I2RS: A Distributed Geo-Textual Image Retrieval and Recommendation System

Lu Chen[†]   Yunjun Gao[†,*]   Zhihao Xing[†,*]   Christian S. Jensen[‡]   Gang Chen[†]

[†]College of Computer Science, Zhejiang University, Hangzhou, China
[*]Innovation Joint Research Center for iCPS, Zhejiang University, Hangzhou, China
[‡]Department of Computer Science, Aalborg University, Aalborg, Denmark
[†]{luchen, gaoyj, xingzh, cg}@zju.edu.cn   [‡]csj@cs.aau.dk

ABSTRACT

Massive amounts of geo-tagged and textually annotated images are provided by online photo services such as Flickr and Zommr. However, most existing image retrieval engines only consider text annotations. We present I2RS, a system that allows users to view geo-textual images on Google Maps, find hot topics within a specific geographic region and time period, retrieve images similar to a query image, and receive recommended images that they might be interested in. I2RS is a distributed geo-textual image retrieval and recommendation system that employs SPB-trees to index geo-textual images, and that utilizes metric similarity queries, including top-m spatio-temporal range and k nearest neighbor queries, to support geo-textual image retrieval and recommendation. The system adopts the browser-server model, whereas the server is deployed in a distributed environment that enables efficiency and scalability to huge amounts of data and requests. A rich set of 100 million geo-textual images crawled from Flickr is used to demonstrate that, I2RS can return high-quality answers in an interactive way and support efficient updates for high image arrival rates.

1. INTRODUCTION

Online photo services such as Flickr and Zommr allow users to share their photos with family, friends, and online communities at large. As the availability of rich media annotations is essential for large-scale image retrieval and recommendation systems to work in practice, numerous web tools ([2, 5, 7], Google, Yahoo! and MSN Live Search) integrate image and text information, to offer additional image retrieval services (e.g., search by text), or to improve the quality of retrieved or recommended results. However, many existing image retrieval and recommendation engines (as mentioned above) only consider text annotations. Due to the proliferation of devices for the capture and sharing of images, e.g., geo-positioned smartphones with cameras, additional metadata is increasingly available for images [6, 8]. Motivated by this development, we support the geo-location, text, and time annotations of images for image retrieval and recommendation.

We demonstrate I2RS, a distributed geo-textual image retrieval and recommendation system that supports two types of metric similarity queries: (1) *top-m spatio-temporal range retrieval*,

where users request a set of hot topics in the specified geographic region and time period, and (2) *k nearest neighbor* (*k*NN) *search*, that returns images similar to a given query image, e.g., an image uploaded by a user or an image in I2RS. In addition, I2RS can utilize *k*NN search to generate recommended images for users, as *k*NN retrieval is an intuitive solution for recommender systems [4]. Hence, I2RS enables users to find out hot topics in a specific spatio-temporal region, to retrieve images similar to query images, and to receive the recommended images that they might be interested in.

A system such as I2RS must support very large collections of images, high image arrival rates, and fast query response. To achieve these, I2RS integrates the following three features.

Scalable distributed deployment, to efficiently manage a very large and growing collection of geo-textual images and to support heavy query workloads. We partition the images to ensure the load balancing and to be able to process queries in parallel. In addition, the distributed environment enables system robustness.

Flexible indexing that is able to efficiently index locations and image features of geo-textual images. We utilize the state-of-the-art SPB-tree [3], a metric index that is applicable to any type of data with a similarity measurement satisfying the triangle inequality. Also, the SPB-tree supports efficient updates so that high image arrival rates can be supported.

Efficient query processing, to support fast query response. We use the triangle inequality property to develop pruning rules [3], and we employ approximation techniques that exploit the proximity preservation property of space-filling curves to improve query efficiency and ensure high quality of results.

