A Novel Approach to Enable Semantic and Visual Image Summarization for Exploratory Image Search

Jianping Fan, Yuli Gao, Hangzai Luo
Dept of Computer Science
UNC-Charlotte
Charlotte, NC 28223, USA
{jfan, ygao, hluo}@uncc.edu

Daniel A. Keim
Dept of Computer Science
University of Konstanz
Konstanz, Germany
keim@inf.uni-konstanz.de

Zongmin Li
Dept of Computer Science
China University of Petroleum
Dongyong, CHINA
lizm@hdpu.edu.cn

ABSTRACT

In this paper, we have developed a novel scheme to incorporate topic network and representativeness-based sampling for achieving semantic and visual summarization and visualization of large-scale collections of Flickr images. First, topic network is automatically generated for summarizing and visualizing large-scale collections of Flickr images at a semantic level, so that users can select more suitable keywords for more precise query formulation. Second, the diverse visual similarities between the semantically-similar images are characterized more precisely by using a mixture-of-kernels and a representativeness-based image sampling algorithm is developed to achieve similarity-based summarization and visualization of large amounts of images under the same topic, so that users can find some particular images of interest more effectively. Our experiments on large-scale image collections with diverse semantics have provided very positive results.

Categories and Subject Descriptors

I.2.6 [**Artificial Intelligence**]: Learning—*concept learning*; I.3.6 [**Computer Graphics**]: Methodolgy and Techniques—*Interaction Techniques*

General Terms

Algorithms, Experimentation

Keywords

Image summarization, topic network, representative images, exploratory search.

1. INTRODUCTION

Flickr has developed an interesting approach to index and search large-scale shared image collections by using manual text annotations to bypass the semantic gap, where the manual text annotations are provided by numerous online users with potential similar interests. To reduce the barries to entry, Flickr has allowed users to use the free text terms from folksonomy to annotate the shared images manually. Because numerous online users can contribute for manual image annotation, Flickr has provided a reasonable solution for bridging the semantic gap. Unfortunately, the users may make mistakes and they may not be well-trained, thus the manual annotations (which are provided by numerous online users) may be in low quality. Many text terms (which may be used by different online users) may have the same word sense, thus Flickr image search engine may seriously suffer from the synonymy problem. On the other hand, one text term may have many word senses under different contexts, Flickr image search engine may also suffer from the homonym problem. In addition, the visual contents of the images are completely ignored, and it is well-accepted that the visual properties of the images are very important for people to search for images. However, the keywords (text terms) for manual image annotation may not be expressive enough for describing the rich details of the visual contents of the images [14]. Thus there is an urgent need to develop new algorithms to support more effective exploration and navigation of large-scale Flickr image collections according to their inherent visual similarity contexts, so that users can look for some particular images of interest interactively.

Based on these observations, we have developed a novel scheme for achieving semantic and visual summarization and visualization of large-scale collections of Flickr images: (a) topic network is automatically generated and used to summarize and visualize large-scale Flickr image collections at a semantic level; (b) a novel algorithm for representativeness-based image sampling is developed to enable visual (similarity-based) summarization and visualization of large amounts of semantically-similar images under the same topic, where the diverse visual similarities between the images are characterized more precisely by using a mixture-of-kernels.

The paper is organized as follows. Section 2 briefly reviews some related work; Section 3 introduces our new scheme to incorporate topic network for summarizing and visualizing large-scale image collections at a semantic level; Section 4 describes our representativeness-based image sampling algorithm to enable visual summarization and visualization of large amounts of semantically-similar images under the same topic, and a mixture-of-kernels algorithm is developed to characterize the diverse visual similarities between the images more precisely; Section 5 introduces our scheme for interactive topic exploration to allow users to select the most relevant keywords for formulating their queries more precisely; Section 6 describes our new scheme for interactive

MIR'08, October 30–31, 2008, Vancouver, British Columbia, Canada.

image exploration to allow users to look for some particular images of interest interactively; Section 7 summarizes our work on algorithm evaluation; We conclude in Section 8.

2. RELATED WORK

When large-sale image collections come into view, it is becoming very important to enable semantic or visual summarization, so that users can get a good global overview of large-sale image collections at the first glance. The image summarization process is formulated as selecting a set of images that efficiently represents the visual contents of large-scale image collections at the semantic or visual levels. The ideal summary should be able to present the most interesting and important aspects of large-scale image collections with minimal redundancy.

