Retrieving and Ranking Unannotated Images through Collaboratively Mining Online Search Results

Songhua Xu
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831, USA

Hao Jiang
Dept. of Computer Science
The University of Hong Kong
Hong Kong S.A.R., P.R. China

Francis C.M. Lau
Dept. of Computer Science
The University of Hong Kong
Hong Kong S.A.R., P.R. China

ABSTRACT

We present a new image search and ranking algorithm for retrieving unannotated images by collaboratively mining online search results, which consist of online image and text search results. The online image search results are leveraged as reference examples to perform content-based image search over unannotated images. The online text search results are utilized to estimate individual reference images' relevance to the search query as not all the online image search results are closely related to the query. Overall, the key contribution of our method lies in its capability to deal with unreliable online image search results through jointly mining visual and textual aspects of online search results. Through such collaborative mining, our algorithm infers the relevance of an online search result image to a text query. Once we estimate a query relevance score for each online image search result, we can selectively use query specific online search result images as reference examples for retrieving and ranking unannotated images. To explore the performance of our algorithm, we tested our algorithm both on the standard public image datasets and several modest sized personal photo collections. We also compared the performance of our method with that of two peer methods. The results are very positive, indicating that our algorithm is superior to existing content-based image search algorithms for retrieving and ranking unannotated images. Overall, the main advantage of our algorithm comes from its collaborative mining over online search results both in the visual and textual domains.

• S. Xu and H. Jiang to be considered as equal first authors. Corresponding author: S. Xu; contact him at xus1 "*at*" ornl "*dot*" gov.

• This manuscript has been authored by UT-Battelle, LLC, under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
CIKM'11, October 24–28, 2011, Glasgow, Scotland, UK.
Copyright 2011 ACM 978-1-4503-0717-8/11/10 ...$10.00.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems; I.4.9 [**Image Processing and Computer Vision**]: Applications; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Query Formulation, Retrieval Models, Search Process*

General Terms

Algorithms, Experimentation, Performance

Keywords

Multimedia information retrieval, web information retrieval, online reference image collection, query processing, retrieving unannotated images, web search mining

1. INTRODUCTION

Today, managing photos is everybody's task, which by no means is an easy one. Using the mainstream image search tools, which mostly rely on traditional text based retrieval methods, is cumbersome and often frustrating because ordinary photos are rarely annotated. If there are indeed annotations, they tend to be quite brief, subjective, and biased.

To enable searching and ranking unannotated images, we introduce a new content-based image retrieval algorithm by leveraging online image search results as reference examples to inform and guide the unannotated image search and ranking process. A key feature of our algorithm is its capability of automatically inferring the relevance of an online image search result to its corresponding textual query through collaboratively mining both visual contents of and text clues associated with the online image search result. Such an image-to-query relevance estimation step enables our method to cope with noise and outliers in the online image search results. With the respective image-to-query relevance estimated for each reference image obtained from the online image search step, we can further derive an arbitrary unannotated image's relevance to a textual query according to the unannotated image's collective visual similarity with the entire reference image set. We calculate the collective visual similarity between an unannotated image and the reference image set by accumulating the individual pairwise visual image similarity between the unannotated image and each example image in the reference image set. During this accumulation process, we also take into account the query relevance score of each reference image example itself. Overall, an unannotated image that has higher visual similarities with more query relevant example images tends to obtain a higher collective visual similarity score. Finally, based on our derived collective visual similarity score for each candidate unannotated image with respect to the

reference image set, we can search and rank all the unannotated images accordingly.

2. RELATED WORK

2.1 Content-Based Image Retrieval

Fergus et al. [9] proposed a method that dynamically derives a visual object model through learning from Google image search results. Their method however does not leverage the annotations or description text of an online image. Popescu et al. [1] proposed a content-based image retrieval method using an ontology driven approach for improving image retrieval precision. Kidambi and Narayanan [17] introduced a human computer integrated approach for content-based image retrieval, taking advantage of user feedbacks. Hsiao et al. [12] represented an image as a visual bag-of-words followed by a relevance feedback scheme for ranking images. Jing and Baluja [15] adapted Google's PageRank algorithm for product image search. Their approach is based on image similarities estimated from low-level visual features proposed in [23]. Compared with our algorithm, their method does not look into the text clues associated with an image. Liu et al. [22] gave a high-level overview of content-based image retrieval research from the perspective of semantics understanding. For more comprehensive reviews on recent progress in content-based image retrieval and web image search, readers are referred to [27, 16, 5, 22, 28, 24]. These also exist abundant surveys on domain specific content-based image retrieval methods and systems, e.g. Muller et al.'s survey [25] on content-based image retrieval for medical applications.

2.2 Image Search and Re-Ranking through Mining Text Clues

Lin et al. contributed a probabilistic model based algorithm [19] for re-ranking image search results according to images' surrounding texts. Our algorithm does more—it also explores the visual content of an image. Cai et al. [3] proposed a hierarchic image clustering method through jointly analyzing visual, textual, and link information of an image. However, their algorithm is proposed for clustering images, which cannot be directly applied to retrieving or re-ranking images. Zhou and Huang's image retrieval algorithm [31] jointly uses the keyword and the visual content of an image for image retrieval. But their method uses reference images in isolation during the image retrieval process whereas our algorithm considers the collective image similarity between a given candidate unannotated image and an entire reference image set. Jia et al. [14] proposed a personal album annotation algorithm. Their algorithm and Feng et al.'s [7] algorithm, similar to ours, use clues from the visual content and HTML text of an image for image content understanding. Jia et al.'s method is a supervised learning approach, which needs a set of pre-labelled images as training samples, whereas our method does not need any human labels. Feng et al.'s method independently trains two separate classifiers, one using text features and the other using visual features of an image. In contrast, our method collaboratively analyzes the visual and text aspects of online image search results in one unified procedure. Through our integral analysis based on image and text data fusion, our method can adaptively prioritize relevant online reference images. As a result, our algorithm can deal with noisy reference image samples more superbly than Feng et al.'s algorithm, which will be proved through our experiments. Recently, Liu et al. [20] [21] developed a textual query based personal photo retrieval system, which also uses web image search results to facilitate querying unannotated personal photos. However, their method focuses on leveraging the relevance feedback strategy to deal with the noisy unreliable online image search results. In contrast, our method collaboratively mines the visual contents and text clues accompanying each online image search result to discriminate reliable online image search results from irrelevant ones. Compared with their relevance feedback based strategy, which requires end user participation, our method can autonomously apply online image search results in a selected manner to guide the unannotated image search and ranking process. The hybrid mining feature of our algorithm allows our method to automatically cope with the noisy and unreliable nature of web image search results. It is likely to further strengthen the image search capability of our algorithm by adopting the relevance feedback strategy proposed in Liu et al.'s work, which is a meaningful topic for future exploration. Fergus et al. [8] proposed an algorithm that learns object categories from Google's image search results. The key part of their algorithm is a model that considers spatial and scalar information of images for learning the potential categories and tags of Google search result images. They also incorporatd an incremental learning algorithm [18] into their method to support relevance feedback by end users. People have also explored the idea of mining web text search results for establishing tagged image databases. For example, Schroff et al. [26] proposed to build image databases by collecting images within webpages of web text search results. In their method, they utilize the textual contents of web search results for image tag identification and selection.