To sum up, our key contributions are as follows:

- I2RS utilizes multiple types of information that is typically attached to the images, including locations, text, and time, to offer a variety of useful image retrieval and recommendation services.
- I2RS employs the state-of-the-art SPB-tree, which is applicable to a broad range of object types and similarity metrics, and that supports efficient updates. I2RS also uses efficient metric similarity search algorithms.
- I2RS is deployed in a distributed environment, to enable scalability in terms of image collection size, image arrival rate, and query workloads.
- We demonstrate I2RS using a collection of 100 million geo-textual images crawled from Flicker [1], with high simulated query and update workloads.

The rest of the demonstration proposal is organized as follows. Section 2 defines the data and queries supported. Section 3 presents the I2RS prototype. Section 4 provides demonstration details. Finally, Section 5 concludes the demonstration.

Figure 1. I2RS Architecture

Figure 2. I2RS Server Distributed Deployment

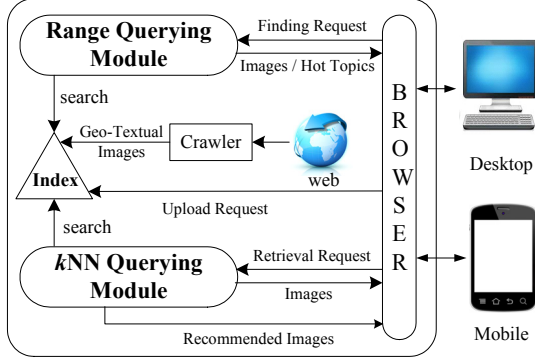## 2. DATA MODEL AND QUERIES

In this section, we define the geo-textual images and the metric similarity queries considered.

DEFINITION 1 (GEO-TEXTUAL IMAGE). *A geo-textual image o is represented as a tuple $o = \langle S, p, t, I \rangle$, where S is a set of text annotations, p is a location, t is a creation time, and I is an image.*

In this demonstration, five MPEG-7 visual descriptors extracted from an image content are used to represent the image I: (1) *scalable color*, a vector of 64 coefficients, where the distance is defined by the L_1 norm; (2) *color structure*, a structuring element of 8×8 pixels, where the L_1 norm is also used; (3) *color layout*, a vector of 12 coefficients, where the L_2 norm is employed; (4) *edge histogram*, a vector of 80 coefficients (5 coefficients for each of 16 sub-images), where the distance is the sum of the weighted sub-sums of absolute differences; and (5) *homogeneous texture*, a vector of 62 coefficients, where the distance is defined as the absolute value of the sum of weighted differences. The similarity between two images can then be defined as a weighted sum of five individual descriptors' distance notions [1].

Since the geo-textual images involve several object types (such as locations, images, and text) and distance metrics (such as the L_P-norm and weighted sum combinations), we introduce a general metric space model that treats the data as unstructured objects together with a function that measures the similarity of object pairs. A metric space is a tuple (M, d), in which M is the domain of objects and d is a distance function that defines the "similarity" between objects in M. In particular, the distance function d has four properties: (1) *symmetry*: $d(q, o) = d(o, q)$; (2) *non-negativity*: $d(q, o) \geq 0$; (3) *identity*: $d(q, o) = 0$ iff $q = o$; and (4) *triangle inequality*: $d(q, o) \leq d(q, p) + d(p, o)$.

We support two types of metric similarity queries, i.e., top-*m* spatio-temporal range retrieval and *k* nearest neighbor (*k*NN) search. To define the top-*m* spatio-temporal range query, we first introduce the spatio-temporal range query.

DEFINITION 2 (SPATIO-TEMPORAL RANGE QUERY). *Given a geo-textual image set O, a query object q, a search radius r, and a time period T, a spatio-temporal range query returns the geo-textual images o in O that are within distance r of q and have a creation time that belongs to T, i.e., $STRQ(q, r, T) = \{o \mid o \in O \wedge d(q, o.p) \leq r \wedge o.t \in T\}$.*