Some pioneer works have been done to enable image summarization [1, 2, 4, 5, 13]. A common summarization method is to select a smaller number of representative images. Some researchers have used the most informative regions and their similarity for image summarization, implicit feedback has also exploited to enable summarization of online images, and bi-directional space-time similarity is integrated to enable scene-based image summarization. One major problem for all these existing image summarization techniques is that they may not be scalable to the sizes of image collections and the diversity of the image semantics.

Concept ontology has recently been used to summarize and index large-scale image collections at the concept level by using some hierarchical inter-concept relationships such as "IS-A" and "part-of" [4, 5, 7, 10]. The image contents at Flickr are highly dynamic and the contextual relationships between the image topics could be more general rather than hierarchy, thus the inter-topic contexts for Flickr image collections cannot be characterized precisely by using only the hierarchical relationships such as "IS-A" or "part-of". Thus there is an urgent need to construct more precise topic network for organizing and summarizing large-scale collections of manually-annotated Flickr images at a semantic level.

Due to the *semantic gap* between the low-level visual features and the high-level human interpretation of image semantics, visualization is becoming very important for users to assess the diverse visual similarities between the images interactively [6, 8, 9, 11]. Visualization is also very important to enable interactive exploration of the image summaries. Some pioneer work have been done by incorporating visualization for navigating and exploring the images interactively [8]. Multi-dimensional scaling (MDS) and feature-based visual similarity have been seamlessly incorporated to create a 2D layout of the images, so that users can navigate the images easily according to their feature-based visual similarity contexts. Recently, isometric mapping has been incorporated to exploit the nonlinear similarity structures for image visualization [11]. Without integrating with a good image summarization scheme, all these existing image visualization techniques can work on only few thousands or even few hundreds of images.

3. SEMANTIC IMAGE SUMMARIZATION

When large-sale Flickr image collections with diverse semantics come into view, it is very important to enable image summarization at the semantic level, so that users can get a good global overview (semantic summary) of large-scale image collections at the first glance. In this paper, we have developed a novel scheme to incorporate a *topic network* to summarize and visualize large-scale collections of Flickr images at a semantic level. The topic network consists of two components: (a) image topics; and (b) their inter-topic contextual relationships (which are very important for supporting interactive exploration and navigation of large-scale image collections at a semantic level).

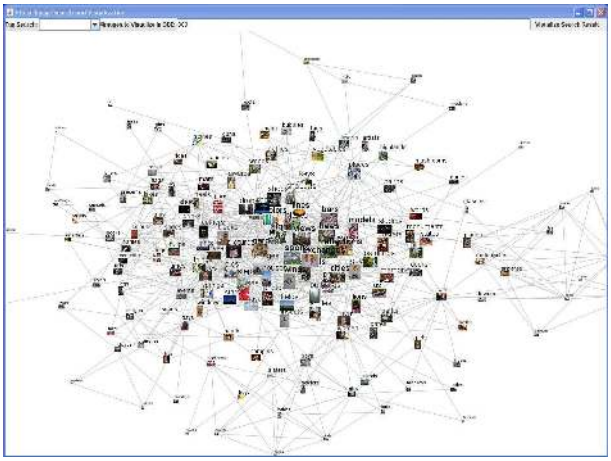

Figure 1: One portion of our topic network for indexing and summarizing large-scale collections of Flickr images at the topic level.

After the images and the associated users' manual annotations are downloaded from Flickr.com, the text terms which are relevant to the image topics (text terms for image topic interpretation) are separated automatically by using standard text analysis techniques, and the basic vocabulary of image topics (i.e., keywords for image topic interpretation) are determined automatically.

Latent Semantic Analysis (LSA) is then used to group the similar image topics (i.e., similar text terms with the same word sense for interpreting similar image topics) and identify the most significant image topics. The results of LSA are the fuzzy clusters of the text terms with the same word sense, where each cluster describes one significant image topic. Because LSA is used to integrate the text terms with the same word sense, the synonymy problem (for folksnomy-based image annotation) can be addressed effectively.

