JustClick: Personalized Image Recommendation via Exploratory Search from Large-Scale Flickr Image Collections

Jianping Fan, Daniel A. Keim, Yuli Gao, Hangzai Luo, Zongmin Li

***Abstract*—In this paper, we have developed a novel framework called *JustClick* to enable personalized image recommendation via exploratory search from large-scale collections of manually-annotated Flickr images. First, a topic network is automatically generated to summarize large-scale collections of manually-annotated Flickr images at a semantic level. Hyperbolic visualization is further used to enable interactive navigation and exploration of the topic network, so that users can gain insights of large-scale image collections at the first glance, build up their mental query models interactively and specify their queries (i.e., image needs) more precisely by selecting the image topics on the topic network directly. Thus our personalized query recommendation framework can effectively address both the problem of query formulation and the problem of vocabulary discrepancy and null returns. Second, a limited number of images are automatically recommended as the most representative images according to their representativeness for a given image topic. Kernel principal component analysis and hyperbolic visualization are seamlessly integrated to organize and layout the recommended images (i.e., most representative images) according to their nonlinear visual similarities, so that users can assess the relevance between the recommended images and their real query intentions interactively. An interactive interface is implemented to allow users to express their time-varying query intentions and to direct the system to more relevant images according to their personal preferences. Our experiments on large-scale collections of Flickr image collections show very positive results.**

***Index Terms*—Topic network, similarity-based image visualization, personalized image recommendation, user-system interaction.**

I. Introduction

THe last few years have witnessed enormous growth in digital cameras and online high-quality digital images, thus there is an increasing need of new techniques to support more effective image search. Because the keywords are more intuitive for users to specify their image needs, keyword-based image retrieval approaches are now becoming more popular than traditional content-based image retrieval (CBIR) ones [1]. However, there are at least three main obstacles for supporting keyword-based image retrieval: (a) Automatic annotation of large sets of images with unconstrained contents and capturing conditions is still an ongoing research challenge because of the semantic gap [8, 14-17]. (b) For the same keyword-based query, different users may look for different types of images with various visual properties and a few keywords for query formulation cannot capture the users' personal preferences effectively and efficiently. Thus most existing systems tend to return the same set of images to all the users (i.e., one size fits all) and users may seriously suffer from the problem of information overload [2-3, 35-39]. (c) Users may not be able to find the most suitable keywords to formulate their image needs precisely or they may not even know what to look for (i.e., *I don't know what I am looking for, but I'll know when I find it*) [4-6]. In addition, there may have a vocabulary discrepancy between the keywords for users to formulate their queries and the text terms for image annotation, and such vocabulary discrepancy may result in null returns for the mismatching queries. Thus users may seriously suffer from both the problem of query formulation and the problem of vocabulary discrepancy and null returns. The keywords for image annotation may not be able to describe the rich details of the visual contents of the images sufficiently, thus most existing keyword-based image retrieval systems cannot support users to look for some particular images according to their visual properties.

Even though the keywords are more intuitive for users to formulate their queries (i.e., image needs), a few keywords cannot describe the users' real query intentions effectively and efficiently, thus there is an uncertainty between the keywords for query formulation and the users' real query intentions. In addition, the users' query intentions may vary with their timely image observations, and such dynamics cannot be characterized precisely by using pre-learned user profiles. The huge number of online users and the uncertainties of the image retrieval environment (i.e., dynamic nature of the users' interests and huge diversity of image semantics) also make it extremely difficult to learn the user profiles precisely. Therefore, it is very important to develop new algorithms that can capture the users' dynamic query intentions implicitly for supporting personalized image recommendation.

Jianping Fan is with the Department of Computer Science, University of North Carolina, Charlotte, NC 28223, USA.
e-mail: jfan@uncc.edu
Daniel A. Keim was with Computer Science Institute, University of Konstanz, Konstanze, Germany.
email: keim@inf.uni-konstanz.de
Yuli Gao was with the Department of Computer Science, University of North Carolina, Charlotte, NC 28223, USA. He is now in HP Labs.
e-mail: yuli.gao@hp.com
Hangzai Luo was with the Department of Computer Science, University of North Carolina, Charlotte, NC 28223, USA. He is now in East China Normal University.
e-mail: hluo@sei.ecnu.edu.cn
Zongmin Li was with the Department of Computer Science, University of North Carolina, Charlotte, NC 28223, USA. He is now in Department of Computer Science, China University of Petroleum.
email: lizm@hdpu.edu.cn.

Manuscript received September 23, 2007; revised January 18, 2008.

It is important to understand that the system alone cannot meet the users' sophisticated image needs effectively, image search requires greater interactivity between the users and the systems [7]. Thus there is an urgent need to enhance the system's ability to allow users to communicate their image needs more precisely and express their time-varying query interests effectively. Visualization and interactive image exploration can offer good opportunities for allowing users to add their background knowledge, powerful pattern recognition capability and inference skills for directing the systems to find more relevant images according to their personal preferences [4-5, 20-28]. Such user-system interaction and exploration process can bring the system perspective of image collections and the users' perspective of image needs together. Thus the interactive user-system interface should aid the users in understanding and expressing their image needs more precisely. Unfortunately, designing interactive user interfaces for the CBIR systems has not received enough attentions [7].

From the users' point of views, such interactive user interfaces should be able to allow them to: (a) communicate their image needs easily to the system; (b) express their time-varying query interests precisely for directing the system to find more relevant images according to their personal preferences; (c) explore large amounts of returned images interactively according to their inherent visual similarity contexts for relevance assessment; and (d) track and visualize their access path (query contexts) and recommend the most relevant image topics or the most representative images for the next search.

From the system's point of view, such interactive user interfaces should be able to: (1) disclose a good global overview (i.e., a big picture) of large-scale image collections to assist users in making better query decisions; (2) visualize large amounts of returned images according to their visual similarity contexts to assist the users in interactive image exploration and relevance assessment; (3) capture the users' time-varying query intentions implicitly and integrate them to find more relevant images according to the users' personal preferences; and (4) provide a good environment to integrate users' background knowledge and pattern recognition capacity for bridging the semantic gap in the loop of image retrieval.

Based on these observations, we have developed a novel framework called ***JustClick*** to achieve personalized image recommendation by supporting interactive image exploration. Our research focuses on large-scale collections of manually-annotated Flickr images to bypass the semantic gap problem for automatic image annotation. In addition, a more effective user-system interface is designed for integrating the users' background knowledge and their pattern recognition capability to bridge the semantic gap interactively in the loop of image retrieval. This paper is organized as follows. Section 2 briefly reviews some existing work on tackling these three obstacles; Section 3 introduces our new scheme to incorporate topic network and hyperbolic visualization for achieving personalized query recommendation (addressing both the problem of query formulation and the problem of vocabulary discrepancy and null returns); Section 4 describes our new algorithm for integrating visualization and interactive image exploration to capture the users' time-varying query intentions implicitly for achieving a user-adaptive image recommendation (addressing the problem of time-varying query intentions and information overload); Section 5 summarizes our evaluation of the algorithms and the system; We conclude in Section 6.

II. Related Work

To tackle the first obstacle (e.g., bridge the semantic gap for image annotation), automatic image annotation via semantic classification has recently received sustantial attentions [14-17]. Unfortunately, supporting automatic annotation of large-scale image collections with unconstrainted contents and capturing conditions is still an ongoing research challenge [1]. Therefore, two alternative approaches are widely used for supporting keyword-based retrieval of large-scale online image collections: (a) *Google image search engine* by indexing large-scale online image collections through the text terms that are automatically extracted from the associated text documents, image file names or URLs; (b) *Flickr image search engine* by indexing large-scale collections of shared images through the taggings that are manually provided by numerous online users. These keyword-based image search engines have offered many good opportunities for the CBIR community while emerging many new challenges.

The two commercial image search engines Google and Flickr have achieved significant progress on supporting keyword-based retrieval of large-scale online image collections by using the manual image taggings or the associated text terms, but their performance (accuracy, efficiency, and effectiveness) is still not acceptable because of the following reasons: (1) The file names, URLs, and the associated text terms may not have exact correspondence with the semantics of the images, and thus the Google image search engine returns large amounts of junk images [40-41]. (2) Different users may use various keywords with ambiguous word senses to annotate the images and one single keyword may have multiple word senses, thus the Flickr image search engine may return inconsistent or incomplete results. In addition, there may have a vocabulary discrepancy between the keywords for users to formulate their queries and the taggings for manual image annotations, and such vocabulary discrepancy may result in null returns for the mismatching queries. Thus only using the manual annotations for image retrieval may be far from people's expectation. (3) The visual contents of the images are completely ignored, thus both Google image search engine and Flickr image search engine cannot allow users to look for some particular images of interest according to their visual properties. However, there are some evidence that the visual properties are very important for people to search for images [20-28]. Unfortunately, the keywords may not be expressive enough for describing the rich details of the visual contents of the images sufficiently. Even the low-level visual features may not be able to carry the semantis of image contents directly [8], they can definitely be used to enhance users' abilities on finding some particular images according to their inherent visual similarity contexts. (4) A few keywords for query formulation may not be able to capture the users' real query intentions effectively, thus users may seriously suffer from the problem of information overload.

Some pioneer work have been done by incorporating relevance feedback to bridge the *semantic gap* in the loop of image retrieval [42-46]. Unfortunately, most existing relevance feedback approaches may bring huge burden on the users and a limited number of labeled images may not be representative enough for learning an accurate prediction model to categorize large amount of unseen images precisely.

To tackle the second obstacle (e.g., the problem of time-varying query intentions and information overload), personalized information retrieval can be treated as one potential solution and there are two well-accepted approaches: *content-based filtering* [38-39] and *collaborative filtering* [35-37]. Unfortunately, both of them cannot directly be extended for supporting personalized image recommendation because of the huge diversity and the time-varying properties of the users' preferences, and it is very hard if not impossible to learn the users' preferences precisely from a few relevance judgments.

The collaborative filtering approach may suffer from the *sparseness* problem when there is a shortage of the users' ratings of the images, and it cannot be used to recommend new images because of the *first rating* problem. On the other hand, the content-based filtering approach cannot be used to achieve serendipitous discovery of unexpected images because the profiles may over-specify the users' interests. An accurate text-based description of image contents is also required to achieve reliable content-based image filtering. Unfortunately, the manual image taggings that are available at Flickr may be incomplete for describing the rich details of the visual contents of the images accurately, thus they cannot be used to support reliable content-based image filtering. On the other hand, achieving automatic annotation of large-scale image collections with unconstrained contents is still an open issue for the CBIR community [1]. The profiles for new users may not be available, thus all these existing personalized information recommendation algorithms cannot support new users effectively.

The interfaces for most existing CBIR systems are designed for users to assess the relevance between the returned images and their real query intentions via page-by-page browsing. Such page-by-page browsing approach may seriously suffer from the following problems: (1) They are not scalable to the sizes of the images and many pages are needed for displaying large amounts of images returned by the same keyword-based query, thus it is very tedious for users to look for some particular images of interest through page-by-page browsing. (2) Because the visual similarity contexts among the returned images are completely ignored for image ranking, the visually-similar images may be separated into different pages and each page may lead the users to new image contents. Such inter-page visual disconnection may divert the users' attentions from their current query contexts and make it very difficult for the users to compare the diverse visual similarities between the returned images. Thus the users cannot assess the relevance between the returned images and their real query intentions effectively. Rather than displaying the returned images page by page, more interactive user interfaces should be developed to allow the users to explore large amounts of returned images according to their inherent visual similarity contexts, so that the users can discover new knowledge from large-scale image collections via exploratory search [3].

Based on these observations, some pioneer works have been done by incorporating visualization to support interactive image navigation and exploration [20-28]. Rubner et al. [20] and Stan et al.[21] have incorporated feature-based visual similarity and multi-dimensional scaling (MDS) to create a 2D layout of the images, so that users can navigate the images easily according to their feature-based visual similarity. Nyuyen and Worring have incorporated isometric mapping to exploit the nonlinear similarity structures for image visualization [23], and Moghaddam et al. have incorporated PCA to enable similarity-based image visualization which focuses on achieving low computational cost and fast convergence rate [25]. Recently, Walter et al. have incorporated MDS with hyperbolic visualization to create the spatial layout of the images based on their pairwise feature-based visual dissimilarity [22]. Unfortunately, all these existing similarity-based image visualization techniques cannot work on large-scale image collections because they may seriously suffer from the following problems: (a) they are not scalable to the sizes of the images because of the overlapping problem; (b) they are not scalable to the diversity of image semantics because of the shortage of the effective techniques for semantic image summarization; (c) the underlying similarity functions may not be able to characterize the diverse visual similarities between the images accurately and the visual similarity structures between the images could be nonlinear; (d) most existing techniques for image projection, such as MDS and PCA, may suffer from low convergence rate, stick in local minima or may not be able to exploit and preserve the nonlinear visual similarity structures between the images precisely.

To tackle the third obstacle (e.g., problem of query formulation and problem of vocabulary discrepancy and null returns), Flickr has provided a list of the most popular taggings (i.e., tagcloud) for assisting users on query formulation. Unfortunately, the contextual relationships between the image topics are completely ignored. However, such inter-topic contextual relationships, which can give a good approximation of the interestingness of the image topics (i.e., like PageRank for characterizing the importance of web pages [34]), may play an important role for the users to make better query decisions. When the most relevant image topics are displayed together according to their association contexts, it is easier for the users to come up with more clear query concepts. Therefore, it is very attractive to exploit both the image topics of interest and their association contexts for supporting personalized query recommendation, so that the users can make better query decisions and formulate their image needs more precisely.

III. Personalized Query Recommendation

Every process for image seeking is necessarily initiated by an image need from the user's side, thus a successful CBIR system should be able to allow the users to obtain a good global overview (i.e., a big picture) of large-scale image collections quickly and communicate their image needs more effectively. In our ***JustClick*** system, we have developed two innovative techniques to support personalized query recommendation for assisting the users in communicating their image needs with the system more effectively: (a) Topic network is automatically generated to summarize large-scale collections of manually-tagged Flickr images at a semantic level; (b) Hyperbolic visualization is implemented to support change of focus effectively, so that the users can gain the significant insights of large-scale image collections via interactive topic network navigation and exploration.

Our topic network consists of two key components: *popular image topics* at Flickr and their *inter-topic contextual relationships*. We have developed an automatic scheme for generating such topic network from large-scale collections of manually-tagged Flickr images. Each image at Flickr is associated with the users' taggings of the underlying image semantics and the users' rating scores. After the images and the associated users' taggings are downloaded from Flickr.com, the text terms for image topic interpretation are identified automatically by using standard text analysis techniques.