3. OUR METHOD

3.1 Finding Reference Images for an Image Query through Online Image Search

Given a textual query Q, we first use a third party commercial image search engine (Google Image Search in our current implementation) to perform an online image search. The purpose is to acquire a set of reference images to perform content-based image search over the candidate unannotated image set. Assume there are $\zeta(Q)$ reference images obtained for the input image query Q, which are organized as a reference image set $\mathbf{G}(Q) \triangleq \{G_1(Q), G_2(Q), \cdots, G_{\zeta(Q)}(Q)\}$, where each $G_i(Q)$ is an online search result image.

It is commonly understood that not all the online image search results are closely related to the query; some are completely irrelevant. Treating these noisy online search result images as equal reference samples will introduce great noise and instability to our content based image search process. To overcome this problem, in our method, we estimate each reference image's relevance to the image search query. The resultant image-to-query relevance scores can help our algorithm differentiate reference images from noisy or irrelevant ones, for improving the overall accuracy and reliability of our method. In Sec. 3.2–Sec. 3.4 below, we will explain how we estimate reference images' query relevance through collaboratively mining online search results, both in the text and image domains.

3.2 Estimating Initial Image-to-Query Relevance for Reference Images

To estimate each reference image's relevance to a text query Q, we first compare the similarity between the text context associated with the reference image and the text context of Q. Our assumption is: the more aligned the two types of text context are, the more likely that the reference image reflects the image search intent expressed by the text query Q. In the following, we will explain details of our text analysis based initial image-to-query relevance estimation procedure.

3.2.1 Characterizing word frequency distribution in online text search result

Given a text query Q, we first submit the query to Google (text) search and retrieve the top N text search results. In our experiments, the typical value assignment for N ranges between $[100, 500]$, which shall never exceed the total number of search result documents returned by Google for the query. Let the j-th text search result document be D_j ($j \in [1, N]$). All these search result documents formulate an online text search result set of Q, denoted as $\mathbf{D}(Q) \triangleq \{D_1, D_2, \cdots, D_N\}$. Next, we derive non-stop words' frequency distribution in $\mathbf{D}(Q)$ to characterize these words' respective importance. To balance between text search result documents with different lengths, we work with *relative word frequency* rather than absolute word frequency. The derivation process for a word's relative frequency is as follows: assume a non-stop word wd_i occurs $y_{i,j}$ times in the document D_j and the total occurrence times of all the non-stop words in D_j is m_j. We then calculate the relative word frequency of wd_i in D_j as $\frac{y_{i,j}}{m_j}$. One step further, we accumulate wd_i's relative word frequencies for all the documents in $\mathbf{D}(Q)$ and treat the sum as wd_i's *accumulated relative word frequency*, $t_i(Q)$, i.e. $t_i(Q) \triangleq \sum_{j=1}^{N} \frac{y_{i,j}}{m_j}$. In addition, we introduce an inverse word frequency term for each non-stop word in $\mathbf{D}(Q)$. Assume n_i is the total number of documents containing the word wd_i in $\mathbf{D}(Q)$. Then, the word wd_i's inverse word frequency term with respect to $\mathbf{D}(Q)$ is formulated as $\ln(1 + \frac{N}{n_i})$. Given wd_i's accumulated relative word frequency term $t_i(Q)$ and its inverse word frequency term $\ln(1 + \frac{N}{n_i})$, we follow the TF–IDF rule and use the product of the two terms, $t_i(Q)\ln(1 + \frac{N}{n_i})$, to indicate the importance of wd_i. Finally, we organize the TF–IDF terms of all the non-stop words occurring in $\mathbf{D}(Q)$ as a vector $\mathbf{\Omega}_{\text{text}}(Q)$, i.e.

$$\mathbf{\Omega}_{\text{text}}(Q) \triangleq \begin{array}{l} \big(t_1(Q)\ln(1 + \frac{N}{n_1}), t_2(Q)\ln(1 + \frac{N}{n_2}), \\ \cdots, t_{V(Q)}(Q)\ln(1 + \frac{N}{n_{V(Q)}})\big), \end{array} \tag{1}$$

where $V(Q)$ is the number of distinct non-stop words in $\mathbf{D}(Q)$.

3.2.2 Characterizing word frequency distribution associated with online image search results

It is noted that each reference image in our method is obtained through an online image search process. A property of images obtained in this way is that every image is linked back to its source webpage where we can obtain surrounding text for the image. Denote a reference image $G_j(Q)$'s source webpage as $P_j(Q)$. The source documents of all the webpages associated with the online image search results $\mathbf{G}(Q)$ form a document set $D'(Q) \triangleq \{P_1(Q), \cdots, P_{\zeta(Q)}(Q)\}$. For each non-stop word wd_i that occurs in the source document of the webpage $P_j(Q)$, we also calculate its relative word frequency among all the non-stop words occurring in the source document and denote the result as $f_{i,j}(Q)$. In a similar procedure as described in Sec. 3.2.1, we also derive a non-stop word wd_i's inverse word frequency with respect to the webpage source document set $D'(Q)$ as $\ln(1 + \frac{\zeta(Q)}{n'_i})$ where n'_i is the total number of webpage source documents containing the non-stop word wd_i. We organize the calculated relative word frequencies and inverse word frequencies for all the non-stop words occurring in $P_j(Q)$ as a $V'_j(Q)$ dimensional vector $\mathbf{\Omega}_{\text{img}}(Q)$, assuming there are $V'_j(Q)$ distinct non-stop words appearing in $P_j(Q)$. That is,

$$\mathbf{\Omega}_{\text{img},j}(Q) \triangleq \begin{array}{l} \big(f_{1,j}(Q)\ln(1 + \frac{\zeta(Q)}{n'_1}), f_{2,j}(Q)\ln(1 + \frac{\zeta(Q)}{n'_2}), \\ \cdots, f_{V'_j(Q),j}(Q)\ln(1 + \frac{\zeta(Q)}{n'_{V'_j(Q)}})\big). \end{array} \tag{2}$$

3.2.3 Measuring the semantic alignment between two types of text context

Previously in Sec. 3.2.1 and Sec. 3.2.2, we have derived two types of text context, one from the online text search results and the other from the webpage source document associated with an online image search result. We have also looked at how to characterize the two types of text context via their respective word frequency distribution vectors $\mathbf{\Omega}_{\text{text}}(Q)$ and $\mathbf{\Omega}_{\text{img},j}(Q)$. Next, we will derive an online reference image $G_i(Q)$'s relevance to its query Q, which is denoted as $r_i(Q)$, by measuring the semantic alignment between the two types of text context through computing the inner product of their respective word frequency distribution vectors. That is,

$$r_i(Q) \triangleq \mathbf{\Omega}_{\text{text}}(Q)\mathbf{\Theta}_i\mathbf{\Omega}^{\text{T}}_{\text{img},i}(Q), \tag{3}$$

where $\mathbf{\Theta}_i$ is a $V(Q) \times V'_i(Q)$ dimensional matrix whose element on the u-th row and v-th column is the semantic relatedness between the u-th non-stop word in the online text search results and the v-th non-stop word in $P_i(Q)$, measured by the algorithm proposed in [10]. We then treat the resulting $r_i(Q)$ as our initially estimated image-to-query relevance for the online reference image $G_i(Q)$. We organize all the estimated initial reference image to query relevance as a $\zeta(Q)$ dimensional vector, i.e. $\mathbf{R}^{\text{init}}(Q) \triangleq \big(r_1(Q), r_2(Q), \cdots, r_{\zeta(Q)}(Q)\big)$. The reason why we call this initial relevance estimation is because not all the reference images acquired through online image search are accompanied with quality annotation text. Such unreliability in the surrounding text of online reference images will surely affect the quality of our text analysis based image to query relevance estimation, a problem commonly suffered by nearly all annotation text based image search methods. Hence we call the above estimation on online reference images' relevance to the textual query *the initial estimation*. In the next, we will discuss how to refine this initial estimation.

