A LOCATION DATA ANNOTATION SYSTEM FOR PERSONAL PHOTOGRAPH COLLECTIONS: EVALUATION OF A SEARCHING AND BROWSING TOOL

Chufeng Chen¹, Michael Oakes², John Tait³
Chufeng.chen@TVU.ac.uk, Michael.oakes@sunderland.ac.uk, John@Johtait.net
Thames Valley University¹, University of Sunderland², Information Retrieval Facility³

ABSTRACT
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. User studies were conducted to compare the two annotation systems with time and location clustering alone, using searching time, recall and precision, and user satisfaction as criteria for success. Our results show that matching based on automatically assigned keywords was significantly better than time and location-based browsing alone, and performed at least as well as matching against user-assigned annotations.

1. INTRODUCTION
The number of people creating personal collections of digital images is continually increasing. Researchers are currently developing indexing and searching tools for such digital image collections. Many of these tools are based on Content-Based Image Retrieval (CBIR), which relies on low-level features of the images such as colour, texture and shape. However, CBIR can produce the problem of the “semantic gap” (Eakins, 1996). This is the mismatch between the high-level concepts humans associate with images, such as “lighthouse” or “pencil”, and the low-level feature descriptions used by CBIR such as “red” or “cylinder”. In addition, Eakins et al. (2004) showed that people do not want to use visual query tools such as drawing shapes and painting colours. Hollink et al. (2004) also found that people prefer to classify images in terms of events (time and place) rather than in terms of their visual features. Many studies (Eakins, 1996) (Fidel, 1997) (Jose et al., 1998) (Rodden et al., 2003) (Tait et al., 2001) (Venters et al., 2001) have suggested that images can be indexed in different ways. Chan & Wang (2004) found that human factors such as visual memory strongly affect the search results of image retrieval. Graham et al. (2002) suggested that GPS (Global Positioning System) location data could be used to assist browsing and searching, and more recently, (Chen et al., 2005) (Chen et al., 2006) (Gurrin et al., 2005) (Naaman et al., 2004) (Pigeau & Gelgon, 2004) (O’Hare et al., 2005) (Sarvas et al., 2004) (Toyama et al., 2003) have incorporated GPS meta-data into photo searching and browsing. O’Hare et al. (2005) believe that within a few years consumer digital cameras will all have GPS capabilities. Chen et al. (2006) developed a personal image browsing system, which relates to episodic memory by using both time and location for the automatic identification of different events. This data is automatically provided by the GPS digital camera which stamps each image according to date and time, latitude and longitude. The starting point of this research is the event clustering model of Chen et al. (2006), which was designed to help users browse their personal photograph collections. In this paper we extend the system to automatically annotate the personal photographs with location names and geographic features. It is also possible for users to supply their own keywords to annotate each image.

In section 2, this paper will review some previous studies related to time and location clustering, browsing and searching. Section 3 describes the development of our location annotation system. Section 4 describes our experimental design, where subjects were asked to perform both general and specific scenario image retrieval tasks using the three different browsing and searching systems (automatic annotation, user annotation and time and location clustering alone) and retrieval performance was assessed by speed of searching, user satisfaction questionnaires and the recall and precision measures. Both user and system-centred evaluation results are given in Section 5, and our analysis in Section 6 concludes that system annotation by location name and geographical features is at least as good as manual user annotation according to our criteria.

2. BACKGROUND AND RELATED WORK
Baddeley (1990) described episodic memory, sometimes referred to as personal memory, as people’s memory of special events, episodes and experiences, such as a birthday. It includes there main factors: time, location and event (when, where, what) (Baddeley, 1990) (Chen et al., 2006). Recently, digital cameras with Global Positioning System (GPS) capabilities have been developed. These cameras record the geographical coordinates at which each picture file is taken. The main factors of episodic memory, time and
location, can be provided by a GPS digital camera. GPS data has been used to support system design in a number of studies (Chen et al., 2005) (Chen et al., 2006) (Gurrin et al., 2005) (Naaman et al., 2004) (Pigeau & Gelgon, 2004) (O’Hare et al., 2005) (sarvas et al., 2004) (Toyama et al., 2003). Abrams et al. (1998) used episodic memory for helping users retrieve web pages and web documents in a bookmark management system. Human memory (Baddeley, 1990), specifically memory about time and location, has been considered in other image searching system design (Chen, 2005) (Chen, 2006) (Cooper, 2005) (Graham, 2002) (Naaman, 2004) (platt, 2002).