Concept ontology has recently been used to index and summarize large-scale image collections at the concept level by using one-direction IS-A hierarchy (i.e., each concept node is linked with only its parent node in one direction) [9-10]. However, the contextual relationships between the image topics at Flickr could be more complex rather than such one-direction IS-A hierarchy, where each image topic may link with multiple related image topics as a network. Thus the inter-topic contexts for Flickr image collections cannot be characterized precisely by using only the one-direction IS-A hierarchy. Based on these observations, we have developed a new algorithm for determining more general inter-topic associations by seamlessly integrating both the semantic similarity and the *information content* gained from the co-occurrence probability of the relevant text terms. The inter-topic association $\phi(C_i, C_j)$ is then defined as:

$$\phi(C_i, C_j) = \nu \cdot \frac{e^{S(C_i,C_j)} - e^{-S(C_i,C_j)}}{e^{S(C_i,C_j)} + e^{-S(C_i,C_j)}} + \omega \cdot \frac{e^{t\cdot G(C_i,C_j)} - e^{-t\cdot G(C_i,C_j)}}{e^{t\cdot G(C_i,C_j)} + e^{-t\cdot G(C_i,C_j)}} \tag{1}$$

where $\nu + \omega = 1$, the first part is used to measure the contribution from the semantic similarity $S(C_i, C_j)$ between the image topics C_j and C_i, the second part indicates the contribution from the information content $G(C_i, C_j)$ gained from the co-occurrence probability of the image topics C_i and C_j, t is an integer to keep $S(C_i, C_j)$ and $t \cdot G(C_i, C_j)$ to be on the same scale, ν and ω are the weighting parameters. In our definition, the strength of the inter-topic association is normalized to 1 and it increases adaptively with the semantic similarity and the information content.

The information content $G(C_i, C_j)$ is defined as:

$$G(C_i, C_j) = -\frac{1}{\log \pi(C_i, C_j)} \tag{2}$$

where $\pi(C_i, C_j)$ is the co-occurrence probability of the text terms for interpreting two image topics C_j and C_i, and it can be obtained from the available image annotation document. From this definition, one can observe that higher co-occurrence probability $\pi(\cdot,\cdot)$ of the image topics corresponds to larger information content $G(\cdot,\cdot)$ and stronger inter-topic association $\phi(\cdot,\cdot)$.

The semantic similarity $S(C_i, C_j)$ is defined as [48]:

$$S(C_i, C_j) = max\left(-\log \frac{Shortest\_Length(C_i, C_j)}{2 \cdot Taxonomy\_Depth}\right) \tag{3}$$

where $Shortest\_Length(C_i, C_j)$ is the length of the shortest path between the image topics C_i and C_j in an one-drection IS-A taxonomy, and $Taxonomy\_Depth$ is the maximum depth of such one-direction IS-A taxonomy. From this definition, one can observe that closer between the image topics (i.e., smaller value of $Shortest\_Length(\cdot,\cdot)$) on the taxonomy corresponds to higher semantic similarity $S(\cdot,\cdot)$ and stronger inter-topic association $\phi(\cdot,\cdot)$.

The information content $G(\cdot,\cdot)$ gained from the co-occurrence probability of the image topics is more representative for characterizing the complex inter-topic associations at Flickr, thus it plays more important role than the semantic similarity $S(\cdot,\cdot)$ on characterizing the strength of the inter-topic associations $\phi(\cdot,\cdot)$, and we set $\nu = 0.4$ and $\omega = 0.6$ heuristically.

Unlike the one-direction IS-A hierarchy in the concept ontology, each image topic can be linked with all the other image topics on the topic network, thus the maximum number of such inter-topic associations is $\frac{n(n-1)}{2}$, where n is the total number of image topics on the topic network. However, the strength of the associations between some image topics may be very weak (i.e., the corresponding text terms for image topic interpretation may not be used simultaneously for manual image annotation), thus it is not necessary for each image topic to be linked with all the other topics on the topic network. Based on this understanding, each image topic is automatically linked with top m ($m \ll n$) most relevant image topics with larger values of the inter-topic associations $\phi(\cdot,\cdot)$, and the potential number of such inter-topic associations is reduced to $\frac{n\times m}{2}$. One small part of our topic network is given in Fig. 1, where the image topics are organized according to the strength of their associations, $\phi(\cdot,\cdot)$. One can observe that such topic network can provide a good global overview (i.e., a big picture) of large-scale collections of manually-tagged Flickr images at a semantic level, thus it can be used to assist users in making better query decisions. When users see how the image topics are related to each other on the topic network, they will have a better understanding of their image needs and come up with more precise query concepts.

To integrate the topic network for supporting personalized query recommendation, it is very attractive to enable graphical representation and visualization of the topic network, so that users can obtain a good global overview of large-scale image collections at the first glance. It is also very attractive to enable interactive topic network navigation and exploration according to the inherent inter-topic contexts, so that the users can easily find the image topics which are more relevant to their mental query models. Thus the queries do not need to be formulated explicitly, but emerge through the interaction of the users with

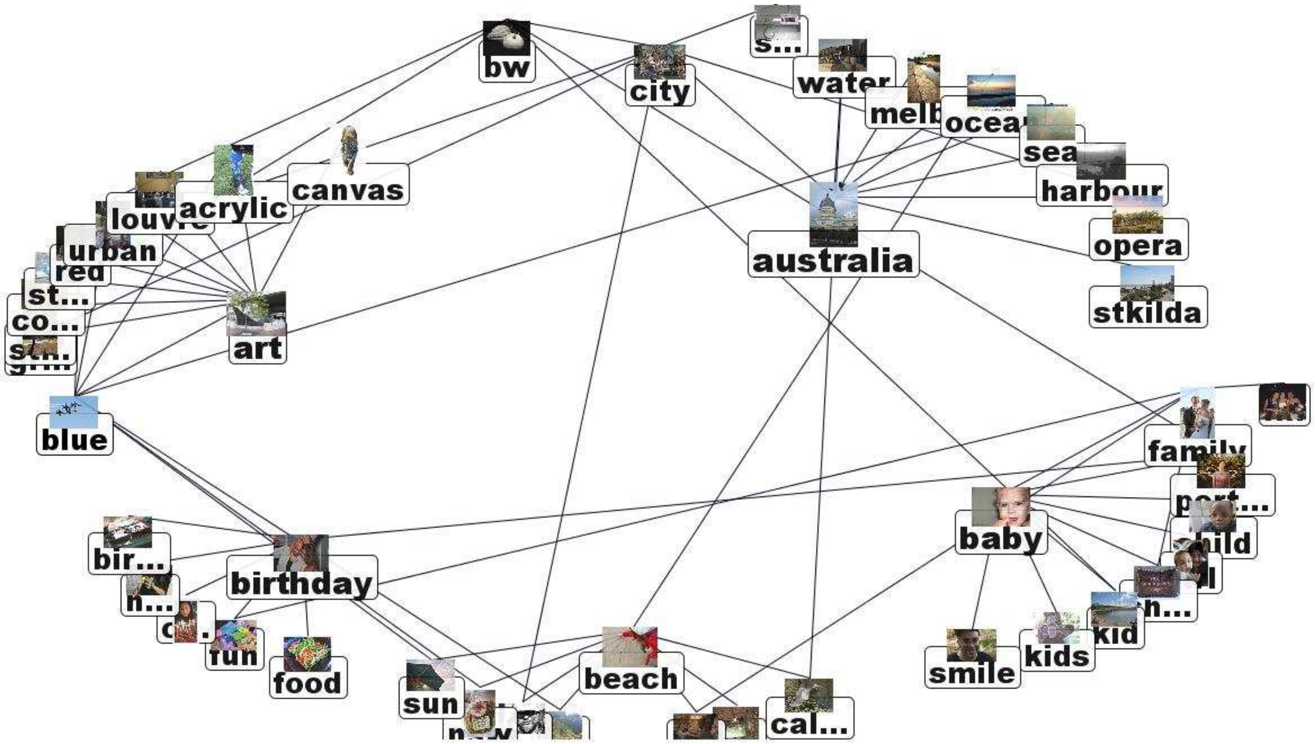


Fig. 1. One portion of our topic network for organizing large-scale manually-annotated Flickr image collections at a semantic level.

the topic network (i.e., users can just choose the visible image topics on the topic network rather than generate the keywords for query formulation, thus we name our system as ***JustClick***). Unfortunately, visualizing large-scale topic network (i.e., the topic network which consists of large amounts of image topics) in a 2D system interface with a limited screen size is not a trivial task because of the overlapping problem [30].

To tackle the overlapping problem for topic network visualization, we have investigated multiple innovative techniques: (a) The interestingness scores are defined for the image topics on the topic network and they are used to highlight the image topics of interest and eliminate some less interesting image topics, so that users can always be acquainted by the most interesting image topics and gain the significant insights of large-scale image collections at the first glance. (b) A new constraint-driven topic clustering algorithm is developed to achieve multi-level representation and visualization of large-scale topic network, so that our topic network visualization algorithm can have good scalability with the number of image topics. (c) A query-driven topic network visualization algorithm is developed to support goal-directed query recommendation, so that users can be acquainted by a limited number of image topics which are most relevant to their current query interests. (d) Hyperbolic visualization is used to enable interactive topic network exploration and tackle the overlapping problem interactively via change of focus.

It is worth noting that the processes for automatic topic network construction, interestingness score calculation, constraint-driven topic clustering, multidimensional scaling (MDS) for topic network projection and visualization, and personalized topic network generation can be performed off-line. Only the processes for interactive topic network exploration and query-driven topic visualization should be performed online and they can be achieved in real time.

A. Query Recommendation via Interactive Topic Network Exploration

We have integrated both the popularity of the image topics and the importance of the image topics to determine their interestingness scores. The popularity for one certain image topic is related to the number of images under the given image topic. If one image topic consists of more images, it tends to be more interesting. The importance of a given image topic is also related to its linkages with other image topics on the topic network. If one image topic is linked with more image topics on the topic network, it tends to be more interesting [34]. Thus the *interestingness score* $\varrho(C_i)$ for a given image topic C_i is defined as:

$$\varrho(C_i) = \varepsilon \cdot \frac{e^{n(c_i)} - e^{-n(c_i)}}{e^{n(c_i)} + e^{-n(c_i)}} + \eta \cdot \frac{e^{r \cdot l(c_i)} - e^{-r \cdot l(c_i)}}{e^{r \cdot l(c_i)} + e^{-r \cdot l(c_i)}} \quad (4)$$

where $\varepsilon + \eta = 1$, $n(c_i)$ is the number of images under C_i, $l(c_i)$ is the number of image topics linked with C_i on the topic network, and r is an integer to keep $n(c_i)$ and $r \cdot l(c_i)$ to be on the same scale. The number of the linked topics is more important than the number of images for characterizing the interestingness for a given image topic [34], e.g., image topics, which consist of a smaller size of images but are linked with many other image topics, are more interesting than the image topics which consist of larger size of images but are linked with few other image topics. Based on this observation, we set $\eta = 0.6$ and $\varepsilon = 0.4$ heuristically. In our definition, the interestingness score $\rho(\cdot)$ is normalized to 1 and it increases

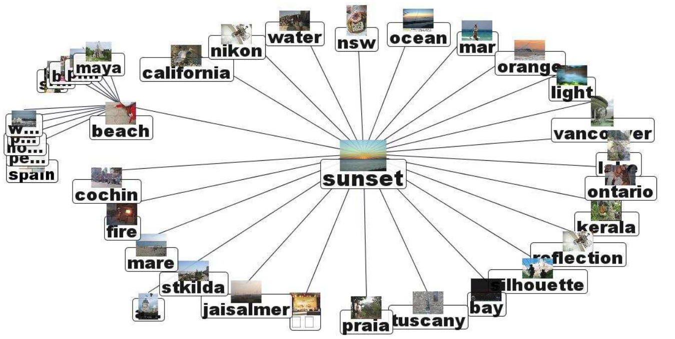


Fig. 2. Query-drievn visualization of the same topic network as shown in Fig. 1, where "sunset" is the query concept from the user.

adaptively with the number of images $n(\cdot)$ and the number of the linked topics $l(\cdot)$.

After such interestingness scores for all the image topics on the topic network are available, the image topics are then ranked according to their interestingness scores, so that the most interesting image topics with larger values of the interestingness scores can be highlighted effectively and the less interesting image topics with smaller values of the interestingness scores can be eliminated automatically from the topic network. Thus users can always be acquainted by the most interesting image topics.

We have also developed a new constraint-driven clustering algorithm to achieve multi-level representation of large-scale topic network, and the pairwise edge weight is defined as:

$$\psi(C_i,C_j)=\phi(C_i,C_j)\times\begin{cases} e^{-\frac{d^2(C_i,C_j)}{\sigma_L^2}}, & if \quad d(C_i,C_j)<\delta \\ 0, & otherwise \end{cases} \tag{5}$$

where the first part $\phi(C_i,C_j)$ denotes the association between the image topics C_i and C_j, the second part indicates their linkage relatedness and constraint, $d(C_i,C_j)$ is the distance between the physical locations for the image topics C_i and C_j on the topic network, σ_L is the variance of their physical location distances, and δ is a pre-defined threshold.

After such pairwise edge weight matrix is obtained, Normalized Cut algorithm is used to obtain an optimal partition of large-scale topic network [29]. Thus the image topics, which are strongly correlated and are connected each other on the topic network, are partitioned into the same cluster and be treated as one super-topic at the higher abstract level. Each super-topic can be expanded into a group of strongly-correlated image topics (i.e., a smaller-size topic network) at the lower level. Thus our constraint-driven topic clustering algorithm is able to achieve multi-level representation and visualization of large-scale topic network. Because the number of image topics at each level are much smaller, our multi-level topic network representation and visualization algorithm can reduce the visual complexity significantly, tackle the overlapping problem effectively, and have good scalability with the number of image topics.

