

WEB-SCALE VISUAL CONTENT MINING
WITH COMPUTER VISION METHODS

Dissertation Proposal

ABSTRACT

Social media provides low-cost and easy access to a large number of images with a wide variety of visual content. With this new source of large-scale image data, there is a critical demand for novel, scalable methods that can understand the information behind the visual content. For example, photos shared on social media are usually accompanied with temporal and location metadata, user profiles and textual annotations. When we integrate the visual content by user profiles or location tags, we can estimate a large variety of information. For example, we can help research in Psychology when we are interested in the human beings who take these photos and in Ecology when we are interested in the natural scenes as the visual content. These attractive properties of this new data source also bring us new challenges. Being able to efficiently processing data with advanced computer vision techniques is an important requirement. To accommodate the graphical structure of metadata in social media, we also need to develop new models and systems.

This thesis introduces three projects: a cloud based robotic system testing the feasibility of a memory and computing consuming task to work in near real time; a data mining application supporting Psychology research in gender differences in preference of color; and a natural event tracking system on a continental scale. We design and implement a cloud based system to apply these algorithms for object detection and recognition with the best precision in near real time. We test our method on an aerial autonomous vehicle. We take advantage of this new data source for psychology research. Based on the color pixels of these photos, we analyze the gender difference in color preference. We study the distribution of color pixels with respect to genders of photographers across geographic locations and content. We find strong sex differences for the predominant reddish and bluish hues, with female users uploading more photographs containing more reddish pixels and male users uploading more photographs containing more bluish pixels. Furthermore, we take Google

Street View to represent the color distribution in the environment, and compare with Flickr photos in three popular outdoor locations. We observe the overall preference of saturated color and reddish color of human compared to the environment. By mining visual content on social media websites web-scale images further benefit studies about the state of nature. We develop and evaluate three models to study ecology phenomena such as snow and vegetation coverage. First, we learn a binomial distribution to model the probability of the appearance of a natural phenomenon. Then, we compute the histogram of the confidence of each image being a positive evidence. Based on this histogram, we learn a classification model for the confidence of each user, and similarly apply this method to predict for each time and location. Finally, inspired by the study of multiple instance learning, we propose an end to end system taking a sequence of images as input and giving predictions as the final output of this holistic system. This process will happen twice, once for users and once for each day and location. The two aggregating processes are optimized simultaneously.

CONTENTS

1	INTRODUCTION	1
----------	---------------------	----------

CHAPTER 1

INTRODUCTION

Social media encourages the development of online social network, connecting billions of users sharing status and comments in text and multimedia format. These increasingly large number of multimedia files such as photographs become a new source of web-scale data with high variety and publicly available. Along with the image files and text annotations, social networking websites and mobile applications also record user profiles as well as time and location metadata of users activity. This new data source is quite noisy but is also easy accessible. It collects data constantly from a broad selection of locations on this earth. To better understand the information behind this large collection of data, there is a critical demand for novel, scalable methods to navigate the visual content together with the textual annotation and user activity. Researchers from computer vision, multimedia, data mining and machine learning become more and more interested in this interdisciplinary area, especially with the progress in computer vision and with the development of computing capacity.

When we integrate the visual content by the metadata of user activities, we will provide a very low-cost, large scale data source for researchers in many areas of social and natural science. For example, we will help research in social science when we are interested in the human behavior of those who take photos and share on social media, and in nature science

when we are interested in the natural scenes and the associated location, time and user profile. These attractive properties of this new data source also bring us new challenges. Being able to efficiently processing data with advanced computer vision techniques is an important requirement. To accommodate the graphical structure of metadata in social media, we also need to develop new models and systems.

This thesis presents our works of solving problems in three aspects to facilitate research in different areas with web-scale image data.

- In **Chapter 2** we design and implement a cloud based system to apply these algorithms for object detection and recognition with the best precision in near real time. We test our method on an aerial autonomous vehicle.
- In **Chapter 3** we take advantage of this new data source for psychology research. Based on the color pixels of these photos, we analyze the gender difference in color preference. We study the distribution of color pixels with respect to genders of photographers across geographic locations and content. We find strong sex differences for the predominant reddish and bluish hues, with female users uploading more photographs containing more reddish pixels and male users uploading more photographs containing more bluish pixels. Furthermore, we take Google Street View to represent the color distribution in the environment, and compare with Flickr photos in three popular outdoor locations. We observe the overall preference of saturated color and reddish color of human compared to the environment.
- In **Chapter 4** we present an example of web-scale images further benefit studies about the state of nature. We develop and evaluate three models to study ecology phenomena such as snow and vegetation coverage. First, we learn a binomial distribution to model the probability of the appearance of a natural phenomenon based on the number of users uploading images with and without the target phenomenon. Then, we compute

the histogram of the confidence of each image being a positive evidence. Based on this histogram, we learn a classification model for the confidence of each user, and similarly apply this method to build a histogram of user confidence in order to predict for each time and location. Finally, inspired by the study of multiple instance learning, we propose an end to end system taking a sequence of images as input and giving predictions as the final output of this holistic system. For the main challenge of multiple instance learning, we use a network to aggregate a bag of evidence. This process will happen twice, once for users and once for each day and location. The two aggregating processes are optimized simultaneously.

BIBLIOGRAPHY