3.3 Measuring Pairwise Reference Image Similarity

Given the initial image-to-query relevance estimation for all the reference images, we then refine the estimated image-to-query relevance through conducting an image content similarity based propagation procedure, which will be presented in Sec. 3.4. A key element involved in performing the procedure is to measure pairwise image similarity between reference images. In our method, we introduce a visual patch set based pairwise image similarity estimation method, which collaboratively mines the online text and image search results to more informedly derive the estimation.

In our method, we calculate pairwise reference image similarity through analyzing image content similarity on the granularity of visual patch sets rather than on the whole image level. Such finer granularity offers us a more elaborate analysis capability, which shall be clear shortly. To extract visual patch sets from the reference image set, we first apply the feature extraction algorithm proposed in [2] to obtain a collection of visual patches for every reference image in the reference image set $\mathbf{G}(Q)$. These visual patches are defined as Speeded-Up Robust Features (SURF), each of which includes a circular effective feature region and a feature descriptor. More concretely, we represent a visual patch v as $v = (C, Des)$, where C represents v's circular effective feature region, and Des is v's feature description vector. We further define the distance between two visual patches as the L_2 distance between the description vectors of the two visual patches. Two visual patches are considered identical if their L_2 distance is below 0.01.

Based on the concept of visual patches introduced above, we further introduce the concept of visual patch sets. A visual patch set is defined as a set of visual patches detected from an image whose

circular regions are overlapping. Formally, we define a visual patch set of size l as:

$$\phi \triangleq (v_1, v_2, \cdots, v_l), \tag{4}$$

where $v_i = (C_i, Des_i)$ and $\forall i, j = 1, 2, \cdots, l$, v_i connects with v_j. Here the relationship connects with is defined as follows:

$$v_i \text{ connects with } v_j \iff \begin{array}{c} C_i \cap C_j \neq \emptyset \\ \text{or} \\ \exists v_k \in \phi,\ s.t.\ C_i \cap C_k \neq \emptyset, C_j \cap C_k \neq \emptyset \end{array}. \tag{5}$$

In the above, $\emptyset$ represents an empty set. In the following, we use the notation $|\phi|$ to denote the size of the visual patch set ϕ.

Finding all the visual patch sets contained in a reference image set incurs an exponential amount of computational overhead. Fortunately, we only need to derive some most frequently occurring visual patch sets in our method as they provide the most essential evidence for estimating pairwise image similarity. Calculating frequently occurring visual patch set is a much more affordable computational task. In our method, we extracted visual patch sets from all the reference images in $\mathbf{G}(Q)$ whose occurrence numbers exceed a certain threshold using the method proposed in [30]. A typical threshold value used in our experiments is 5. Two visual patch sets are identical if they contain the same set of constituent visual patches. Let $\phi_i(Q)$ denote a frequent visual patch set detected from the reference image set $\mathbf{G}(Q)$. All the frequently occurring visual patch sets detected from the reference image set $\mathbf{G}(Q)$ are organized as a collection, which is denoted as $\mathbf{\Phi}(Q) \triangleq \{\phi_1(Q), \phi_2(Q), \cdots, \phi_{z(Q)}(Q)\}$, assuming there are $z(Q)$ frequent visual patch sets contained in the reference image set $\mathbf{G}(Q)$.

It is noted that in [30], the authors also proposed to use visual patch sets to estimate pairwise image similarity. Different from our method, their algorithm only detects visual patch sets containing no more than two patches; while our algorithm explores visual patch sets of all sizes. The reason why their algorithm tightly limits the size of visual patch sets considered is because their algorithm relies on extracting visual patch sets from a large number of sample images to perform content-based image search, an image collection that aims to work universally for all the queries. Such an algorithm design feature incurs a very expensive amount of computational overhead in visual patch set extraction. In contrast, our algorithm is designed to use a much smaller set of sample images acquired from online image searches in a query specific manner. The query-specific nature of our algorithm's reference image acquisition process allows us to work with a very limited number of sample images each time for performing content-based image retrieval. Such careful control of reference image set size successfully restricts the time spent on extracting visual patch sets from the reference image set. This algorithm feature enables us to explore visual patch sets of larger sizes in our method, leading to a richer set of image features for image content understanding.

Once we detect visual patch sets from all the reference images, we can then analyze those patch sets commonly occurring in a pair of reference images for estimating pairwise image similarity. This practice is inspired by the classic $TF \times IDF$ based document similarity measurement method [29]. Borrowing their idea for document similarity estimation in the text domain, in our method, we attempt to estimate pairwise image similarity through counting the common visual patch sets existing among two images and in a weighted way, prioritizing co-occurrence of salient visual patch sets. The assumption is, the larger number of salient visual patch sets commonly existing among two reference images, the more similar the two images are.

3.4 Refining Image-to-Query Relevance Estimation for Reference Images

Based on the pairwise reference image similarity estimation method introduced in the previous subsection, we will now introduce a multi-step propagation procedure to refine the initial image-to-query relevance estimation for reference images. We call the resultant estimation the *refined image-to-query relevance*. Our work is inspired by the recent work on product image ranking by Jing and Baluja, who modified the traditional Pagerank algorithm for ranking images through an iterative process [15].

First, we construct a pairwise image similarity matrix $\mathbf{S}_{\zeta(Q)\times\zeta(Q)}$, where the element on the i-th row and the j-th column of $\mathbf{S}_{\zeta(Q)\times\zeta(Q)}$ is the image similarity between the i-th and the j-th reference images, $G_i(Q)$ and $G_j(Q)$. The pairwise image similarity is estimated using the method introduced in Sec. 3.3. Noticing that the more steps we take in propagating a belief, the less reliable the propagated belief tends to be, we penalize the results obtained with more steps of propagation using an attenuation factor β. To formally define this multi-step propagation process, we employ the matrix exponent notation [11], as follows:

$$\begin{aligned} \mathbf{R}^{\text{refined}} &\triangleq e^{\beta\mathbf{S}}\mathbf{R}^{\text{init}} \\ &= (\mathbf{I}_{\zeta(Q)\times\zeta(Q)} + \beta\mathbf{S} + \tfrac{\beta^2\mathbf{S}^2}{2!} + \tfrac{\beta^3\mathbf{S}^3}{3!} + \cdots)\mathbf{R}^{\text{init}}, \end{aligned} \tag{6}$$

where $\mathbf{S}_{\zeta(Q)\times\zeta(Q)}$ is the pairwise image content similarity matrix; $\mathbf{R}^{\text{init}}$ and $\mathbf{R}^{\text{refined}}$ are respectively the initial and refined estimated image-to-query relevance for all the reference images; β is the propagation attenuation constant. A typically value assignment for β, as used in all our experiments reported in this paper, is 0.3.

