A Location Data Annotation System for Personal Photograph Collections: Evaluation of a Searching and Browsing Tool

Chufeng Chen, Michael Oakes and John Tait
University of Sunderland
{chufeng.chen, michael.oakes, john.tait}@sunderland.ac.uk

Extended Abstract

Categories and Subject Descriptors: H.3.3 Information Search and Retrieval: Clustering and Search Process.

General Terms: Algorithms, Design, Human Factors.

Keywords. Episodic Memory Annotation, Clustering, Browsing.

1. INTRODUCTION

In this paper we describe a system for storing and retrieving digital images from personal collections. The images can either be manually annotated with a set of keywords chosen by the owner of the collection, or keywords can be automatically inferred from the time and location stamps associated with the image and the Geographic Names Data Base gazetteer. User queries are matched against the image annotations using the Cosine Similarity Measure, and the best matching images are displayed. To enable browsing of the image collection, the images are clustered according to time and location, the two main factors of episodic memory (Baddeley, 1990). The main factors of episodic memory, time and location, can be provided by a GPS digital camera. GPS data has been used to support digital image browsers in a number of studies (Naaman, 2004) (Pigeau, 2004) (O'Hare, 2005). We aim to overcome the problem of the semantic gap (Eakins, 1996) inherent in content-based image retrieval (CBIR) by making use of these high level features of the events of human episodic memory, and semantically-rich keywords.

2. DEVELOPMENT OF THE IMAGE ANNOTATION SYSTEM FOR PERSONAL PHOTOS

The original system of Chen et al. (2006), derived from Platt et al.'s (2002) system of clustering digital images according to the time they were taken, enabled the clustering of personal image collections into separate events using numeric time and location (latitude and longitude) values extracted from GPS metadata. The system was extended to allow the automatic annotation of each image with keywords extracted from a gazetteer, corresponding to place names and other geographical features. The gazetteer we used was the UK file of the Geographic Names Data Base, maintained by the National Geospatial Intelligence Agency (http://earth-info.nga.mil/gns/html/index.html). This gazetteer (see figure 2) contains entries for locations throughout the UK, with data for latitude, longitude, sort name, full name and DSG-code (the DSG-code stands for location features such as rock, beach or harbor) of the specific location. We compared the longitude and latitude stamps on each image with the longitudes and latitudes of the places in the gazetteer, using Euclidian distance:

$\sqrt[2]{[Lat\,(i) - Lat\,(k)]^2 + [Long\,(i) - Long\,(k)]^2}$, to find the closest entry. However, this is only an approximation to calculate the Euclidian distance that in order to find out the closest place. The image was then annotated with the full place name and DSG-code for this location (see figure 1). For example an image might be annotated with "London Buckingham Palace" and "Palace".

ID	NAME	CREATE_TI	LOCATI(	Lat	Long	EVEN'	
1	London-plane-wing.JPG	09:20:00	London	51	-0.5	1	London Heathrow populated place 1
2	London-plane-sky.JPG	09:22:00	London	51	-0.5	1	London Heathrow populated place 2
3	London-house-wife-sis	06:09:00	London	52	-0.1	2	London Buckingham Palace palace 3
4	London-house-wife-me.	06:09:00	London	52	-0.1	2	London Buckingham Palace palace 4
5	London-statuary-man-w	07:15:00	London	52	-0.1	2	London Buckingham Palace palace 5
6	London-statuary-man-m	07:16:00	London	52	-0.1	2	London Buckingham Palace palace 6
7	London-buckingham-wif	07:40:00	London	52	-0.1	2	London Buckingham Palace palace 7
8	London-buckingham-me.	07:41:00	London	52	-0.1	2	London Buckingham Palace palace 8
9	London-statuary-white	07:42:00	London	52	-0.1	2	London Buckingham Palace palace 9
10	London-statuary-white	07:42:00	London	52	-0.1	2	London Buckingham Palace palace 10

cord: 5 of 200

Figure 1. Example of database data

ID	LAT	LONG	DSG	CC1	SORT_NAME	FULL_NAME	DSGC
1	56.367	-4.53	MT	UK	AAN BEN	Ben A'an	mountain
2	56.233	-4.42	MT	UK	AAN BEN	Ben A'an	mountain
3	60.167	-1.43	RK	UK	AASKERRY	Aa Skerry	rock
4	52.3	-2.37	PPL	UK	ABBERLEY	Abberley	populated place
5	51.833	0.917	PPL	UK	ABBERTON	Abberton	populated place
6	52.183	-2.02	PPL	UK	ABBERTON	Abberton	populated place
7	51.783	0.267	PPL	UK	ABBESSRODING	Abbess Roding	populated place
8	52.333	-3.4	PPL	UK	ABBEYCWMHIR	Abbey-Cwmhir	populated place

