fakulti Sains Komputer dan Teknologi Maklumat
Universiti Malaysia Sarawak

Received for examination by: ____________________

(Name of the supervisor)

Date: 22/6/16

I declare this 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 declared that this said Project/Thesis shall be placed officially in the Centre for Academic Information Services with the abide 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 digitise the content to 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 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's signature ____________ (Date) 22 June 2016

Supervisor's signature: ____________ (Date) 22/6/16

Pro. Dr. Narayanan Kulathuramaiyer
Fakulti Sains Komputer dan Teknologi Maklumat
Universiti Malaysia Sarawak

Current Address:

____

____

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 was duly prepared by The Centre for Academic Information Services]

ACKNOWLEDGEMENT

First of all, I would like to take this opportunity to express my gratitude towards my main supervisor, Prof. Dr. Narayanan Kulathuramaiyer for his invaluable guidance and support throughout this research. Thanks to my co-supervisors, Dr. Dayang Nur Fatimah Awang Iskandar and Dr. Chiew Kang Leng for their priceless advices and recommendations.

I would also like to acknowledge the scholarship provided by Zamalah Pascasiswazah UNIMAS. Not forgetting staff and technician from Faculty of Computer Science and Information Technology for doing what they do best.

I would like to thank all the participants who kindly contributed to the surveys.

Last but not least, I would like to thank my beloved family for their encouragement and unconditional support.

ABSTRACT

The advance of digital media technologies has led to the creation of massive amounts of mix quality digital images, making conventional image-related processing applications inadequate. Research to gain insight from these massive numbers of images for education or business purposes is crucial and challenging. The challenges include image collection, analysis, organisation, visualisation and search. Research work in tackling these challenges started from the utilisation of visual content to gradually move towards using semantic features. This research has demonstrated the design and the creation of semantic attributes from visual descriptors that are useful to describe objects. In order to achieve the goals and the objectives of this research, a three-phase framework was proposed. The first phase consists of the sequential steps in data sampling and collection where three datasets of different sample sizes (292, 593 and 610 images) and the perspectives (in generic and specific domains) have been obtained. Images have been transformed into a matrix format before pre-processing for the main statistical analysis. The average reductions of 14% and 20% of image and visual descriptors have been achieved using univariate descriptive statistics. Exploratory factor analysis has been conducted iteratively in order to discover patterns from the pre-processed data before a finalised factor structure is obtained in the second phase (a 3-factors, 4-factors and 2-factors for three datasets respectively). The third phase mainly assimilates the patterns discovered for evaluation and interpretation. The ranking of images on factors has been quantitatively experimented and translated into semantic attributes. Three visual descriptors (extent, solidity and circular variance) loaded onto a semantic attribute (*Symmetrical*) have been experimented and successfully mapped to three instances of butterfly wings from an ontology. Instance characterisations from the ontology have also been transferred onto semantic attributes for object description at the abstract level.

Penemuan Sifat-sifat Semantik dari Deskriptor Visual

ABSTRAK

Perkembangan dalam teknologi media digital membawa kepada penciptaan imej digital yang berkualiti campuran dalam jumlah yang besar setiap hari, menyebabkan aplikasi pemprosesan imej dengan kaedah konvensional dalam kekurangan yang ketara. Penyelidikan untuk mendalami maksud daripada jumlah imej yang besar, tidak kira untuk tujuan pendidikan atau perniagaan adalah penting dan mencabar. Cabaran yang dimaksudkan termasuk pengumpulan imej, analisis, organisasi, visualisasi dan carian. Kebanyakan penyelidikan untuk menyahut cabaran ini bermula dengan penggunaan kandungan visual dan secara beransur-ansur menyusuri ke arah berciri semantik disebabkan isu jurang semantik. Kajian ini menunjukkan hasil reka bentuk dan penciptaan sifat-sifat semantik dari deskriptor visual di mana ia berguna dalam penerangan objek. Rangka kerja tiga fasa telah dicadangkan dalam usaha untuk mencapai matlamat dan objekti penyelidikan ini. Fasa pertama merangkumi langkah-langkah berjujukan dalam persampelan dan pengumpulan data di mana tiga set data yang berlainan dalam saiz (292, 593 dan 610 imej) dan perspektif (domain generik dan khusus) telah diperolehi. Imej-imej telah diubah kepada format matriks untuk pra-proses sebelum diguna dalam analisis statistik utama. Purata pengurangan imej dan deskriptor visual sebanyak 14% dan 20% telah dicapai dengan menggunakan statistik diskriptif univariat. Dalam fasa kedua, analisis faktor berasaskan penerokaan telah dijalankan secara berulang-ulang agar corak daripada data pra-proses sebelum struktur faktor dimuktamadkan (3-faktor, 4-faktor dan 2-faktor struktur bagi ketiga-tiga set data). Fasa ketiga merangkumi proses mengasimilasikan corak yang ditemui untuk penilaian dan tafsiran. Kedudukan imej pada faktor-faktor telah diuji melalui eksperimen berasaskan kuantitatif dan diterjemahkan kepada sifat-sifat semantik. Tiga deskriptor visual (extent,