The propagation procedure described in (6) can be concisely represented as $\mathbf{R}^{\text{refined}} \triangleq e^{\beta S}\mathbf{R}^{\text{init}}$, where the matrix exponent can be efficiently computed by observing the fact that $\mathbf{S}$ is symmetric. Therefore, we can decompose $\mathbf{S}$ as $\mathbf{S} \triangleq \mathbf{QDQ}^{-1}$, where $\mathbf{Q}$ is an orthogonal $\zeta(Q) \times \zeta(Q)$ matrix containing the eigenvectors of $\mathbf{S}$, and $\mathbf{D}$ is an $\zeta(Q) \times \zeta(Q)$ diagonal matrix containing the eigenvalues of $\mathbf{S}$. We can then rewrite $e^{\beta\mathbf{S}}$ as follows:

$$\begin{aligned} e^{\beta\mathbf{S}} &= \mathbf{I} + \beta\mathbf{QDQ}^{-1} + \tfrac{\beta^2(\mathbf{QDQ}^{-1})^2}{2!} + \tfrac{\beta^3(\mathbf{QDQ}^{-1})^3}{3!} + ... \\ &= \mathbf{I} + \beta\mathbf{QDQ}^{-1} + \tfrac{\beta^2\mathbf{QD}^2\mathbf{Q}^{-1}}{2!} + \tfrac{\beta^3\mathbf{QD}^3\mathbf{Q}^{-1}}{3!} + ... \\ &\approx \mathbf{I} + \beta\mathbf{QDQ}^{-1} + \tfrac{\beta^2\mathbf{QD}^2\mathbf{Q}^{-1}}{2!} + \tfrac{\beta^3\mathbf{QD}^3\mathbf{Q}^{-1}}{3!} \\ &\quad + \tfrac{\beta^4\mathbf{QD}^4\mathbf{Q}^{-1}}{4!} + \tfrac{\beta^5\mathbf{QD}^5\mathbf{Q}^{-1}}{5!}, \end{aligned} \tag{7}$$

According to this new representation, the matrix exponentiation can be efficiently computed by an eigenvalue decomposition process followed by a few matrix multiplication and addition operations. As mentioned earlier, the constant β here is the propagation attenuation parameter that penalizes results obtained through multiple steps of propagation.

Combining (6) and (7), we have:

$$\mathbf{R}^{\text{refined}} = e^{\beta\mathbf{S}}\mathbf{R}^{\text{init}} \approx (\mathbf{I} + \sum_{i=1}^{5} \frac{\beta^i\mathbf{QD}^i\mathbf{Q}^{-1}}{i!})\mathbf{R}^{\text{init}}. \tag{8}$$

Each time when a query Q is submitted, the initial image-to-query relevance score $\mathbf{R}^{\text{init}}$ is first estimated and then the corresponding $\mathbf{R}^{\text{refined}}$ can be derived through matrix multiplications and additions according to (8). The i-th component of $\mathbf{R}^{\text{refined}}$ is denoted

as $r_j^{\text{refined}}(Q)$, which represents the j-th reference image $G_j(Q)$'s relevance to the query Q.

3.5 Retrieving and Re-Ranking Unannotated Images

Given an arbitrary unannotated image I_x, to estimate I_x's relevance to the textual query Q, we refer to a set of highly similar reference images to I_x and use these reference images' query relevance for estimating I_x's relevance to the query. Our idea is to leverage the "intelligence of the crowd," i.e., if multiple reference images that are highly relevant to the query exhibit a high visual similarity with I_x, I_x is then more likely to be closely related to the query too.

More concretely, given a textual query Q, we perform an online image search to obtain a collection of reference images $\mathbf{G}(Q) = \{G_1(Q), G_2(Q), \cdots, G_{\zeta(Q)}(Q)\}$ (Sec. 3.1). In Sec. 3.2–Sec. 3.4, we have explained our method for estimating an arbitrary reference image $G_j(Q)$'s relativeness to the query Q, which is denoted as $r_j^{\text{refined}}(Q)$. We have also explained our method to extract frequent visual patch sets from $\mathbf{G}(Q)$, which are denoted as $\mathbf{\Phi}(Q) = \{\phi_k\}$, where each ϕ_k is an extracted frequent visual patch set (see Sec. 3.3).

For each visual patch set ϕ_k, we first estimate its relevance to the query Q. Since we have estimated each reference image G_i's relevance to the query Q, our method then estimates each visual patch set's relevance to the query by distributing the image to query relevance to the individual visual patch sets' relevance to the query. To realize the above idea, we first estimate the salience of each visual patch set contained in a reference image G_i. Again, we borrow the traditional TF–IDF scheme in the text processing domain for characterizing a visual patch set's salience. Let $m_j(\phi_k)$ be the number of times the visual patch set ϕ_k occurs in the j-th reference image $G_j(Q)$. Then we can mimic the TF term for the visual patch set ϕ_k as $(1 + \ln m_j(\phi_k))$. Let $m(\phi_k)$ be the number of images in the reference image collection $\mathbf{G}(Q)$ that contain the visual patch set ϕ_k. Then we can emulate the IDF term for ϕ_k as $\ln(1 + \frac{\zeta(Q)}{m(\phi_k)})$. Recall that $\zeta(Q)$ is the total number of images in the reference image collection. Overall, we define the TF–IDF term for the visual patch set ϕ_k as $(1 + \ln m_j(\phi_k)) \ln(1 + \frac{\zeta(Q)}{m(\phi_k)})$. We can then calculate the relative salience of the visual patch set ϕ_k among all the visual patch set contained in the reference image G_i as follows:

$$\rho(\phi_k) = \frac{(1 + \ln m_j(\phi_k)) \ln(1 + \frac{\zeta(Q)}{m(\phi_k)})}{\sum_{\phi_t \in \mathbf{\Phi}(Q)} (1 + \ln m_j(\phi_t)) \ln(1 + \frac{\zeta(Q)}{m(\phi_t)})}. \quad (9)$$

According to the relative salience of each visual patch set contained in G_i, we then proportionally attribute G_i's relevance score to the query Q to individual visual patch sets present in the image. That is, the fraction of query relevance score allocated for the visual patch set Φ_k is $r_j^{\text{refined}}(Q)\rho(\phi_k)$. Recall that $r_j^{\text{refined}}(Q)$ is the j-th reference image $G_j(Q)$'s refined estimated relevance score to the query Q. Finally, we derive ϕ_k's overall relevance score to the query, denoted as $\theta(\phi_k, Q)$, by summing up the fractional relevance scores ϕ_k acquired from all the reference images, i.e.:

$$\theta(\phi_k, Q) = \sum_{j=1}^{\zeta(Q)} r_j^{\text{refined}}(Q)\rho(\phi_k). \quad (10)$$

With every visual patch set's query relevance estimated through (10), we can now estimate an unannotated image I_x's image-to-query relevance, denoted as $\widehat{\theta}(I_x, Q)$, based on the query relevance of all the visual patch sets carried in I_x, as follows:

$$\widehat{\theta}(I_x, Q) = \sum_{\phi_k \in I_x} \theta(\phi_k, Q), \quad (11)$$

Overall, via (11), we can estimate an unannotated image's relativeness to a textual query through weighted sum of the query relevance of all the visual patch sets present in the image.

