AUTOMATIC ANNOTATION OF LOCATION INFORMATION FOR WWW IMAGES

Zhigang Hua[2*], *Chuang Wang*[3*], *Xing Xie*[1], *Hanqing Lu*[2], *Wei-Ying Ma*[1]
[1]Microsoft Research Asia, 5/F Sigma Center, No. 49, Zhichun Road, Beijing, 100080, China
[2]Inst. of Automation, Chinese Academy of Sciences, P.O. Box 2878, Beijing, 100080, China
[3]Dept. of Computer Science, Huazhong Univ. of Science and Technology, Wuhan, 430074, China

ABSTRACT

Currently, a crucial challenge is raised on how to manage a large amount of images on the Web. Due to a real synergy between an image and its location, we propose an automatic solution to annotate contextual location information for WWW images. We construct an image importance model to acquire the dominant images in a page that comprise contextual surrounding text. For each acquired image, we develop an effective algorithm to compute location from its contextual text. We apply our approach to 1,000 pages from various websites for image location annotation. The experiments demonstrated that more than 30% WWW images are related with geographic location information, and our solution can achieve the satisfactory results. Finally, we present some potential applications involving the utilization of image location information.

1. INTRODUCTION

Nowadays, World Wide Web (WWW) has brought about challenges for organizing and searching a large volume of available images. The traditional image retrieval techniques, such as those content-based image retrieval (CBIR) systems, are usually not scalable for the use in WWW images to handle the large number of images. Moreover, different from traditional image retrieval and browsing, there is a lot of additional information on the Web, such as contextual texts and image link information.

As we know, the majority of images would be accompanied by time stamps, and image owner can also be also easily determined. But, very little other information would be provided with the images. Search engines such as Google Image Search [3] make good use of surrounding keywords when available, but searching images through keywords may be frustrating – keywords have linguistic and person-dependent components that can make them difficult to use [9]. Among various features, the location where a photo was shot is important because it says much about its semantic content.

There exist different ways to acquire location information for image media, such as manual annotation, location-aware devices and so on. However, the thing becomes more favorable on the Web, where image location data is becoming increasingly available from their contextual information. In this paper, we propose an automatic solution to annotate location information for WWW images. Our approach comprises two steps: 1) developing an image importance model to extract dominant images; and 2) computing image location through an algorithm based on the analysis of contextual texts.

The rest of this paper is organized as follows. In Section 2, we present related works. We provide the details of our approach in Section 3. In section 4, we present experimental results. In Section 5, we present some future applications. Finally, we conclude this paper in Section 6.

2. RELATED WORK

Current studies have shown that users associate their photos with event, location, subject, and time [9]. Among them, event is usually determined by time and location, and subject is often defined by combinations of who, what, when, and where. Based on analyzing semantic info, the image organization or retrieval results will better consist with user perception. Rodden [6] proposed to organize images according to the event or subject. Naaman [5] developed a system named PhotoCompas that uses the time and location metadata to cluster photos into events. Tollmar [8] built a system named IDeixis, through which users can search a captured photo on the Web to find out related information to the location where the user locates. These methods mainly consider utilizing location data rather than how to acquire such data.

Toyama [9] proposed an end-to-end system called WWMX (i.e. World Wide Media eXchange) that indexes image media by timestamp, owner, and critically, location stamp. The location metadata is mainly acquired by various forms such as manual entry, in the image header (from camera), or from a separate location-aware device. However, these methods are not suitable for WWW images, for they may require innumerable manual efforts.

In this paper, we propose a novel solution to automatically annotate WWW images with location information, which is computed from contextual data on the Web. We believe this automatic approach can be scalable for the use in the large image collection on the Web.

[*] This work was done while the authors were doing research internship at Microsoft Research Asia.

0-7803-9332-5/05/$20.00 ©2005 IEEE

3. OUR APPROACH

3.1. System Framework

We develop an automatic solution to annotate location for WWW images. As shown by a number of existing studies, the majority of web pages contain noise info, such as ad, logo and etc. In our system, we utilize page segmentation technique to partition a page into blocks, and then extract contextual information for each image in block granularity. In Figure 1, we present the work flow including two parts, i.e. *dominant image extraction* and *location computing*.

