EnjoyPhoto—A Vertical Image Search Engine for Enjoying High-Quality Photos

Lei Zhang[1], Le Chen[2], Feng Jing[1], Kefeng Deng[2], Wei-Ying Ma[1]
[1]Microsoft Research Asia, 49 Zhichun Road, Beijing 100080, China
[2]Tsinghua University, Beijing 100084, China
{leizhang, fengjing, wyma}@microsoft.com, {chenle02, dkf03}@mails.tsinghua.edu.cn

ABSTRACT
In this paper, we propose building a vertical image search engine called EnjoyPhoto that leverages rich metadata from various photo forum web sites to meet users' requirements for enjoying high-quality photos, which is virtually impossible in traditional image search engines. To solve the ranking problem when aggregating multiple photo forums, we propose a novel rank fusion algorithm that uses duplicate photos to normalize rating scores. To further improve user experiences in enjoying photos, we design an in-place image browsing interface, and compare it with several other interfaces in a user study. With rich metadata and rating information, more attractive user interfaces are enabled, including slideshow authoring and photo recommendations. We conducted experiments and user studies on a 2.5-million image database to evaluate the proposed rank fusion algorithm, investigate the rationale behind building a vertical image search engine, and study user interfaces and preferences for the purpose of enjoying high-quality photos. The experimental results demonstrate the effectiveness of the proposed ranking algorithm. The results also show that the 2.5-million high-quality image database in EnjoyPhoto performs comparably with Google's 1-billion image database for queries related to location, nature, and daily life categories. Finally, our results show that the in-place browsing interface—called Force-Transfer view—is much more convenient for users than traditional interfaces.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *Search Process.* H.5.0 [**Information Storage and Retrieval**]: Information Interfaces and Presentation – *General.*

General Terms
Algorithms, Design, Performance

Keywords
User interface, high-quality photo, vertical search, image browsing interface, user study

1. INTRODUCTION
In the past decade, the Internet has become an incredible resource, allowing users to easily access a huge number of images. However, compared to the more than 1 billion images indexed by commercial search engines [7][25], queries submitted to image search engines are relatively minor, and occupy only 8–10 percent of total image and text queries submitted to commercial search engines [21]. This is partially because user requirements for image search are far less than those for general text search. On the other hand, current commercial search engines still lack various user requirements because there is no effective and practical solution to allow an understanding of image content.

To better understand user needs in image search, we conducted a query log analysis based on a commercial search engine. The results show that more than 20% of image search queries are related to location-, nature-, and daily life-related categories. Users apparently are interested in enjoying high-quality photos or searching for beautiful images of locations (or seasons and nature). However, such needs are not well supported by current image search engines because of the difficulty of the quality assessment problem.

Therefore, in this work, we particularly focus on the location-, nature-, and daily life-related categories and attempt to build a vertical[1] image search engine specifically for high-quality photos. We will concern ourselves with other category requirements in future endeavors.

Ideally, the most critical part of a search engine, the ranking function, mainly consists of two key factors: relevance and quality. For the relevance factor, a search in current commercial image search engines provides images that are quite relevant to queries, except for some ambiguity. However, with the quality factor, there is still no way to give an optimal rank to an image. Although content-based image quality assessment has been investigated over many years [22][18][23], it is still far from ready to provide an artistic quality measure in the immediate future.

Because of the difficulty of ranking photos in terms of quality, it seems pessimistic to build an image search engine that can fulfill the potentially large requirement of searching high-quality photos. Despite the existence of several online stock photo galleries, such as www.corbis.com and www.gettyimages.com, the coverage of

[1] "Vertical" is a common term in Web search, and means that the search engine is built for a specific domain and is expected to perform better than general search engines in this domain, because domain knowledge can be leveraged.

the images and their update speeds are usually limited due to the need of manual efforts in labeling. More challengingly, there is still no effective solution to further rank these professional photos.

However, within a growing number of Web communities, we notice that people are creating and sharing a lot of high-quality images on the web on virtually any topic, which provides a rich source for building a better image search engine. These web communities include personal photo sharing websites like photos.yahoo.com, photo blogs like www.photoblogs.org, and photo forums like www.photosig.com. In general, photos from various photo forums are of higher quality than personal photos, and are also more appealing to public users than personal photos. In addition, photos uploaded to photo forums require rich metadata about the title, camera setting, category, and description provided by photographers. These metadata are the most precise descriptions for photos and undoubtedly can be indexed to solve the relevance problem in a search engine. More importantly, there are volunteer users in each web community actively providing valuable ratings for these photos. The rating information is of great value in solving the photo quality ranking problem.