Finally, given an input query Q, each search candidate image I_x will be assigned a query relevance score $\widehat{\theta}(I_x, Q)$ via (11). Those candidate images whose estimated query relevance scores are above a certain threshold will be returned as image search results for the query. All the search result images are also ranked in a descendent order according to each image's estimated query relevance score.

4. EXPERIMENTAL RESULTS

We invited eight volunteers to participate in our experiments conducted on two standard image datasets: the Caltech101 image collection [6] and MIRFLICKR collection [13]. We compared the measured performance of our algorithm with two peer content-based image retrieval algorithms. All our experiments were performed on a desktop computer with 2.66GHz Intel Core 2 Duo CPU and 2GB main memory, running Windows XP.

4.1 Evaluation Metric

We evaluate the quality of an image search and ranking result by adopting the normalized discounted cumulative gain (NDCG) metric. The participating users would assign a numerical label to each of the top N_{top} search result images to indicate their opinions on these images' relevance to the corresponding queries. The numeric label, in $[0, 1]$, represents the image's relevance with the query (1: "fully relevant"; 0: "fully irrelevant"). For those image search experiments performed on the standard image collections, since all the images in the collection have been carefully classified and labeled, we automatically assign a binary value of either 0 or 1 to indicate whether a search result image is relevant to the query or not. In all the experiments reported in this paper, we confine N_{top} to be no more than 50.

4.2 Execution Efficiency

Much of the execution time of our algorithm is spent on downloading online reference images through Google image search. For a typical query performed in our experiments for the Caltech101 data set, we use the top 50 images retrieved from Google image search as the reference image collection, i.e. $\zeta(Q) = 50$. Table 1 shows the run-time breakdown statistics for each part of our algorithm. The exact duration of downloading time needed varies from query to query, which also heavily depends on the quality of a particular user's Internet connection condition. For real world deployment, our algorithm can run on the server side in order to avoid the reference image downloading time. Therefore, in our analysis, we exclude the online reference image downloading step.

4.3 Image Retrieval on Standardized Image Sets

The Caltech101 image collection contains 8677 images, pre-labelled with 101 textual tags, which has been widely used as benchmark for image retrieval and object recognition. We used each textual tag as the image query keyword and took all the 8677 images in the entire Caltech101 image collection as the candidate image search result set. Therefore, all together, we performed 101 sessions of image retrieval experiments over the collection. At the end of each image querying experiment, we evaluate the performance of our algorithm

Algorithm Step		Execution Time
Step 1:	Download reference images	ISP-dependent
Step 2:	Estimate initial image-to-query relevance for reference images	$20 \sim 30$ sec.
Step 3:	Detecting visual patch sets	$10 \sim 40$ sec.
Step 4:	Refine image-to-query relevance for reference images	< 1 sec.
Step 5:	Retrieve and rank unannotated images	< 1 sec.

Table 1: Run-time breakdown of different components of our algorithm for performing a typical image search query in our experiments. We report the time consumed by every step of our algorithm individually.

by calculating its NDCG score for the image querying session. If the search result image comes from the then chosen image class, it is considered a correct image retrieval result; otherwise, it is a false retrieval result. We also perform the same image query sessions on MIRFLICKR collection, which consists of 25000 pre-tagged images that were used in recent years' ImageCLEF corss language image retrieval tasks [4].

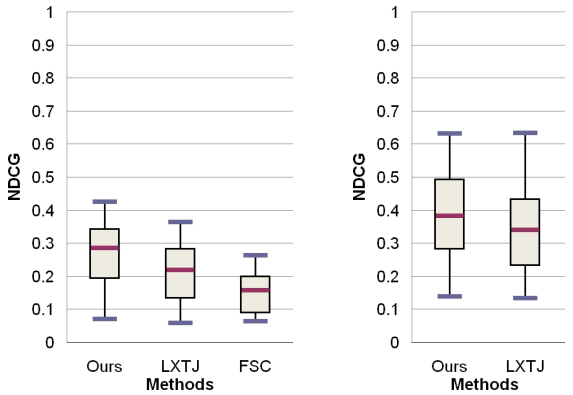
For comparison purpose, we also implemented two recent image retrieval algorithms respectively proposed in [21] and [7], abbreviated as "LXTJ" and "FSC" in our following discussions. We chose these two algorithms as peer methods for our comparative studies because both prior algorithms share the same idea of mining online information for image search as our method does; the two peer algorithms also achieve today's leading performance among all existing methods for image search. During our comparison, we applied the three algorithms respectively over the Caltech101 and MIRFLICKR collections for image retrieval experiments. Note that the "FSC" algorithm requires a labelled training set. We employed the 10-fold cross validation as our evaluation strategy.

4.4 Comparing Algorithm Performance without Relevance Feedback

In Figure 1(a) and Figure 2(a), we report the NDCG scores obtained by three algorithms respectively over the image querying sessions. All these results are generated by following a fully automatic way, i.e. no user feedback is provided for algorithms. From these results, we can clearly see that our algorithm consistently outperforms the other two algorithms.

4.5 Comparing Algorithm Performance with Relevance Feedback

In practice, a frequent strategy for improving the performance of an image retrieval system is to incorporate relevance feedback. For example, "LXTJ" proposed a relevance feedback based approach to enhance their image classifiers' performance interactively via cross-domain learning. To explore our algorithm's image search capability and potential more thoroughly, we also conduct experiments with relevance feedback. In our comparative study, we perform three rounds of relevance feedback when executing each query. In each feedback round, a user randomly selected from our participation group is asked to mark one relevant image and one irrelevant image among the top N_{top} (usually 50) images in the query result. For the algorithm of "LXTJ," their method is designed with a relevance feedback module and hence can directly use the provided relevance feedback data to improve its image search result. Our algorithm however is not equipped with a relevance module because we focus on providing a fully automatic mechanism for



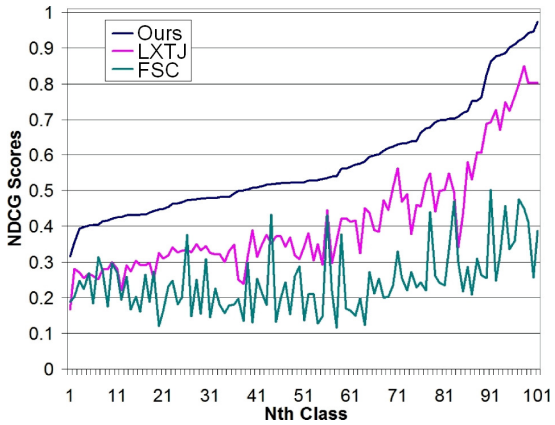
(a) without relevance feedback (b) with relevance feedback

Figure 1: NDCG scores of image search experiments conducted over the MIRFLICKR image set using "LXTJ," "FSC," and our algorithm respectively. We illustrate the key statistics of all the NDCG scores using boxplots.