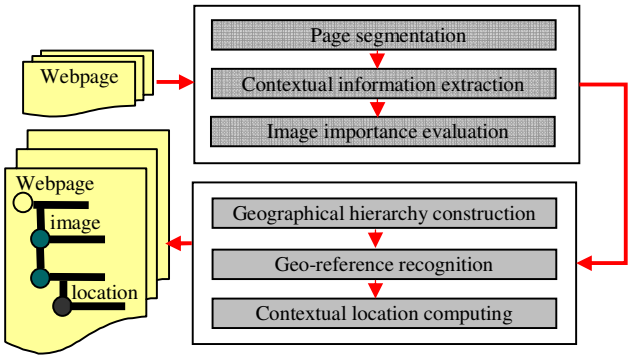

Figure 1. Our system framework

3.2. Dominant Image Extraction

It is not necessary to compute location for each image in a web page, such as the ad or logo images. We are thus setting out to construct an *image importance model* to extract dominant images that indicate content information of a page, comprising a three-step process as follows.

Page segmentation. Among various methods that segment a page into blocks, vision-based page segmentation (VIPS) excels in both an appropriate partition granularity and coherent semantic aggregation [10], which can efficiently keep related content together while separating semantically different blocks from each other. Figure 2 shows an example of using VIPS to segment a page into blocks labeled from 1 to 4. Further, song et al. [7] proposed an importance model to map from the spatial features to importance for a block, and it also pointed out that, among various forms of feature representations, *relative spatial features* prove to be more effective for block importance modeling, which can be described as follows:

<Block features>➔IMP(Block)

Contextual information extraction. After a web page is segmented into blocks, for each image in a block, we set out to extract the appropriate surrounding text as its contextual information. Instead of using the whole text passages in the host block as image context, we propose to utilize visual cue information in a block to allocate appropriate contextual text for an image.

According to the structure layout, explicit separators of various images in a block can be detected by analyzing the properties of the tags. The following two types of explicit separators are widely used:

1) <*HR*> is the most frequently used explicit separator for representing a horizontal line in a web page.

2) The tags like <*TABLE*>, <*TR*>, <*TD*>, and <DIV> have border properties. When their border properties are set, there would be separators at corresponding borders.

For example, in Figure 2, the <*TR*> and <*TD*> tags that separate all the eight images into 2x4 layout can be detected by a DOM-tree based analysis. By analyzing the explicit separators, we extract contextual information for the images in the block 3 in Figure 2, as shown in Figure 3.

Give each block $B_i \in$ a page P, for each image $I_{ij} \in B_i$, its feature is represented as follows:

$$I_{ij}=\{\text{ImgLink, ImgText, ImgSize,}$$

$$\text{ImgCenterX, ImgCenterY, ImgWidth, ImgHeight}\}$$

where ImgLink refers to the page that the image links to, usually designed to illustrate more details on the image; ImgText stands for the surrounding text of the image; the features ImgSize, ImgCenterX, ImgCenterY, ImgWidth and ImgHeight indicate position and shape information.

Image importance evaluation. For an image I_{ij} in a block B_i, we extend the following factors to consider the image weight:

1) Image area coverage. The area percentage speaks the importance of an image in a block.

$$P_s(I_{ij})=\text{ImgSize/BlockSize} \quad (1)$$

2) Image contextual text length. As we know, an important image will be illustrated with more text. Surrounding text and linked page are two forms to illustrate an image:

$$\text{TextLen} = \max(|\text{ImgText}|, |\text{ImgLink}|) \quad (2)$$

where |ImgText|, |ImgLink| respectively indicates the length of the surrounding text and linked page. Thus,

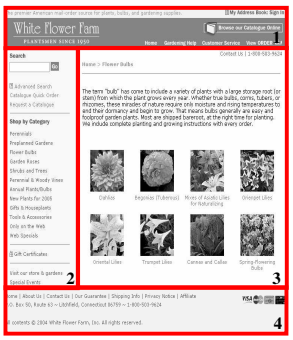

Figure 2. An example for page segmentation

Figure 3. The contextual information extraction

$$P_l(I_{ij})=\begin{cases}1 & \text{TextLen}>500\\ 0.5 & \text{TextLen}>=150, \text{ and TextLen}<=500\\ 0 & \text{TextLen}>150\end{cases} \quad (3)$$

3) Width/height ratio. Image shape reflects importance:

$$P_r(I_{ij})=\begin{cases}1 & 0.4<\text{ImgWidth}/\text{ImgWidth}<4\\ 0 & \text{Otherwise}\end{cases} \quad (4)$$

Combining the features, we formulate the weight of I_{ij} in its host block B_i as:

$$W(I_{ij}) = w_1 \times P_s(I_{ij}) + w_2 \times P_l(I_{ij}) + w_3 \times P_r(I_{ij}) \quad (5)$$