cord: 3 of 31708

Figure 2. UK Gazetteer, last modified in 1994

Our user interface (see figure 3) has two parts: a command panel and a display panel. The command panel allows users to submit their queries and use the function buttons: search, browse and exit. The display panel displays the searching results. Users can choose either query keyword searching or to browse the entire photo collection. For keyword searching, users submit their query keywords into the text input area, and the 30 best matching images (according to the cosine similarity measure) are displayed in the browsing panel, following the time line (the earliest taken photo in the top left corner, the last taken in the bottom right hand corner). Fewer than 30 images are displayed if fewer than 30 images have a similarity > 0 with respect to the query. All the images are clustered into events. If the user opts for the browsing approach, the display panel will display the whole photo collection following the time line and clustered into events, an updated version of (Chen et al, 2006).

Figure 3. User interface

3. MAIN EXPERIMENT

The previous study by Chen et al. (2006) indicated that factors related to human episodic memory, time and location, could be used to help users browsing their personal photograph collections more easily. In these experiments we wished to find out whether the automatic image annotations helped users to search for their images, and how this compared with annotations provided by the users themselves.

The hypotheses of this study are: 1. The user's own annotation of images will not support image retrieval from personal collections significantly better than automatic image annotation for personal images, according to our qualitative and quantitative criteria. 2. Automatic image annotation will provide significantly better support for image retrieval than no annotation (browsing images clustered by time and location). Nine volunteers participated into this study, six male and three female, all of whom were staff and students at the University of Sunderland with experience in managing their own digital photos. Each subject was asked to provide a personal collection of about 200 manually-annotated images. Subjects were instructed to provide a maximum of five keywords. Generally they used personal names, event names, place names or object names such as red car, small cat as their own annotation.

3.1 Scenario tasks

We gave each subject four general and four specific tasks for image searching. They were each asked to read each scenario description beforehand, and then in the timed phase of the experiment, find out the image which best matched each scenario description.

The **General search tasks** involved non-specific scenarios, so that every collection would contain at least one image answering that description. No time limit was set for any of the searching tasks, but for each task the time spent searching was recorded. The general search tasks were as follows:

a) Please find a photo where you were with one or two of your friends (or family members), at an outdoor sunny place on your last holiday.
b) Please find a photo where you were standing in a crowded city centre.
c) Please find a photo where you were with your family in an indoor environment.
d) Please find a photo of the most famous building of the city which you last visited.

The **Specific search tasks** were based on the contents of each subject's personal collection, so were different for each subject. For example, one subject was asked to perform the following four search tasks:

a) Please find the photo of a light house.
b) Please find the photo of a Christmas tree in the city centre.
c) Please find the photo of a black swan in the water.
d) Please find the photo of the front view of the British Museum.

All the subjects performed four general scenario tasks and four specific scenario tasks for each of the three browsers.

The experiment used a repeated measures design, and the order in which the subjects used each browser was determined by a Latin square to compensate for learning effects. Half of the subjects did the general tasks first followed by the specific tasks, and half did the specific tasks first followed by the general tasks. Questionnaires were used to determine the level of user satisfaction with each browser.

3.2 Recall and Precision

In the Recall and Precision experiment, the general and specific tasks were similar to those used in section 3.1, four general tasks and four specific tasks, with the difference that there was no single correct photo, but rather a set of photographs in the collection relevant to the query. Thus an example of a specific scenario was: Please find all the photos of a lake. After each task we required the subjects to answer following questions:

a) How many photos were displayed on the browser?
b) How many of the photos displayed on the browser match your query?
c) Please look at the whole photo collection. How many matching photos are there which were not displayed on the browser?

Recall is b / (b + c), Precision is b / a, and F1 = (recall + precision) / (2 * recall * precision).

4. RESULTS

4.1 Searching time

The comparison of searching times in seconds (averaged over all 9 subjects) for the time and location and user-annotation browsers is shown in Table 1. Specifically, each subjects' general search time on figure 3, each subjects' specific searching time on figure

4 and total searching time on figure 5. A two-tail matched pairs t-test with the Bonferroni correction for multiple comparisons ($\alpha = 0.05/3 = 0.017$, so significance would be achieved if $p < 0.017$) was used to analyse the results. The system searching time for user annotation was significantly less than that required for browsing alone for both the general scenario tasks and total finishing time, but this difference was not significant for the specific scenario tasks.