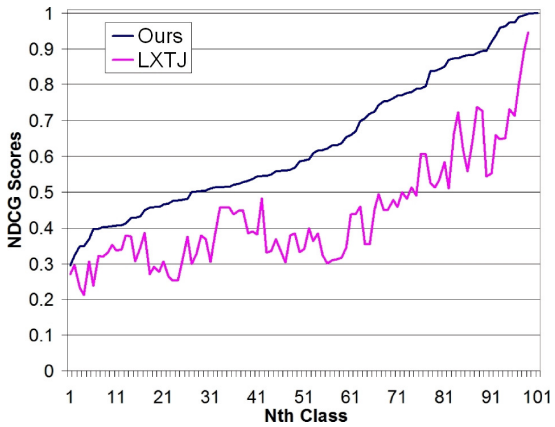
image search, aiming at the most convenient user experience. For a fair comparison, when comparing search results of our algorithm with those of LXTJ under the relevance feedback setting, we introduce a trivial relevance feedback step for our algorithm — if a user indicates a negative search result item, i.e. the item is not related to the query, we will simply remove the item from the returned search image list. If a user indicates a positive search result item, our algorithm will simply ignore the user input. Figure 1(b) and Figure 2(b) report the respective NDCG scores of image search experiments performed over two public image datasets by our algorithm and LXTJ with a three-round user feedback. The results show that our algorithm still outperforms the LXTJ algorithm with three rounds of user feedback.

4.6 Comparing Algorithm Performance with Query Keywords Extended

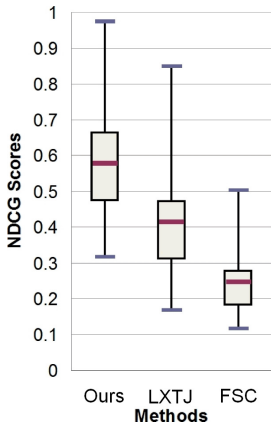
We aim to more realistically emulate users' image search behaviors and scenarios in reality. This time, when doing an image retrieval, instead of using the original query keyword, we use a synonym of the keyword as the query text. For example, for the query keyword of "lotus," we use "lily" or "lily lotus" or "lotus flower" or "water-lily" as the query text in our image retrieval experiments. For this new set of experiments, we randomly selected 10 image tags from the Caltech101 collection as our retrieval target objects to perform ten image retrieval experiments, one experiment for each tag. In each experiment, we invited five volunteers. We asked each volunteer to give a few words that could best describe the common content of a few randomly selected images from the same image tag in the Caltech101 collection. In this process, we hide the original image tag to these volunteers; and the answer given by each volunteer is kept hidden from other volunteers. Therefore, each volunteer needs to come up with a name for the image tag independently. For each image tag, if there is a word that appears in at least three volunteers' description texts, we will accept the word as a commonly agreed *user nominated search query* for the image query. If there are more than one word commonly named by more than three volunteers during our image querying experiments, we will randomly choose one word as the keyword to compose the image query. We designed this set of experiments to emulate image search scenarios in reality when the query text submitted by the end user does not use exactly the same word as the labels of an online



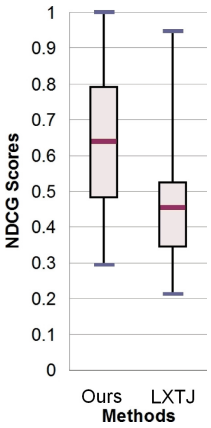
(a) Experiments without relevance feedback on Caltech101



(b) Experiments with relevance feedback on Caltech101



(c) Box-plots of (a)

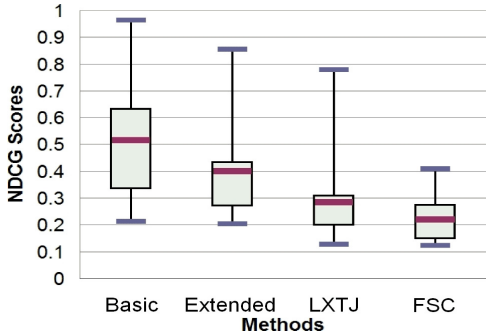


(d) Box-plots of (b)

Figure 2: NDCG scores of 101 image search experiments conducted over the Caltech101 image set using "LXTJ," "FSC," and our algorithm respectively. We report the NDCG scores for all 101 image search sessions (a), (b) and also illustrate their key statistics using boxplots (c), (d). For easy reading, all the NDCG scores are sorted in the ascending order according to the NDCG score of our method. In (a) & (c) we compare our algorithm's performance with that of LXTJ and FSC, all without using any relevance feedback measurement; in (b) & (d) we compare the performance of our algorithm and that of LXTJ, both using three rounds of relevance feedback.

Image Tag	User Nominated Query Keywords	
brain	neuron	nervous
cannon	gun	weapon
dollar bill	dollar	money
ferry	ship	boat
laptop	notebook	computer
menorah	candle	candelabrum
motorbike	motorcycle	pedal bicycle
pyramid	Egypt	triangle
stop sign	sign	car sign
sunflower	flower	flora

(a)



(b)

Figure 3: Statistics of the NDCG scores attained by "LXTJ," "FSC," and our method respectively, in our experiments simulating real-world user search scenarios. (a) lists user nominated query keywords. In (b), we illustrate NDCG scores using box-plot diagrams, where the box-plots "Basic" and "Extended" respectively indicate search scores achieved by our method when we use the exact image class names (basic) and the user nominated keywords (extended) as image search queries to perform image searches. We also provide search scores achieved by "LXTJ" and "FSC" for image searches using user nominated search keywords for comparison.

image. For the "FSC" algorithm, again half of images in each class of Caltech101 were used as the training set and the rest half were used as the testing set. Figure 3 shows key statistics on the results of our image search experiments achieved by all three algorithms. These results again demonstrate the superiority of our algorithm to the other two peer methods.

4.7 Image Retrieval on Personal Photo Collections

Eight volunteers were invited to use the prototype image search system that implements our algorithm to retrieve images from their respective personal photo albums. The number of photos in a volunteer's personal photo album is between 500 and 4000. The number of photos in the eight volunteers' personal photo albums are 1172, 729, 2305, 671, 1380, 2768, 3884, 3374 respectively.

We asked every volunteer in our study to perform 30 image searches. For each volunteer, his or her 30 image search queries were formulated by the volunteer himself or herself by freely choosing 30 search keywords. We then ranked all the image search query words suggested by the eight volunteers according to the frequency that a query word was commonly suggested by multiple users. Once all the query words were ranked, we selected the top 20 most fre-

quently suggested query keywords. Figure 4(a) lists these twenty image query keywords used in our experiments. Figure 5 shows some example image search results by our algoirthm in our personal photo retrieval experiments. We performed image searches using these query words over each user's personal photo collection. After that, we manually evaluated the NDCG score for each query performed over a user's personal photo collection, as introduced earlier in Sec. 4.1. For each user, we performed twenty image search sessions, where each time we use one of the twenty image search keywords as the image search query and performed the search session over the user's personal photo collection. Hence for each user, we obtained twenty NDCG scores, one for each image query session performed. Figure 4(b) reports the key statistics on the distributions of these twenty NDCG scores attained by our method for personal photo search experiments performed for every user respectively.

