

LOGO RECOGNITION USING TEXTUAL AND VISUAL SEARCH

Ximing Hou^{1,2} and Hao Shi²

¹School of Computer Science, College of Engineering and Computer Science, Australian
National University, Canberra, Australia

ximing.developer@gmail.com

²School of Engineering and Science, Victoria University, Melbourne, Australia

hao.shi@vu.edu.au

ABSTRACT

The amount of digital data transmitting via internet has reached an enormous level. In order to conduct efficient web data analysis, effective web mining tools are needed. Logos, which represent companies' brands, are highly regarded in a business world. These logos embedded in ordinary pictures could give an indication of popularity of the companies and their products in a region. Therefore, it is imperative to build a computer system to extract company logos from these pictures. In this paper, a Logo on Map (LoM) system is proposed, which consists of three modules: picture extraction module (PEM), logo matching module (LMM) and web mapping module (WMM). Only the first two modules are covered in this paper. The PEM is based on a keyword textual search while the LMM is a visual search using SIFT (Scale-Invariant Feature Transform) algorithm. The three experiments are conducted using different sets of pictures extracted from the Flickr® website. The experimental results have proven that visual search is more accurate than textual search and also demonstrated that LoM could be used to discover hidden knowledge beyond logos.

KEYWORDS

Web Mining, Logo, Flickr API, Textual Search, SIFT, Visual Search, LoM (Logo on Map).

1. INTRODUCTION

In the contemporary world, the internet as well as the mobile devices has been widely used around the world. As a matter of fact, the volume of digital data transmitted through internet has reached an extremely high level. According to a study conducted by International Data Corporation, the global digital data transmitting in 2010 has grown to 1.2 zettabytes and the number will hit at 35 zettabytes in the year of 2020 [1]. Picture, as one of the typical data type, in the virtual digital world, is playing an increasingly important role due to its better presentation and visual impression on people's mind. According to statistics on the number of pictures from the famous social networking websites such as Facebook® and Flickr®, there were 100 billion pictures having been uploaded to Facebook® by mid-2011 and 6 billion pictures to Flickr® by August in 2011 [2, 3]. Hence, it is imperative to conduct web mining on different source of pictures to analyse a large amount of data.

In this paper, a new system called Logo on Map System (LoM) has been developed for picture collection, search and matching. The background about Flickr® website and SIFT algorithm are given in Section 2. The LoM architecture is described in Section 3 while the experimental results are present in Section 4. The conclusion is drawn in the last section.

2. BACKGROUND

2.1. Flickr® Website

Flickr® is one of the most reliable and famous image hosting website which can provide adequate picture resources for research. Moreover, on Flickr® website, pictures can be viewed by millions of people and the viewer can tag [4] the pictures with their impressions on the pictures and more importantly, a considerable proportion of the Flickr® pictures has geographical location information as shown in Figure 1.



(a) A picture from Flickr®

new • york • nyc • city • skyline • sky •
line • scraper • america • landscape •
cityscape • buildings • hdr • high • dynamic •
range • isherwood • chris • fuji • finepix •
S5700 • S700 • fujifilm • breath • taking •
view • breathtaking • vertigo • skyscrapers •
rockefeller • new york • new york skyline • 紐約

(b) Tags on the picture



(c) Geo-graphical location of the picture

Figure 1. Pictures from Flickr® [6]

2.2. SIFT Algorithm

SIFT (Scale-Invariant Feature Transform) algorithm was developed by David Lowe in 1999 [6]. It is an algorithm in computer vision to detect and describe local features in images [6]. For any object in the image, there are key points called interesting points, which are always located in the high-contrast regions of the image as illustrated in Figure 2, for example, the edges. The interesting points can be extracted as the “feature description”, which does not change with the position or scale of the object. It provides a reliable method to detect object features in an image regardless of the scale, noise and illumination.

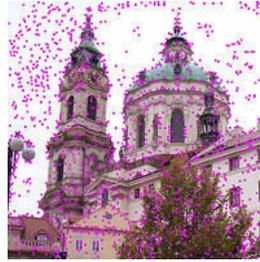


Figure 2. Pink dots as the key points of a picture [7]

During implementation of SIFT algorithm, key feature points are first extracted from the reference images [6] and then stored in a database. In order to recognize the object in the testing image, the algorithm extracts all the interesting points from the testing image, and individually compares them with the interesting points in the database, and selects the candidate matching features based on Euclidean Distance of their feature vectors. From the matching results, the subsets of the key points that agree on the object and its location, scale, and orientation in the new image are identified to be a good match. SIFT algorithm determines the consistent clusters using an efficient hash table implementation of the generalized Hough transform [6]. To wipe out the outliers, for each cluster, if more than 3 features agree on an object and its position, the cluster is subjected to further detailed model verification. The algorithm calculates the accuracy of fit and the number of false matches. Finally, the object passing all the tests is regarded as matched with high confidence [8].

