Generating Summaries for Large Collections of Geo-Referenced Photographs

Alexander Jaffe[1], Mor Naaman[1], Tamir Tassa[2], Marc Davis[1,3]

[1]Yahoo! Research Berkeley
1950 University Ave., Suite 200
Berkeley, CA 94704-1024
USA

[2]Department of Mathematics and Computer Science
The Open University of Israel
Ra'anana, Israel

[3]School of Information
University of California, Berkeley
Berkeley, CA 94720-4600
USA

ajaffe@yahoo-inc.com, mor@yahoo-inc.com, tamir_tassa@yahoo.com, marcd@yahoo-inc.com

ABSTRACT

We describe a framework for automatically selecting a summary set of photographs from a large collection of geo-referenced photos. The summary algorithm is based on spatial patterns in photo sets, but can be expanded to support social, temporal, as well as textual-topical factors of the photo set. The summary set can be biased by the user, the content of the user's query, and the context in which the query is made. An initial evaluation on a set of geo-referenced photos shows that our algorithm performs well, producing results that are highly rated by users.

Categories and Subject Descriptors

H.3.3 [**Information Systems**]: Information Storage and Retrieval – *Information Search and Retrieval*

General Terms

Algorithms, Design, Human Factors

Keywords

Geo-referenced photographs, photo collections, semantic zoom, photo browsing, collection summary, geo-referenced information

1. INTRODUCTION

With the popularization of digital photography, people are now capturing and storing far more photographs than ever before. These large collections of photographs are inherently difficult to navigate, due to their size and the inability of computers to understand the content. Yet, viewing and interacting with such collections has a broad social and practical importance. Today, we are moving towards Susan Sontag's 1977 vision [1] of a world where "everything exists to end up in a photograph". As a result, users are experiencing a content overload, in which the prospect of viewing some photo collections becomes genuinely infeasible.

Some steps forward have been made through geo-referencing of digital photographs [2,3]. An increasing number of photographs on the web are associated with the exact coordinates at which they were taken. Such geo-referenced photos can be categorized geographically or displayed on a digital map, providing a rich spatial context in which to view subsets of a collection. Yet even then, the viewable space inevitably becomes obscured after the data set has surpassed a certain size, with overlapping photographs making viewing impossible.

Copyright is held by the author/owner(s).
WWW 2006, May 23–26, 2006, Edinburgh, Scotland.
ACM 1-59593-323-9/06/0005.

Our goal is thus to facilitate a system which can automatically select representative and relevant photographs from a particular spatial region. This system will enable users to navigate through large-scale geo-referenced photo collections using map visualizations (and other spatial browsers) in a manner that *improves* rather than degrades with the addition of more photos.

Figure 1. Most representative photos of San Francisco, as chosen by our system. One photo is enlarged for illustration.

Selecting the most representative photos from a given region is a difficult task for several reasons. For instance:

- Image analysis alone is historically poor at capturing the semantic concepts in an image, making purely visual relevance insufficient for summarization.
- In multi-user sets, the biases of one user's data may weigh the data towards generally insignificant subjects.

We have designed and implemented a simple algorithm that attempts to satisfy the goals stated above. We use metadata-facilitated heuristics that capitalize on patterns in photography behavior. Foremost among these heuristics is the premise that photographs taken at a location imply the presence of something interesting. Previous work has suggested that frequencies of photography correspond to levels of interest over time [4]. However, our algorithm analyzes geographical rather than temporal frequency, and looks at a multitude of spatial-temporal-social as well as textual-topical patterns. Furthermore, the algorithm can be tuned to bias the results using various factors, such as the social network proximity of the photographers.

Our method can be used in a number of applications. A key example is that of *semantic zoom*. This is an interface for browsing photos from a global geo-referenced collection on a map. At any zoom level, panned to any region, the user sees a subset of photos that best represent that region. For example, Figure 1 shows a map of San Francisco, in which a limited set of eight photos selected by our system are marked on the map.

2. OUR METHOD

We first describe the insight into how we approach the problem of selecting summary photos from a region. We then describe an algorithm that implements these principles.

2.1 Basic Framework

To choose the most representative photos from a region we follow a few simple heuristics that try to model and capture human attention, as reflected in the set of photos taken in that region. Among these heuristics are the notions that:

- More photographs are taken at locations that provide views of some interesting object or landmark.
- A location is more relevant if the photos around it were taken by a large number of photographers.
- If available, textual tags can reflect the presence of interesting landmarks in a location, for example via a tag that appears frequently in that location, but rarely elsewhere.