When the users have clear query concepts in mind, they can directly submit their keyword-based queries to our system and we have developed a novel algorithm to enable ***query-drievn topic network visualization***. As shown in Fig. 2, large-scale topic network is re-organized according to the user's query concept and more spaces are arranged automatically for the image topics which are more relevant to the given query concept. The *query-driven interestingness score* $\varrho^q(C_k)$ for a given image topic C_k on the topic network is defined as:

$$\varrho^q(C_k)=\varrho(C_k)\times\phi(C_k,C_{match})\times\begin{cases} 1, & d(C_k,C_{match})\le 2 \\ 0, & otherwise \end{cases} \tag{6}$$

where the C_{match} is used to denote the image topic on the topic network which best matches with the user's query concept, $d(C_k,C_{match})$ is the location distance between the given image topic C_k and the best matching image topic C_{match} by counting the image topic nodes between them on the topic network. In our current implementation, we just consider the most relevant image topics which the distance is no more than 2 (i.e., second-order nearest neighbors on the topic network). Our query-driven topic network visualization algorithm can offer at least two advantages: (a) it can significantly reduce the overlapping between the image topics by focusing on only a small number of image topics which are most relevant to the user's query concept (e.g., $\varrho^q(\cdot)\neq 0$); (b) it can guide the user on which image topics they can access for next search according to the inter-topic contexts.

We have investigated multiple solutions to layout the topic network: (1) A string-based approach is incorporated to visualize the topic network with a nested view, where each image topic is displayed closely with the most relevant image topics according to the strength of their inter-topic associations. The interestingness score for each image topic can also be visualized, so that users can get the sense of the interestingness of the image topics at the first glance. (2) The inter-topic contextual relationships are represented as the weighted undirected edges, and the length of such weighted undirected edges are inversely proportional to the strength of the corresponding inter-topic association $\phi(\cdot,\cdot)$, e.g., closer image topics on the topic network are more relevant with stronger inter-topic associations. Thus the geometric closeness between the image topics is strongly related to their associations, so that such graphical representation of the topic network can reveal a great deal about how these image topics are correlated and how they are intended to be used jointly for manual image tagging. (3) An iconic image is selected automatically for each image topic and it is visualized simultaneously with the corresponding image topic node. Such combination of the iconic images and the text terms for multi-modal image topic interpretation and visualization can provide more intuitive format for human cognition.

Our approach for topic network visualization has also exploited hyperbolic geometry [30]. The hyperbolic geometry is particularly well suited for achieving graph-based layout of the topic network. The essence of our approach is to project the topic network onto a hyperbolic plane according to the strength of the associations between the image topics, and layout the topic network by mapping the relevant image topic nodes onto a circular display region. Thus our hyperbolic topic

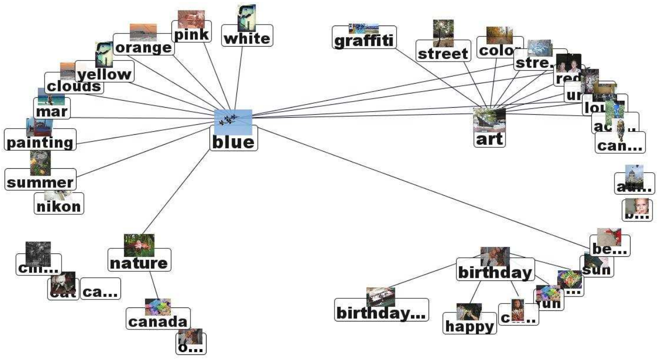


Fig. 3. Interactive exploration of the topic network (shown in Fig. 1) via change of focus.

network visualization scheme takes the following steps: (a) The image topic nodes on the topic network are projected onto a hyperbolic plane according to the strength of their associations by minimizing Sammon's cost function [47]:

$$E(\{C_i\}) = \sum_{i=1}^{n} \sum_{j>i}^{n} w_{ij} |\delta(C_i, C_j) - \phi(C_i, C_j)|^2 \qquad (7)$$

where $\phi(C_i, C_j)$ is the strength of the inter-topic association between the image topics C_i and C_j, the factors w_{ij} are used to weight the disparities individually and also to normalize the Sammon's cost function E to be independent to the absolute scale of the inter-topic association $\phi(C_i, C_j)$, and $\delta(C_i, C_j)$ is the location distance between the image topics C_i and C_j on the hyperbolic plane.

$$\delta(C_i, C_j) = 2 \cdot arctanh \left(\frac{|X_{c_i} - X_{c_j}|}{|1 - X_{c_i} \bar{X}_{c_j}|} \right) \qquad (8)$$

where X_{c_i} and X_{c_j} are the locations of the image topics on the hyperbolic plane. In our current implementation, Sammon's intermediate normalization factor is used to calculate w_{ij}:

$$w_{ij} = \frac{1}{\sum_{k=1}^{n} \sum_{l>k}^{n} \phi(C_l, C_k)} \frac{1}{\phi(C_i, C_j)} \qquad (9)$$

(b) After such context-preserving projection of the image topics is obtained, Poincaré disk model [30] is used to map the image topic nodes on the hyperbolic plane onto a 2D display coordinate. Poincaré disk model maps the entire hyperbolic space onto an open unit circle, and produces a non-uniform mapping of the image topic nodes to the 2D display coordinate.

After the hyperbolic visualization of the topic network is available, it can be used to enable interactive exploration and navigation of the semantic summary (i.e., topic network) of large-scale collections of manually-annotated Flickr images via *change of focus*. The *change of focus* is implemented by changing the mapping of the image topic nodes from the hyperbolic plane to the unit disk for display, and the positions of the image topic nodes in the hyperbolic plane need not to be altered during the focus manipulation [30]. As shown in Fig. 3, users can change their focuses of image topics by clicking on any visible image topic node to bring it into focus at the screen center, or by dragging any visible image topic node interactively to any other screen location without losing the semantic contexts between the image topic nodes, where the rest of the layout of the topic network transforms appropriately. Users can directly see the *topics of interest* in such interactive topic network navigation and exploration process, thus they can build up their mental query models interactively and specify their queries precisely by selecting the visible image topics on the topic network directly as shown in Fig. 4. By supporting interactive topic network exploration, our hyperbolic topic network visualization scheme can support personalized query recommendation interactively, which can address both the problem of query formulation and the problem of vocabulary discrepancy and null returns more effectively. Such interactive topic network exploration process does not require the user profiles, thus our ***JustClick*** system can also support new users effectively.

When large-scale topic network comes into view, visualizing all the image topics and their contextual relationships in one sceen with size limitation is impractical. Thus our hyperbolic visualization scheme focuses on assigning more spaces for the image topic node in current focus and ignoring the details for the residual image topic nodes on the topic network, which can tackle the overlapping problem interactively. Through change of focus, users can always be acquainted by the image topics of interest according to their time-varying query interests. At the same time, users can also see the local inter-topic contexts which are embedded in a global structure (i.e., topic network), so that they can easily perceive and recognize the appropriate directions for future search as shown in Fig. 4. Our ***JustClick*** system can also track and visualize the users' access path (query contexts over the topic network) as shown in Fig. 4(a), so that the users can see a "big picture" about which image topics they have visited, which image topics they are visiting now, and which image topics are recommended by our system for next search. Thus the users can always be acquainted by the image topics which are most relevant to their current query interests. By tracking and visualizing the users' access path, the users can see a good global structure of their query contexts and keep track of the progress of their sequential searches. Through examining and comparing the results (see section 4) obtained by these sequential searches (i.e., using different image topics) on the query context map, the users can easily discover the boundary of the meaning for their query concepts (i.e., which image topic starts to return the images with completely different visual properties). Such boundary can also allow the users to know which image topics on the query context map (embedded on the topic network) can give them more relevant images according to their personal preferences. The query context maps, which can also be used to record the users' dynamic query actions, are further exploited to generate personalized topic network for query recommendation.

B. Personalized Topic Network Generation for Query Recommendation

The users' query interests have two significant properties: (a) dynamic (current and time-varying interests); (b) consistency (long-term and general interests) [35-39]. Our user-system interaction interface focuses on capturing the users'

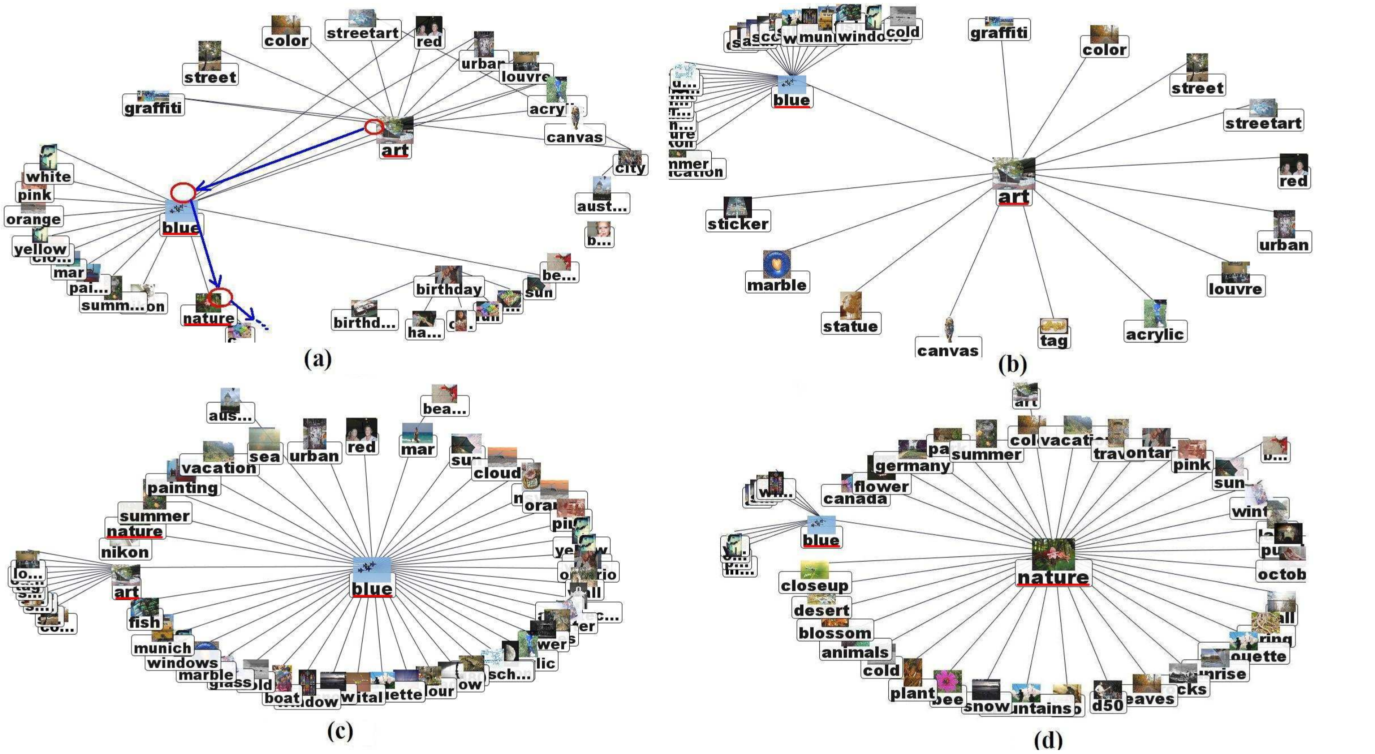


Fig. 4. Our JustClick system can achieve a good balance between global context and local details: (a) user's access path on a global context of the topic network; (b), (c) and (d) the local details of the semantic contexts for query recommendation.

dynamic query interests adaptively for supporting personalized topic recommendation. On the other hand, it is also very attractive to learn the users' long-term query interests (i.e., user profiles) for personalized topic recommendation.

In our ***JustClick*** system, we have developed a new algorithm for learning the user profiles automatically from the collection of the user's dynamic query actions (i.e., query context maps) for updating the system's knowledge of the user's image needs and personal preferences. Thus the *personalized interestingness score* $\varrho^p(C_i)$ for a given image topic C_i on the topic network can be defined as:

$$\varrho^p(C_i) = \varrho(C_i) + \varrho(C_i)\left(\beta_v \frac{e^{v(C_i)} - e^{-v(C_i)}}{e^{v(C_i)} + e^{-v(C_i)}} + \right.$$
$$\left. \beta_s \frac{e^{s(C_i)} - e^{-s(C_i)}}{e^{s(C_i)} + e^{-s(C_i)}} + \beta_d \frac{e^{d(C_i)} - e^{-d(C_i)}}{e^{d(C_i)} + e^{-d(C_i)}} \right) \qquad (10)$$

where $\varrho(C_i)$ is the original interestingness score of the given image topic C_i, $v(C_i)$ is the visiting times of the given image topic C_i from the particular user, $s(C_i)$ is the staying seconds for the particular user to stick on the given image topic C_i, $d(C_i)$ is the access depth for the particular user to interact with the image topic C_i and the images under C_i, β_v, β_s, and β_d are the normalization parameters, $\beta_v + \beta_s + \beta_d = 1$. Thus the personalized interestingness scores of the image topics can be determined immediately when such user-system interaction happens, and they will converge to the stable values for characterizing the user's long-term query interests (i.e., user profiles).

After the personalized interestingness scores for all these image topics are learned from the collection of the user's dynamic query actions, they can be used to highlight the image topics for generating a personalized topic network to represent the user profiles. Thus the image topics with smaller values of the personalized interestingness scores can be eliminated automatically, so that each user can be acquainted by the most interesting image topics according to his/her personal preferences.

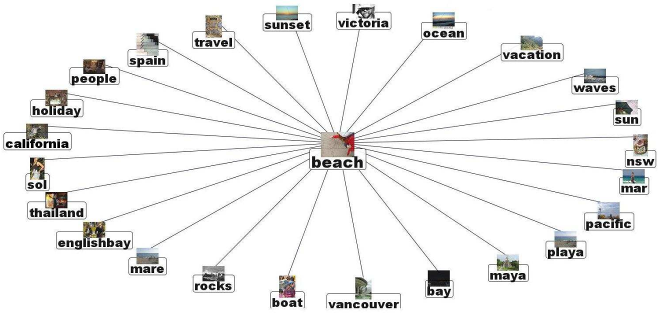


Fig. 5. The first-order nearest neighborhood and inter-topic correlations for automatic preference propagation, where the image topic "beach" in the center is to be propagated.