3. LOGO RECOGNITION

3.1 LoM Architecture

LoM (Logo on Map) consists of three main modules as shown in Figure 3:

- Textual search, Picture Extraction Module (PEM)
- Visual search, Logo Matching Module (LMM)
- GPS-based Web Mapping Module (WMM)

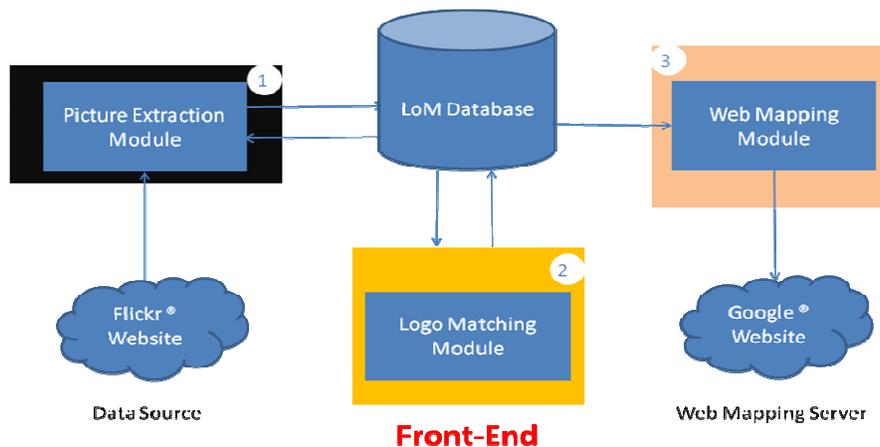


Figure 3. LoM (Logo on Map) architecture [9]

In this paper, the first two modules are presented, i.e. Picture Extraction Module (PEM) and Logo Matching Module (LMM). The Web Mapping Module (WMM) will be described in a separated follow-up paper due to the limit on the number of pages.

3.2. Picture Collection

In order to download large amount of good quality pictures, it is essential to use a reliable picture repository containing large amount of data. As discussed in the previous section, Flickr® is qualified and chosen as the data source for this project.

In Flickr®, a picture can be simply obtained from <http://www.flickr.com>. For example, once a keyword “KFC” is entered in the search box, over 100,000 pictures can be viewed as shown in Figure 4.

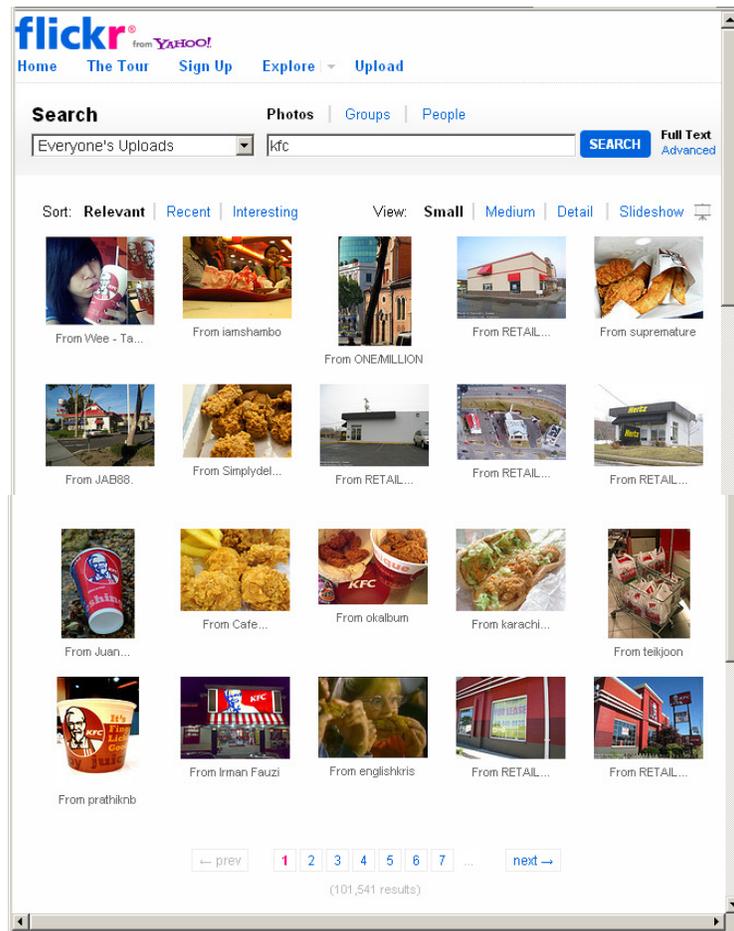


Figure 4. Pictures from Flickr® by using keywords “KFC” [8]