DEFINITION 3 (TOP-*M* SPATIO-TEMPORAL RANGE QUERY). *Given a geo-textual image set O, a query object q, a search radius r, and a time period T, a top-*m* spatio-temporal range query*
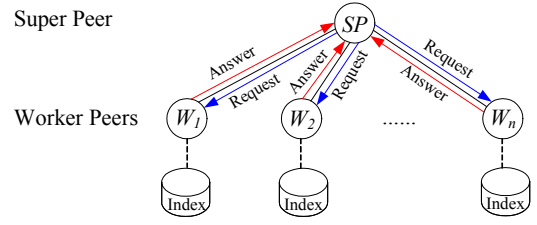
finds *m* most frequent text annotations contained in the geo-textual images returned by $STRQ(q, r, T)$, i.e., $mSTRQ(q, r, T) = \{S_r \mid |S_r| = m \wedge S_r \subseteq S_{RQ} \wedge \forall s_i \in S_r, \forall s_j \in S_{RQ} - S_r, F(s_i) \geq F(s_j)\}$, where $F(s_i)$ denotes the occurrence frequency of s_i and $S_{RQ} = \{o.S \mid o \in STRQ(q, r, T)\}$.

For a spatio-temporal range query $STRQ(q, r, T)$ and a top-*m* spatio-temporal range query $mSTRQ(q, r, T)$, the query object is actually a location, and $d(q, o.p)$ is Euclidean distance (L_2-norm).

DEFINITION 4 (*K*NN QUERY). *Given a geo-textual image set O, a query object q, and an integer k, a kNN query finds from O the k geo-textual images most similar to q, i.e., $kNN(q, k) = \{O_r \mid O_r \subseteq O \wedge |O_r| = k \wedge \forall o_i \in O_r, \forall o_j \in O - O_r, d(q, o_j.I) \geq d(q, o_i.I)\}$.*

For a *k*NN query $kNN(q, k)$, the query object is actually an image and $d(q, o.I)$ denotes image similarity, which is computed by using the weighted sum of distance between MPEG-7 features.

3. I2RS PROTOTYPE

We cover the framework of I2RS and then present the indexing and query processing in I2RS.

3.1 I2RS Framework

I2RS adopts the browser-server model. The architecture is depicted in Figure 1, which includes two querying modules, centered around metric indexes. Geo-textual images are crawled from the web (e.g., Flicker) and are then indexed by the index module. Users submit their requests through a web browser on a desktop or mobile device. The requests are sent to the index module, the range or *k*NN querying module accordingly. For a geo-textual image upload request, I2RS involves the index module; for a hot topics finding request, I2RS involves the range querying module; and for a geo-textual image retrieval request, I2RS involves the *k*NN querying module. In addition, the *k*NN querying module can automatically generate personalized recommended images for users. Finally, query results are sent back to the user who issues the query, and are displayed on Google Maps in the user's browser.

In order to achieve the scalability and robustness, we deploy the server side of I2RS in a distributed environment, as illustrated in Figure 2. The I2RS server distributed deployment is a two layer structure on a peer-to-peer (P2P) overlay network, which consists of two classes of peers, i.e., super peers and worker peers. For simplicity, we only use one super peer in this demonstration, although it can be easily extended to multiple super peers. Each worker peer maintains its own data. For every request (e.g., upload, finding or retrieval request) sent to a super peer, the super peer distributes it to the qualified worker peers, and then waits for the answers returned by the worker peers. Finally, the super peer ranks all the returned answers and sends them back to the users.
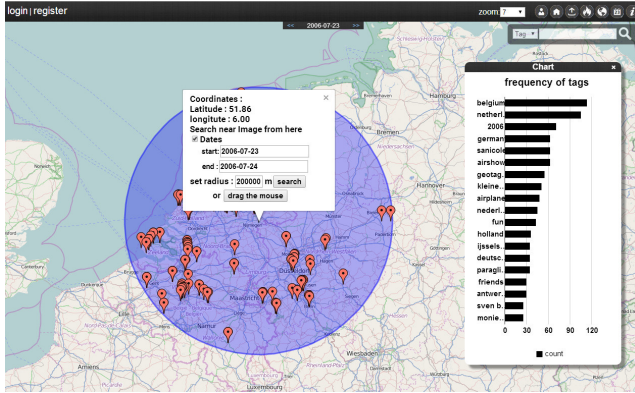
Figure 3. Geo-Textual Image Browsing