The user's interests may be changed according to his/her timely image observations, and one major problem for integrating the pre-learned user profiles for query recommendation is that such pre-learned user profiles may over-specify the user's interests. Thus the pre-learned user profiles may hinder the user to access other interesting image topics on the topic network. Based on this observation, we have developed a novel algorithm for propagating the user's interests over other relevant image topics on the topic network. Thus the *personalized interestingness score* $\varrho^p(C_j)$ for the image topic

C_j to be propagated can defined as:

$$\varrho^p(C_j) = \varrho(C_j)\varphi^p(C_j) \tag{11}$$

where $\varrho(C_j)$ is the original interestingness score for the image topic C_j to be propagated, $\varphi^p(C_j)$ is the propagation strength of the image topic C_j and it is defined as:

$$\varphi^p(C_j) = \sum_{l \in \Omega} \bar{\phi}(C_l)$$

$$\bar{\phi}(C_l) = \begin{cases} \phi(C_l, C_j), & d(C_l, C_j) = 1 \\ \\ 0, & otherwise \end{cases} \tag{12}$$

where Ω is the set of the accessed image topics in the first-order nearest neighborhood of the image topic C_j as shown in Fig. 5, $\phi(C_l, C_j)$ is the inter-topic association between the image topic C_j to be propagated and the image topic C_l that has been accessed by the particular user. Thus the image topics, which have larger values of the personalized interestingness scores $\varrho^p(\cdot)$, can be propagated adaptively.

It is worth noting that the image topics, which are less interesting in the large pool of image topics and are eliminated at the beginning for reducing the complexity for large-scale topic network visualization, can be recovered and be presented to the particular user when their strongly-related image topics (which are strongly related to these eliminated image topics) are accessed by the particular user and thus the values of their personalized interestingness scores may become larger. Therefore, our interestingness propagation algorithm can allow users to access some less interesting image topics (which are not significant in the pool of large amounts of image topics and are eliminated at the beginning for reducing visualization complexity) according to their personal preferences.

By integrating the inter-topic correlations for automatic propagation of the user's preferences, our ***JustClick*** system can precisely predict the user's hidden preferences from the collection of his/her dynamic query actions. Thus the personalized topic network can be used to represent the user profiles precisely, where the interesting image topics can be highlighted according to their personalized interestingness scores. The personalized topic network, which is treated as the user profiles to interpret the user's personal preferences, can recommend the topics of interest to each individual user directly without requiring him/her to make an explicit query.

IV. Personalized Image Recommendation

Multiple keywords may simultaneously be used to tag the same image, thus one single image may belong to multiple image topics on the topic network. On the other hand, the same keyword may be used to tag many semantically-similar images, thus each image topic at Flickr may consist of large amount of semantically-similar images with diverse visual properties (i.e., some topics may contain more than 100,000 images at Flickr). Unfortunately, most existing keyword-based image retrieval systems tend to return all these images to the users without taking their personal preferences into consideration. Thus query-by-topic via keyword matching will return large amounts of semantically-similar images under the same topic and users may seriously suffer from the problem of information overload. In order to tackle this problem in our ***JustClick*** system, we have developed a novel framework for personalized image recommendation and it consists of three major components: (a) *Topic-Driven Image Summarization and Recommendation*: The semantically-similar images under the same topic are first partitioned into multiple clusters according to their nonlinear visual similarity contexts, and a limited number of images are automatically selected as the most representative images according to their representativeness for a given image topic. Our system can also allow users to define the number of such most representative images for relevance assessment. (b) *Context-Driven Image Visualization and Exploration*: Kernel PCA and hyperbolic visualization are seamlessly integrated to enable interactive image exploration according to their inherent visual similarity contexts, so that users can assess the relevance between the recommended images (i.e., most representative images) and their real query intentions more effectively. (c) *Intention-Driven Image Recommendation*: An interactive user-system interface is designed to allow the user to express his/her time-varying query intentions easily for directing the system to find more relevant images according to his/her personal preferences.

It is worth noting that the processes for kernel-based image clustering, topic-driven image summarization and recommendation (i.e., most representative image recommendation) and context-driven image visualization can be performed off-line without considering the users' personal preferences. Only the processes for interactive image exploration and intention-driven image recommendation should be performed on-line and they can be achieved in real time.

A. Topic-Driven Image Summarization and Recommendation

The visual properties of the images and their visual similarity contexts are very important for users to assess the relevance between the images and their real query intentions. Unfortunately, Flickr image search engine has completely ignored such important characteristics of the images. To characterize the diverse visual principles of the images efficiently and effectively, both the global visual features and the local visual features should be extracted for image similarity characterization [11-13]. To avoid the pitfalls of the image segmentation tools, segmentation is not performed for feature extraction. In our current implementations, 16-bin color histogram [11] and 62-dimensional texture features from Gabor filter banks [12] are used to characterize the global visual properties of the images, a number of interest points and their SIFT (scale invariant feature transform) features are used to characterize the local visual properties of the images [13]. As shown in Fig. 6, one can observe that our feature extraction operators can effectively extract the principal visual properties of the images.

To achieve more accurate approximation of the diverse visual similarities between the images, different kernels should be designed for various feature subsets because their statistical properties of the images are very different. Unfortunately, most existing machine learning tools use one single kernel

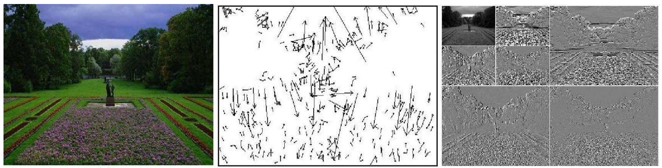

(a) (b) (c)

Fig. 6. Visual feature extraction for image similarity characterization: (a) original images; (b) interest points and SIFT vectors; (c) wavelet transformation.

for characterizing the diverse visual similarities between the images and completely ignore the heterogeneity of the statistical properties of the images in the high-dimensional multi-modal feature space. Based on these observations, we have studied the particular statistical property of the images under each feature subset, and the gained knowledge is then used to design the most suitable kernel for each feature subset [19]. Thus three particular image kernels (color histogram kernel, wavelet filter bank kernel, interest point matching kernel) are first constructed to characterize the diverse visual similarities between the images, and a linear combination of these three particular image kernels (i.e., mixture-of-kernels) can further form a family of mixture-of-kernels for characterizing the diverse visual similarities between the images more accurately [19]. Because multiple kernels are seamlessly integrated to characterize the heterogeneous statistical properties of the images in the high-dimensional multi-modal feature space, our mixture-of-kernels algorithm can achieve more accurate image clustering and can also provide a natural way to add new feature subsets and their particular kernels incrementally.

For a given image pair I and J under the same topic, their visual similarity context is characterized by using a mixture of three basic image kernels (i.e., mixture-of-kernels) [19]:

$$\kappa(I,J) = \sum_{h=1}^{3} \alpha_h \kappa_h(I,J), \qquad \sum_{h=1}^{3} \alpha_h = 1 \tag{13}$$

where α_h is the importance factor for the hth basic image kernel. Our mixture-of-kernels algorithm can achieve more accurate approximation of the diverse visual similarities between the images and produce nonlinear separation hypersurfaces between the images. Thus it can achieve more accurate image clustering and exploit the nonlinear visual similarity contexts for image visualization.

The semantically-similar images (which belong to the same image topic) are automatically partitioned into multiple clusters according to their kernel-based visual similarity contexts. The optimal partition of the images under the same topic is obtained by minimizing the trace of the within-cluster scatter matrix, S_w^ϕ. The scatter matrix is given by:

$$S_w^\phi = \frac{1}{N} \sum_{l=1}^{\tau} \sum_{i=1}^{N} \beta_{li} \left(\phi(x_i) - \mu_l^\phi\right) \left(\phi(x_i) - \mu_l^\phi\right)^T \tag{14}$$

where $\phi(x_i)$ is the mapping function of the image with the visual features x_i, N is the number of images and τ is the number of clusters, β_{li} is the membership parameter, $\beta_{li} = 1$ if $x_i \in \hat{C}_l$ and 0 otherwise, μ_l^ϕ is the center of the lth cluster $\hat{C}_l$ and it is given as:

$$\mu_l^\phi = \frac{1}{N_l} \sum_{i=1}^{N} \beta_{li} \phi(x_i), \qquad N_l = \sum_{i=1}^{N} \beta_{li} \tag{15}$$

where N_l is the number of images in the lth cluster. The trace of the scatter matrix S_w^ϕ can be computed by:

$$T_r\left(S_w^\phi\right) = \frac{1}{N} \sum_{l=1}^{\tau} \sum_{i=1}^{N} \beta_{li} \left(\phi(x_i) - \mu_l^\phi\right)^T \left(\phi(x_i) - \mu_l^\phi\right) \tag{16}$$

and it can further be re-written as:

$$T_r\left(S_w^\phi\right) = \frac{1}{N} \sum_{l=1}^{\tau} \sum_{i=1}^{N} \beta_{li} \Delta^2(x_i, \mu_l^\phi) \tag{17}$$

where $\Delta^2(x_i, \mu_l^\phi)$ is defined as:

$$\Delta^2(x_i, \mu_l^\phi) = \kappa(x_i, x_i) - \frac{2}{N_l} \sum_{j=1}^{N} \beta_{lj} \kappa(x_i, x_j) +$$

$$\frac{1}{N_l^2} \sum_{j=1}^{N} \sum_{m=1}^{N} \beta_{li} \beta_{lm} \kappa(x_j, x_m) \tag{18}$$

$$\kappa(x, x_i) = \phi(x)^T \phi(x_i) = \sum_{h=1}^{3} \alpha_h \kappa_h(x, x_i) \tag{19}$$

$\Delta^2(x_i, \mu_l^\phi)$ can further be refined as:

$$\Delta^2(x_i, \mu_l^\phi) = \sum_{h=1}^{3} \alpha_h \bar{\Delta}^2(x_i, \mu_l^\phi) \tag{20}$$

$$\bar{\Delta}^2(x_i, \mu_l^\phi) = \kappa_h(x_i, x_i) - \frac{2}{N_l} \sum_{j=1}^{N} \beta_{lj} \kappa_h(x_i, x_j) +$$

$$\frac{1}{N_l^2} \sum_{j=1}^{N} \sum_{m=1}^{N} \beta_{li} \beta_{lm} \kappa_h(x_j, x_m) \tag{21}$$

The trace of the scatter matric S_w^ϕ can be refined as:

$$T_r\left(S_w^\phi\right) = \sum_{h=1}^{3} \alpha_h \left(\frac{1}{N} \sum_{l=1}^{\tau} \sum_{i=1}^{N} \beta_{li} \bar{\Delta}^2(x_i, \mu_l^\phi) \right) \tag{22}$$

To achieve geometrical interpretation of the image clusters, we can define a cluster sphere for describing the images in the same cluster (i.e., the sphere with minimal radius containing all the images in the same cluster) [33]. The images, which locate on the boundary of the cluster sphere, are treated as the support vectors for the corresponding image cluster. Thus the distance between a given image with the visual features x and the center μ_l^ϕ of the best matching cluster $\hat{C}_l$ can be defined as:

$$R_l^2(x) = \|\phi(x) - \mu_l^\phi\|^2 = \Delta^2(x, \mu_l^\phi), \qquad x \in \hat{C}_l \tag{23}$$

The radius of the cluster sphere is given by the distance between a support vector and the center of the cluster sphere, thus the radius of the cluster sphere for the lth image cluster $\hat{C}_l$ can be defined as:

$$R_l = \{R_l(x_i) = \Delta(x_i, \mu_l^\phi) | x_i \in \Theta_l\} \tag{24}$$

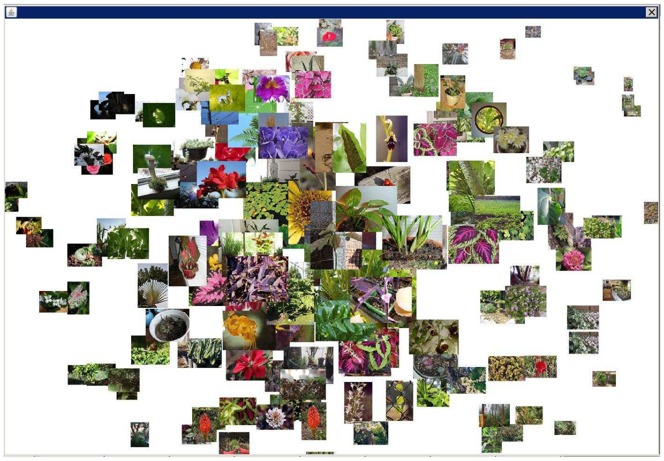


Fig. 7. Our representativeness-based sampling technique can tackle the overlapping problem by selecting 200 most representative images to represent and preserve the visual similarity contexts between 28365 semantically-similar images under the same topic "plant".

where Θ_l is the set of support vectors of the lth cluster $\hat{C}_l$.

The image with the visual features y can be detected as the outlier (i.e., $O(y) = 1$):

$$O(y) = \begin{cases} 1, & if \;\; R_l^2(y) > R_l^2 \;\; for \;\; all \;\; l \in [1, \tau] \\ 0, & otherwise \end{cases} \tag{25}$$

Based on such geometrical interpretation of the image clusters, searching the optimal values of the elements τ and α that minimizes the expression of the trace in Eq. (22) can be achieved effectively by using two iterations: (a) outer α iteration; and (b) inner τ iteration.

Ideally, a good combination of these three basic image kernels (i.e., with optimal values for these three α parameters) should be able to achieve more accurate characterization of the nonlinear visual similarity contexts between the images and result in better image clustering with less overlapping between the spheres for different image clusters. Thus the optimal values of the parameters α for combining three basic image kernels are obtained by minimizing the *cluster sphere overlapping*:

$$min \left\{ \sum_{i=1}^{N} \Big(R_l(x_i) \leq R_l \cap R_k(x_i) \leq R_k | l, k \in [1, \tau] \Big) \right\} \tag{26}$$

where $R_l(x_i) = \Delta(x_i, \mu_l^\phi)$ and $R_k(x_i) = \Delta(x_i, \mu_k^\phi)$ are used to determine the distances between the given image with the visual features x_i and the centers for the image clusters $\hat{C}_l$ and $\hat{C}_k$. To reduce the computational cost for the outer α iteration, we have pre-defined a set of the potential combinations of these three α parameters with different values. Such pre-defined set of α parameters can be obtained from a small set of images via semi-supervised learning. Thus the problem for finding an optimal combination of these three basic image kernels (i.e., finding optimal values of α) is simplified to search an optimal unit sequentially over the pre-defined set of the potential combinations of these three α parameters.

For the inner τ iteration (each iteration picks one integer from $[\tau_{min}, \tau_{max}]$ sequentially), the membership parameters β are determined automatically by minimizing the trace of the scatter matric S_w^ϕ in Eq.(22). In this inner τ iteration procedure, our algorithm uses a K-means-like strategy [31], i.e., updating all the cluster centers repeatedly according to the image memberships (i.e., the parameters β) which are obtained by minimizing the trace of the scatter matric S_w^ϕ. This inner τ iteration procedure will stop until the cluster centers become stable (no change anymore).