Thus, we model image importance as the product of the importance of its hosting block in the page, and the image weight in its hosting block:

$$\text{Importance}(I_{ij}) = \text{IMP}(B_i) \times W(I_{ij}) \quad (6)$$

We thus take the images whose importance beyond a fixed threshold ε as the dominant images in the page P. Let the set of the dominant images in P be IMG(P):

$$\text{IMG(P)} = \{I_{ij} \mid \forall I_{ij} \in P, \text{Importance}(I_{ij}) > \varepsilon\} \quad (7)$$

3.3. Image Location Computing

After acquiring the dominant images in a page, we are thus beginning to compute contextual location for each of them. We adopted the algorithm to compute location as follows:

Input: P, a web page for image location calculation
Algorithm ComputeImageLoc(P)
  For each image I∈ IMG(P)
    LocCompute (I.ImgText)→loc
    If loc is non-location **Then**
      LocCompute (I.ImgLink)→loc
    Loc→I.Location

where LocCompute() is a function to compute location from a text passage, comprising three steps as follows.

Geographical hierarchy construction. To recognize and extract geo-references in a text passage, we prepare a gazetteer in advance. In our approach, the forms of geo-references include three types, i.e. telephone number, postal code and geographical place name. The gazetteer can constitute a complete three-level hierarchy view of a geographical scope (USA in our practice), where location nodes distribute in the three geographical levels such as country, state and city. This structure is shown in Figure 4.

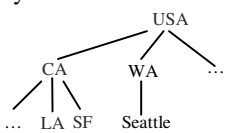

Figure 4. Geographical hierarchy view for USA location

Geo-reference recognition. When image contexts are scanned, we identify each geo-reference that is included by our gazetteer. As aforementioned, a location node can have various representation forms. Thus, we place each geo-reference to its node in the hierarchical tree.

Contextual location computing. Since it is common that multiple location nodes may coexist in a text passage, it is thus crucial to estimate representative location. Assume LA and SF available in Figure 4, there may exist two representations: 1) LA and SF; or 2) CA. To determine the representative location, we borrow two measures from the CGS/EGS approach [2], i.e. *power* and *spread*. We extend the power concept of a location node, by comprehensively considering the influences of its offspring and ancestors. Given a text passage t, the power of a location node *l* is:

$$Power(t,l) = f(t,l) + \sum_{l_i \in offspring(l)} f(t,l_i) \quad (8)$$

where $f(t, l)$ refers to the frequency that *l* occurs in *t*. We adopted the Entropy definition of the spread concept, which can achieve the best performance according to [2]. Once the power and spread values are computed, the final location fall into the location nodes with spread and power values that are both beyond the given thresholds. More implementation details are available in [2].

4. EXPERIMENTS

4.1. Data Preparation

We used three sources to prepare the gazetteer: 1) USA Postal Services for USA postal code; 2) North American Numbering Plan for telephone number; and 3) Geographic Names Information System for geographic name. The images are crawled from Yahoo (http://dir.yahoo.com/) with 14 categories. For each category, we crawled the top 200 pages, and we calculate location for the image with the largest importance per page. The 993 images that meet the condition regulated in Equation 7 are listed in Table 1.

Further, we list in Table 1 the number of images that are related with location (*Loc-image*) and its corresponding ratio (*Loc-ratio*) in each category. Totally, 299 images are manually labeled to be related with location, covering a ratio of about 30% in the testing images. As for all of the categories, the highest location-related ratios (about 45%) are among the Arts, Government and Regional categories, and the least (about 10%) is falling into the Recreation category. To our surprise, the majority of the Arts category includes museum information highly related with location.

4.2. Experimental Results

We use precision, recall and F-measure to evaluate the performance of our approach. Assume the 14 categories in our test to be $C=\{C_i \mid i:1\rightarrow 14\}$, and let $P(C_i)$ and $R(C_i)$ be the precision and recall for each category. The F-measure for each category can be computed as follows:

$$F(C_i) = \frac{2P(C_i) \times R(C_i)}{P(C_i) + R(C_i)} \quad (9)$$

Table 2 presents the precision and recall in each category. The maximal precision (84.4%) occurs in the Government category, and the least precision falls into the Computers & Internet category. Our approach can achieve a favorable recall, ranging from 72.0% to 100.0%. The F-

measure value ranges from 0.35 to 0.86 for all categories. Further, we give the average precision and recall results in the last row of Table 2. It achieves a precision of about 60% and a recall of about 85% over all testing data, and the overall F-measure value reaches about 0.71.