solidity dan circular variance) yang dikumpulkan sebagai salah satu sifat semantik (Symmetrical) telah berjaya dipetakan kepada tiga ciri-ciri sayap rama-rama dari ontologi. Pencirian terperinci dari ontologi juga telah dipindahguna ke atas ciri-ciri semantik untuk penerangan objek di peringkat abstrak.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1. Introduction

As the focus of this thesis is in the field of computer science, the definition of *semantic gap* which is made famous by Smeulders et al. (2000), who reviewed 200 references in content-based image retrieval with over 5000 citation as of today as

> "*…lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation…*"

is adapted in this research. The approaches used in solving the semantic gap issue in image-related research are grouped into two categories based on the approach used. Top-down approach is originated from information system field and favoured by researcher working on textual information where ontologies are the famous knowledge representation scheme used. Bottom-up approach is originated from computer vision field and hence, image processing methods are usually used to extract visual descriptors from image content. Due to limitations in each approach in solving the semantic gap problem, the trend gradually shifted to object descriptions rather than object detection.

Although the focus of current research shifted from object detection, the various image visual descriptors created for object detection are inevitably important in forming object descriptions. Semantic attributes are created for describing an object from patterns in visual descriptors groupings or factors, which are determined using exploratory factor analysis. These factors are hypothetical variables which explain why a number of visual descriptors are correlated with each other – it is because they have one or more visual descriptors in common, and provides better insight about the original data.

1.2. Motivation

The advance of digital media technologies has drastically influenced the explosion of multimedia data (“Big data”) created at every second. On average, 2476 Instagram photos are uploaded in every second (Stats, 2015), and approximately 200,000 pictures make their way to Facebook every minute (Horaczek, 2013). If a picture is worth a thousand words, approximately 2.5 million words are required to describe the average 2476 images uploaded on Instagram every second, and this is only one out of numerous image sharing networks available. Majority of these pictures are of mixed quality and unstructured, which are disproportionate to the rich information they contain with the available volume. Hence, the process of extracting this rich and semantically meaningful information from these huge volumes of images is challenging but is essential for image-based applications.

One of the fundamental techniques in understanding and learning from huge volumes of images is to discover the natural groupings of images the application of clustering algorithms (Pan et al., 2005; Song et al., 2006; Torres et al., 2006; Liu et al., 2007; Han et al., 2007). The semantic meaning in an image has shifted from representing image as a whole to sub regions, thus localising the search space in image retrieval. Semantic image region clustering before image retrieval (Zhang & Chen, 2007; Liu et al., 2009) or query results clustering (Wang et al., 2007; Wang et al., 2008; Leuken et al., 2009) both achieved massive search space reduction. However, image analysis and clustering schemes are faced with a major issue of semantic gap (Smeulders et al., 2000; Smith, 2007), where the formation of clusters is solely on the basis of numerical attributes, which led to semantically different images being perceived as close to each other in the same feature space. Such challenges are believed to be imposed by the technique itself (Everitt, 1993; Jain et al., 1999; Xu & Wunsch, 2005).

Semantic clustering originated from the information system field to solve text classification problems and was first used by Gotlieb & Kumar (1968) to indicate

> "*the association measure between index terms is drawn from the vocabulary of a structured indexing system*".

The association measure is later used for grouping index terms into clusters or concepts, while a more recent definition by Lippincott & Passonneau (2009) as

> "*the incorporation of semantic/lexical information in WordNet into a clustering process*".

More recently, semantic clustering was used as indication of the adaptation of semantic relationship into a clustering algorithm (Cheng, 2008; Patino-Escarcina & Ferreira Costa, 2008; Chimlek et al., 2010; Sun et al., 2012; Bdiri et al., 2013).