The contextual relationships among the image topics are obtained automatically, where both the semantic similarity and the co-occurrence probability for the relevant text terms are used to formulate a new measurement for determining the inter-topic associations effectively. The inter-topic association $\phi(C_i, C_j)$ is determined by:

$$\phi(C_i, C_j) = -\nu \cdot \log \frac{l(C_i, C_j)}{2D} + \omega \cdot \gamma(C_i, C_j), \quad \nu + \omega = 1 \quad (1)$$

where the first part denotes the semantic similarity between the image topics C_j and C_i, the second part indicates their co-occurrence probability, ν and ω are the weighting parameters, $l(C_i, C_j)$ is the length of the shortest path between the image topics C_i and C_j by searching the relevant keywords for interpreting the corresponding image topics from WordNet, D is the maximum depth of WordNet, $\gamma(C_i, C_j)$ is the co-occurrence probability between the relevant image topics

(i.e., keywords which are simultaneously used for annotating the same image). The co-occurrence probability $\gamma(C_i, C_j)$, between the keywords for interpreting two image topics C_j and C_i, is directly obtained from the available image annotation document. Because the co-occurrence is more general for characterizing more complex inter-topic contextual relationships in Flickr, the co-occurrence probability plays more important role than the semantic similarity on characterizing the more general associations between the image topics, and thus we set $\nu = 0.4$ and $\omega = 0.6$ heuristically. Because the inter-topic contextual relationships on the topic network can be used to constraint the word senses for image topic interpretation, the homonym problem (for folksnomy-based image annotation) can be addressed effectively.

Each image topic is automatically linked with multiple relevant image topics with the higher values of the associations $\phi(\cdot, \cdot)$. One portion of our topic network for summarizing large-scale Flickr image collections is given in Fig. 1, where the image topics are connected and organized according to the strength of their associations, $\phi(\cdot, \cdot)$. One can observe that such topic network can provide a good global overview (semantic summary) of large-scale collections of Flickr images at a semantic level.

4. VISUAL IMAGE SUMMARIZATION

When the topic network is available, large-scale collections of Flickr images are automatically categorized into the image topics on the topic network and one single image could be categorized into multiple image topics because multiple keywords may simultaneously be used for annotating the same image. Thus each image topic may consist of large amounts of semantically-similar images with diverse visual properties. For example, some topics in Flickr may contain more than 100,000 images. Exploring such amount of semantically-similar images under the same topic may seriously suffer from the problem of information overload. In addition, some of these semantically-similar images may share similar visual properties, and thus they may not be able to provide any additional information to the users. Therefore, it is very attractive to develop new algorithms to select the most representative images for summarizing and visualizing large amounts of semantically-similar images under the same topic, so that users can find some particular images of interest more effectively.

To achieve visual image summarization according to their diverse visual similarities, the most critical problem is to define more suitable similarity functions to characterize the diverse visual similarities between the images more precisely. Recently, kernel methods have recently attracted sustaining attentions to characterize the nonlinear data similarities. Based on these observations, we have incorporated three basic image kernels to characterize the diverse visual similarities between the images, and a linear combination of these basic image kernels can further form a family of various kernels to achieve more accurate characterization of the diverse visual similarities between the images. In this paper, we focus on three basic kernel functions: (a) histogram kernel for global color histogram; (b) texture kernel for texture histograms of wavelet filter banks; (c) interest point matching kernel for local invariant feature point set.

We calculate a coarse color histogram as a rough approximation of the color distribution in an image. Specifically, we quantize the color channel uniformly into 16 bins. Given

Figure 2: Our representativeness-based sampling technique can automatically select 200 most representative images to achieve precise visual summarization of 48386 **semantically-similar images under the topic "orchids".**

this image representation, a kernel function is designed to construct the kernel matrix for characterizing the visual similarity between the images according to their color principles. Given two color histograms $I(u)$ and $J(v)$ with equal length (16 bins) for two images I and J, their kernel-based color similarity $K_c(I, J)$ is defined as:

$$K_c(I, J) = e^{-\chi^2(I(u), J(v))/\delta} \tag{2}$$

where δ is set to be the mean value of the χ^2 distances between all the image pairs in our experiments.

The kernel-based texture similarity $K_t(I, J)$ between two images I and J is then defined as:

$$K_t(I, J) = \prod_{i=1}^{m} e^{-\chi_i^2(h_i(I), h_i(J))/\omega_i} \tag{2}$$

where ω_i is set to be the mean value of the distance between the histograms $h_i(I)$ and $h_i(J)$ of the i^{th} wavelet filtering channel.

For two images I and J, their interest point sets $I(Q)$ and $J(P)$ may have different numbers of interest points, their kernel-based point-matching similarity $K_p(I, J)$ is defined as:

$$K_p(I, J) = e^{-D(I(Q), J(P))/\rho} \tag{4}$$

where $D(I(Q), J(P))$ is the Earth Mover's distance (EMD) between these two interest point sets $I(Q)$ and $J(P)$, ρ is set as the mean value of $D(I(Q), J(P))$ of the image pairs in our experiments.