Table 1. System searching time for time and location browser vs. user annotation

	Time & location browser	User Annotation	t	p<0.017 significant
1. AVE/STDEV searching time general scenario tasks	51/18.1	35.9/11.7	t= 4.31	**p=0.0026**
2. AVE/STDEV searching time specific scenario tasks	42.2/16.2	32.3/10.2	t=2.24	p=0.059
3. AVE/STDEV total finish time	93.2/28.2	68.2/20	t=4.18	**p=0.0031**

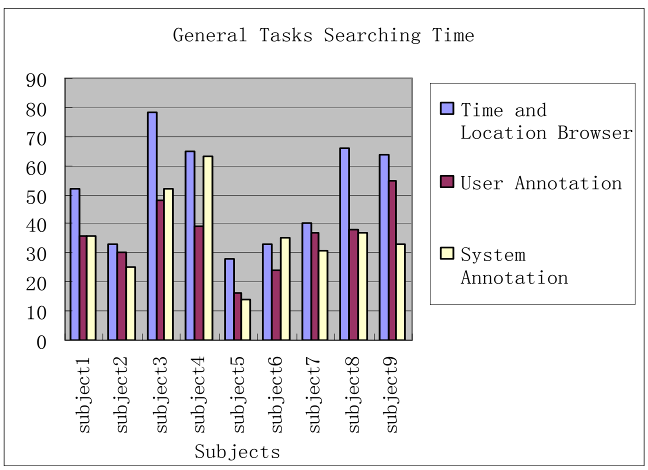

Figure 3. General tasks searching time for each subject

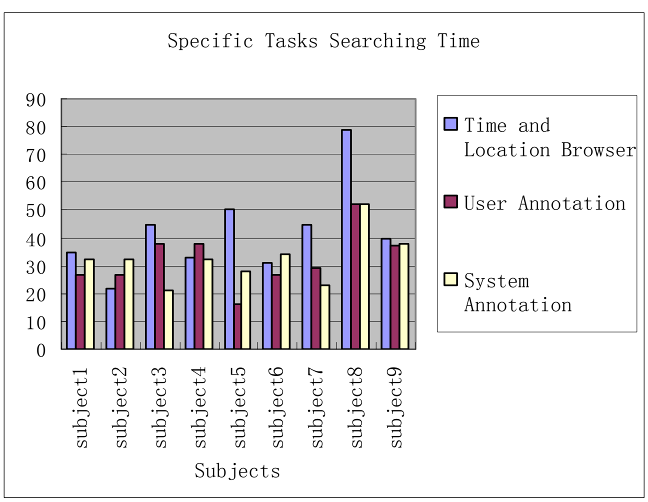

Figure 4. Specific tasks searching time for each subject

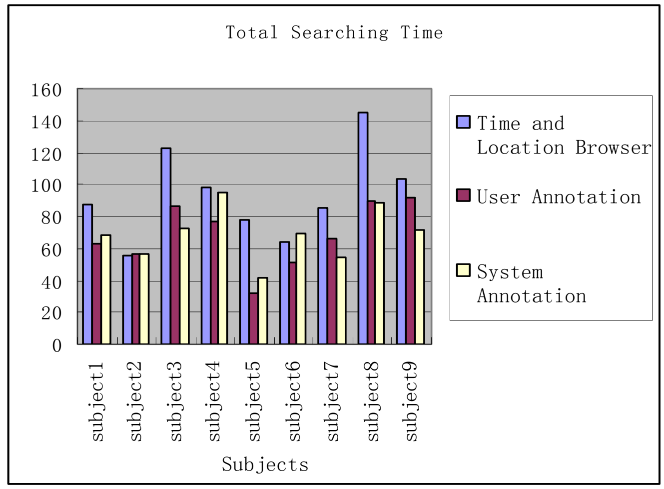

Figure 5. Total searching time for each subject

Table 2. System searching time for time and location browser vs. system annotation

	Time & location browser	System Annotation	t	p<0.017 significant
1. AVE/STDEV searching time general scenario tasks	51/18.1	36.2/14.3	t= 3.75	**p=0.0056**
2. AVE/STDEV searching time specific scenario tasks	42.2/16.2	32.4/9.1	t=2.12	p=0.067
3. AVE/STDEV total finish time	93.2/28.2	68.7/16.6	t=3.31	**p=0.011**

Table 3. User satisfaction for the three different systems.

	Time and location browser	User annotation	System annotation	ANOVA F (2, 24) =
1. I like this image browser	3.56	**3.89**	3.44	1.24, p= 0.31
2. This browser is easy to use	3.44	**4.11**	3.56	2.95, p=0.071
3. This browser feels familiar	3.44	3.78	**4**	1.73, p= 0.199
4. It is easy to find the photo I am looking for	3.22	**4.11**	3.67	2.59, p=0.095
5. A month from now, I would still be able to find these photos	3.56	**4**	3.78	1.09, p= 0.35
6. I was satisfied with how the pictures were organized	3.56	**3.89**	3.67	0.36, p=0.7
Average	3.46	**3.96**	3.69	1.31, p= 0.29

Table 2 shows the comparison of time and location browsing alone and system annotation. Similar results were obtained as in Table 1. The average searching time with system annotation was significantly shorter than that for browsing alone for both the general scenario tasks and the total searching time, but not significantly shorter for the specific scenario tasks. There was no significant difference between the time required for searching with system annotation, than for user annotation. For the searching time of each individual subject see also figures 3, 4 and 5.