When there are N images under the given topic, the computational cost for obtaining their kernel matrix is approximated as $O(N^2)$. Therefore, the total computational cost for clustering these N images into τ clusters can be approximated as $O(\tau N^3)$. Thus supporting kernel-based image clustering may require huge memory space to store the kernel matrix when the given topic consists of large amount of semantically-similar images. To address this problem, we have developed a new algorithm for reducing the memory cost by seamlessly integrating parallel computing with global decision optimization. Our new algorithm takes the following key steps: (a) To reduce the memory cost, the images under the same topic are randomly partitioned into multiple smaller subsets. (b) Our kernel-based image clustering algorithm is performed parallelly on all these image subsets to obtain a within-subset partition of the images according to their diverse visual similarity contexts. (c) The support vectors for each image subset are validated by other image subsets through testing Karush-Kuhn-Tucker (KKT) conditions. The support vectors, which violate the KKT conditions, are integrated to update the decision boundaries for the corresponding image subset incrementally. This process is repeated until the global optimum is reached and an optimal partition of large amount of images under the same image topic is obtained.

Our kernel-based image clustering algorithm has the following advantages: (1) It can seamlessly integrate multiple kernels to characterize the diverse visual similarities between the images more accurately. Thus it can provide a good insight of large amount of images by determining their global distribution structures (i.e., image clusters and their similarity contexts) accurately, and such global image distribution structures can further be used to achieve more effective image visualization by selecting the most representative images automatically from all these image clusters. (2) Only the most representative images (which are the support vectors) are stored and validated by other image subsets, thus it requests far less memory space. The redundant images (which are not the support vectors) are eliminated early, thus the kernel-based image clustering process can be accelerated significantly. (3) The support vectors for each image subset are validated by other image subsets, thus our algorithm can handle the outliers and noise effectively and it can generate more robust clustering results.

To allow users to assess the relevance between the images returned by the keyword-based query and their real query intentions, it is very important to visualize the semantically-similar images under the same topic according to their inherent visual similarity contexts. Because each topic may relate to large amounts of semantically-similar images, visualizing such large-size of images on a size-limited display screen

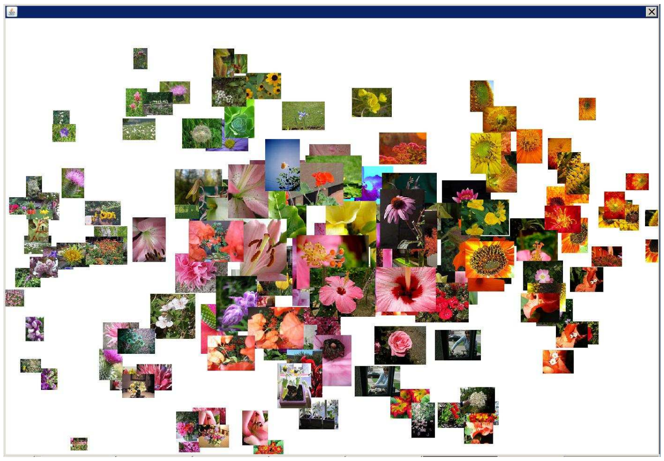


Fig. 8. Our representativeness-based sampling technique can tackle the overlapping problem by selecting 200 most representative images to represent and preserve the visual similarity contexts between 26858 semantically-similar images under the same topic “flower”.

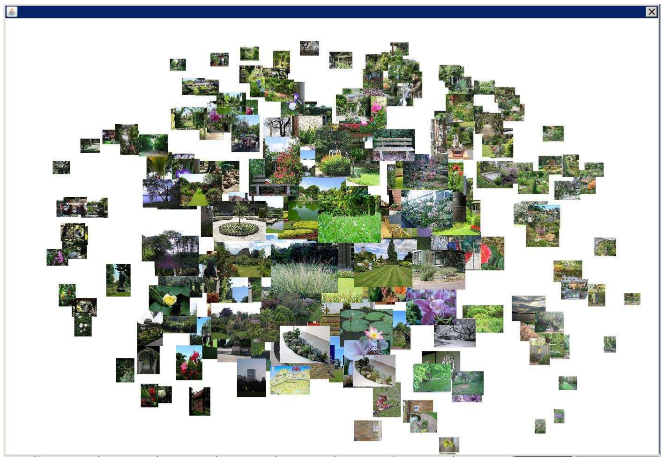


Fig. 9. Our representativeness-based sampling technique can tackle the overlapping problem by selecting 200 most representative images to represent and preserve the visual similarity contexts between 30887 semantically-similar images under the same topic “garden”.

may seriously suffer from the overlapping problem (e.g., overlapping between the images may decrease the visibility of the images significantly and the overlapped images will compete each other visually for human attention). On the other hand, displaying large amounts of redundant images with similar visual properties to the users cannot provide them with any additional information. Obviously, simply selecting only one iconic image from each image cluster is unable to represent and preserve the nonlinear visual similarity contexts among large amounts of semantically-similar images under the same topic [18], thus more effective techniques are strongly expected to achieve representativeness-based image summarization. Based on these understandings, we have developed a novel algorithm to achieve representativeness-based image summarization and overlapping reduction by selecting the most representative images automatically according to their representativeness for a given image topic.

Different clusters are used to interpret different groups of the images with various visual properties and different important aspects for a given image topic, thus the most representative images should be selected from all these clusters and the outliers to preserve the nonlinear visual similarity contexts between the images. Obviously, different clusters may contain different numbers of images, thus larger number of such most representative images should be selected from the clusters with bigger coverage percentages. On the other hand, some representative images should prior be selected from the outliers for supporting serendipitous discovery of unexpected images. The optimal number of such most representative images depends on their effectiveness and efficiency for representing and preserving the nonlinear visual similarity contexts among large amounts of semantically-similar images under the same topic.

For the visually-similar images in the same cluster, we have seamlessly integrated both the subjective measure and the objective measure to quantify their representativeness scores. The subjective measure of an image depends on the users’ rating scores that are available at Flickr. The objective measure of an image depends on its representativeness for the underlying nonlinear visual similarity contexts between the images. Thus three types of images can be selected to achieve representativeness-based summarization of the visually-similar images in the same cluster: (1) The images, which locate on the cluster sphere and are treated as the support vectors for the corresponding image cluster, can effectively capture the essentials (i.e., principal visual properties) of the images in the same cluster; (2) The images, which locate at the center of the cluster and are far away from the cluster sphere, are more representative for the popular images located in the areas with higher densities; and (3) The images, which have higher rating scores from numerous online users, are more interesting to the users. Thus the *representativeness score* $\rho(x)$ for a given image with the visual features x can be defined as:

$$\rho(x) = \bar{\rho}(x) + \bar{\rho}(x) \times \frac{e^{UR(x)} - e^{-UR(x)}}{e^{UR(x)} + e^{-UR(x)}}, \qquad \bar{\rho}(x) = e^{d^2(x, \hat{C}_k)} \tag{27}$$

where $\bar{\rho}(x)$ is the objective measure of the representativeness score for the image with the visual feature x, $UR(x)$ is the users’ rating score for the given image, and $d(x, \hat{C}_k)$ is the context-oriented correlation between the given image and the corresponding image cluster $\hat{C}_k$.

The context-oriented correlation $d(x, \hat{C}_k)$ can be defined as:

$$d(x, \hat{C}_k) = \begin{matrix} max \\ l \end{matrix} \left\{ max \left\{ - \sum_{x_i \in \Theta_l} \beta_{li} \kappa(x, x_i), -\beta_{li} R_l^2(x) \right\} \right\} \tag{28}$$

where $-\sum_{x_i \in \Theta_l} \beta_{li} \kappa(x, x_i)$ is used to characterize the correlation between the given image and the decision boundary (i.e., support vectors) for the image cluster $\hat{C}_l$, Θ_l is the set of the support vectors for the image cluster $\hat{C}_l$, and $-\beta_{li} R_l^2(x)$ is used to characterize the correlation between the given image and the center for the image cluster $\hat{C}_l$. Thus the images, which are closer to the decision boundary or closer to the cluster center for one of these τ image clusters or have higher users’ rating scores, will have larger values of the representativeness scores $\rho(\cdot)$. The images in the same cluster can be ranked precisely according to their representativeness scores, and the most representative images with larger values of $\rho(\cdot)$ can be selected and be recommended to the users for relevance assessment.

The user profiles are not required for selecting the most representative images, thus our topic-driven image recommendation scheme can support new users effectively. Only the most representative images are recommended, and large amount of redundant images, which have similar visual properties with the most representative images, are eliminated automatically. Through supporting topic-driven image recommendation, our ***JustClick*** system can significantly reduce both the information overload for relevance assessment and the visual complexity for image visualization and exploration.

B. Context-Driven Image Visualization and Exploration

To assist users in assessing the relevance between the most representative images (i.e., recommended images) and their real query intentions, it is very important to develop new visualization algorithms that are able to preserve the nonlinear visual similarity contexts between the images in the high-dimensional feature space. We have incorporated kernel PCA [32] to project the most representative images onto a hyperbolic plane, so that the nonlinear visual similarity contexts between the images can be preserved precisely for interactive image exploration.

The most representative images for a given image topic are first centered according to their center in the feature space. For a given most representative image with the visual features x, its feature-based representation can be centered by using the center μ^{ϕ} of these L most representative images:

$$\bar{\phi}(x) = \phi(x) - \mu^{\phi}, \qquad \mu^{\phi} = \frac{1}{L}\sum_{i=1}^{L}\phi(x_i), \qquad \sum_{i=1}^{L}\bar{\phi}(x_i) = 0 \tag{29}$$

The covariance matrix $\bar{K}$ and its component is defined as the dot product matrix:

$$\bar{K}_{ij} = \bar{\phi}(x_i)^T\bar{\phi}(x_j) = \kappa(x_i, x_j) \tag{30}$$

Then the kernel PCA is obtained by solving the eigenvalue equation:

$$\bar{K}\overrightarrow{V} = \lambda\overrightarrow{V} \tag{31a}$$

or

$$\bar{K}\overrightarrow{\omega} = \lambda L\overrightarrow{\omega} \tag{31b}$$

where $\overrightarrow{V}$ are the eigenvectors, L is the number of the most representative images, $\lambda = [\lambda_1, \cdots, \lambda_L]$ denotes the eigenvalues, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L$, and $\overrightarrow{\omega} = [\overrightarrow{\omega}_1, \cdots, \overrightarrow{\omega}_L]$ denotes the expansion coefficients of an Eigenvector,

$$\overrightarrow{V} = \sum_{i=1}^{L}\overrightarrow{\omega}_i\bar{\phi}(x_i), \qquad \overrightarrow{V}^k = \sum_{j=1}^{L}\overrightarrow{\omega}_j^k\bar{\phi}(x_j) \tag{32}$$

where $\overrightarrow{V}^k$ is used to interpret the eigenvectors with non-zero values of eigenvalues, $\overrightarrow{\omega}_j^k$ is the expansion coefficients for the top k eigenvectors with non-zero eigenvalues, $k < L$.

For a given image with the visual features x, its projection $P(x, \overrightarrow{V}^k)$ on the selected top k eigenvectors with non-zero eigenvalues can be defined as:

$$P(x, \overrightarrow{V}^k) = \sum_{j=1}^{L}\overrightarrow{\omega}_j^k\bar{\phi}(x_j)^T\bar{\phi}(x) = \sum_{j=1}^{L}\overrightarrow{\omega}_j^k\kappa(x, x_j) \tag{33}$$

The most representative images, which are projected on the top k principal components, are further mapped onto the hyperbolic plane. After such context-preserving projection of the most representative images is obtained, Poincaré disk model is used to map the most representative images on the hyperbolic plane onto a 2D display coordinate to support change of focus and interactive image exploration.

Even the low-level visual features may not be able to carry the image semantics directly, supporting similarity-based image visualization can significantly leverage humans' powerful capabilities on pattern recognition for interactive relevance assessment. Through change of focus, users can easily control the presentation and visualization of the recommended images for interactive relevance assessment.

Our hyperbolic visualization of the most representative images recommended for the image topics "plant", "garden" and "flower" are shown in Fig. 7, Fig. 8 and Fig. 9, where 200 most representative images for the image topics "plant", "garden" and "flower" are recommended and visualized. One can observe that such 2D hyperbolic visualization of the most representative images can provide an effective interpretation and summarization of the original visual similarity contexts among large amounts of semantically-similary images under the same topic. Visualizing the images according to their visual similarity contexts can allow users to find interesting visual similarity contexts between the images and discover more relevant images according to their nonlinear visual similarity contexts. Through selecting the most representative images for image summarization and visualization, our ***JustClick*** system can provide larger coverage of the diverse image contents in a sized-limited screen and save the users' efforts on relevance assessment significantly.

The change of focus is implemented for allowing users to navigate and explore the most representative images according to their nonlinear visual similarity contexts. Users can change their focuses of the images by clicking on any visible image to bring it into focus at the screen center, or by dragging any visible image interactively to any other screen location without losing their visual similarity contexts, where the rest of the images can transform appropriately. With the power of high interaction and rapid response for exploring and navigating the recommended images (i.e., most representative images) according to their nonlinear visual similarity contexts, our ***JustClick*** system can support more effective solution for users to assess the relevance between the recommended images and their real query intentions interactively.