For a given image pair I and J, their kernel-based similarity $\kappa(I, J)$ is finally characterized by using a mixture of three basic image kernels (i.e., mixture-of-kernels):

$$\kappa(I, J) = \sum_{j=1}^{3} \alpha_j K_j(I, J), \quad \sum_{j=1}^{3} \alpha_j = 1 \tag{5}$$

where α_j is the importance factor for the jth basic image kernel and it can be learned automatically from the available images. Such mixture-of-kernels can achieve more accurate approximation of the diverse visual similarities between the images, produce nonlinear separation hypersurfaces between the images, result in more accurate image clustering,

Figure 3: Our representativeness-based sampling technique can automatically select 200 most representative images to achieve precise visual summarization of 53829 **semantically-similar images under the topic "rose".**

Figure 4: Our representativeness-based sampling technique can automatically select 200 most representative images to achieve precise visual summarization of 31482 **semantically-similar images under the topic "bugs".**

and exploit the nonlinear visual similarity contexts for image summarization.

The optimal partition of the semantically-similar images under the same topic is then obtained by minimizing the trace of the within-cluster scatter matrix, S_w^ϕ. The scatter matrix is given by:

$$S_w^\phi = \frac{1}{N}\sum_{l=1}^{k}\sum_{i=1}^{N}\beta_{li}\left(\phi(x_i)-\mu_l^\phi\right)\left(\phi(x_i)-\mu_l^\phi\right)^T \quad (6)$$

where $\phi(x_i)$ is the mapping function and $\kappa(x_i,x_j)=\phi(x_i)^T\phi(x_j)$ $=\sum_{j=1}^{3}\alpha_j K_j(x_i,x_j)$, N is the number of images and k is the number of clusters, μ_l^ϕ is the center of the lth cluster and it is given as:

$$\mu_l^\phi = \frac{1}{N_l}\sum_{i=1}^{N}\beta_{li}\phi(x_i) \quad (7)$$

The trace of the scatter matrix S_w^ϕ can be computed by:

$$T_r\left(S_w^\phi\right) = \frac{1}{N}\sum_{l=1}^{k}\sum_{i=1}^{N}\beta_{li}\left(\phi(x_i)-\mu_l^\phi\right)^T\left(\phi(x_i)-\mu_l^\phi\right) \quad (8)$$

Searching the optimal values of the elements β that minimizes the expression of the trace in Eq. (8) can be achieved effectively by an iterative procedure.

Because our mixture-of-kernels image clustering algorithm can seamlessly integrate multiple kernels to characterize the diverse visual similarities between the images more accurately, it can provide a good insight of large amounts of semantically-similar images (under the same image topic) by determining their inherent visual similarity structures precisely, and such inherent visual similarity structures can further be used to achieve more effective visual image summarization by selecting the most representative images adaptively according to their representativeness scores.

Our representativeness-based image sampling technique has exploited three criteria for selecting the most representative images: (a) *Image Clusters*: Our kernel-based image clustering algorithm has provided a good global distribution structure (i.e., image clusters and their relationships) for large amounts of semantically-similar images under the same topic. Thus adaptive image sampling can be achieved by selecting the most representative images to summarize the visually-similar images in the same cluster. (b) *Coverage Percentage*: Different clusters may contain various numbers of images, and thus more images should be selected from the clusters with bigger coverage percentages. Obviously, the relative numbers of their most representative images can be optimized according to their coverage percentages. (c) *Outliers*: Even the outliers may have much smaller coverage percentages, some representative images should prior be selected from the outliers for supporting serendipitous discovery of unexpected images.

For the visually-similar images in the same cluster, the representativeness scores of the images depend on their closeness with the cluster centers. The representativeness score $\rho(x)$ for the given image with the visual features x can be defined as:

$$\rho(x) = max\left\{e^{-\beta_l\left(\phi(x)-\mu_l^\phi\right)^T\left(\phi(x)-\mu_l^\phi\right)}, \quad l\in C_j\right\} \quad (9)$$

where μ_l^ϕ is the center of the lth cluster of the image topic C_j. Thus the images, which are closer to the cluster centers, have larger values of $\rho(\cdot)$. 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 to generate the similarity-based summary of the images for the corresponding image topic.