In addition to following these heuristics, a desired solution could also follow these guidelines:

- In a summary, it is preferable to represent a broad range of subjects than to thoroughly display a few.
- Personal or query bias should be easily introduced to modify the algorithm's results.

Indeed, our system is flexible enough to allow users to specify a personal bias, or preference, towards any of the features above or to utilize implicit bias in the query context and history. We allow for biasing by any quantifiable feature of the photographs, such as the social proximity of the photographer to the user, preference for weekend photos or tourist photos, or the externally determined quality of the image.

2.2 Algorithm

We formalize our problem as that of producing a ranking on the collection in question. In other words, we summarize a set of photos by ordering the set and selecting the top ranked photos. Producing an ordering is a two-step process.

1. We apply a modified version of the Hungarian Clustering Method [4] to our collection of photographs. The clustering assigns every photo to one of several clusters, based solely on the geographic locations of the photos. The clustering algorithm generates a hierarchical cluster structure; other hierarchical clustering methods could potentially work here as well.

2. We recursively rank the photos of a cluster based on orderings of its subclusters. At the leaf cluster level, photos are assigned a basic relevance score that could reflect any parameterized bias. For clusters in the higher level of the hierarchy, the scores incorporate several factors, including the summed weights of their subclusters, inter-cluster variance among tags and users, and cluster density. An ordering of photos in a cluster is generated by interleaving photos from its subclusters. We first select a small number of photographs from each of the subclusters. We then go into depth, selecting photos in proportion to the subclusters' scores.

Though the Hungarian Clustering Method runs in worst case $O(n^3)$ time, the resultant hierarchy can be cached, allowing new queries to be run in $O(n \log n)$ time.

2.3 Initial Evaluation

The goal of our evaluation was to determine whether our system generates a satisfactory selection of photos from a given spatial region. Our dataset consisted of approximately 1100 photos from Flickr [5], taken in San Francisco and associated with exact location coordinates. This number appears to be sufficient for effectively ranking photos at the city scale. We focus on the user experience of the results, and compare them to results achieved by three baseline systems: a) recency, b) random selection and c) Flickr 'interestingness'. Recency simply selects the most recent photographs from a region. Flickr interestingness is the website's measure of the attention given by other users to photographs. We used two different random selections for a total of five conditions, including our system.

The implementation of our algorithm used for the evaluation calculates cluster weights based on three criteria: number of photos, local repetition of tags, and local diversity of photographers. We performed an inter-subject evaluation of the results from the different methods. We presented 15 users - each having some familiarity with San Francisco - with the top nine photographs from each of the five summarizations. Users were asked to rate each result on a 1—9 Leikert scale for a number of different criteria: relevance to the city, attractiveness of the photos, usefulness for showing the city to a friend, and the extent to which the entire city is represented.

2.4 Results

For lack of space, we supply a brief overview of the results here. Our method's summary received better scores than all others, on each of the criteria, in a statistically significant manner. The only exception is 'attractiveness', for which the 'interestingness' condition did equally well. Most importantly, for the 'relevance' criterion, the mean score for our algorithm was 6.79. The closest condition was 'interestingness', with 5.53, ($n=15$, $p<.05$). The score for the two random conditions and recency were 4.8, 5.27, and 2.93 respectively.

3. CONCLUSIONS

The phenomenal growth of personal and shared digital photo collections presents considerable challenges in building navigation and summarization applications. By utilizing our hierarchical clustering algorithm, which can be parameterized by user and contextual bias, we enable users to view the most relevant samples from large-scale photo collections, with little to no effort.

4. REFERENCES

[1] Susan Sontag. *On Photography*. Picador, 1977.

[2] Toyama, K., Logan, R., and Roseway, A.. *Geographic location tags on digital images*. In Proc., ACM MM 2003.

[3] Naaman, M., Song, Y.J., Paepcke, A., and Garcia-Molina, H. *Automatic Organization for Digital Photographs with Geographic Coordinates*. In Proc., JCDL 2004.

[4] Nair, R., Reid, N., and Davis, M. *Photo LOI: Browsing Multi-User Photo Collections.* In Proc., ACM MM 2005.

[5] Goldberger, J. and Tassa, T. *The Hungarian Clustering Method*. Submitted for publication.

[6] Flickr. http://www.flickr.com