Interactive Social, Spatial and Temporal Querying for Multimedia Retrieval

Gamhewage C. de Silva and Kiyoharu Aizawa
Interfaculty Initiative in Information Studies, the University of Tokyo, Japan
{chamds,aizawa}@hal.t.u-tokyo.ac.jp

Yuki Arase and Xing Xie
Microsoft Research Asia, China
{yukiar,xing.xie}@microsoft.com

Abstract

We propose a scheme for faster and more effective retrieval of temporal, spatial and social multimedia from large collections. We define interactive multimedia queries that allow simultaneous query refinement on multiple search dimensions. User interaction techniques based on line and iconic sketches allow specifying queries based on the above definition. We prototype a multi-user travel media network and implement the proposed user interaction techniques for retrieving locomotion patterns of the users. The proposed queries facilitate easy input and refinement of queries, and efficient retrieval.

1. Introduction

Online multimedia collections can be divided into three main categories according to the dimension along which they are distributed. Temporal multimedia collections such as blogs and news archives are distributed over time. Spatial multimedia collections such as online Maps and *Street View* panoramas span a large geographical area. The third and the latest category is Social multimedia collections, such as multimedia hosted on social networking services, that are distributed among groups of users based on their relationships. By combining collections from different categories, Internet services that provide the users with rich multimedia experiences can be created.

The large domains and steady growth of these multimedia collections call for efficient search strategies for retrieval of relevant multimedia. However, the existing user interaction techniques are not sufficiently rich for effectively querying and interacting with multiple types of media. For example, consider the query "find photos that I and my friends took around *German Corner* last year." Given the search domains this includes, it can be termed a socio-spatio-temporal query. While this is a natural query from the user's viewpoint, the existing user interaction techniques are unable to capture the imprecise and subjective nature of this query (for instance, there is no method to precisely and quickly specify what is meant by "around"). Therefore, it will require several iterations of search with a number of refinements, and manually browsing the results, to retrieve the right images. Designing interactive socio-spatio-temporal queries to quickly retrieve relevant content from large multimedia collections therefore solves an important research problem.

In this paper, we present an interactive querying strategy for querying a large collection of spatial, temporal and social multimedia. First, we define multimedia queries with support for simultaneous refinement in multiple dimensions. Thereafter, we design user interaction to combine spatial, temporal and social queries according to the above definition. Finally, we implement a prototype multimedia application that demonstrates the effectiveness of the proposed query definition and interaction techniques.

The rest of this paper is organized as follows. Section 2 contains a brief review of related work. Section 3 formally defines the proposed interactive queries. Section 4 describes the design of user interaction for entering and refining queries. Section 5 presents *TrailNet*, the prototype application. Section 6 concludes the paper, outlining future research directions.

2. Related work

There has been a large amount of research on multimedia information retrieval, during the past few years [2]. Most of these researches focus on the algorithmic aspects of retrieval. However, research related to improved interaction techniques for searching has been growing in the last couple of years. In addition to commonly used text-based queries, different types of visual queries based on time lines [7], images [11] and sketches [4, 3] have been successfully used for direct search or refining search results.

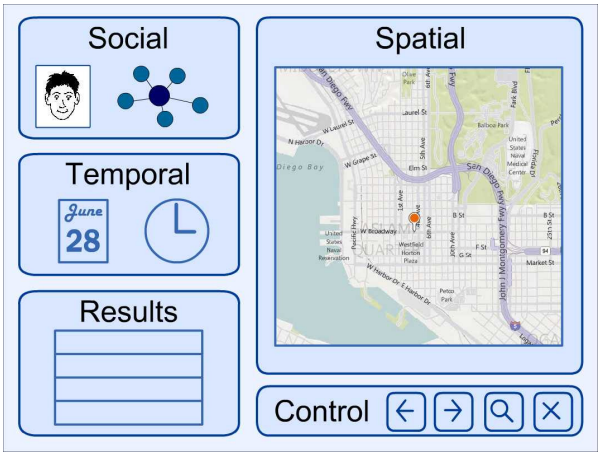

Figure 1: Interface layout for facilitating social, spatial, and temporal queries.