Motivated by such observations, we attempt in this work to build a vertical image search engine by aggregating high-quality photos from multiple forums. Essentially, building vertical search engines includes data crawling, information extraction, object identification and integration, and object-level Web information retrieval (or Web object ranking)[13]. Ranking is one of the most important factors, as it deals with the core problem of how to combine and rank objects coming from multiple communities. Moreover, the user interface design is also an important factor in such a high-quality photo search engine.

Therefore, in this paper, we focus on investigating the rationale behind building a vertical image search engine by extracting rich metadata and integrating information from various photo forums. We also address the problem of ranking photos coming from multiple forums, and study user interfaces and preferences for the purpose of searching and browsing high-quality photos.

The contributions of this paper can be highlighted as follows:

1. We proposed and built a vertical image search engine by leveraging rich metadata from various photo forum web sites to meet user requirements of searching for and enjoying high-quality photos, which is impossible in traditional image search engines.

2. We proposed a novel rank fusion algorithm for photos from multiple forums, which can automatically and efficiently integrates as many Web communities as possible that have rating information.

3. To further improve the user's experience, we proposed a new browsing interface and compared it to several other interfaces in a user study. With rich metadata and rating information, more attractive user interfaces (UIs) are enabled, including slideshow authoring, photo recommendations, and similar photo browsing.

The rest of the paper is organized as follows. In Section 2, we briefly review the related works. In Section 3, we introduce the data collection, present the algorithm for how to rank photos from multiple forums, and describe the user interface design. Then, in Section 4, we present the user study and experimental results. Finally, in Section 5, we provide the concluding remarks.

2. RELATED WORK

2.1 General and Vertical Image Search

Web image search has been actively explored and developed in academic as well as commercial areas. There are commercial search engines available such as Google Image Search [7], Yahoo Image Search [25], and PicSearch [15]. There are also systems developed by academic researchers including WebSeek [20], WebSeer [5], Image Rover [19], and iFind [2].

As Web images usually come with HTML source code including textural descriptions, most Web image search engines are text-based [7][25]. However, as we discussed in the introduction, the main drawback is that there is no effective solution to provide an optimal ImageRank like PageRank. Therefore, the image search result usually looks noisy and unattractive. Therefore, we expect that a relatively small-scale (but well-ranked) high-quality image database might perform better than a large-scale (but poorly ranked) image database.

Some researchers have attempted to build vertical image search engines for specific domains. Yee et al. **Error! Reference source not found.** built a prototype system for art image search by explicitly leveraging faceted metadata. Yet it is mainly for art history, and the data collection is a closed database.

2.2 Object Ranking in Vertical Search

Although object-level ranking has been well studied in building vertical search engines, there are still some kinds of vertical domains in which objects cannot be effectively ranked. For example, algorithms such as PopRank [13] (that evolved from PageRank [14]) were proposed to rank objects coming from multiple communities, but can only work on well-defined graphs, in which heterogeneous data are well linked by identifying all the relationships (e.g., cited-by, authored-by, and published-by) among multiple communities.

However, we cannot assume this would work for high-quality photo search, because different photo forums seem to lack any relationships and generally contain no cited-by relationships. Therefore, although each photo has a rating score in a forum, it is nontrivial to rank photos coming from different photo forums.

The ranking problem can also be viewed as a rank aggregation problem [3][4], since we deal with the problem of how to combine multiple rank lists. However, there are fundamental differences between them. The number of duplicate photos from different Web forums are small relative to the number of whole photo sets. In other words, the top K rank lists of different Web forums are almost disjointed for a given query. In this case, both the algorithms proposed in [3] and its measurements—Kendall's tau distance or Spearman's footrule distance—will degenerate.

Another category of ranking methods is based on machine learning algorithms, such as RankSVM [10], RankBoost [6], and RankNet [1]. All of these methods require labeled datasets to train a model. In current settings, it is difficult or even impossible to get these datasets labeled as professional and popular, since the photos are too vague and subjective to rank.

A different approach we also considered was manually tuning the rating scores for each forum. The movie discussion web site, IMDB [9], proposes using a Bayesian ranking function to normalize rating scores within one community. But it requires manual tuning of three parameters for each Web forum, which becomes laborious and impractical as more and more Web forums are crawled and indexed.