Only the most representative images are selected to generate the visual summary of the images for each image topic, and large amounts of redundant images, which have similar visual properties with the most representative images, are eliminated automatically. By selecting the most representative images to summarize large amounts of semantically-similar images under the same topic, the inherent visual similarity contexts between the images can be preserved accurately and thus it can provide sufficient visual similarity contexts to enable interactive image exploration.

Our visual summarization (i.e., the most representative images) results for the image topics "orchids", "rose" and "bugs" are shown in Fig. 2, Fig. 3 and Fig. 4, where 200 most representative images for the image topics "orchids",

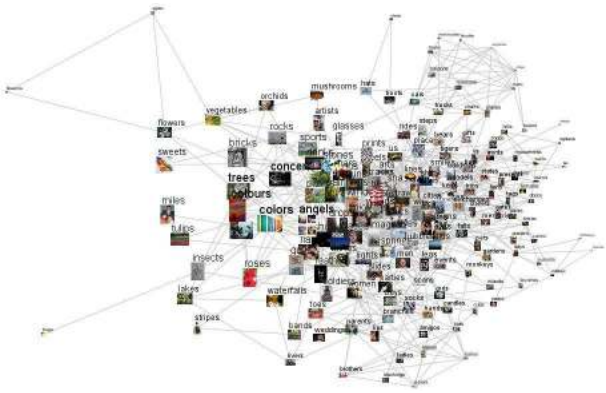

Figure 5: The visualization of the same topic network as shown in Fig. 1 via change of focus.

"rose" and "bugs" are selected for representing and preserving the original visual similarity contexts between the images. One can observe that these 200 most representative images can provide an effective interpretation and summarization of the original visual similarity contexts among large amounts of semantically-similar images under the same topic. The underlying the visual similarity contexts have also provided good directions for users to explore these most representative images interactively.

5. INTERACTIVE TOPIC EXPLORATION

To integrate the topic network for exploring large-scale collections of Flickr images at a semantic level, it is very attractive to support graphical representation and visualization of large-scale topic network (i.e., topic network with large amounts of image topics), so that users can obtain a good global overview (semantic summary) of large-scale image collections at the first glance and select more suitable keywords (image topics) for formulating their queries more precisely. Unfortunately, visualizing large-scale topic network in a 2D system interface with a limited screen size is not a trivial task. On the other hand, displaying all these available image topics at one compact view may not be meaningful to the users or may even confuse them. To address these issues effectively, we have developed multiple innovative techniques for topic network visualization: (a) highlighting the interestingness of the image topics for allowing users to obtain the most significant insights of large-scale image collections at the first glance; (b) integrating hyperbolic geometry to create "more" spaces and support change of focus to reduce the overlapping interactively.

We have integrated both the popularity of the image topics and the importance of the image topics to determine their interestingness scores. The popularity of a given 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 linkage structure with other image topics on the topic network. If one image topic is linked to more image topics on the topic network, it tends to be more interesting [3]. Thus the interestingness score $\varrho(C_i)$ for a given image topic C_i is defined as:

$$\varrho(C_i) = \varepsilon \cdot \frac{e^{2n(c_i)} - 1}{e^{2n(c_i)} + 1} + \eta \cdot \frac{e^{2l(c_i)} - 1}{e^{2l(c_i)} + 1}, \quad \varepsilon + \eta = 1 \tag{10}$$

where $n(c_i)$ 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. Such interestingness scores can be used to highlight the most interesting image topics (i.e., eliminate the less interesting image topics with smaller values of $\varrho(\cdot)$), thus the visual complexity for topic network visualization can be reduced significantly. For a given image topic, the number of images is more important than the number of linked topics on characterizing its interestingness, thus we set $\eta = 0.4$ and $\varepsilon = 0.6$ heuristically. In our definition, the interestingness score $\varrho(\cdot)$ is normalized to 1 and it increases adaptively with the number of images $n(c_i)$ and the number of linked topics $l(c_i)$.

We have investigated multiple solutions for topic network visualization: (a) A string-based approach is incorporated to visualize the topic network with a nested view, where each image topic node is displayed closely with the most relevant image topic nodes according to the values of their associations $\phi(\cdot, \cdot)$. The underlying contextual relationships between the image topics are represented as the linkage strings. (b) The geometric closeness of the image topic nodes is related to their semantic associations $\phi(\cdot, \cdot)$, so that such graphical representation of the topic network can reveal a great deal about how these image topics are connected and how the relevant keywords are intended to be used jointly for manual image annotation. (c) The *change of focus* is used to adjust the levels of visible details automatically according to the users' preferences of the image topics of interest.