There exist a large number of multi-user Internet services that allow storing and retrieval of spatial multimedia. Geocaching [8] and EveryTrail [6] are two examples. Most of these services facilitate retrieval using text based search followed by browsing a map with results overlaid on it. This becomes time consuming and tedious once the geographical area or the amount of data is large.

There has been some research on interaction techniques that allow faster searching within the social dimension. Lee et al. [12] used a tree-like metaphor to represent social relationships for browsing and querying online communities. Zhang et al. [13] proposed creation of face clusters for image navigation in personal photo collections. However, this method is mostly used as a visualization technique that speeds up browsing intermediate results, not as a means of specifying initial queries.

3. Query Formulation

3.1 Conventional multimedia queries

Multimedia retrieval from a large collection is usually composed of three steps. They are:

1. Querying: the user enters the search criteria
2. Response: the system extracts the best matches for the criteria, orders them by relevance, and displays the results to the user
3. Searching/Surfing/Browsing [2]: the user goes through the results to find what he/she was looking for, or interesting content

The above three steps are quite distinct when it comes to conventional document retrieval based on text queries. The user makes a query by typing some keywords in a text box and clicking a button. The results are displayed as a list of links (often augmented with thumbnails, summaries and other metadata) to relevant documents. The user can scroll through the result pages, and click on the links to look at the results. The visualization of the response usually includes the text box to allow the user to repeat/refine the search.

However, using text for querying is not always efficient for location-based multimedia retrieval. Queries such as "photos of me and my friends, taken around Paris last year"(with social, spatial and temporal dimensions) are lengthy and ambiguous if only text is used as an input. Further, the responses to such queries span multiple dimensions (i.e. space, time, people), making it harder to browse the results. The current solution for this problem is to start with a shorter text query (like "Paris") and then browse the results.

A conventional multimedia database query Q can be defined as $Q(Y_1, Y_2, ...Y_n)$ and returns an ordered list of results S_Q of the form $S_Q(X_1, X_2, ...X_n)$ [1]. Here, Y_i are the dimensions of the query and X_i are the values of those dimensions for each item in S_Q. While this formulation is useful for the design and analysis of retrieval algorithms, it is not sufficient for defining user interaction for searches over multiple dimensions, for the following reasons:

1. For a visualization of results with multiple search dimensions (for example, a map, time line, and users), an ordered list (one-dimensional) is inadequate.
2. The formulation does not allow us to define a query within the search results. If a user submits multiple queries to retrieve a particular document, each query is treated independent of the others.

To allow more efficient retrieval, we suggest forming multimedia queries that allow interactive refinement. Such refinement should help the user to narrow down his/her search in each query dimension independent of the others. The following subsection defines and describes *interactive multimedia queries* that we propose.

3.2 Interactive multimedia queries

We define an interactive multimedia query Q as $Q(S_1, S_2, ...S_n)$ where each S_i is a subset of a set ζ_i that corresponds to each search dimension. This query returns an ordered list of multimedia $L_Q(X_1, X_2, ...X_n)$ and a sub-domain $R(S'_1, S'_2, ...S'_n)$ where $\cup S'_i$ is the smallest subset of S_i that contains L_Q. A refined query can now be applied on R, instead of the universal set $\cup\zeta_i$. The interface for querying can be designed to show each ζ_i separately, and be updated with S'_i when results are returned. The steps for a multimedia retrieval task using this query definition are:

1. Querying: the user enters search criteria in one or more query dimensions
2. Response: the system shows the results matching the criteria to the user, together with the sub-domain of each query dimension