Rodden (2003) and savakis et al. (2000), suggest that the most convenient browsing criteria are time, geographical location, and photograph keyword descriptions such as people’s names, object names, and place names. On the other hand, classical features such as color, shape, layout, and texture cues are rated of little relevance. Platt et al. (2002) developed an image browsing system called Phototoc which clusters images based on their time stamps when these are available and on their colour histograms when they do not have time stamps. Platt et al. (2002) and Loui & Savakis (2000) used a similar technique for time clustering which used the variety of time scales present in the image collection to partition it into separate time gaps. Graham et al. (2002) developed two photo browser systems for collections of thousands of time-stamped digital photographs. Their first browser included only the date line and time line to enable browsing and summarisation. Their second system is a hierarchical browser which allowed users to browse a hierarchy of year, month, date, and time. Their experimental results showed that time can improve both the retrieval performance and user satisfaction of image searching. Cooper et al. (2005) are also interested in the specific events associated with images, but because they feel that GPS cameras are still not widespread, their image browsing system concentrates on time and visual content clustering. Shneiderman et al. (2002) introduced a photo browsing system called PhotoFinder, which allowed users to annotate information to personal photos such as people’s names and events. This system has been used to record the photographic history of SIGCHI. However, these annotations cannot be produced automatically. Toyama et al. (2003) showed that the location information is very important, because this provides high level semantic concepts, which can be easily acquired, indexed and searched, helping to close the semantic gap. O’Hare et al. (2005) showed that the key benefit of location annotation is that it enables users to support a number of access methodologies; search by location (for example, street, town, city or country), and search by proximity to a location or to other photos. By using location information the searching results can be drastically reduced. Pigeau & Gelgon (2004) developed a statistical technique of automated organization of a personal image collection, based on the location where the photos were taken and their timestamps. The focus of their study is on temporal and spatial meta-data, especially for camera mobile phone applications, but they believe that geographic information can be used on a wide variety of applications. Naaman et al. (2004) developed a time and location hierarchy clustering system. They also discuss human memory, particularly the location component, but did not perform any user studies of their interface. More recently, Shneiderman et al. (2006) suggested that a combination of annotation, browsing and sharing in personal image management systems can fulfill the users’ needs.

3. Development of annotation system for personal photos

The original system of Chen et al. (2005) (2006) 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. This enables users to submit keyword queries which can be matched against each image annotation, using the cosine similarity measure, to retrieve the best matching images. 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).

3.1 EXIF meta-data extraction

EXIF (Exchangeable Image File Format) is the most commonly used type of metadata. It is a standard controlled by the Japan Electronics and Information Technologies Industries Association (JEITA). The EXIF records the time, date, f-stop, shutter speed, focal length, type of camera, and many other details. There are about 20 tags relating to GPS, including latitude, longitude, altitude, speed, course, and direction of image. When the image was taken by the GPS digital camera, all this information would be stocked into the EXIF format. We extracted these data (including photo creation time, latitude, longitude and name of the place where the photo was taken) into the database for later processing.

3.2 Time and location clustering model

The time clustering model sorts the images in the collection by time, then announces the start of a new event whenever the time gap between two successive images is greater than the average time gap by a prespecified amount, as shown in formula 1, where gn+i is the time gap between image i and image i+1, gN is the local time gap between two successive images, d is the width of a sliding window of successive
images (e.g. \(d = 10\)). \(K\) is a suitable threshold (Platt et al. (2002) used \(K = \log_{17}\)).

\[
\sum_{i=0}^{d-1} \log(g^N+i)
\]

is the average time gap between successive images in the window. Our location clustering model has two components: a latitude clustering algorithm and a longitude clustering algorithm. We use formulas which are analogous with that of Platt et al. (2002), except in that the images are initially sorted by latitude and longitude respectively rather than by time, and instead of \(g^N\) and \(g^{N+1}\) being time differences, they are differences between adjacent images in latitude or longitude. For the latitude-based algorithm (see formula (2)), a new event is announced whenever the difference in latitude between two successive images is greater than the average latitude gap between successive images in the window, and for the longitude-based algorithm see formula (3), a new event is announced whenever the difference in longitude between two successive images is greater than the average for the window. In our initial implementation, \(K = 0\) for latitude and longitude-based clustering, and raw differences in latitude and longitude are used rather than their logarithms. The three variants of the event clustering formula are shown in Figure 1. We combine the results of the time, latitude and longitude event clustering as follows:

Two images are considered to represent the same overall event if and only if they have been found to be in the same time event, the same latitude event, and the same longitude event.