Yet, initial reviews of terms such as image clustering and semantic clustering tend to be interchangeably used, especially when the problem to be solved concerns the issue of semantic gap. Semantic gap arises due to the subjectivity of human perception (which is a top-down perception problem) instead of considering semantic gap from a bottom-up perspective, as a computer vision problem. Therefore, research efforts are required to bring closer the image visual descriptors with high-level human concepts in order to create a representation that matches human perception. This research has taken up the challenge of introducing a three-phase framework to derive semantic attributes from measurable characteristics from images.

1.3. Problem Statement

When trying to solve the issue of semantic gap for a vast number of images, one would expect that the definition of image clustering and semantic clustering is interchangeable, but in fact, there is great variation among the two. Image clustering targets on solving image classification and categorisation issue, while semantic clustering focuses on the extraction and transformation of low-level image visual descriptors into semantic space before being mapped to semantically meaningful terms. An existing work succeeded in representing symbolic terms such as wing ratio and tailed-wings et cetera from mapping image clusters from feature clustering and visual knowledge acquisition (Lim & Kulathuramaiyer, 2007) on domain specific images. As single feature clustering was applied, visual clusters were limited in representing abstract perception in annotation and categorisation. Hence, the need of deriving abstract feature is crucial. This work focuses on the framework and accompanying procedures in detecting groupings of image descriptors that conform to abstract description, which are later mapped to semantic labels acquired from human, in addition to detecting some semantic structure from data. Figure 1.0 illustrates the proposed conceptual framework with three phases enclosed in bolded boxes. The inputs and outputs are denoted with rectangles while solid arrows denoted as the process flow. Note should be taken that the dotted line arrows are simply used to indicates the output from a previous phase is used as input in the next phase.

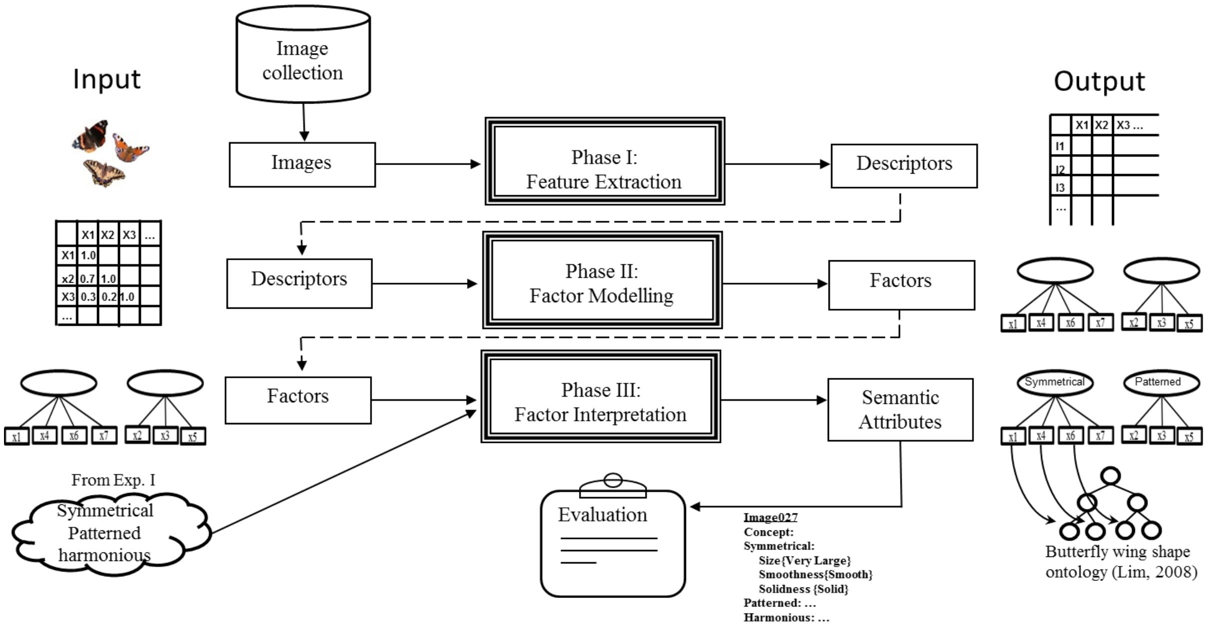

Figure 1.1: Research conceptual framework.