For comparison purpose, we also report the NDCG scores of personal image retrieval experiments, performed using "LXTJ" and "FSC" respectively, under the same experiment setup. For the "FSC" algorithm, we notice that the algorithm can't be run in a straightforward way on a personal photo album because it needs a labelled training set. To enable the "FSC" algorithm to run in our comparison, we first retrieved the top 500 images for each query keyword using Google image search as its training set. In Figure 4, we use yellow, red, and blue boxplots to respectively illustrate the statistic distributions of search scores obtained using "LXTJ," "FSC," and our algorithm when performing twenty personal image retrieval sessions for every user in our study. According to results shown in the figure, for all the subjects in our user study, the image search scores obtained using our algorithm are noticeably higher than the image search scores of the other two peer methods, which demonstrates the general superiority of our algorithm to the two existing methods for retrieving unannotated personal photos.

5. CONCLUSION

We introduced a new content-based image search and ranking algorithm for unannotated images. Our algorithm infers the relevance of an unannotated image to a text query through a set of reference images from online image search. To deal with noisy reference image samples, we introduced an elaborate joint text analysis and image visual content analysis procedure to set apart the good and the bad (i.e., irrelevant) reference images. Then through a weighted visual phrase based scheme, we transfer the image-to-query relevance information on the reference images to an arbitrary unannotated image. The transfer relevance is then used to retrieve and rank unannotated images. The experimental results convincingly show the effectiveness and advantages of our new algorithm for searching unannotated images given an input textual query.

In the future, we plan to explore advanced machine learning algorithms to more reliably and comprehensively understand semantics of an unannotated images for content-based image search. We also plan to design some intelligent and friendly user interface that can leverage user feedback or utilizes some social tags to improve the retrieval quality through online learning. We plan also to extend our algorithm to automatically and efficiently annotate unlabelled images on the Internet. This new problem is closely related to the task tackled by our algorithm in this paper—for our current algorithm presented in this paper, we intend to find relevant images for a given text query; while in the automatic image labeling task, the goal is to look for text term(s) that can maximize their relevance with a given image. We suspect many steps of our algorithm can be readily used to address that new problem.

Acknowledgement

Songhua Xu performed this research as a Eugene P. Wigner Fellow and staff member at the Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract DE-AC05-00OR22725. This research is also partially sponsored by Centers for Medicare & Medicaid Services of United States Department of Health and Human Services.

6. REFERENCES

[1] P. Adrian, M. Christophe, and M. Pierre-Alain. Ontology driven content based image retrieval. In *CIVR '07: Proceedings of the 6th ACM International Conference on Image and Video Retrieval*, pages 387–394, New York, NY, USA, 2007. ACM.

[2] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 2008.

[3] D. Cai, X. He, Z. Li, W.-Y. Ma, and J.-R. Wen. Hierarchical clustering of www image search results using visual, textual and link information. In *MULTIMEDIA '04: Proceedings of the 12th Annual ACM International Conference on Multimedia*, pages 952–959, New York, NY, USA, 2004. ACM.

[4] P. Clough, M. S, and H. Mlzller. H.: A proposal for the clef cross language image retrieval track (ImageCLEF) 2004. In *CIVR '04: The Challenge of Image and Video Retrieval*, page 3115. Springer LNCS, 2004.

[5] R. Datta, J. Li, and J. Z. Wang. Content-based image retrieval: approaches and trends of the new age. In *MIR '05: Proceedings of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval*, pages 253–262, New York, NY, USA, 2005. ACM.

[6] L. Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. In *CVPRW '04: Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop, Volume 12*, page 178, Washington, DC, USA, 2004. IEEE Computer Society.

[7] H. Feng, R. Shi, and T.-S. Chua. A bootstrapping framework for annotating and retrieving www images. In *MULTIMEDIA '04: Proceedings of the 12th Annual ACM International Conference on Multimedia*, pages 960–967, New York, NY, USA, 2004. ACM.

[8] R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman. Learning object categories from google's image search. In *10th IEEE International Conference on Computer Vision*, ICCV '05, pages 1816–1823. IEEE, 2005.

[9] R. Fergus, P. Perona, and A. Zisserman. A visual category filter for google images. In *ECCV '04: 8th Proceedings of the European Conference on Computer Vision*, pages 242–256, London, UK, 2004. Springer-Verlag.

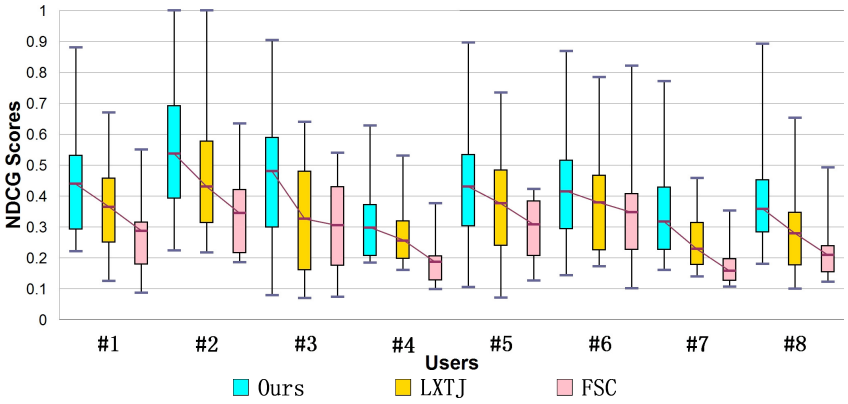
[10] E. Gabrilovich and S. Markovitch. Computing semantic relatedness using wikipedia-based explicit semantic analysis. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, pages 1606–1611, San Francisco, CA, USA, 2007. Morgan Kaufmann Publishers Inc.

[11] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1991.

[12] J.-H. Hsiao, C.-S. Chen, and M.-S. Chen. A novel language-model-based approach for image object mining and

apple	baby
beach	chair
cloud	desert
everest	flower
girl	green trees
house	jokul
lotus	mountain
park	people
pyramid	railway
river	rocket

(a)



(b)

Figure 4: (a) Twenty image queries performed over each user's personal photo collection in our experiments. (b) Distributions of NDCG scores of three algorithms for searching unannotated images over eight users' personal photo collections in box-plots. Twenty image queries were performed over each user's personal photo collection using "LXTJ" (without relevance feedback), "FSC," and our algorithm respectively.

re-ranking. In *ICDM '08: Proceedings of the 8th IEEE International Conference on Data Mining*, pages 243–252, Washington, DC, USA, 2008. IEEE Computer Society.

[13] M. J. Huiskes and M. S. Lew. The mir flickr retrieval evaluation. In *Proceeding of the 1st ACM International Conference on Multimedia Information Retrieval*, MIR '08, pages 39–43, New York, NY, USA, 2008. ACM.

[14] J. Jia, N. Yu, and X.-S. Hua. Annotating personal albums via web mining. In *Proceeding of the 16th ACM International Conference on Multimedia*, MM '08, pages 459–468, New York, NY, USA, 2008. ACM.

[15] Y. Jing and S. Baluja. Pagerank for product image search. In *WWW '08: Proceedings of the 17th International Conference on World Wide Web*, pages 307–316, New York, NY, USA, 2008. ACM.

[16] M. L. Kherfi, D. Ziou, and A. Bernardi. Image retrieval from the world wide web: Issues, techniques, and systems. *ACM Computing Survey*, 36(1):35–67, 2004.

[17] P. Kidambi and S. Narayanan. A human computer integrated approach for content based image retrieval. In *ICCOMP'08: Proceedings of the 12th WSEAS International Conference on Computers*, pages 691–696, Stevens Point, Wisconsin, USA, 2008. World Scientific and Engineering Academy and Society (WSEAS).