C. Intention-Driven Image Recommendation

Through such interactive image exploration process via change of focus, users can easily build up their mental query models about which types of images they really want to look

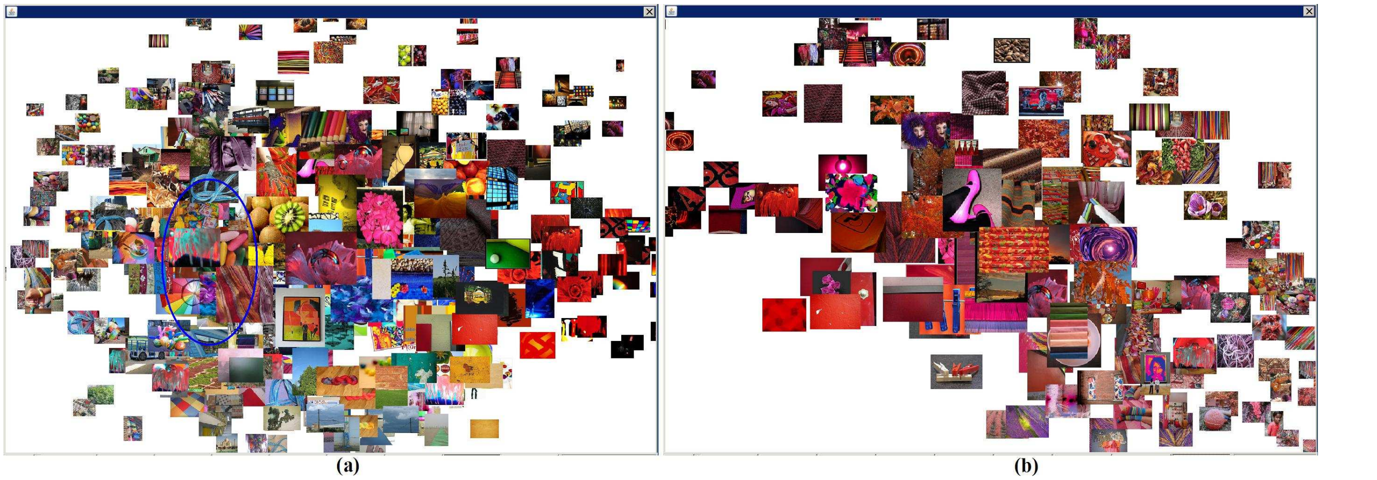


Fig. 10. (a) 500 most representative images recommended for the image topic "art work"; (b) Geometric zooming into the area of interest in the blue circle.

for and gain the insights about what are the most significant visual properties of the recommended images. After the users find some areas of interest via interactive exploration, our system can allow the users to zoom into the area of interest to look for some local visual similarity contexts between the images as shown in Fig. 10. Through zooming into the area of interest, the users may obtain some additional images of interest that may not be found from traditional page-by-page top ranking list, e.g., some interesting images, which belong to the outliers but are semantically relevant to the users' query intentions, may have low ranking scores and cannot be listed on the first few pages. Therefore, fortunate discoveries of some unexpected images can be achieved effectively by selecting the most representative images from the outliers autonomously and incorporating hyperbolic visualization to allow the users to zoom into the area of interest interactively. Thus our ***JustClick*** system can facilitate discovery of new knowledge through the interactive image exploration process.

Through interactive exploration of the recommended images (i.e., most representative images) according to their nonlinear visual similarity contexts, the users can easily find some particular images according to their personal interests. We have developed a new scheme to achieve user-adaptive image recommendation by autonomously adjusting the recommendation and visualization of the most representative images according to the users' time-varying query interests. After the users find some images of interest via interactive image exploration, our ***JustClick*** system can allow the users to click these images of interest to express their time-varying query interests interactively for directing the system to find more relevant images according to their personal preferences.

After such the user's time-varying query interests are captured, the personalized interestingness scores for the images under the same topic are calculated automatically, and the *personalized interestingness score* $\rho^p(x)$ for a given image with the visual feature x is defined as:

$$\rho^p(x) = \rho(x) + \rho(x) \times e^{-\kappa(x,x_c)} \tag{34}$$

$$\kappa(x, x_c) = \sum_{h=1}^{3} \alpha_h \kappa_h(x, x_c)$$

$$R_l(x) \leq R_l \cap R_l(x_c) \leq R_l \quad for \quad all \quad l \in [1, \tau] \tag{35}$$

where $\rho(x)$ is the original representativeness score for the given image, $\kappa(x, x_c)$ is the kernel-based visual similarity correlation between the given image with the visual features x and the clicked image with the visual features x_c which belong to the same image cluster. Thus the redundant images with larger values of the personalized interestingness scores, which have similar visual properties with the clicked image (i.e., belonging to the same cluster) and are initially eliminated for reducing the visual complexity for image summarization and visualization, can be recovered and be recommended to the users adaptively as shown in Fig. 11, Fig. 12, Fig. 13 and Fig. 14. One can observe that integrating the visual similarity contexts for personalized image recommendation can significantly enhance the users' ability on finding some particular images of interest even the low-level visual features may not be able to carry the semantics of the image contents directly. Thus integrating the visual similarity contexts between the images for personalized image recommendation can significantly enhance the users' ability on finding some particular images of interest. With a higher degree of transparency of the underlying image recommender, users can achieve their image retrieval goals (i.e., looking for some particular images) with a minimum of cognitive load and a maximum of enjoyment [3]. By supporting intention-driven image recommendation, users can maximize the amount of relevant images while minimizing the amount of irrelevant images according to their personal preferences.

It is worth noting that our ***JustClick*** system for personalized image recommendation is significantly different from traditional relevance feedback approaches for image retrieval [42-46]: (a) The relevance feedback approaches require users to label a reasonable number of returned images into the positive or negative classes for learning a reliable model to predict the user's query intentions, and thus they may bring huge burden on the users even active learning has recently been proposed for label propagation. On the other hand, our personalized image recommendation framework can allow the users to express their time-varying query intentions easily and thus it can lessen the burden on the users significantly. (b) When

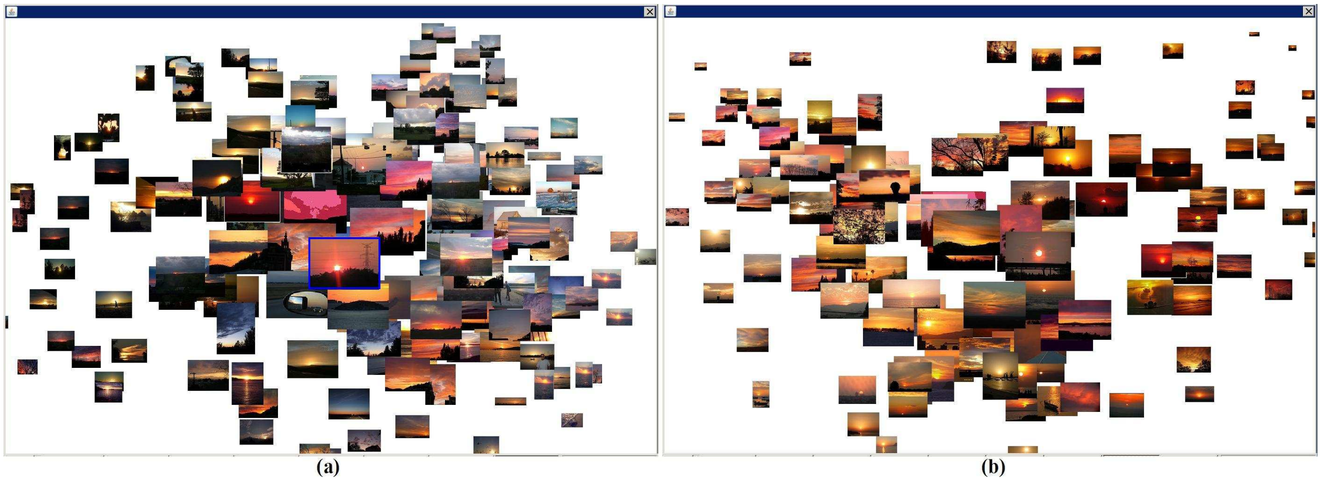


Fig. 11. Our *JustClick* system for personalized image recommendation: (a) the most representative images recommended for the topic-based query "sunset", where the image in blue box is clicked by the user (i.e., query intention); (b) more images which are similar with the accessed image are recommened adaptively according to the user's query intentions of "blight sunset".

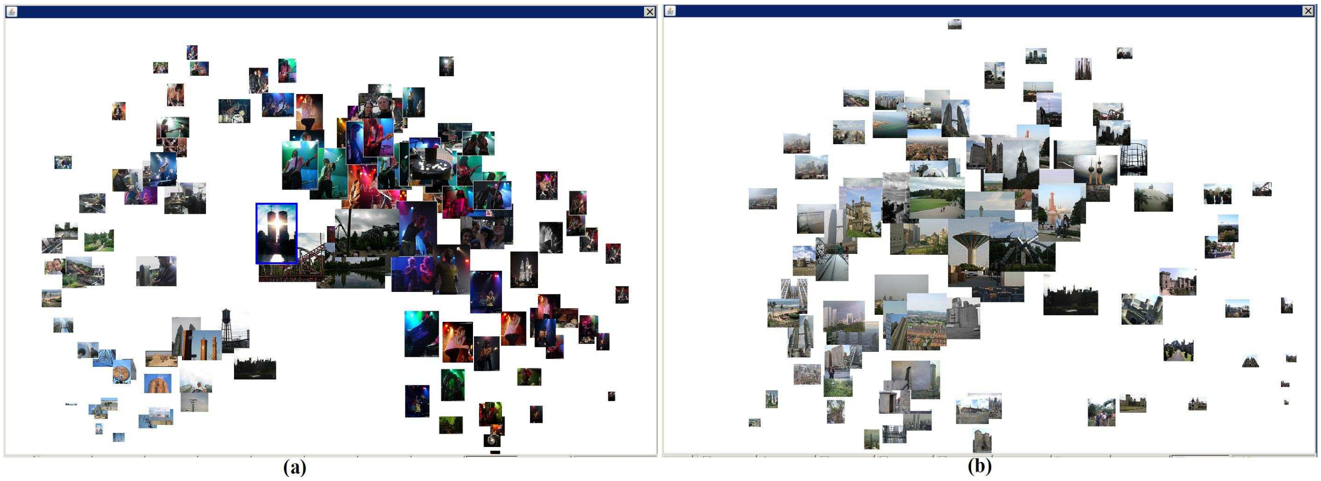


Fig. 12. Our *JustClick* system for personalized image recommendation: (a) the most representative images recommended for the topic-based query "towers", where the image in blue box is clicked by the user (i.e., query intention); (b) more images which are similar with the accessed image are recommened adaptively according to the user's query intentions of "tower building".

large-scale image collections come into view, a limited number of labeled images may not be representative for large amounts of unseen images and thus a limited number of labeled images are insufficient for learning an accurate model to predict the user's query intentions precisely. On the other hand, our personalized image recommendation framework can select a reasonable number of most representative images according to their representativeness of the nonlinear visual similarity contexts between the images. Thus the users can always be acquainted by the most representative images according to their personal preferences. (c) Most existing relevance feedback approaches use page-by-page ranked list to display the query results, and the nonlinear visual similarity contexts between the images are completely ignored. On the other hand, our personalized image recommendation framework can allow users to see the most representative images and their nonlinear visual similarity contexts at the first glance, and thus the users can obtain more significant insights and make better query decisions and assess the image relevance more effectively.

V. Algorithm and System Evaluation

We carry out our experimental studies by using large-scale collections of manually-tagged Flickr images with unconstrained contents and capturing conditions. We have downloaded more than 1.5 billions Flickr images and their tagging documents. We have learned a large-scale topic network with more than 4000 most popular image topics (i.e., most popular taggings along the time) at Flickr. Creating a new search engine which scales to such size of image collections may emerge many new challenges while offering many good opportunities for the CBIR community.

Our work on algorithm evaluation focus on: (1) evaluating the response time for supporting change of focus in our ***JustClick*** system, which is critical for supporting interactive exploration of large-scale topic network and large amounts of recommended images; (2) evaluating the performance (efficiency and accuracy) of our ***JustClick*** system for achieving personalized image recommendation according to the users' personal preferences; (3) evaluating the benefits for integrating topic network (i.e., a global overview of large-scale image collections), hyperbolic visualization and interactive image

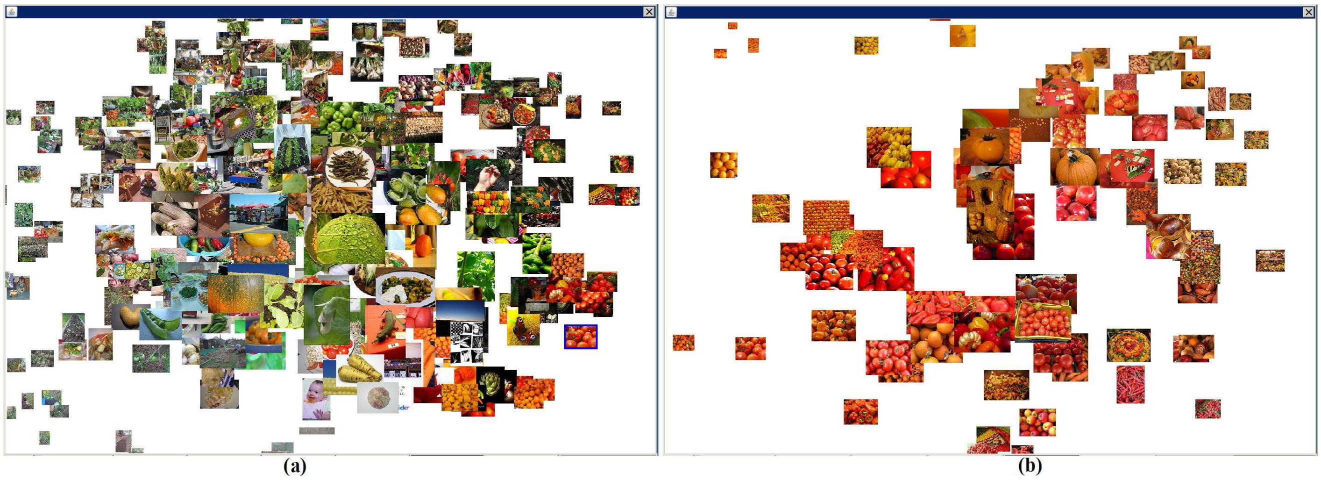


Fig. 13. Our *JustClick* system for personalized image recommendation: (a) The most representative images recommended for the keyword-based query "vegetables", where the image in blue box is clicked by the user; (b) The most relevant images recommended according to the user's query intention of "tomato".

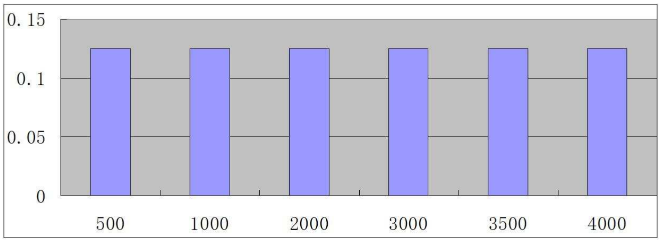


Fig. 15. The empirical relationship between the computational time T_1 (seconds) and the number of image topic nodes.