**Formula 1 Time clustering:**

\[
\log(g^N) \geq K + \frac{1}{(2d+1)} \sum_{i=0}^{d-1} \log(g^N+i)
\]

**Formula 2 Latitude clustering:**

\[
\text{Lat}g^N \geq \frac{1}{(2d+1)} \sum_{i=0}^{d-1} \text{Lat}g^N + i
\]

**Formula 3 Longitude clustering:**

\[
\text{Long}g^N \geq \frac{1}{(2d+1)} \sum_{i=0}^{d-1} \text{Long}g^N + i
\]

The formulas for time, latitude and longitude is shown above. If any one of the above conditions (formula 1 to 3) occurs, a new event has been located. Thus events are not only located by time, but also by location (see figure 1).

### 3.3 Location Name and Feature Annotation

The location name and feature annotation model annotates each image in the database by location name and other geographical features. These keywords are looked up in the gazetteer (see figure 4), which 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:

\[
\frac{1}{3} [ (\text{Lat}(i) - \text{Lat}(k))^2 + (\text{Long}(i) - \text{Long}(k))^2 ]
\]

(4)

to find the closest entry.

The image was then annotated with the full place name and DSG-code for this location. For example an image might be annotated with “London Buckingham Palace” and “Palace”.

---

**Figure 1. Location annotation indexing model**

### 3.4 Indexing and Cosine Similarity Matching

A small search engine was implemented to retrieve the images with the best-matching annotations with respect to a user query in the form of string of keywords. Stopwords (frequently occurring words such as “the”, “a”, or “of” which do not reflect the semantic content of the query) were removed, and Porter’s stemming rules were applied, so that different grammatical variants of a word would be regarded as equivalent. We used the Cosine Similarity Measure (Salton & McGill, 1983: p121) to determine the degree of match between the query and each image annotation, as follows:

\[
\text{COSINE(Image}_i, \text{Query}_j) = \frac{\sum_{i=0}^{d}(\text{Keyword}_{K} \cdot \text{Query}_{K})}{\sqrt{\sum_{i=0}^{d}(\text{Keyword}_{K})^2 \cdot (\text{Query}_{K})^2}}
\]

(5)

Here each image \(\text{Image}_i\) (see figure 1), is identified by a collection of Keywords \(\text{Keyword}_{1}, \text{Keyword}_{2} \ldots \text{Keyword}_{d}\), where \(\text{Keyword}_{i}\) is assumed to represent the weight, or importance, of Keyword (i) assigned to Image (i). A particular Query can be similarly identified as a vector \(\text{Query}_{K_{1}}, \text{Query}_{K_{2}} \ldots \text{Query}_{K_{d}}\), where \(\text{Query}_{K_{i}}\) represents the weight, or importance, of the Keyword (i) assigned to...
query (j). In this implementation, we took the weight of each keyword in each image to be the number of times it appeared in the annotation of that image.

3.5 User interface
Our user interface (see figure 2) 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. These query keywords are compared with each image annotation using the cosine similarity measure, and the 30 best matching images 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. During a previous experiment, Chen et al. (2006), some subjects suggested that the event separation needed to be more clear, and so we updated the event separation symbol to include “time”, “location” and “event” rather than only using the symbol “event”. 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 2. User interface

4. MAIN EXPERIMENT
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” or “small cat” as their own annotation. Free choice of vocabulary was allowed for the users’ annotation. We did not control the vocabulary to make the situation more real. The previous study 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 experiment compared three different systems: 1, automatic key word annotation system (described in section 3); 2, time and location clustering browser (developed by Chen et al. (2006)) and 3, user annotation systems (same structure and interface as automatic annotation system, but we used users own annotation instead of geography information).