2.3 User Interface

There are also other works on improving search result presentation forms. Liu et al. [12] proposed employing a similarity-based organization to present the search results with a cluster-based interface. Users can interact with interested images to generate a Fisheye-style local view. However, as there are too many (about 200) thumbnail images embedded in one overview, it is too crowded to browse the interested images. Li et al. [11] proposed grouping image search results into clusters. However, each group does not have a clearly semantic meaning. As users are required to click twice to reach the final list view and find the images they are interested in, they might be reluctant to further interact with the search results.

For the purpose of enjoying high-quality photos, it is of great importance to make users' browsing experience smooth and avoid requiring users to repeatedly open and switch to new windows. Therefore, we propose to adopt an in-place browsing mode by employing and improving the Force-Transfer algorithm [8]. In such a browsing mode, users can interact with the interested images to dynamically get a detailed high-resolution view without the need to switch between multiple windows. Furthermore, other integrated UI features, such as photo recommendation and slideshow authoring, also greatly improve user experiences in searching and browsing high-quality photos.

3. BUILDING A HIGH-QUALITY PHOTO SEARCH ENGINE

3.1 Metadata & Collection

Nowadays, people around the world use photos to visually communicate with others and present their feelings about a vacation, a party, a news event, or about virtually any topic related to their daily lives. Photo forums provide an energetic environment for people to share and discuss photography. To attract more attention, most photographers are very enthusiastic about providing metadata such as a title, category, location, camera setting, and description to the uploaded photos. And numerous volunteer users provide ratings and critiques to photos that they particularly like and enjoy.

For example, a photo entitled "*early morning*" at http://www.photosig.com/go/photos/view?id=733881 has the following metadata:

> Description: *I found this special light one early morning in the Pyrenees along the Vicdessos river near our house...*
>
> Category: *landscape, nature, rural*
>
> Camera: *Nikon Coolpix 5700*
>
> One of the critiques: *Wow … I like this picture very much … I guess the light has to do with everything ... the light is great on the snow and on the sky (strange-looking sky, by the way) … greatly composed … nice crafted border … a beauty.*

Interested readers may go to www.photosig.com or any other photo forums to see more examples.

To study the rationale of building a vertical image search engine for the purpose of enjoyment and recreation, we extracted from a commercial search engine a subset of photos coming from various photo forums all over the world, and explicitly parsed web pages containing these photos. The number of photos in the data collection is about 2.5 million. After the explicit parsing, each photo was associated with a title, category, description, camera, EXIF data (when available for digital images), location (when available in some photo forums), and rating.

Then the whole data collection was indexed by general text-based search technologies. Given a query q, the ranking function for each image I_i is a weighted combination of relevance and quality factors as follows:

$$Score(I_i, q) = w_1 \cdot R(I_i, q) + w_2 \cdot Q(I_i)$$

where $R(I_i, q)$ measures the relevance score between image I_i and the query q, and $Q(I_i)$ is the quality score that is query-independent. w_1 and w_2 were empirically chosen as 0.67 and 0.33.

The relevance factor $R(I_i, q)$ is further a linear combination of five relevance scores in title, location, category, description and critique fields.

$$R(I_i, q) = \alpha_1 R_{title}(I_i, q) + \alpha_2 R_{location}(I_i, q) + \alpha_3 R_{category}(I_i, q) \\ + \alpha_4 R_{description}(I_i, q) + \alpha_5 R_{critique}(I_i, q)$$

Each relevance score is calculated using BM25 formula [16] in each field. The weighting parameters α_1, α_2, α_3, α_4, and α_5 were empirically chosen as 1.0, 2.0, 2.0, 0.5, and 0.05, so that title, location and category can be emphasized in the final ranking function.

The quality factor $Q(I_i)$ is the rating score which is normalized according to the algorithm presented in the following subsection.

3.2 Ranking Photos from Multiple Forums

The difficulties of integrating multiple Web forums lie in their different rating systems, where there are generally two kinds of freedom.

The first kind of freedom is the rating interval or rating scale, including the minimal and maximal ratings for each photo. For example, some forums use a 5-point rating scale whereas other forums use a 3-point or 10-point rating scale. It seems easy to fix this freedom, but detailed analysis of the data and experiments show that it is not a trivial problem.

The second kind of freedom is the different rating criteria in different Web forums. That is, the same score does not mean the same quality in different forums. Intuitively, if we can detect the same photographer or same photographs, we can build relationships between any two photo forums and therefore, can standardize the rating criterion by score normalization and transformation.