[18] L.-J. Li and L. Fei-Fei. Optimol: Automatic online picture collection via incremental model learning. *International Journal of Computer Vision*, 88:147–168, 2010.

[19] W.-H. Lin, R. Jin, and A. Hauptmann. Web image retrieval re-ranking with relevance model. *Proceedings of IEEE/WIC International Conference on Web Intelligence*, pages 242–248, 2003.

[20] Y. Liu, D. Xu, I. W. Tsang, and J. Luo. Using large-scale web data to facilitate textual query based retrieval of consumer photos. In *Proceedings of the 17th ACM International Conference on Multimedia*, MM '09, pages 55–64, New York, NY, USA, 2009. ACM.

[21] Y. Liu, D. Xu, I. W. Tsang, and J. Luo. Textual query of personal photos facilitated by large-scale web data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33:1022–1036, May 2011.

[22] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma. A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, 40(1):262–282, 2007.

[23] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[24] R. Marée, P. Denis, L. Wehenkel, and P. Geurts. Incremental indexing and distributed image search using shared randomized vocabularies. In *Proceedings of the International Conference on Multimedia Information Retrieval*, MIR '10, pages 91–100, New York, NY, USA, 2010. ACM.

[25] H. Muller, N. Michoux, D. Bandon, and A. Geissbuhler. A review of content-based image retrieval systems in medical applications–clinical benefits and future directions. *International Journal of Medical Informatics*, 73(1):1–23, 2004.

[26] F. Schroff, A. Criminisi, and A. Zisserman. Harvesting image databases from the web. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33:754–766, April 2011.

[27] 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 Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.

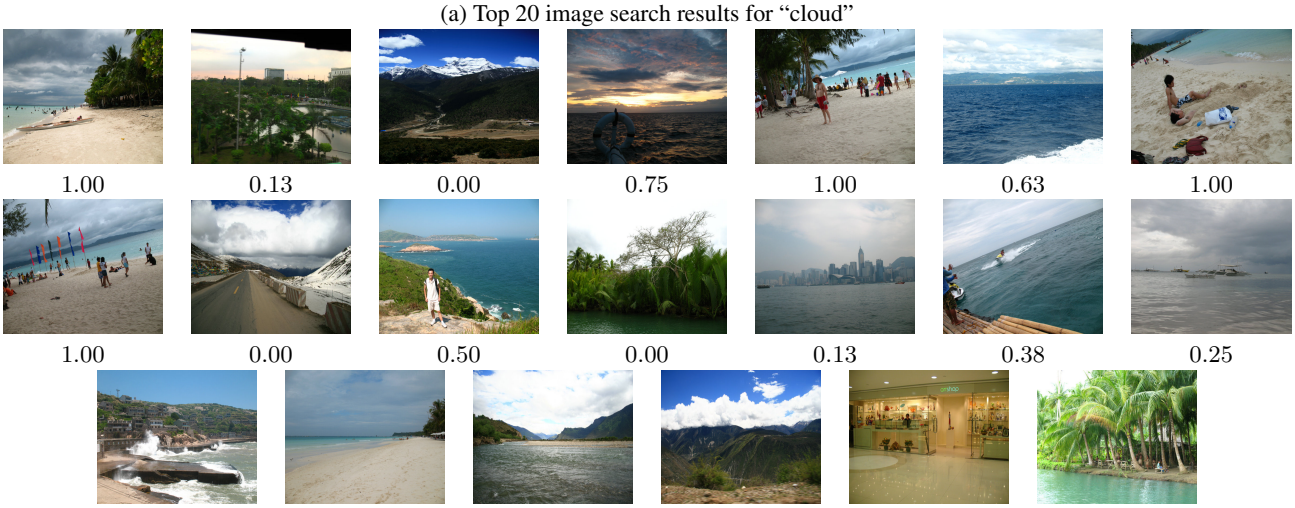
[28] L. Torresani, M. Szummer, and A. Fitzgibbon. Learning query-dependent prefilters for scalable image retrieval. In *CVPR '09: IEEE Conference on Computer Vision and Pattern Recognition*, pages 2615–2622, Washington, DC, USA, 2009. IEEE Computer Society.

[29] I. H. Witten, A. Moffat, and T. C. Bell. *Managing gigabytes: compressing and indexing documents and images*. Morgan Kaufmann, 1999.

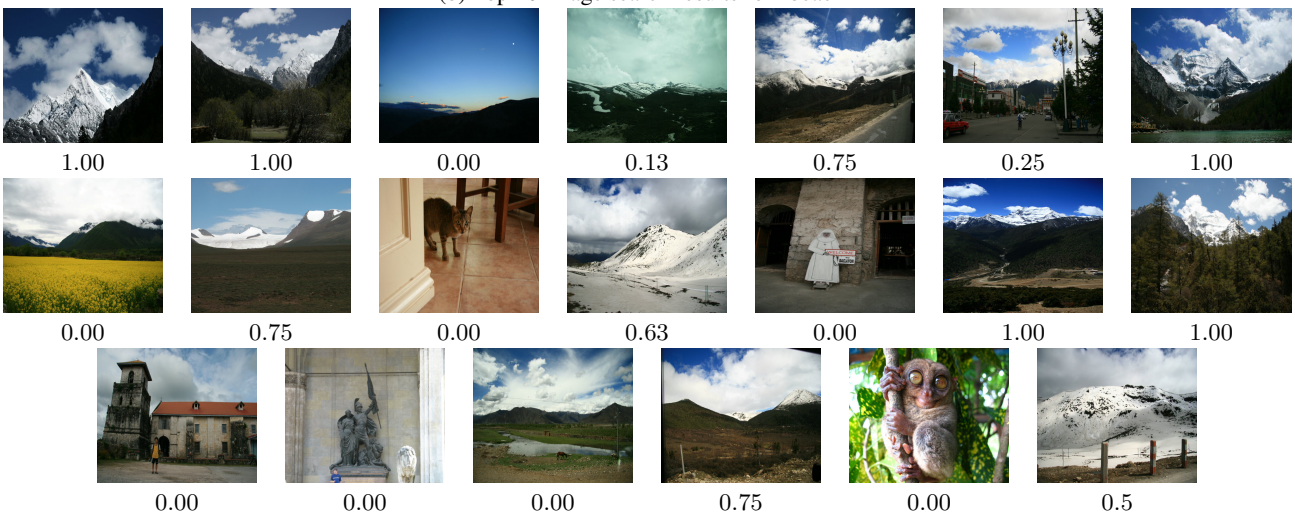
[30] Q.-F. Zheng and W. Gao. Constructing visual phrases for effective and efficient object-based image retrieval. *ACM Transactions on Multimedia Computing, Communications and Applications*, 5(1):1–19, 2008.

[31] X. S. Zhou and T. Huang. Unifying keywords and visual contents in image retrieval. *IEEE Multimedia*, 9(2):23–33, 2002.

(a) Top 20 image search results for “cloud”



(b) Top 20 image search results for “beach”



(c) Top 20 image search results for “jokul”

Figure 5: Three personal photo album search experiments using the keywords “cloud,” “beach,” and “jokul” respectively. For each experiment, we show its top 20 search result images, along with each image’s relevance to the query (beneath the image).