Table 1. The categories and numbers of testing data set

No.	*Category*	*Image*	*Loc-image*	*Loc-ratio*
1	Business & Economy	47	9	19.1%
2	Computers & Internet	75	10	13.3%
3	News & Media	56	16	28.6%
4	Entertainment	187	56	29.9%
5	Recreation	39	4	10.3%
6	Health	82	29	35.4%
7	Society & Culture	24	8	33.3%
8	Education	91	39	42.9%
9	Arts	55	25	45.5%
10	Government	69	31	44.9%
11	Regional	112	50	44.6%
12	Science	85	11	12.9%
13	Social Science	43	9	20.9%
14	Reference	28	2	7.1%
15	**Total**	993	299	30.1%

Table 2. The precision and recall of our approach

No.	**Precision**	**Recall**	*F-Measure*
1	46.7%	77.8%	0.58
2	22.0%	90.0%	0.35
3	55.6%	93.8%	0.70
4	59.1%	92.9%	0.68
5	23.1%	75.0%	0.35
6	75.8%	86.2%	0.81
7	70.0%	87.5%	0.78
8	81.0%	87.2%	0.84
9	83.3%	80.0%	0.82
10	84.4%	87.1%	0.86
11	62.1%	72.0%	0.67
12	40.9%	81.8%	0.55
13	66.7%	88.9%	0.76
14	66.7%	100.0%	0.80
AVG	60.5%	84.9%	0.71

After analyzing the reasons that cause the location computing error, we found that it was mainly caused by the ambiguities in the location identifications. For example, a place name can refer to different location nodes or even non-geographical senses, e.g. Washington corresponding to a number of places and also can refer to characters. However, in our current work we have not eliminated these ambiguities. In the future, we are planning to adopt some heuristic rules proposed by [1] to solve this matter. We believe this will help us to achieve higher precisions.

5. FUTURE APPLICATIONS

We are planning to explore some potential applications.

Location-based image organization. According to WWMX [9], it is fine to organize large image collection on a geographical map. Different from image clustering [4] based on low-level features, we are planning to cluster image search results through location information. In this way, users can easily locate images from a large image list.

Location-based image search. Image search engines may also benefit from our proposed approach. The search quality can be refined according to the relevance with user contextual location. Thus, image search results will be tailored according to user location.

6. CONCLUSIONS

Due to a real association between image and its location, in this paper, we propose a novel solution to automatically annotate location information for WWW images. By using an image importance model, we develop an effective algorithm to extract contextual location for the dominant images in pages. The results demonstrated its efficiency in automatically annotating location information for WWW images. We will employ more sophisticated methods to improve the precision of our approach in future. Finally, we have presented some potential applications involving the utilization of image location.

7. ACKLOWNEDGEMENTS

We would like to give our special thanks to Ruihua Song and Zhiwei Li for their generous helps.

8. REFERENCES

[1] E. Amitay, N. Har'El, R. Sivan and A. Soffer. Web-a-where: geotagging web content. The 27th Annual International Conference on Research and Development in Information Retrieval, Sheffield, UK, Jul 2004, pp. 273-280.

[2] J. Ding, L. Gravano and N. Shivakumar. Computing geographical scopes of web resource. The 26th International Conference on Very Large Data Bases, Cairo, Egypt, Sep 2000.

[3] Google image search engine: http://images.google.com.

[4] H. Liu, X. Xie, X.-O. Tang, Z.-W. Li and W.-Y. Ma. Effective browsing of web image search results. The 6th ACM SIGMM International Workshop on Multimedia Information Retrieval, in conjunction with ACM Multimedia 2004, New York, USA, Oct 2004.

[5] M. Naaman, Y. J. Song, A. Paepcke and H. Garcia-Molina. Automatic organization for digital photographs with geographic coordinates. The 4th ACM/IEEE-CS joint Conference on Digital Libraries, Tuscon, USA, Jun 2004.

[6] K. Rodden and K. R. Wood. How do people manage their digital photographs? The CHI 2003 Conference on Human Factors in Computing Systems, Florida, USA, Apr 2003.

[7] R. Song, H. Liu, J.-R. Wen and W.-Y, Ma. Learning block importance models for web pages. The 13th World Wide Web Conference, New York, USA, May 2004.

[8] K. Tollmar, T. Yeh and T. Darrell. IDeixis: image-based deixis for finding location-based information. The 6th Intl Conference on Human Computer Interaction with Mobile Devices and Services, Glasgow, Scotland, Sep 2004.

[9] K. Toyama, R. Logan, A. Roseway and P. Anandan. Geographic location tags on digital images. ACM Multimedia 2003 Conference, CA, USA, Nov 2003.

[10] S. Yu, D. Cai, J.-R.Wen and W.-Y. Ma. Improving pseudo-relevance feedback in web information retrieval using web page segmentation. The 12th World Wide Web Conference, Budapest, Hungary, May 2003.