Based on these observations, we propose a novel rank fusion approach that can effectively rank photos from multiple forums. First, we analyze the rating distribution in each forum, and normalize the scores to a standard scale. Then we adopt an efficient algorithm [24] to identify duplicate photos between any two forums and build implicit relationships among forums. Finally, we formulate the ranking problem as an optimization problem which attempts to make the transformed scores of

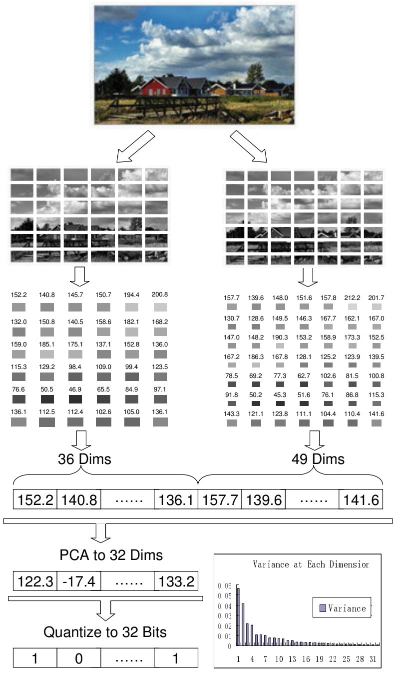


Figure 1. Duplicate photo detection hashing procedure.

duplicate photos as close as possible, and propose a linear transformation model to effectively solve the ranking problem.

3.2.1 Score normalization

Because different photo forums on the Web usually have different rating criteria, it is necessary to normalize them before applying any score transformation. In addition, with varied rating systems, including some that rate novelty, aesthetics, or other factors, it is reasonable to choose a common one—total score or average score—that can always be extracted in any Web forum or calculated by corresponding ratings. This allows the normalization method on the total score or average score to be viewed as an impartial rating method between different Web forums.

It is straightforward to normalize average scores by linearly transforming them to a fixed interval. The problem of using this normalization method is that if there are only a few users to rate a photo, the average score for the photo may not be trustable and is easy to be spammed.

A total score can avoid such drawbacks, since it contains more information, including a Web object's quality and popularity. So the problem is how to normalize different total scores. The simplest way may be the normalization by the maximal and minimal scores. The drawback of this normalization method is its non-robustness, which is sensitive to the outliers. For example, an exceptionally highly scored photo will push other photos to lower scores under this transformation.

To make the normalization not so sensitive to the extreme data, we propose *the Mode and the 90th Percentile* normalization method. This normalization method utilizes the mode and the 90th percentile as two reference points to align two rating systems. The underlying assumption is that, even the quality of the best photos in different forums may vary greatly and depend less on the overall quality of a forum, the distribution of mid-level quality photos (from the mode to the 90th percentile) should be almost the same, which reflects the rating criterion (strictness) of Web forums. Usually, more than 70% of the photos are of this mid-level quality in each photo forum.

3.2.2 Duplicate Photo Detection

We use Dedup [24] to find duplicate photos common to any two photo forums. This algorithm uses a PCA-based Hash function to map a high-dimensional feature to a 32-bits hash code. Its computational complexity to find all the duplicate images among n images is about $O(n\log(n))$. If a hash table is used to detect hash collision, the computational complexity can be reduced to a constant level which is far less than n. The low-level visual feature for each photo is extracted on $k \times k$ regular grids. Based on all the features extracted from the image database, a PCA model is built. The visual features are then transformed to a low-dimensional and zero mean PCA space, which is a 29-dimensional space in our system. Then the hash code for each photo is built as follows: each dimension is transformed to one if the current value in this dimension is greater than zero, and zero otherwise. Photos in the same bucket that are deemed potential duplicates are further filtered by a threshold in terms of Euclidean similarity in visual feature space.

Figure 1 illustrates the hashing procedure, where visual features—mean gray values—are extracted on both 6×6 and 7×7 grids. The 85-dimensional features are transformed to a 32-dimensional vector, and the hash code is generated according to the signs.

3.2.3 Score fusion by duplicate photos

The proposed methods below are based on the following considerations. As we are solving a ranking problem, a standardized rating criterion rather than a reasonable rating criterion is needed. Therefore, we can take a large scale forum as the reference one, and align other forums to it by taking duplicate photos into account.

Ideally the transformed scores of duplicate photos should be equal despite the fact that they are in different forums. Because it is difficult to satisfy this criterion in practice, we formulate the ranking problem as an optimization problem that attempts to make the transformed scores of duplicate photos in non-reference forums as close as possible to those in a reference forum. Then, we can effectively solve the ranking problem.