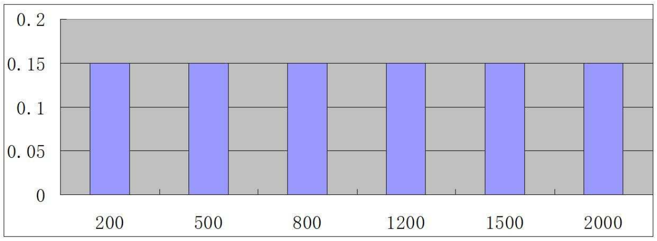


Fig. 16. The empirical relationship between the computational time T_1 (seconds) and the number of recommended images.

exploration for improving image search.

One critical issue for evaluating our personalized image recommendation system is the response time for supporting change of focus. In our ***JustClick*** system, the change of focus is used for achieving interactive exploration and navigation of large-scale topic network and large amounts of recommended images. The *change of focus* is implemented by changing the Poincaré mapping of the image topic nodes or the recommended images from the hyperbolic plane to the display unit disk, and the positions of the image topic nodes or the recommended images in the hyerbolic plane need not to be altered during the focus manipulation. Thus the response time for supporting change of focus depends on two components: (a) The computational time T_1 for re-calculating the new Poincaré mapping of large-scale topic network or large amounts of recommended images from a hyperbolic plane to a 2D display unit disk, i.e., re-calculating the Poincaré position for each image topic node or each recommended image; (b) The visualization time T_2 for re-layouting and re-visualizing large-scale topic network or large amounts of recommended images on the display disk unit according to their new Poincaré mappings.

Because the computational time T_1 may depend on the number of image topic nodes, we have tested the performance differences for our system to re-calculate the Poincaré mappings for different numbers of image topic nodes. Thus our topic network with 4000 image topic nodes is partitioned into 5 different scales: 500 nodes, 1000 node, 2000 nodes, 3000 nodes, 3500 nodes and 4000 nodes. We have tested the computational time T_1 for re-calculating the Poincaré mappings of different numbers of image topic nodes when the focus is changed. As shown in Fig. 15, one can find that the computational time T_1 is not sensitive to the number of image topics, and thus re-calculating the Poincaré mapping for large-scale topic network can almost be achieved in real time.

Following the same approach, we have also evaluated the empirical relationship between the computational time T_1 and the number of the recommended images. By computing the Poincaré mappings for different numbers of the recommended images, we have obtained the same conclusion, i.e., the computational time T_1 for re-calculating the new Poincaré mappings is not sensitive to the number of the recommended images as shown in Fig. 16, and thus re-calculating the Poincaré mapping for large amounts of recommended images can almost be achieved in real time.

We have also evaluated the empirical relationship between the visualization time T_2 and the number of image topic nodes and the number of recommended images. In our experiments, we have found that re-visualization of large-scale topic network and large amounts of recommended images is not sensitive to the number of image topics and the number of recommended images, and thus our system can support re-visualization of large-scale topic network and large amounts of recommended images in real time.

From these evaluation results, one can conclude that our ***JustClick*** system can support change of focus in real time, and thus our ***JustClick*** system can achieve interactive navigation and exploration of large-scale image collections effectively.

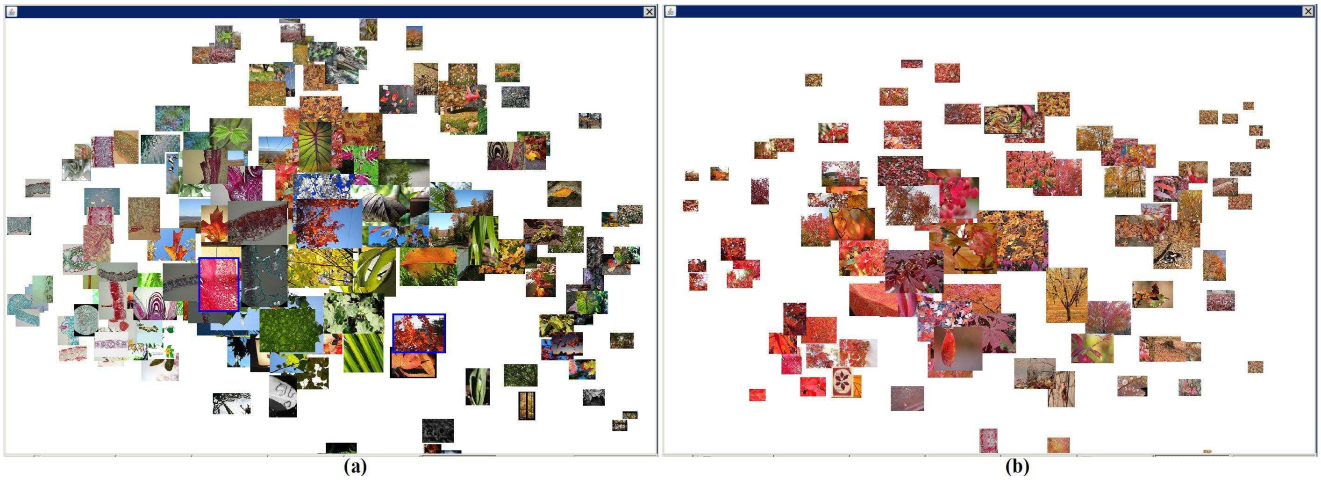


Fig. 14. Our *JustClick* system for personalized image recommendation: (a) The most representative images recommended for the keyword-based query “leaves”, where the image in read box is clicked by the user; (b) The most relevant images recommended according to the user’s query intention of “red leaves”.

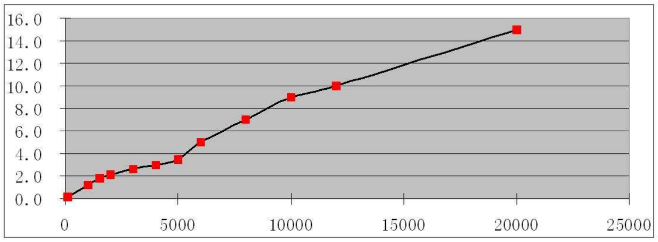


Fig. 17. The empirical relationship between the computational cost Ω_1 (seconds) and the number of images.

To support similarity-based visualization of large amount of most representative images recommended for each image topic, the computational cost depends on two issues: (a) The computational cost Ω_1 for supporting kernel-based image clustering and achieving representativeness-based image summarization; (b) The computational cost Ω_2 for performing kernel PCA on the most representative images to obtain their similarity-preserving projections on the hyperbolic plane.

To achieve kernel-based image clustering, the kernel matrix for the images should be calculated and the computational cost largely depends on the number of images for the given image topic. The computational cost for achieving kernel-based image clustering is approximated as $O(\tau N^3)$, where N is the total number of the images and τ is the number of image clusters. Because each image topic in Flickr may consist of large amount of images, we have obtained the empirical relationship between the computational cost Ω_1 (CPU time) and the number of images as shown in Fig. 17. One can observe that the computational cost Ω_1 increases exponentially with the number of images. Based on this observation, the images under the same topic are first partitioned into multiple subsets, parallel computing and global decision optimization are integrated to reduce the computational cost significantly from $O(\tau N^3)$ to $O(\frac{\tau N^3}{\mu^3})$, where μ is the number of image subsets. It is also worth noting that such cost-sensitive process for kernel-based image clustering can be performed off-line.

After the images under the same topic are partitioned into multiple clusters via kernel-based clustering, our system can select the most representative images automatically and perform kernel PCA to obtain their similarity-preserving projections on the hyperbolic plane. The computational cost for performing kernel PCA is approximated as $O(L^3)$, where L is the number of the most representative images recommended for the given image topic. As shown in Fig. 18, we have obtained the empirical relationship between the computational cost Ω_2 and the number of most representative images. One can observe that the computational cost Ω_2 exponentially increases with the number of the most representative images. In our ***JustClick*** system, the number of the most representative images is normally less than 500, thus the computational cost Ω_2 is acceptable for supporting interactive image exploration and navigation.

When the most representative images for the given image topic are recommended and visualized, our system can further allow users to click one or multiple images of interest for expressing their query intentions and directing our system to find more relevant images according to their personal preferences. For evaluating the effeciency and the accuracy of our personalized image recommendation system, the *benchmark metric* includes *precision* θ and *recall* ϑ. The precision θ is used to characterize the accuracy of our system for finding more relevant images according to the user’s personal preferences, and the recall ϑ is used to characterize the efficiency of our system for finding more relevant images. They are defined as:

$$\theta = \frac{\zeta}{\zeta + \varsigma}, \qquad \vartheta = \frac{\zeta}{\zeta + \xi} \tag{36}$$

where ζ is the set of true positive images that are visually-similar with the images accessed by the users and are recommended correctly, ς is the set of fause positive images that are visually-similar with the accessed images and are not recommended, and ξ is the set of true negative images that are visually-dissimilar with the accessed images but are recommended incorrectly.

Table 1 gives the precision and recall of our *JustClick* system for personalized image recommendation. From these experimental results, one can observe that our system can support personalized image recommendation effectively. Thus the visual properties of the images (i.e., characterized by

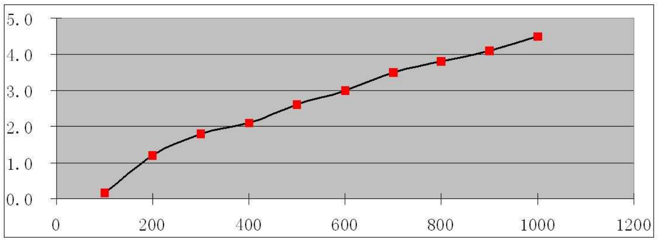


Fig. 18. The empirical relationship between the computational cost Ω_2 (seconds) and the number of the most representative images.

TABLE I

The precision and recall for our *JustClick* system to look for some particular images via intention-driven image recommendation.

query	apples	flowers	garden	glasses
θ/ϑ	90.3% /94.2%	93.3% /92.8%	91.2% /92.6%	91.3% /91.8%
query	stockings	mushrooms	cats	tigers
θ/ϑ	82.5% /83.6%	86.3% /84.2%	86.2% /82.9%	80.6% /79.8%
query	vegetables	roses	feathers	socks
θ/ϑ	85.8% /86.3%	86.5% /86.2%	81.2% /81.3%	83.2% /84.1%
query	trees	shoes	woods	planes
θ/ϑ	94.6% /94.3%	90.1% /89.8%	83.5% /83.6%	90.2% /90.7%
query	stars	rivers	paintings	shops
θ/ϑ	89.3% /86.8%	89.8% /89.6%	89.6% /90.2%	91.2% /91.9%
query	flags	cities	signs	buildings
θ/ϑ	90.6% /91.2%	88.6% /84.8%	90.2% /91.2%	89.3% /90.2%
query	mountaints	parks	sunsets	doors
θ/ϑ	90.6% /90.4%	90.8% /91.7%	91.5% /92.3%	92.8% /92.8%
query	parties	springs	windows	walls
θ/ϑ	80.6% /79.8%	80.6% /80.7%	82.6% /80.8%	91.8% /91.5%
query	temples	holidays	weddings	vacations
θ/ϑ	90.6% /91.6%	89.6% /90.2%	85.8% /87.4%	82.6% /90.5%
query	leaves	birthdays	streets	plants
θ/ϑ	81.3% /81.5%	82.8% /80.3%	86.8% /86.5%	80.6% /80.9%
query	mars	men	lakes	waterfall
θ/ϑ	91.2% /91.5%	87.5% /88.2%	85.6% /86.7%	89.4% /91.5%
query	hands	bubbles	stones	faces
θ/ϑ	78.5% /79.3%	82.5% /85.6%	84.3% /85.2%	75.6% /74.2%
query	monkeys	girls	lions	bears
θ/ϑ	76.3% /77.4%	81.6% /82.8%	80.5% /81.6%	79.3% /79.8%
query	motorcycles	cars	bicycles	trains
θ/ϑ	83.8% /84.5%	87.9% /88.8%	89.6% /90.7%	90.3% /89.8%
query	victoria	baby	california	holiday
θ/ϑ	89.2% /87.3%	81.6% /81.7%	91.3% /92.5%	89.6% /89.4%
query	art	food	urban	boats
θ/ϑ	84.5% /84.8%	83.6% /86.7%	83.9% /90.2%	90.8% /86.9%
query	ocean	waves	summer	spain
θ/ϑ	86.6% /87.8%	85.4% /88.9%	80.5% /80.3%	80.8% /80.6%

using the low-level visual features) are very important for achieving more effective image retrieval, even the low-level visual features may not be able to carry the semantics of image contents directly. It is also worth noting that such interactive process for intention-driven image recommendation can be achieved in real time, and thus our *JustClick* system can support interactive image exploration on-line.

Our evaluation of the benefits from similarity-based image visualization on assisting users to access large-scale image collections focuses on three issues: (a) Do our topic network visualization and exploration tools allow users to communicate their image needs more effectively and precisely? (b) Do our hyperbolic image visualization and interactive exploration tools allow users to direct the system for finding more relevant images effectively? (c) Do our hyperbolic image visualization and interactive exploration tools allow users to assess the relevance between the recommended images and their real query intentions more effectively?

When large-scale collections of shared Flickr images come into view, it is reasonable to assume that users are unfamiliar with the image contents (which is significantly different from personal image collections [26-27]). Thus query formulation (specifying the image needs precisely) is one critical issue for users to access large-scale image collections. On the other hand, users may expect to formulate the image needs intuitively not just type the keywords. By incorporating topic network to summarize and visualize large-scale image collections at a semantic level, our ***JustClick*** system can make all these image topics to be visible to the users as shown in Fig. 1, Fig. 2 and Fig. 3, so that they can have a good global overview of large-scale image collections at the first glance. Our hyperbolic topic network visualization algorithm can achieve a good balance between the local detail around the users' current focus which is embedded in the global contexts of the topic network, thus the users can see not only the image topic in current focus but also the appropriate directions for future search as shown in Fig. 4. Through such user-system interaction process, users can easily communicate their image needs by selecting the visible image topics on the topic network directly.

Our context-driven image visualization and exploration algorithm can help human beings understand the image contents and the visual similarity contexts between the images at the first glance. Our interactive user-system interface can also allow users to express their time-varying query interests easily for directing the system to find more relevant images according to their personal preferences. Thus our ***JustClick*** system for personalized image recommendation can significantly improve the users' ability on locating some particular images of interest or a group of visually-similar images as shown in Fig. 11, Fig. 12, Fig. 13 and Fig. 14.