4.1 Subjects
Nine volunteers participated in this study, six male and three female, all of whom participated in the previous study (Chen 2006). The image collections were also the same as those used in the previous study. They subjects were all staff and students at the University of Sunderland, and all had experience in managing their own digital photos. Each subject was asked to provide a personal collection of about 200 images, as shown in Table 1. On this occasion we also asked the subjects to manually annotate each image by no more than five key words.

Table 1. Collection size of each subject

<table>
<thead>
<tr>
<th>Subject</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collections</td>
<td>202</td>
<td>199</td>
<td>200</td>
<td>202</td>
<td>208</td>
<td>199</td>
<td>214</td>
<td>200</td>
<td>214</td>
</tr>
</tbody>
</table>

4.2 Hypotheses & Aim
The Hypotheses of this study were:
1. The user’s own annotation will not provide better support than automatic location name and feature annotation for personal image browsing and searching.
2. The location name annotation will provide better support than no annotation (time and location clustering browser).

4.3 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: 1. 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. 2. Please find a photo where you were standing in a crowded city centre. 3. Please find a photo where you were with your family in an indoor environment. 4. 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: 1. Please find the photo of a light house. 2. Please find the photo of a Christmas tree in the city centre. 3. Please find the photo of a black swan in the water. 4. 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 user annotation searching browser, system annotation searching browser and time and location clustering alone browser.

### 4.4 The Experimental method

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 for determine the level of user satisfaction with each browser. The subjects were asked to rate each browser on a 5-point Likert scale according to the following six criteria: 1. I like this image browser. 2. This browser is easy to use. 3. This browser feels familiar. 4. It is easy to find the photo I am looking for. 5. A month from now, I would still be able to find these photos. 5. I was satisfied with how the pictures were organised. This questionnaire had been used in the Platt et al. (2002) and Chen et al. (2006) user studies.

### 4.5 Recall and Precision experiment

We evaluated the retrieval performance for both system-selected keywords for annotation and user-selected keywords for annotation, according to the information retrieval measures of recall and precision (Salton & McGill, 1983: pp. 104-109), which can be combined as the F1 measure (van Rijsbergen, 1979). The tasks, general and specific, were similar to those used in section 4.3, with the difference that there was no single correct photo, but rather a set of photographs in the collection relevant to the query. Examples of general scenarios were: 1. Please find out all the photos where you were with one or two of your friends (or family members), outdoors on a sunny day during the last holiday. 2. Please find all the photos where you were standing in a very crowded city centre. 3. Please find all the photos where you were with your family in an indoor environment. 4. Please find photos of all the most famous buildings of a city which you have visited. Examples of specific scenarios were: 1. Please find all the photos of a lake. 2. Please find all the photos of Sunderland city centre during Christmas. 3. Please find all the photos of swans in the water. 4. Please find all the photos of the British Museum.

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 $F1 = (recall + precision) / (2 * recall * precision)$.

### 5. RESULTS

We compared three different searching browsers: time and location browser, user annotation and system annotation (location name and geographical feature annotation). We separated these three systems into 3 groups for comparison: group one, time and location browser vs. User annotation; group two, time and location browser vs. system annotation; group three, user annotation vs. system annotation. 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. Figure 3, 4, 5, is shown each subject’s general, specific and total searching time.

#### 5.1 Time and location browser vs. user annotation

The comparison of searching times in seconds (averaged over all 9 subjects) for each pair of browsers is shown in Table 2. 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 2. System searching time for time and location browser vs. user annotation

<table>
<thead>
<tr>
<th>Time &amp; location browser</th>
<th>User Annotation</th>
<th>t</th>
<th>p &gt; 0.017 significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. AVE/STDEV</td>
<td>51/18.1</td>
<td>4.31</td>
<td>p = 0.0026</td>
</tr>
<tr>
<td>searching time general scenario tasks</td>
<td>35.9/11.7</td>
<td>2.44</td>
<td>p = 0.059</td>
</tr>
<tr>
<td>2. AVE/STDEV</td>
<td>42.2/16.2</td>
<td>2.24</td>
<td>p = 0.059</td>
</tr>
<tr>
<td>searching time specific scenario tasks</td>
<td>32.3/10.2</td>
<td>1.84</td>
<td>p = 0.067</td>
</tr>
<tr>
<td>3. AVE/STDEV total finish time</td>
<td>93.2/28.2</td>
<td>4.18</td>
<td>p = 0.0031</td>
</tr>
</tbody>
</table>

5.2 Time and location browser vs. system annotation

Table 3 shows the comparison of time and location browsing alone and system annotation. Similar results were obtained as in Table 2. 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.