4.2 Questionnaires

The user satisfaction questionnaires were filled in immediately after the timed searching tasks had been performed, and the average satisfaction ratings for the nine subjects are shown in Table 3. The user annotation system was rated more highly than the other browsers according to all six criteria, except for question 3, where it was rated less than the system annotation. The system annotation was rated a little more highly than time and location browser according to all six criteria. The average satisfaction ratings were best for the user annotation, next best for system annotation and poorest for the time and location alone browser third. However, one way ANOVA tests for each question showed that there was no significant difference in the reported user satisfaction ratings between the three browsers at $p < 0.05$.

4.3 Recall and Precision

The results show that recall and precision for system annotation is better than for user annotation for both the general scenario tasks and the specific scenario tasks. The overall effectiveness, as estimated by the F1 measure, was therefore also better for system annotation for both the specific and general scenario tasks. For both types of annotation, recall, precision and F1 were better for the specific scenario tasks than for the general scenario tasks. The results of the recall and precision experiment are shown in table 4.

Table 4. Results of the recall and precision experiment

	System Annotation			User Annotation		
T4	Recall	Precision	F_1 measure	Recall	Precision	F_1 measure
General tasks	0.723	0.448	0.552	0.677	0.192	0.3
Specific tasks	0.963	0.643	0.771	0.765	0.488	0.596
Overall	0.842	0.545	0.662	0.721	0.340	0.462

5. DISCUSSIONS AND CONCLUSION

Our results show that 1. There were no significant differences between user annotation and system annotation for either the user-centered or system-centered evaluations. 2. User annotation and system annotation produced significantly faster searching times than time and location browsing only. 3. The system annotation produced greater retrieval effectiveness, as measured by recall and precision, than user annotation. 4. There were no significant differences between user annotation, system annotation and the time and location browser in the user satisfaction evaluation. In fact the subjects were polarized, with three clearly preferring the experience of browsing, and six much preferring the experience of keyword searching.

Overall, we have demonstrated the feasibility of using GPS data with gazetteers in automatically assigning annotation keywords to images, so that these images can be retrieved in response to user queries. Not only does the automatic approach spare the collection owner the arduous task of annotating a large number of images manually, but in our experiments retrieval performance was at least as good for system-assigned annotations as for the user-assigned annotations. The recall and precision results suggested that searching images by location keyword would provide more relevant results. One limitation of the study is that some types of queries are more easily retrieved by the location annotation system, because they ask for items included in the gazetteer such as “lake”. Queries which asked about non-permanent events or features such as “graduation day” or “red car” would be less easily retrieved by the system. One reviewer suggested that we should examine the effect of adding the names of nearby places rather than just the very closest. This would be useful to distinguish two distinct places with the same name.

6. REFERENCES

Baddeley A.: Human Memory, Theory and Practice, LEA Publishers (1990).

Chen, C., Oakes, M. and Tait, J.: Browsing Personal Images Using Episodic Memory (Time + Location). Proceedings of the European Conference on Information Retrieval, Volume 1, (2006) 362-372

Eakins, J. P.: Automatic image content retrieval--are we getting anywhere? In: Proceeding of Third International on Conference of Electronic Library and Visual Information Research, Research, De Montfort University. (1996) 123-135.

Eakins, J.P., Graham, M.E.: Content-Based Image Retrieval: A Report to the JISC Technology Applications Programme. Institute for Image Data Research, University of Northumbria at Newcastle, UK (1999). Available at http://www.unn.ac.uk/iidr/report.html.

Naaman, M., Song, Y. J., Paepcke, A., and Garcia-Molina, H.: Automatic organization for digital photographs with geographic coordinates. In Proceedings of the Joint Conference on Digital Libraries. ACM Press, New York, NY, (2004) 326--35.

O’Hare, N., Gurrin, C., Lee, H., Murphy, N., Smeaton, A., Jones., G., J.F.: My Digital Photos: Where and When? SIGMM (2005) Singapore.

Pigeau, A., Gelgon, M. Organizing a personal image collection with statistical model-based IVL clustering on spatio-temporal camera phonemeta-data, J. Vis. Commun. Image R. 15 (2004) 425-445.

Platt, J.C., Czerwinski, M. & Field, B.: PhotoTOC: Automatic clustering for browsing personal photographs, Microsoft Research Technical Report MSR-TR-2002-17 (2002).