Figure 4. Geo-Textual Image Retrieval

3.2 Indexing Geo-textual Images

We utilize the SPB-tree [3] to index geo-textual images crawled from the web. A SPB-tree used to index an object set in a generic metric space contains three parts, i.e., the pivot table, the B^+-tree, and the random access file (RAF). A pivot table stores selected objects to map a metric space into a vector space. A B^+-tree is employed to index the integer values of objects after a pivot mapping and a space-filling curve mapping. A RAF keeps objects separately and supports both random access and sequential scan. In this demonstration, we build SPB-trees on locations and images of the geo-textual images, respectively.

In order to ensure the load balancing of I2RS, we need to partition the geo-textual image set into multiple equal-size disjoint parts. Hence, for locations of the geo-textual images, we build one location SPB-tree for every time unit (one day is chosen as a time unit in this demonstration), and then deploy those location SPB-trees uniformly among the worker peers of I2RS, to guarantee that each worker peer maintains similar number of location SPB-trees. In addition, since users usually prefer searching images in a continuous time period, the time units of the location SPB-trees stored in the same worker peer are continuous.

For the geo-textual images to be indexed, as image features extracted can be mapped into one-dimensional integers when constructing the SPB-tree, we can split the whole one-dimensional space into several disjointed segments, to ensure the numbers of the corresponding images contained in every segment are similar. Then, we build one image SPB-tree for each segment. Note that, according to the proximity preservation of space-filling curves, images contained in the same segment and the nearby segments are more likely to be similar than those in others.

Index update. For a new coming geo-textual image $o = \langle S, p, t, I \rangle$ uploaded by a user, we find the worker peer W_i that maintains the location SPB-tree L with its time unit containing $o.t$, and then, we send o to W_i and insert $o.p$ into L. Note that, if there does not exist such location SPB-tree, we take a lazy strategy to just find a worker peer W_i that has free space and maintains the location SPB-tree with its time unit closest to $o.t$. Then, we create a new location SPB-tree L on W_i for the time unit containing $o.t$ and insert $o.p$ into L. After a certain time period (e.g., a month), we can re-distribute the location SPB-trees to ensure that each worker peer maintains similar number of SPB-trees and the time units of the SPB-trees stored in the same worker peer are continuous.

To insert $o.I$ of the new coming geo-textual image o into the corresponding image SPB-tree, we first compute the integer v of
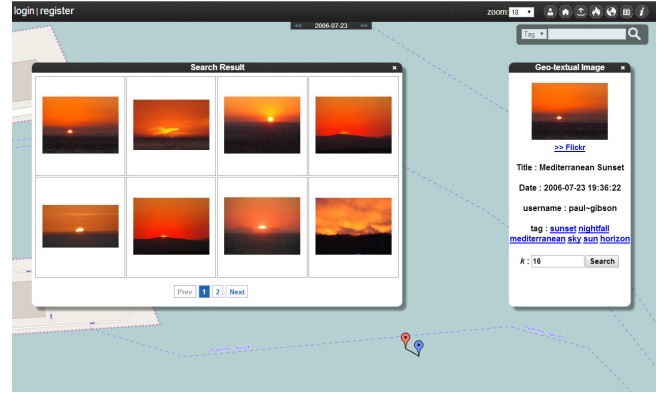
$o.I$ according to the pivot and space-filling curve mappings, and then find the worker peer W_j that maintains the segment containing v. Thereafter, we send o to W_j, and add $o.I$ to the corresponding image SPB-tree.

3.3 Query Processing

IR2S supports two types of metric similarity queries.