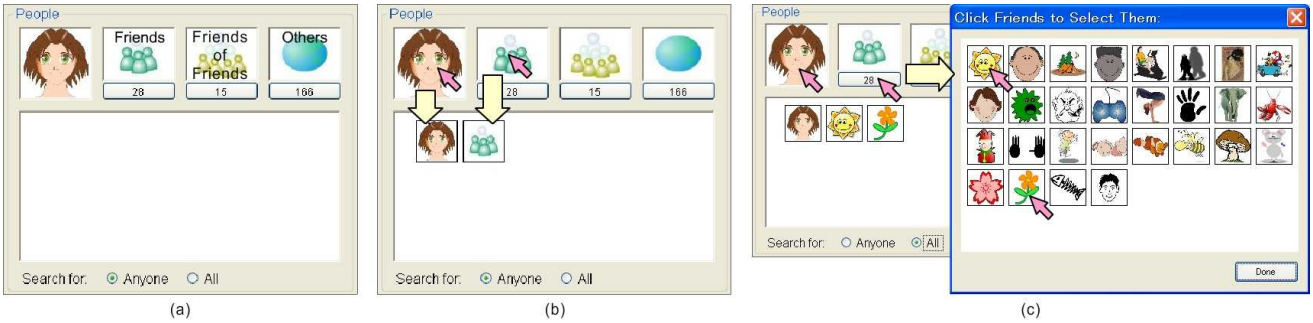

Figure 2: User interaction for social querying: (a) interface layout; (b) coarse selection; (b) fine selection.

3. Refinement: if the set of results is too large for manual search, the user further refines search criteria for sub-domains iterating between steps 2 and 3

4. Searching/Surfing/Browsing

The following is a possible scenario of using the above steps to answer the query "photos of me and my friends, taken around Paris last year." Let us assume that interaction mechanisms for querying spatial, temporal and social dimensions exist. The user first selects "last year", and "around Paris", as temporal and spatial criteria respectively. This returns a list of photos, with the following:

- Spatial sub-domain: a neighborhood around Paris, covering the area where the photos were taken
- Temporal sub-domain: a time interval in the previous year, during which the photos were taken
- Social sub-domain: persons appearing in the photos

The user should now be able to refine the neighborhood and the time interval as necessary, and also specify the relevant persons using the sub-domains. This can be repeated if necessary, until the list of results is sufficiently small.

The concept of interactive refinement, as described above, is quite simple. However, its usability for multimedia retrieval depends on the availability of user interaction mechanisms for querying each dimension. There should be intuitive and nonrestrictive ways to repeatedly refine different types of search criteria. The returned sub-domains should be easy to understand. Therefore, user interaction design is an essential component for implementing interactive multimedia queries defined above. The following section presents the interaction techniques we propose for querying social, spatial and temporal dimensions.

4. User Interaction Design

We propose a design where the social, spatial and temporal components of a query are entered in separate regions of the user interface. The sub-domains are displayed on the same regions, resulting in a retrieval process with an "enhanced browsing" experience. The user can iteratively refine the search in each dimension, instead of browsing all the results. Ability to navigate "back" and "forward", along the iterative query - with the same effect as in web browsing - is provided to make querying easier.

Figure 1 illustrates the proposed interface layout. The user can enter each type of query in the corresponding region. The regions with no input will be updated to show the domain of results. For example, specifying a set of users and a time interval and executing a search will result in changing the map to show the area where results are available, with thumbnails and markers for visualization. The following subsections describe the design of interaction for each query dimension.

4.1. Social Querying

We propose an icon-based interaction technique for forming social queries. A user is represented with an icon. This is intuitive due to its similarity to the use of profile pictures for visualizing users in social networking sites. We also added three extra icons to represent groups of users according to their social relationship with the current user; *friends*, *friends of friends*, and *others*.

Figure 2a illustrates the main components of this interface. The four icons on the top row, from left to right, represent the current user("myself"), friends, friends of friends and other users respectively. The button below each icon indicates the number of users in that group. The blank *canvas* below these icons is used for forming a visual query.