An individual picture is accessible via a specific URL address as shown below:

```
http://farm{farm-id}.static.flickr.com/{server-id}/{id}_{secret}_{mstzb}.jpg
```

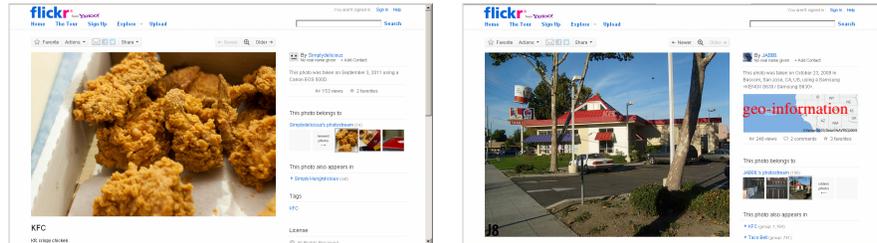
where *farm-id*, *server-id*, *id*, *secret* and *mstzb* are the parameters defining a unique picture on Flickr®.

With Flickr® API, a large quantity of pictures can be collected.

3.3 Textual Search - Picture Extraction Module (PEM)

Even though Flickr® can provide us millions of pictures, not all the pictures contain geo-information as shown in Figure 5. These two specific pictures can be downloaded via URL from Flickr® API:

- (a) http://farm7.static.flickr.com/6061/6108275577_117feb251c_m.jpg
- (b) http://farm3.static.flickr.com/2471/4038987444_ea3caccb03_m.jpg



(a) picture without geo-information

(b) picture with geo-information

Figure 5. Pictures from Flickr® with and without geo-information [5]

Another useful URL provided by the Flickr® API is to download pictures by scope as follows:

```
http://api.flickr.com/services/rest/?method=flickr.photos.search&api_key={api_key}&tags={tags}&per_page={per_page}&has_geo=1&in_gallery=true&sort=interestingness-desc
```

The parameters inside the curly braces are defined as below:

- api_key: the unique key with a combination of letters and numbers that gives developers permissions to use the Flickr® API
- tags: tags of pictures
- per_page: the number of pictures listed in a return XML
- has_geo: pictures containing the geo-location information
- in_gallery: pictures from a high quality gallery
- sort: pictures orders

After above URL request is sent, an XML file is generated by Flickr® Server which contains the following information:

```
<?xml version="1.0" encoding="utf-8" ?>
<rsp stat="ok">
<photos page="1" pages="5238" perpage="2" total="10475">
  <photo id="4054542431" owner="7434726@N07" secret="7f888c73da"
    server="2716" farm="3" title="The eyes of a fly" ispublic="1"
    isfriend="0" isfamily="0" />
  <photo id="4412545871" owner="7434726@N07" secret="47b1b54a85"
    server="4069" farm="5" title="Jumping spider" ispublic="1"
    isfriend="0" isfamily="0" />
</photos>
</rsp>
```

The unique parameters such as id, secret, server and farm, which can be used to retrieve any individual picture on Flickr®, can be easily collected and stored in a database.

Picture Extraction Module (PEM), a keyword-based textual search, is developed with Matlab® and Flickr® APIs as illustrated in Figure 6. The PEM first downloads geo-information pictures from Flickr® and then saves the extracted picture information into an LoM database. For simplicity of future reference, these extracted pictures are called PEM-pictures.

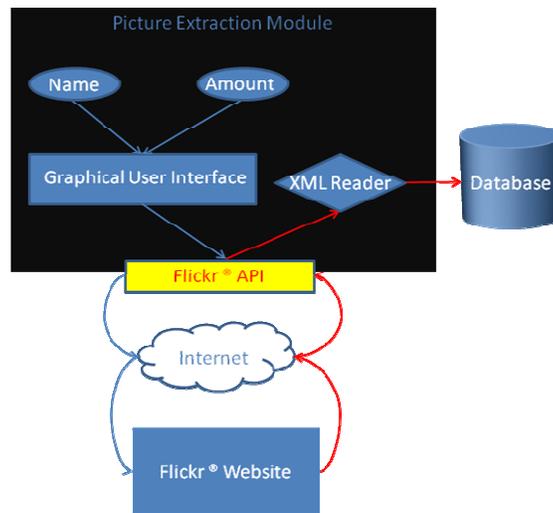


Figure 6. Picture Extraction Module (PEM) structure

For example, after a keyword search using ‘KFC’ from Flickr®, 50 PEM-pictures can be automatically collected and stored in the LoM database as shown in Figure 7.