Top-m Spatio-Temporal Range Query. A top-m spatio-temporal range query $mSTRQ(q, r, T)$ request returns m most frequent text annotations contained in a specified spatio-temporal region. In order to efficiently answer the request $mSTRQ (q, r, T)$ issued by the user and sent to the super peer, we first find qualified worker peers that maintain location SPB-trees with their time units contained in T, and send the request to those worker peers. After that, a range query [3] is performed on every location SPB-tree with its time unit contained in T, and the query results are then returned to the super peer. Finally, the super peer sums up all the geo-textual images returned by the quailed worker peers, and sends them back to the user. In particular, the super peer sorts the text annotations of the returned geo-textual images based on their occurrences' frequency, and sends the top-m text annotations to the user (m is set to 20 in this demonstration).

kNN Query. A kNN query kNN(q, k) request retrieves k geo-textual images o that are most similar to a given query image q. In order to efficiently answer the retrieval request kNN(q, k) issued by the user and sent to the super peer, we first compute the integer v of q using the pivot and space-filling curve mappings, and then find the segment containing v and w nearest segments (w is set to 10 in this demonstration). According to the proximity preservation of the space-filling curve, we only send the kNN(q, k) request to the $w + 1$ worker peers that maintain image SPB-trees of the corresponding segments. Thereafter, kNN queries [3] are performed on the SPB-trees of the worker peers who receive the retrieval request. Finally, the super peer ranks all the query results returned by $(w + 1)$ worker peers based on the similarity between images, and sends the top-k geo-textual images back to the user.

In this demonstration, kNN search can also be used to generate personalized geo-textual images for users according to the images browsed most recently by each particular user. In order to efficiently answer this recommendation request issued by the system itself, we conduct multiple kNN queries for each of n images that the user viewed most recently (n is set to 20 and k is set to 5 in this demonstration), with the kNN query processing discussed as above. Finally, the super peer ranks all the kNN
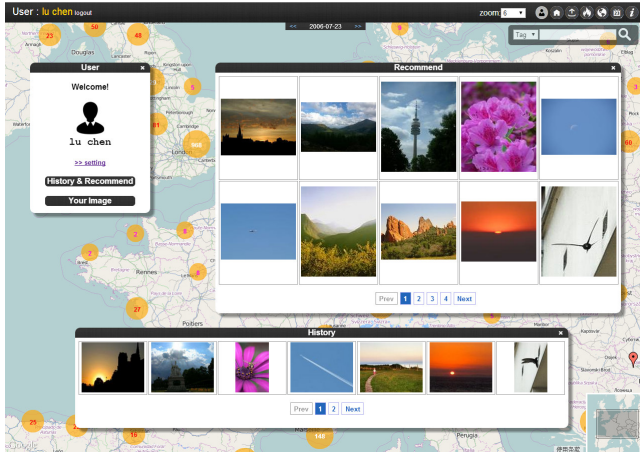
Figure 5. Geo-Textual Image Recommendation

Figure 6. Geo-Textual Image Upload

query results returned by the qualified worker peers according to the similarity between images, and sends them to the user.

4. DEMONSTRATION

I2RS is demonstrated by using 100 million geo-textual images crawled from Flicker [1]. The demo uses a simulated distributed environment, with one super peer and 10K worker peers, on an Intel Core i7 3.6GHz PC of a 16GB memory. The demo participants can interact with I2RS through the following five scenarios in an interactive manner.

Geo-Textual Image Browsing. I2RS compacts the geo-textual images into multiple geographic regions (yellow circle shown in Figures 5 and 6). Participants can browse images on Google Maps.

Hot Topics Finding. Participants can issue a hot topics finding request, as exemplified in Figure 3. To issue a hot topics finding request, a participant (1) selects a center of a geographic area on the Google Maps and inputs a radius, (2) specifies a time period, and (3) clicks the search button. Then, the browser submits this request to the server side of I2RS. Once the hot topics and the geo-textual images are returned, the browser shows them to the user on the map and in the right-side panel. As depicted in Figure 3, each red position icon indicates an image result, and users can browse this image by clicking the icon. Hot topics are shown in the right-side panel. In this example, the hot topics (e.g., "belgium", "germany", "airshow") mean the countries and actives of the images' potential contents.