Figure 2b demonstrates how a simple query can be formed using the above interface. The user clicks on the icons corresponding to him/herself and his/her friends, to add the icons to the canvas. The radio button is set to "anyone". The query thus formed can be interpreted as "myself or any of my friends."

The query shown in Figure 2a is coarse, in the sense that it deals only with preset groups of users. Figure 2c shows how the interface is used to form a more specific query. After selecting him/herself, the user clicks on the button under the "friends" icon. This opens a window showing all the

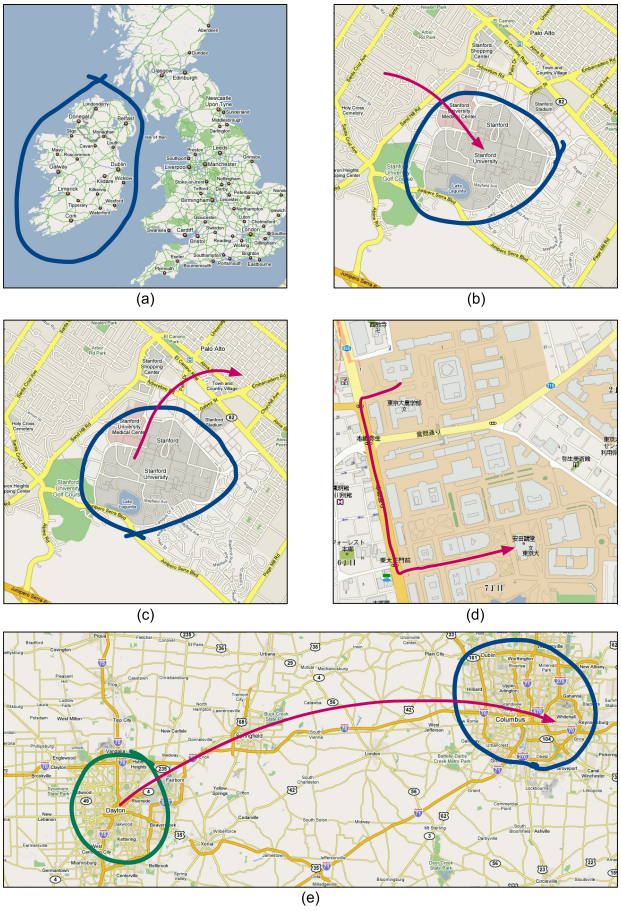

Figure 3: Different types of spatial queries.

friends as a set of icons. The users clicks on two of them to add them to the canvas. Together with the radio button setting, the query now represents all three users.

Right clicking on an icon on the canvas deletes it, removing the corresponding user from the query. This interface is not able to generate a complete set of queries that cover the social space (for instance "*A* and *B* but not *C*"). However, we do not consider it as a limitation at this stage of our research. We intend to modify the queries and interaction strategy based on user studies, for which the current interface serves as a starting point.

4.2. Spatial querying

We selected a sketch-based technique, developed in our earlier work [3] for specifying spatial queries. This choice is justified by the effectiveness of sketch-based spatial queries, as demonstrated by our user studies and other related work [5, 9]. The user makes sketches on a map displayed on the interface, to specify spatial queries.

Figure 3 outlines the formation of spatial queries. Two basic sketch primitives are used for making sketches. A closed curve represents a region; a line segment (to which

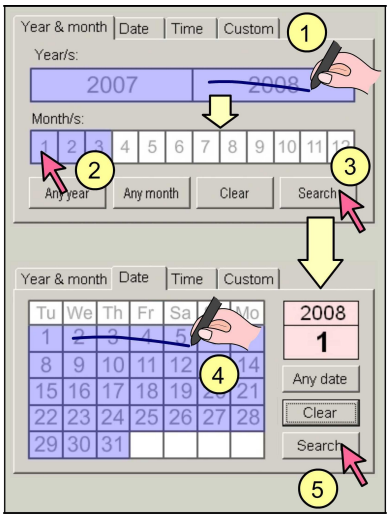

Figure 4: An example temporal query.