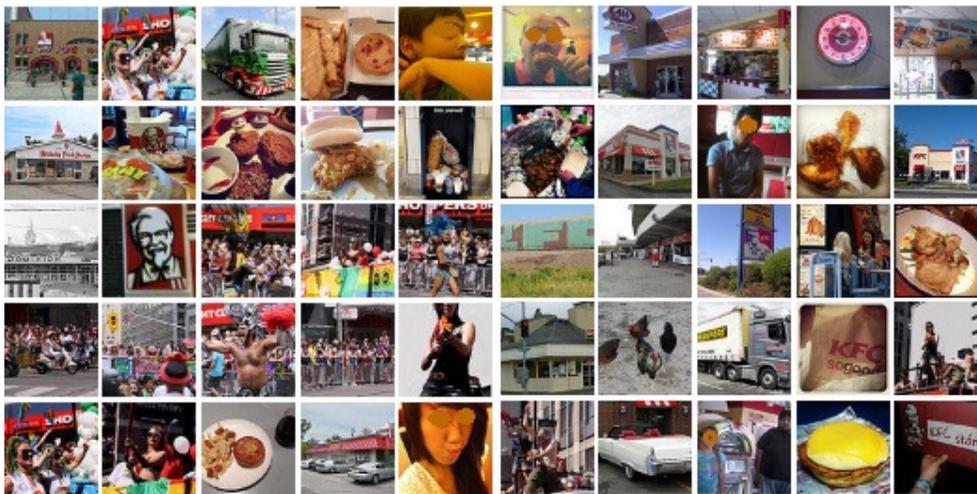


Figure 7. 50 PEM-pictures from Flickr® [5]

3.4. Visual Search - Logo Matching Module (LMM)

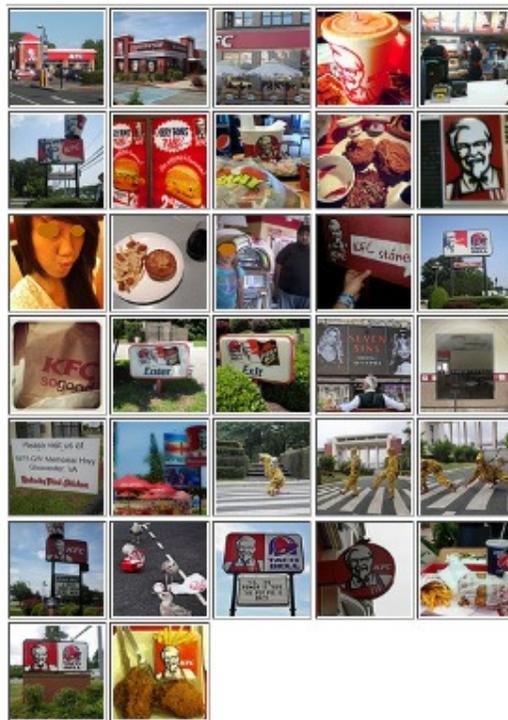
It is obvious that many PEM-pictures do not contain the KFC® logo picture. In order to screen out the PEM-pictures containing ‘KFC’ logo for this instance, Logo Matching Module (LMM), a visual-based search module, is implemented based on SIFT algorithm. The detail of LMM is shown in Figure 8. For simplicity, the pictures processed after LMM are named as LMM-pictures.



(a)



(b)



(c)

Figure 10. LMM-pictures obtained from 50,100 and 200 PEM-pictures, respectively

The recognition rates are recorded in Table 1. However, it is obvious that some pictures without logos exist in the LMM-pictures as shown in Figure 10. To calculate accuracy, LMS (Least Mean Squares) algorithm is used to obtain an error matrix. For the case of KFC® 200 pictures, Table 2 is obtained. The normalized error matrix is obtained as shown in Table 3.