1.4. Research Challenges

The task of deriving, grouping of image visual descriptors, and associating with high-level concepts for object description is a challenging problem. In this work, focus is on three research challenges in describing objects by their semantic attributes which revolve around these research questions:

1. What is feature extraction? How to extract visual descriptors from object images?
2. What is exploratory factor analysis? How to extract factors from visual descriptors?
3. What is factors interpretation? How to obtain semantic attributes from factors?

This research attempts to address the challenging problem on derivation of meaningful semantic attributes in object description by scaling the scope into three research questions. The first question is addressed by employing image processing methods in object segmentation and visual descriptors extraction. For the second question, iterative processes of exploratory factor analysis are conducted in order to detect factors or constructs (in psychology) that cannot be measured directly but is inferred through the relationship (or common variance) of a subset of visual descriptors. Selection on factor extraction and rotation methods with appropriate number of factors to retain is the major concern in answering the second research question. The third research question is answered by designing and development of user studies to select common descriptive words and to associate factors to the selected descriptive words.

1.5. Objectives

This research aims to design, develop and evaluate the Symbolic-level Abstraction for object interpretation. This aim can be achieved by fulfilling the following research objectives:

1. To devise the extraction of groups of commonly used visual descriptors from sets of images.
2. To conduct, test and evaluate Exploratory Factor Analysis for summarising the pattern from visual descriptors.
3. To map groups of visual descriptors (factors) to abstract words in creating semantic attributes.

1.6. Scope

In this research domain, complete object images are used in image datasets to perform empirical evaluation of the proposed method. Image datasets are extracted from two image databases. The LabelMe (Russell et al., 2008) image databases (hereafter denoted as LM) included traced silhouette of objects and annotations by public users using the provided LM online annotation tool. The second image database was created by scanning from a book (Etsuzo & Yasusuke, 1982) without annotation information. Since object segmentation and recognition is not the focus in this work, available segmentation tools were utilised. Focus objects of ‘butterfly’ and ‘bird’ are chosen based on heuristic reasons.

Representation of image characteristics is of significance to any object description, but most visual descriptors are tailored to specific applications (Andreopoulos & Tsotsos, 2013). Each and every one of the available visual descriptor has its own strengths and weaknesses. There is no perfect visual descriptor that is able to represent the visual perception of an object, therefore, more visual descriptors are used. Since it will be difficult to include all the available image visual descriptors and this work is not meant to review an exhaustive list of image visual descriptors, available tools and code are utilised in the extraction of computationally simple shape, colour and texture visual descriptors from the extracted image datasets.

Increases or decreases of the number of images in the image datasets may affect the interpretability of factors created. Visual descriptors of additional images that fall within the existing data distribution will not affect the factors model derived. On the contrary, visual descriptors of additional images that are outliers cause the factors derived to be unstable and provide misleading interpretation. The decreasing in the number of images for a small sample size image dataset may lead to either statistical failure or less representative model being derived.

The extracted visual descriptors are used as input in the Exploratory Factor Analysis to create semantic attributes. Input structure of image visual descriptors initially represented in vector space is transformed into factor space through Exploratory Factor Analysis. Various factor extraction and rotation method are available and each has its own strengths and limitations which are reviewed in Chapter 2. Note should be taken that there is no guideline in determining the best configuration in factor analysis but there are best practices that guide the most suitable configuration for an exploratory approach.

1.7. Significance of Research

Introducing a symbolic-level abstraction on top of visual-level abstraction is crucial to derive meaningful semantic attributes for object description. This study attempts to design and derive semantic attributes that conform to human visual perception systematically to fill the semantic gap between low-level abstraction and high-level abstraction. By refining the steps in modelling factors from image visual descriptors and then associating factors to a subset of human pre-selected descriptive words, meaningful semantic attributes are discovered from images of two different objects. These semantic attributes are common yet controlled vocabularies for describing object's visual characteristics. The systematic approach lays the groundwork which is expandable to meaningful semantic attributes derivation for other object classes.

1.8. Thesis Organisation

The overall thesis structure is illustrated in Figure 1.1 where chapters are organised in the following manner.

Chapter 2 presents reviews of relevant researches to this study. These related researches include (a) semantic gap; (b) image visual descriptors; and (c) factor analysis. Summarisation of reviewed techniques and methods are discussed at the end of each section before situating the proposed study at the end of this chapter.