Geo-Textual Image Retrieval. Participants are able to issue geo-textual image retrieval requests, as exemplified in Figure 4. To issue a request when viewing an image, a participant (1) specifies a required number in the textbox, and (2) clicks the search button in the right-side panel. The left-side panel shows the results. In Figure 4, 16 similar images are returned. The user can also view detailed information by clicking the images. Since an approximate technique is used for the retrieval request, the accuracy of the returned answer set $O_r = \{r_i \mid 1 \leq i \leq k\}$ can be computed as $\frac{1}{k} \sum_{i=1}^{k} |d(q, r_i) - d(q, r'_i)| / d(q, r'_i)$ where q is a query image and O_r' = $\{r'_i \mid 1 \leq i \leq k\}$ denotes the real answer set, and the average accuracy of a retrieval request is 96%.

Geo-Textual Image Recommendation. I2RS demo is able to perform a geo-textual image recommendation for registered users according to the images browsed most recently. Figure 5 gives a recommended result example for the user named as 'lu chen',
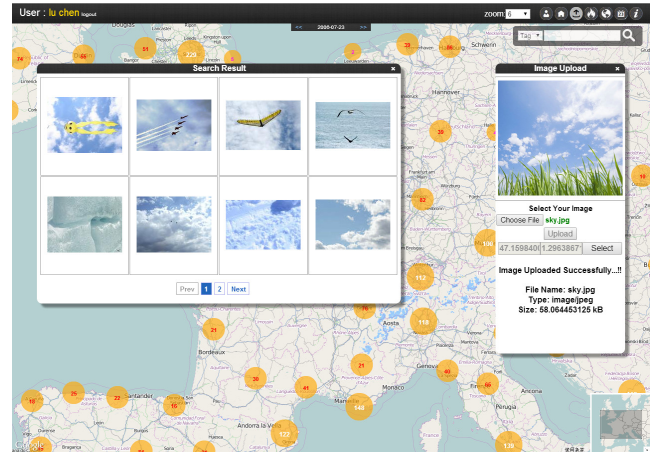
where the lower panel shows the images viewed most recently by the user, and the upper panel shows the recommended images.

Geo-Textual Image Upload. Demo participants can upload new geo-textual images, as depicted in Figure 6. To upload a geo-textual image, a user (1) chooses the image to be uploaded, and (2) selects an upload location on the map. The upload time can be automatically obtained by the system. After the image upload, the system can also find similar images for the uploaded image, which is illustrated in the left-side panel.

5. CONCLUSION

In this demonstration, we present I2RS, a distributed geo-textual image retrieval and recommendation system, which utilizes the images and their corresponding geo-location, text, and time annotations crawled from Flickr. In the future, we try to extend I2RS to provide more useful services.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] P. Bolettieri, A. Esuli, F. Falchi, C. Lucchese, R. Perego, T. Piccioli, and F. Rabitti. CoPhIR: A test collection for content-based image retrieval. CoRR abs/0905.4627, 2009.

[2] M. Batko, F. Falchi, C. Lucchese, D. Novak, R. Perego, F. Rabitti, J. Sedmidubsky, and P. Zezula. Building a web-scale image similarity search system. *Multimedia Tools Appl.*, 47(3):599–629, 2010.

[3] L. Chen, Y. Gao, X. Li, C. S. Jensen, and G. Chen. Efficient metric indexing for similarity search. In: *ICDE*, 519–602, 2015.

[4] D. Conway and J. M. White. Machine learning for hackers. O'Reilly Media, February 2012.

[5] J. Fan, D. A. Keim, Y. Gao, H. Luo, and Z. Li. JustClick: Personalized image recommendation via exploratory search from large-scale Flickr images. *IEEE Trans. Circuits Syst. Video Techn.*, 19(2):273–288, 2009.

[6] M. H. Memon, A. Khan, J.-P. Li, R. A. Shaikh, I. Memon, and S. Deep. Content based image retrieval based on geo-location driven image tagging on the social web, In: *ICCWAMTIP*, 280–283, 2014.

[7] SAPIR, http://sapir.isti.cnr.it/.

[8] Y. Yang, P. Wu, C. Lee, K. Lin, W. H. Hsu, and H. Chen. ContextSeer: Context search and recommendation at query time for shared consumer photos. In: *MM*, 199–208, 2008.