Table 1. Recognition Rate (LMM)

<i>KFC</i>	<i>Logo Recognized</i>	<i>Percentages (%)</i>
50	8	16
100	15	15
200	32	16

Table 2. The error matrix

<i>200 PEM-pictures</i>	<i>Predicted Logo</i>	<i>Predicted Non-Logo</i>
<i>Logo picture</i>	18	8
<i>Non-logo picture</i>	14	160

Table 3. The normalized error matrix (%)

<i>200 PEM-pictures</i>	<i>Predicted Logo</i>	<i>Predicted Non-Logo</i>
<i>Logo picture</i>	9	4
<i>Non-logo picture</i>	7	80

According to Table 3, the accuracy of PEM can be calculated as below:

$$Accuracy = (0.09+0.04)/1.0 = 13\%.$$

For LMM, its accuracy, specificity, precision and recall of the system are calculated as follows:

$$Accuracy = (0.09+0.80)/1.0 = 89\%.$$

$$Specificity = 0.80/(0.035+0.80) = 95\%$$

$$Precision = 0.09/(0.09+0.035) = 72\%$$

$$Recall = 0.09/(0.09+0.04) = 69\%$$

5. CONCLUSIONS

In this paper, the proposed LoM system has been successfully demonstrated. The system comprises three major components: Picture Extraction Module (PEM), Logo Matching Module (LMM) and Web Mapping Module (WMM) but only the first two modules are described in this paper. During the experiments, three sets of PEM-pictures containing 50, 100 and 200 pictures are extracted from Flickr® website. For the set of 200 PEM-pictures, the textual-based PEM accuracy is 13% and the visual-based LMM accuracy is 89% which indicate a dramatic improvement in accuracy. The specificity is 95%, which means that the visual search can effectively remove the pictures without any relevant search contents. The 72% precision shows a small possibility that the visual search keeps some pictures without the target logos. The recall of 69% means a possibility that some pictures with the target logos have been filtered out. In summary, it can be concluded that visual search is more efficient and accurate than textual search. It is expected that Web Mapping Module (WMM) will be presented in a follow-up paper to fulfil completeness of the proposed system.

ACKNOWLEDGEMENTS

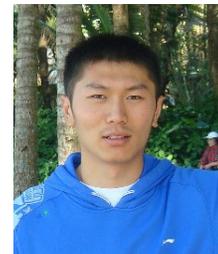
The first author would like to thank Dr. Lexing Xie for her introduction to data mining and her guidance during his study of COMP 8780 Information and Human Centred Computing Project at Australian National University in 2011.

REFERENCES

- [1] Trent Nouveau, "Digital Data to Earth: You have run out of memory", 3 May 2010, <http://www.tgdaily.com/hardware-features/49611-digital-data-to-earth-you-have-run-out-of-memory>
- [2] Eric Schonfeld, "Who Has The Most Photos Of Them All? Hint: It Is Not Facebook", 7 April, 2009, <http://techcrunch.com/2009/04/07/who-has-the-most-photos-of-them-all-hint-it-is-not-facebook/>
- [3] Pingdom.com, "Internet 2011 in numbers", January 17th, 2012, <http://royal.pingdom.com/2012/01/17/internet-2011-in-numbers/>
- [4] Screenshot of tags on del.icio.us in 2004 and Screenshot of a tag page on del.icio.us, also in 2004, both published by Joshua Schachter on July 9, 2007
- [5] Flickr ® (2012), <http://www.flickr.com>.
- [6] Lowe, D. G., "*Object recognition from local scale-invariant features*", Proceedings of International Conference on Computer Vision, Corfu, Greece, September 1999, pp. 1150-1157.
- [7] Lukas Mach, "Sift keypoints filtering.jpg", 13 February 2008, http://en.wikipedia.org/wiki/File:Sift_keypoints_filtering.jpg
- [8] Lowe, D. G., "*Distinctive Image Features from Scale-Invariant Keypoints*", International Journal of Computer Vision, 60(2): 91-110 (2004).
- [9] Ximing Hou, "*Visual matching on user-generated photo collections*", Subject Report, 33 pages, School of Computer Science, College of Engineering and Computer Science, Australian National University, June, 2011.

AUTHORS

Ximing Hou completed his Master of Computing at Australian National University in 2011 and obtained his Bachelor of Computer Science degree at Victoria University in 2008 and Bachelor of Computer Science in Henan University, China in 2010. His research interests include Data Mining, Pattern Recognition and Geographical Information.



Hao Shi is an Associate Professor in School of Engineering and Science at Victoria University, Australia. She completed her PhD in the area of Computer Engineering at University of Wollongong and obtained her Bachelor of Engineering degree at Shanghai Jiao Tong University, China. She has been actively engaged in R&D and external consultancy activities. Her research interests include Location-Based Services, Web Services, Computer/Robotics Vision, Visual Communications, Internet and Multimedia Technologies, and Data Mining.