When the hyperbolic visualization of the topic network is available, it can be used to enable interactive exploration and navigation of large-scale collections of Flickr images at a semantic level 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. As shown in Fig. 5, users can change their focuses of the 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 contextual relationships between the image topic nodes. In such interactive topic network navigation and exploration process, users can easily obtain the *topics of interest*, build up their mental search models and specify their queries more precisely by selecting the image topics on the topic network directly. Through change of focus, our hyperbolic topic network visualization framework has also provided an interactive approach to allow users to explore large-scale image collections at the semantic level.

6. INTERACTIVE IMAGE EXPLORATION

To support interactive exploration of the most representative images for a given image topic, it is very important to enable similarity-based image visualization by preserving the nonlinear similarity structures between the images in the high-dimensional feature space. Thus the most representative images are projected onto a hyperbolic plane by using the kernel PCA to preserve their nonlinear similarity structures [12]. The kernel PCA is obtained by solving the eigenvalue equation:

$$K\vec{v} = \lambda M\vec{v} \tag{11}$$

where $\lambda = [\lambda_1, \cdots, \lambda_M]$ denotes the eigenvalues and $\vec{v} = [\overrightarrow{v_1},$

Figure 6: Our interactive image exploration system: (a) the most representative images for the image topic "star" with blue box is selected; (b) more images which are relevant to the user's query intentions of "space star".

$\cdots$, $\overrightarrow{v_M}$] denotes the corresponding complete set of eigenvectors, M is the number of the most representative images, K is a kernel matrix and its component can be defined as $K_{ij} = \kappa(x_i, x_j)$.

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:

$$P(x, \overrightarrow{v}^k) = \sum_{j=1}^{L} \overrightarrow{\omega}_j^k \phi(x_j)^T \phi(x) = \sum_{j=1}^{L} \overrightarrow{\omega}_j^k \kappa(x, x_j) \quad (12)$$

If two images have similar visual properties in the high-dimensional feature space, they will be close on the hyperbolic plane for image display.

After such similarity-preserving image projection on the hyperbolic plane is obtained, we have used Poincaré disk model to map the most representative images on the hyperbolic plane onto a 2D display coordinate. By incorporating hyperbolic geometry for image visualization, our framework can support change of focus more effectively, which is very attractive for interactive image exploration and navigation. Through change of focus, users can easily control the presentation and visualization of large amounts of images according to the inherent visual similarity contexts.

It is important to understand that the system alone cannot meet the users' sophisticated image needs. Thus user-system interaction plays an important role for users to express their image needs, assess the relevance between the returned images and their real query intentions, and direct the system to find more relevant images adaptively. Based on these understandings, our system can allow users to zoom into the images of interests interactively and select one of these most representative images to express their query intentions or personal preferences. Once our system captures such personal preferences automatically (with minimal extra efforts from users), it can further allow users to look for some particular images of interest effectively as shown in Fig. 6, Fig. 7 and Fig. 8. Thus the redundant images, which have similar visual properties with the accessed image (which is clicked by the user to express his/her personal preference) and are initially eliminated for visual image summarization, are recovered automatically and displayed to the user according to his/her personal preference. Through such simple user-system interaction, the users can express their personal preferences easily to direct our system for obtaining more relevant images.

By focusing on a small number of images which are most relevant the users' personal preferences, our interactive image exploration technique can help users to obtain better understanding of the visual contents of the images, achieve better assessment of the inherent visual similarity contexts between the images, and make better decisions on what to do next according to the inherent visual similarity contexts between the images. Through such user-system interaction process, users can explore large-scale collections of images interactively and discover some unexpected images serendipitously.

7. ALGORITHM EVALUATION

We carry out our experimental studies by using large-scale collections of Flickr images with unconstrained semantic and visual contents. The topic network which consists of 4000 most popular topics is generated automatically from large-scale collections of Flickr images (more than 1.5 billions of Flickr images).

Our evaluation of the benefits from semantic and visual image summarization and visualization on assisting users in interactive exploration and navigation of large-scale image collections focuses on three issues: (a) Does our tool for semantic image summarization and visualization allow users to express their image needs more effectively and precisely? (b) Does our tool for visual image summarization and visualization allow users to find some images of interest more effectively according to their diverse visual similarity contexts?