an arrowhead is automatically added to show direction) represents a path. These two sketch primitives form the basis of the following detailed spatial queries:

- **Staying within a region:** The user specifies the region by sketching a closed curve around it (Figure 3a)
- **Entering a region:** The user specifies the region, and draws a path into the region from outside (Figure 3b)
- **Leaving a region:** The user specifies the region, and draws a path from inside of the region to the outside (Figure 3c)
- **Moving from one region to another, irrespective of the actual path taken:** The user specifies the two regions (in any order), and then draws a path from the originating region to the destination (Figure 3e)
- **Specific path:** The user draws the path that he/she wishes to retrieve, on the map. The path has to be drawn as an open curve, to prevent misdetection as a region (Figure 3d)

In addition to querying using sketches, the user can preset and retrieve frequenty searched regions using a combo-box. The user can navigate the map to a given area and assign it a label. Examples are "around home" and " Japan." Upon selecting such an entry, the map automatically adjusts to show the corresponding area.

4.3. Temporal querying

We adopted an interface developed in our earlier work [3] for specifying temporal queries. The user sketches on a calendar-like interface to select a duration to retrieve data from. Figure 4 shows how a user queries for the duration "from the 2^{nd} to the 5^{th} of January, 2008". Where

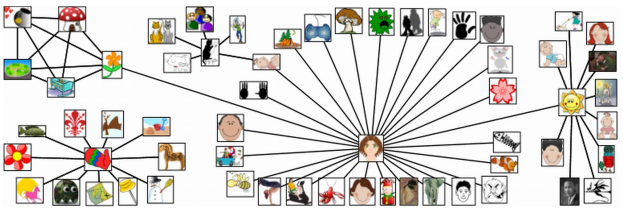

Figure 5: Social graph for TrailNet prototype.

only one item is selected, clicking can be used in place of sketching, facilitating faster interaction. The user can choose some frequently-used time intervals (such as "last week", and "this year") directly from a combo-box.

5. Prototype implementation

5.1 System overview and data

We implement the proposed interactive multimedia queries on a multi-user network of travel media named *TrailNet*. A user can login to TrailNet to enter and retrieve his/her travel media such as GPS traces, photos and videos. They can browse data shared by other users. The users can also become friends with other users.

At the current state, TrailNet is a prototype version with limited functionality. Only GPS traces are available for retrieval as travel media. We imported GPS data from the "GeoLife GPS Trajectory Dataset" [14, 10] to create the network. This dataset consists of travel data captured by 165 persons during a period of two years. Since these persons are not actually using TrailNet, it is used as a read-only system. However, this is not a big limitation for this work since our focus is on querying and retrieval.

We used a clustering algorithm [3] to segment the GPS data in to two categories of locomotion. *Navigating* segments represent the movement of the user from one place to another. *Non-navigating* segments correspond to the user staying at a given location. The users can submit queries to retrieve these segments. We assigned profile pictures (icons) to visually represent users. We also created "friendships" among users, by manually assigning friends for each user. Figure 5 shows the social graph of the network. A line connecting two icons indicate that the two users are friends. The icons for the users who do not have friends have been omitted for clarity.

Figure 6a shows the appearance of the user interface of TrailNet, when a user logs in. The profile icon and the lists of friends, friends of friends and other users have been loaded for querying. The canvas for social querying is empty. The map has been initialized to show the entire Earth, and the calendar to represent all years from which data are available. The user can now make queries to retrieve locomotion segments.

Figure 6: User interface, querying and results.

5.2. Queries and Results

The user can specify queries by combining the interaction techniques described in Section 4. Figure 6b illustrates a socio-temporal query for the movement of two users, during the year 2007. Figure 6c is a screenshot of the results returned for this query. The calendar shows the tempral sub-domain of the results (September and October, 2007), with highlighting. The map has changed to show the region where the results are coming from. It shows a summary of result with a graph-like visualization. The non-navigating segments are shown as circles with the mean location of the user as the center. The radius of the circle visualizes the confidence of the location estimation. The navigating segments are visualized with arrows.