The assessment of the relevance between the images and the users' query intentions is strongly influenced by the inherent visual similarity contexts. Our ***JustClick*** system can exploit and preserve the inherent visual similarity contexts between the images effectively. Through change of focus, our ***JustClick*** system can also allow users to assess the nonlinear visual similarity contexts between the images interactively via simple mouse dragging without losing the nonlinear visual similarity contexts between the images.

VI. Conclusions and Future Works

In this paper, we have developed a novel framework called ***JustClick*** to enable personalized image recommendation via exploratory search from large-scale collections of Flickr images. The topic network is automatically generated to summarize and visualize large-scale image collections at a semantic level. The nonlinear visual similarity contexts between the images are exploited to select a limited number of most representative images to summarize large amounts of semantically-similar images under the same topic according to their representativeness scores. Kernel PCA and hyperbolic visualization are used to exploit and preserve the nonlinear similarity structures between the images more effectively, so that users can navigate and explore the most representative images according to their nonlinear visual similarity contexts and assess the relevance between the recommended images and the real query intentions interactively. An interactive user-system interface is designed to allow users to express their time-varying query intentions easily for directing our system to find more relevant images according to their personal preferences. Our *JustClick* system has provided a novel framework to integrate keyword-based image retrieval with content-based image retrieval seamlessly for enhancing image search. Our experiments on large-scale image collections (1.5 billions Flickr images) with diverse semantics (4000 image topics) have obtained very positive results.

Our future works will focus on: (a) integrating space optimization to improve the space utilization efficiency of our hyperbolic algorithm for topic network visualization and supporting theoretical analysis of our context-preserving image projection algorithm; (b) developing new algorithms to assess the quality of the topic network, such as its representativeness for a given set of annotated images or whether it can cover all potential queries from large group of users with diverse query interests; (c) investigating the algorithms for visualizing and exploring large-scale image collections when text annotations are not available; (d) developing more effective machine learning algorithm to generate the semantic tags automatically for all the image clusters under the same topic, and integrate such semantic tags to help users understand the underlying visual similarity structures; (e) developing new algorithm for selecting more suitable kernels for different image topics and selecting more discriminative feature subsets for different image clusters under the same topic; (f) achieving more effective visualization of time-varying image collections (i.e., topic network with the appearances of new topics and inter-topic associations, context-preserving image projection and visualization with the appearances of new visual similarity contexts, covariance matrix calculation with the exponential growth and rapid change of online image collections, and incremental kernel PCA without re-solving the eigenvalue problem); and (g) integrating social network for supporting personalized image recommendation.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their insightful comments and suggestions to make this paper more readable. We also want to thank the Associate Editor Prof. Ling Guan for his handling the review process of our paper. This work is supported by National Science Foundation under 0601542-IIS and 0208539-IIS. Part of this work is performed when Jianping Fan visits Konstanz University.

REFERENCES

[1] M.S. Lew, N. Sebe, C. Djeraba, R. Jain, “Content-based multimedia information retrieval: State of the art and challenges”, *ACM Trans. on Multimedia Computing and Applications*, vol. 2, no.1, pp.1-19, 2006.

[2] N. Sebe, Q. Tian, “Personalized multimedia retrieval: the new trend?”, ACM Workshop on Multimedia Information Retrieval, pp.299-306, 2007.

[3] G. Marchionini, “Exploratory search: from finding to understanding”, *Commun. of ACM*, vol. 49, no.4, pp. 41-46, 2006.

[4] S.-F. Chang, J. R. Smith, M. Beigi, A. B. Benitez, “Visual information retrieval from large distributed on-line repositories”, *Comm. of the ACM*, vol.40, no.12, pp.63-71, 1997.

[5] S. Santini, A. Gupta, R. Jain, “Emergent semantics through interaction in image databases”, *IEEE Trans. Knowledge and Data Engineering*, vol.13, no.3, pp.337-351, 2001.

[6] J. Vendrig, M. Worring, A.W.M. Smeulders, “Filter image browsing: Interactive image retrieval by using database overviews”, *Multimedia Tools and Applications*, vol.15, pp.83-103, 2001.

[7] A. Jaimes, “Human factors in automatic image retrieval system design and evaluation”, Proc. SPIE, vol.6061, 2006.

[8] J.S. Hare, P.H. Lewis, P.G.B. Enser, C.J. Sandom, “Mind the gap: another look at the problem of the semantic gap in image retrieval”, Proc. SPIE, vol. 6073, 2006.

[9] M. Naphade, J.R. Smith, J. Tesic, S.-F. Chang, W. Hsu, L. Kennedy, A. Hauptmann, J. Curtis, “Large-scale concept ontology for multimedia”, *IEEE Multimedia*, 2006.

[10] A. B. Benitez, J. R. Smith, S.-F. Chang, “MediaNet: A multimedia information network for knowledge representation”, Proc. SPIE, vol. 4210, 2000.

[11] Y. Deng, B.S. Manjunath, C. Kenney, M.S. Moore, H. Shin, “An efficient color representation for image retrieval”, *IEEE Trans. on Image Processing*, vol.10, pp.140-147, 2001.

[12] T. Chang, C.Kou, “Texture analysis and classification with tree-structured wavelet transform”, *IEEE Trans. on Image Processing*, vol.2, 1993.

[13] D. Lowe, “Distinctive image features from scale invariant keypoints”, *Intl Journal of Computer Vision*, vol.60, pp.91-110, 2004.

[14] J. Fan, Y. Gao, H. Luo, “Multi-level annotation of natural scenes using dominant image compounds and semantic concepts”, ACM Multimedia, 2004.

[15] K. Barnard and D. Forsyth, “Learning the semantics of words and pictures”, IEEE ICCV, pp.408-415, 2001.

[16] A. Vailaya, M. Figueiredo, A.K. Jain and H.J. Zhang, “A Bayesian framework for semantic classification of outdoor vacation images”, Proc. SPIE, vol.3656, 1998.

[17] E. Chang, K. Goh, G. Sychay, G. Wu, “CBSA: Content-based annotation for multimodal image retrieval using Bayes point machines”, *IEEE Trans. CSVT*, 2002.

[18] Y. Chen, J. Wang, R. Krovetz, “CLUE: cluster-based retrieval of images by unsupervised learning”, *IEEE Trans. on Image Processing*, vol.14, no.8, pp.1187-1201, 2005.

[19] J. Fan, Y. Gao, H. Luo, “Integrating concept ontology and multi-task learning to achieve more effective classifier training for multi-level image annotation”, *IEEE Trans. on Image Processing*, vol. 17, no.3, pp.407-426, 2008.

[20] Y. Rubner, C. Tomasi, L. Guibas, “A metric for distributions with applications to image databases”, IEEE ICCV, pp.59-66, 1998.

[21] D. Stan, I. Sethi, “eID: A system for exploration of image databases”, *Information Processing and Management*, vol.39, pp.335-361, 2003.

[22] J.A. Walter, D. Webling, K. Essig, H. Ritter, “Interactive hyperbolic image browsing-towards an integrated multimedia navigator”, ACM SIGKDD, 2006.

[23] G.P. Nyuyen, M. Worring, “Interactive access to large image visualizations using similarity-based visualization”, *Journal of Visual Languages and Computing*, 2006.

[24] D. Heesch, A. Yavlinsky, S. Ruger, “NN^k networks and automated annotation for browsing large image collections from the world wide web”, demo at ACM Multimedia, 2006.

[25] B. Moghaddam, Q. Tian, N. Lesh, C. Shen, T.S. Huang, “Visualization and user-modeling for browsing personal photo libraries”, *Intl. Journal of Computer Vision*, vol.56, pp.109-130, 2004.

[26] K. Rodden, W. Basalaj, D. Sinclair, K. Wood, “Evaluating a visualization of image similarity as a tool for image browsing”, IEEE InfoVis, 1999.

[27] R. Torres, C. Silva, C. Medeiros, H. Rocha, “Visual structures for image browsing”, ACM CIKM, 2003.

[28] P. Janecek, P. Pu, “Searching with semantics: An interactive visualization technique for exploring an annotated image collection”, OTM Workshops, pp.185-196, 2003.

[29] J Shi, J Malik, “Normalized cuts and image segmentation”, *IEEE Trans. on PAMI*, 2000.

[30] J. Lamping, R. Rao, “The hyperbolic browser: A focus+content technique for visualizing large hierarchies”, *Journal of Visual Languages and Computing*, vol.7, pp.33-55, 1996.

[31] M. Girolami, “Mercer kernel-based clustering in feature space”, *IEEE Trans. on Neural Networks*, vol.13, no.3, pp.780-784, 2002.

[32] B. Scholkopf, A. Smola, K.R. Muller, “Nonlinear component analysis as a kernel eigenvalue problem”, *Neural Computation*, vol.10, no.5, pp.1299-1319, 1998.

[33] A. Ben-Hur, D. Horn, H.T. Siegelmann, V. Vapnik, "Support vector clustering", *Journal of Machine Learning Research*, vol.2, pp.125-137, 2001.

[34] S. Brin, L. Page, "The anatomy of a large-scale hypertextual web search engine", WWW, 1998.

[35] K. Goldberg, T. Roeder, D. Gupta, C. Perkins, "Eigentaste: A constant time collaborative filtering algorithm", *Information Retrieval*, vol.4, no.2, pp.133-151, 2001.

[36] B. Miller, I. Albert, S. Lam, J. Konstan, J. Riedl, "MovieLens unplugged: Experiences with an occasionally connected recommender system", ACM Intl. Conf. on Intelligent User Interfaces, 2003.

[37] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, "Item-based collaborative filtering recommendation algorithms:, WWW, pp.285-295, 2001.

[38] M. Balabanovic, Y. Shoham, "Fab: Content-based, collaborative recommendation", *Comm. of ACM*, vol.43, no.3, pp.66-72, 1997.

[39] R. Mooney, L. Roy, "Content-based book recommending using learning for text categorization", ACM Conf. on Digital Libraries, pp.195-204, 2000.

[40] R. Fergus, P. Perona, A. Zisserman, "A visual category filter for Google images", ECCV, 2004.

[41] D. Cai, X. He, Z. Li, W.-Y. Ma, J.-R. Wen, "Hierarchical clustering of WWW image search results using visual, textual, and link information", ACM Multimedia, 2004.

[42] Y. Rui, T.S. Huang, M. Ortega, S. Mehrotra, "Relevance feedback: A power tool in interactive content-based image retrieval", *IEEE Trans. on CSVT*, vol.8, no.5, pp.644-655, 1998.

[43] S. Tong and E. Chang, "Support vector machine active learning for image retrieval", ACM Multimedia, 2001.

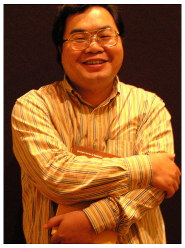
[44] X. Zhou, T. Huang, "Small sample learning during multimedia retrieval", Proc. IEEE CVPR, pp.11-17, 2001.

[45] X. He, W.-Y. Ma, O. King, M. Li and H.J. Zhang, "Learning and inferring a semantic space from user's relevance feedback", ACM Multimedia, 2002.

[46] J. Chai, C. Zhang, R. Jin, "An expirical investigation of user term feedback in text-based targeted image search", *ACM Trans. on Information Systems*, vol.25, no.1, 2007.

[47] M.F. Cox, M.A. Cox, *Multidimensional Scaling*, Chapman and Hall, 2001.

[48] C. Leacock, M. Chodorow, G.A. Miller, "Combining local context and wordnet similarity for sense identification", *WordNet: An electronic lexical database*, pp.265-283, 1998.



Jianping Fan received his MS degree from in theory physics from Northwestern University, Xian, China in 1994 and his PhD degree in optical storage and computer science from Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai, China, in 1997.

He was a Postdoc Researcher at Fudan University, Shanghai, China, during 1998. From 1998 to 1999, he was a Researcher with Japan Society of Promotion of Science (JSPS), Department of Information System Engineering, Osaka University, Osaka, Japan. From September 1999 to 2001, he was a Postdoc Researcher in the Department of Computer Science, Purdue University, West Lafayette, IN. At 2001, he joined the Department of Computer Science, University of North Carolina at Charlotte as an Assistant Pofessor and then become Associate Professor. His research interests include image/video analysis, semantic image/video classification, personalized image/video recommendation, surveillance videos, and statistical machine learning.

Daniel A. Keim received the PhD degree in computer science from the University of Munich in 1994. He is a full professor in the Computer and Information Science Department, University of Konstanz. He has been an assistant professor in the Computer Science Department, University of Munich, and an associate professor in Computer Science Department, Martin Luther University Halle. He also worked at AT&T Shannon Research Labs, Florham Park, New jersey. In the field of information visualization, he developed several novel techniques, which use visualization technology for the purpose of exploring large databases. Dr. Keim has published extensively on information visualization and data mining, he has given tutorials on related issues at several large conferences, including Visualization, SIGMOD, VLDB, and KDD, he was program cochair of the IEEE Information Visualization Symposia in 1999 and 2000, the ACM SIGKDD Conference in 2002, and the Visual Analytics Symposium in 2006. Currently, he is on the editorial board of the IEEE Transactions on Knowledge and Data Engineering, the Knowledge and Information System Journal, and the Information Visualization Journal. He is a member of IEEE Computer Society.

Yuli Gao received his BS degree in computer science from Fudan University, Shanghai, China, in 2002. At the same year, he joined University of North Carolina at Charlotte to pursue his PhD degree on Information Technology. He got his PhD degree at 2007 and then joined HP Labs. His research interests include computer vision, image classification and retrieval, and statistical machine learning. He got an award from IBM as emerging leader in multimedia at 2006.

Hangzai Luo received the BS degree in computer science from Fudan University, Shanghai, China, in 1998. At the same year, he joined Fudan University as Lecturer. At 2002, he joined University of North Carolina at Charlotte to pursue his PhD degree on Information Technology. He got his PhD degree at 2006 and joined East China Normal University as Associate Professor at 2007. His research interests include computer vision, video retrieval, and statistical machine learning. He got second place award from Department of Homeland Security at 2007 for his excellent work on video analysis and visualization for homeland security applications.

Zongmin Li received the M.S. degree from Nanjing University of Aeronautics and Astronautics in CAGD in 1992 and the Ph.D. degree in Institute of Computing Technology, Chinese Academy of Sciences in 2005. From 2007 to 2008, he was a visiting professor at University of North Carolina at Charlotte. He is currently a Professor of Computer Science at China University of Petroleum. His research interests include image processing, pattern recognition, and computer graphics.