When large-scale online Flickr image collections come into view, it is reasonable to assume that users are unfamiliar with the image contents (which is significantly different from personal image collections). Thus query formulation (specifying the image needs precisely) is one critical issue for users to access large-scale Flickr image collections. On the other hand, users may expect to formulate their image needs easily rather than typing keywords. By incorporating topic network to summarize and visualize large-scale image collections at a semantic level, our system can make all these

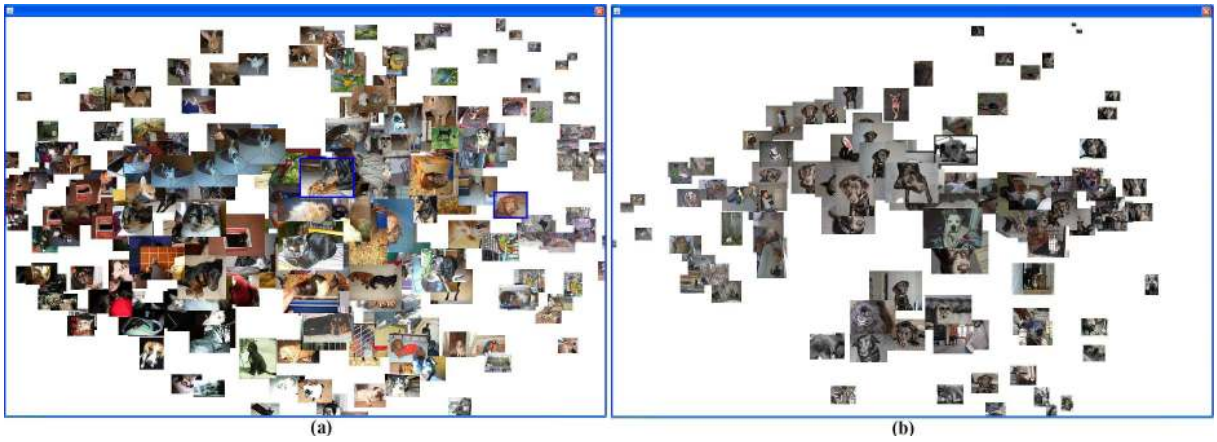

Figure 7: Our interactive image exploration system: (a) the most representative images for the image topic "pets", where the image in blue box is selected; (b) more images which are relevant to the user's query intentions of "dog".

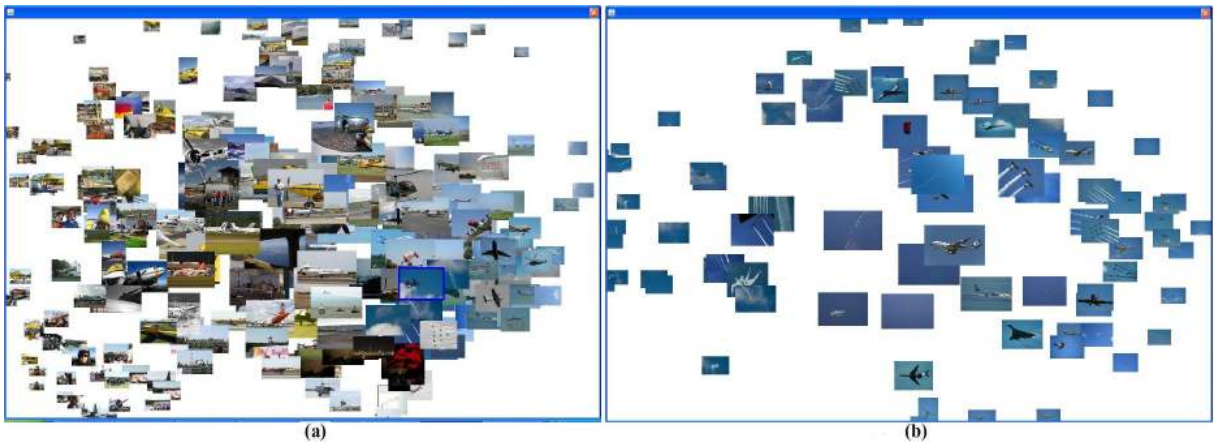

Figure 8: Our interactive image exploration system: (a) the most representative images for the image topic "planes", where the image in blue box is selected; (b) more images which are relevant to the user's query intentions of "plane in blue sky".

image topics to be visible to the users as shown in Fig. 2, so that they can have a good global overview (semantic summary) 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 and the contexts in the global structure of the topic network through change of focus as shown in Fig. 5, thus users can make better query decisions and formulate their queries more precisely by selecting the visible keywords (image topics) directly according to their personal preferences. Through such user-system interaction process, users can easily formulate their image needs by clicking the visible image topics on the topic network directly, and thus our system can offer more flexibility in specifying the image needs more precisely and intuitively.