Figure 7: Detailed visualization of a locomotion segment.

The user can refine queries in all three domains repeatedly, until the desired results are returned. The bottom right region of the window contains the list of results. Each segment is shown as a clickable square panel. The icon on the panel represents the user who owns the segment. The circle or arrow on the panel indicates whether the segment is navigating (arrow) or non-navigating (circle). Clicking on a panel gives a detailed visualization of the corresponding segment(Figure 7). The actual GPS data points are plotted on the map, joined by lines. The color of the data points and the line segments change from blue to red with time, to indicate direction.

6. Conclusion and future directions

We proposed a scheme for interactive multimedia querying to allow faster and more effective retrieval of temporal, spatial and social multimedia from a large collection. We designed an example set of social queries and user interaction to specify them. We combined the proposed interaction technique with some previously designed interactions to design a prototype system for retrieving data from a social multimedia network that stores GPS traces from users who have social relationships. The queries facilitated fast, interactive retrieval of the users' locomotion patterns.

We plan to enhance TrailNet to allow storage and retrieval of other types of travel related multimedia such as photos, videos and street views. We are currently designing user studies to evaluate interactive multimedia querying and social querying interaction techniques, so that they can be further improved.

7. Acknowledgment

This research has been supported by a faculty-specific research grant from Microsoft Research Asia.

References

[1] K. S. Candan, W. Li, and M. L. Priya. Similarity-based ranking and query processing in multimedia databases. *Data Knowl. Eng.*, 35(3):259–298, December 2000.

[2] R. Datta, D. Joshi, J. Li, and J. S. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40(2):1–60, 2008.

[3] G. de Silva, T. Yamasaki, and K. Aizawa. Sketch-based spatial queries for retrieving human locomotion patterns from continuously archived gps data. *IEEE Transactions on Multimedia*, 11(7):1240–1253, November 2009.

[4] G. C. De Silva and K. Aizawa. Visual querying with iconic sketches for face image retrieval. In *HCI International 2009*, San Diego, CA, USA, July 2009.

[5] M. J. Egenhofer. Query processing in spatial-query-by-sketch. *Journal of Visual Languages and Computing*, 8:403–424, 1997.

[6] Google. Google maps with street view. In *http://www.google.com/intl/en_us/help/maps/streetview/*, 2011.

[7] Google. Options panel on the search results page. In *http://www.google.com/support/websearch/bin/answer.py?hl=en&answer=142143*, 2011.

[8] Groundspeak. Geocaching - the official global gps cache hunt site. In *http://www.geocaching.com*, 2011.

[9] M. Kopczynski. Efficient spatial queries with sketches. In *Ifroml06*, pages 597–600, 2006.

[10] Z. Yu, Z. Lizhu, X. Xie, and W.-Y. Ma. Mining interesting locations and travel sequences from gps trajectories. In *Proceedings of International conference on World Wild Web (WWW 2009)*, pages 791–800, Madrid, Spain, 2009.

[11] Z. Zha, L. Yang, T. Mei, M. Wang, and Z. Wang. Visual query suggestion. In *Seventeenth ACM international Conference on Multimedia*, pages 15–24, Beijing, China, October 2009.

[12] J. Zhang and A. Lee. etree: A browse and query interface for online communities. *http://www.scientificcommons.org/43518033*, 2008.

[13] T. Zhang, J. Xiao, D. Wen, and X. Ding. Face based image navigation and search. In *Seventeenth ACM international Conference on Multimedia*, pages 597–600, Beijing, China, October 2009.

[14] Y. Zheng, L. Liu, L. Wang, and X. Xie. Learning transportation modes from raw gps data for geographic application on the web. In *Proceedings of International conference on World Wild Web (WWW 2008)*, pages 247–256, Beijing, China, 2008.