Table 3. System searching time for time and location browser vs. system annotation

<table>
<thead>
<tr>
<th>Time &amp; location browser</th>
<th>System Annotation</th>
<th>t</th>
<th>p &gt; 0.017 significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. AVE/STDEV</td>
<td>51/18.1</td>
<td>3.75</td>
<td>p = 0.0056</td>
</tr>
<tr>
<td>searching time general scenario tasks</td>
<td>36.2/14.3</td>
<td>2.84</td>
<td>p = 0.047</td>
</tr>
<tr>
<td>2. AVE/STDEV</td>
<td>42.2/16.2</td>
<td>2.12</td>
<td>p = 0.067</td>
</tr>
<tr>
<td>searching time specific scenario tasks</td>
<td>32.4/9.1</td>
<td>1.84</td>
<td>p = 0.067</td>
</tr>
<tr>
<td>3. AVE/STDEV total finish time</td>
<td>93.2/28.2</td>
<td>3.31</td>
<td>p = 0.011</td>
</tr>
</tbody>
</table>

5.3 User annotation vs. system annotation

There was no significant difference between the time required for searching with system annotation, than for user annotation. The results are shown in Table 4.

Table 4. System searching time for user annotation vs. system annotation

<table>
<thead>
<tr>
<th>User Annotation</th>
<th>System Annotation</th>
<th>t</th>
<th>p &gt; 0.017 significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. AVE/STDEV</td>
<td>35.87/18.1</td>
<td>-0.08</td>
<td>p = 0.938</td>
</tr>
<tr>
<td>searching time general scenario tasks</td>
<td>36.22/14.3</td>
<td>2.84</td>
<td>p = 0.047</td>
</tr>
<tr>
<td>2. AVE/STDEV</td>
<td>32.33/16.2</td>
<td>-0.04</td>
<td>p = 0.970</td>
</tr>
<tr>
<td>searching time specific scenario tasks</td>
<td>32.44/9.1</td>
<td>1.84</td>
<td>p = 0.067</td>
</tr>
<tr>
<td>3. AVE/STDEV</td>
<td>68.22/28.2</td>
<td>-0.1</td>
<td>p = 0.925</td>
</tr>
<tr>
<td>total finish time</td>
<td>68.67/16.6</td>
<td>2.84</td>
<td>p = 0.047</td>
</tr>
</tbody>
</table>

5.4 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 5. 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.

Table 5. User satisfaction for the three different systems.

<table>
<thead>
<tr>
<th>5.5 Recall and Precision Experimental Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time and location browser</td>
</tr>
<tr>
<td>---------------------------</td>
</tr>
<tr>
<td>1. I like this image browser</td>
</tr>
<tr>
<td>2. This browser is easy to use</td>
</tr>
<tr>
<td>3. This browser feels familiar</td>
</tr>
<tr>
<td>4. It is easy to find the photo I am looking for</td>
</tr>
<tr>
<td>5. A month from now, I would still be able to find these photos</td>
</tr>
<tr>
<td>6. I was satisfied with how the pictures were organized</td>
</tr>
<tr>
<td>Average</td>
</tr>
</tbody>
</table>

The results for Recall, Precision and the F1 measure for both the system annotation and user annotation are shown in Table 6.

Table 6. Recall and precision results

<table>
<thead>
<tr>
<th>5.5 Recall and Precision Experimental Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>System Annotation</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>Recall Precision</td>
</tr>
<tr>
<td>Recall Precision</td>
</tr>
<tr>
<td>General tasks</td>
</tr>
<tr>
<td>Specific tasks</td>
</tr>
<tr>
<td>Overall</td>
</tr>
</tbody>
</table>

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.

6. DISCUSSIONS & CONCLUSION

Our experimental 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. In the future we will 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.

7. REFERENCES


541