Our visual image summarization algorithm can represent and preserve the diverse visual similarities between the images effectively as shown in Fig. 2, Fig. 3 and Fig. 4. Because human beings can understand the image contents at the first glance, visual image summarization and similarity-based image visualization can significantly improve their ability on locating some particular images of interest or a group of visually-similar images by using our interactive image exploration technique as shown in Fig. 6, Fig. 7 and Fig. 8, where the search space for looking for some images of interest can be narrowed down dramatically.

One critical issue for evaluating our interactive image exploration system is the response time for supporting change of focus. In our system, the change of focus is used for achieving interactive exploration and navigation of large-scale topic network and large amounts of most representative images. The *change of focus* is implemented by changing the Poincaré mapping of the image topic nodes or the most representative images from the hyperbolic plane to the display unit disk, and the positions of the image topic nodes or the most representative images in the hyperbolic 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 most representative 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 of these most representative images; (b) The visualization time T_2 for re-layouting and re-visualizing large-scale topic network or large amounts of most representative 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

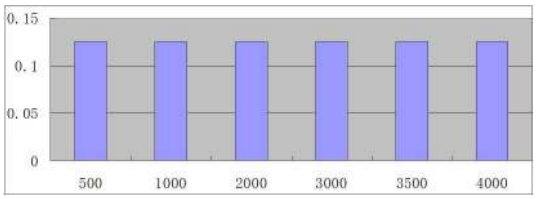

Figure 9: The empirical relationship between the computational time T_1 (seconds) and the number of image topic nodes.

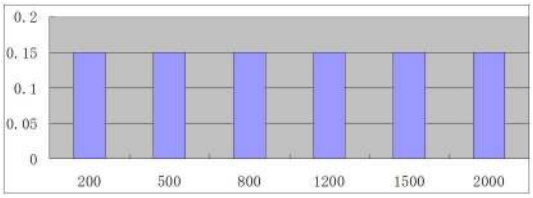

Figure 10: The empirical relationship between the computational time T_1 (seconds) and the number of most representative images.

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. 9, 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 most representative images. By computing the Poincaré mappings for different numbers of the most representative 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 most representative images as shown in Fig. 10, and thus re-calculating the Poincaré mapping for large amounts of most representative 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 most representative images. In our experiments, we have found that re-visualization of large-scale topic network and large amounts of most representative images is not sensitive to the number of image topics and the number of most representative images, and thus our system can support re-visualization of large-scale topic network and large amounts of most representative images in real time.

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

8. CONCLUSIONS

In this paper, we have developed a novel scheme to enable semantic and visual summarization of large-scale Flickr image collections, which is critical for supporting more effective image visualization and interactive image exploration. Our experiments on large-scale image archives (1.5 billions of Flickr images) with diverse semantics (4000 image topics) have provided very positive results.

9. REFERENCES

[1] S. Ahern, S. King, M. Naaman, and R. Nair. Summarization of online image collections via implicit feedback. *WWW*, pages 1325–1326, 2007.

[2] N. Bouguilaa and D. Ziou. Unsupervised learning of a finite discrete mixture: Applications to texture modeling and image databases summarization. *Journal of Visual Communication and Image Representation*, 18(4):295–309, 2007.

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

[4] J. Fan, Y. Gao, and 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*, 17(3):407–426, 2008.

[5] J. Fan, H. Luo, Y. Gao, and R. Jain. Incorporating concept ontology to boost hierarchical classifier training for automatic multi-level video annotation. *IEEE Trans. on Multimedia*, 9(5):939–957, 2007.

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

[7] L. Hollink, M. Worring, and G. Schreiber. Building a visual ontology for video retrieval. *ACM Multimedia*, 2005.

[8] G. Marchionini. Exploratory search: from finding to understanding. *Commun. of ACM*, 49(4):41–46, 2006.

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

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

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

[12] B. Scholkopf, A. Smola, and K.R. Muller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5):1299–1319, 1998.

[13] I. Simon, N. Snavely, and S.M. Seitz. Scene summarization for online image collections. *IEEE ICCV*, 2007.

[14] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Trans. on PAMI*, 2000.