For convenience, the following notations are employed. S_i^k denotes the total score of the ith Web object (photo) in the kth Web site. There are a total of K Web sites. We further use

$$\{S_i^{kl} \mid i = 1,..., I_{kl}; k, l = 1,..., K; k \neq l\}$$

to denote the set of scores for Web objects (photos) in kth Web forums that are duplicate with the lth Web forums, where I_{kl} is the total number of duplicate Web objects between these two Web sites. For simplicity, we also use $\vec{S}^{kl} = (S_1^{kl},..., S_{I_{kl}}^{kl})^T$ to denote these scores. In general, score fusion can be seen as the procedure of finding K transforms

$$\psi_k(S_i^k) = \tilde{S}_i^k, \quad k=1,\ldots,K$$

such that $\tilde{S}_i^k$ can be used to rank Web objects from different Web sites. The objective function described in the above paragraph can then be formulated as

$$\min_{\{\psi_k|k=2,\dots,K\}} \sum_{k=2}^{K}\sum_{i=1}^{I_{k1}} \overline{w}_i^k \left(S_i^{1k} - \psi_k(S_i^{k1})\right)^2 \tag{1}$$

where we use k = 1 as the reference forum, and thus $\psi_1(S_i^1) = S_i^1$. $\overline{w}_i^k$ (≥0) is the weight coefficient that can be set heuristically according to the numbers of voters (reviewers) in both the reference forum and the non-reference forum. The more reviewers, the more popular the photo is and the larger the corresponding weight $\overline{w}_i^k$ should be, because usually users are inclined to rate only those photos they like[1]. In this implementation, we simply set $\overline{w}_i^k$ to one, and use a linear transformation model to solve the optimization problem.

In the linear model, we assume ψ_k has the following form

$$\psi_k(S_i^k) = \alpha_k S_i^k + t_k\,, \qquad k = 2,\dots,K \tag{2}$$

$$\psi_1(S_i^1) = S_i^1 \tag{3}$$

which means that the scores of the k(≠1)th forum should be scaled by α_k relative to the center $t_k/(1-\alpha_k)$.

Then, if we substitute the above ψ_k to Equation 1, we get the following objective function:

$$\min_{\{\alpha_k,t_k|k=2,\dots,K\}} \sum_{k=2}^{K}\sum_{i=1}^{I_{k1}} \overline{w}_i^k \left(S_i^{1k} - \alpha_k S_i^{k1} - t_k\right)^2 \tag{4}$$

By solving the following set of functions,

$$\begin{cases} \dfrac{\partial f}{\partial \alpha_k} = 0 \\ \dfrac{\partial f}{\partial t_k} = 0 \end{cases}, \qquad k = 1,\dots,K$$

where f is the objective function defined in Equation 4, we get the closed form solution as

$$\begin{pmatrix} \alpha_k \\ t_k \end{pmatrix} = A_k^{-1} L_k \tag{5}$$

where

$$A_k = \begin{pmatrix} \sum_i \overline{w}_i (S_i^{k1})^2 & \sum_i \overline{w}_i S_i^{k1} \\ \sum_i \overline{w}_i S_i^{k1} & \sum_i \overline{w}_i \end{pmatrix} \tag{6}$$

$$L_k = \begin{pmatrix} \sum_i \overline{w}_i S_i^{1k} S_i^{k1} \\ \sum_i \overline{w}_i S_i^{1k} \end{pmatrix} \tag{7}$$

and $k = 2,\dots,K$.

The linear fusion method enjoys simplicity and shows good performance in the following experiments.

[1] Though a photo with many low scores does exist, such cases are rare compared with high scores, and therefore a larger $\overline{w}_i^k$ has little impact in the objective function.

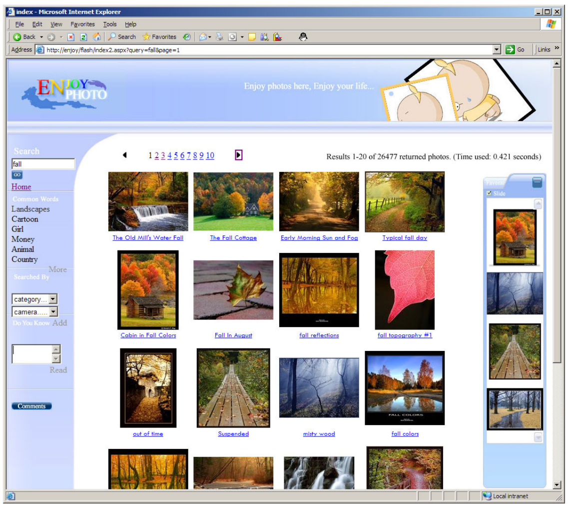


Figure 2. Search interface.

3.2.4 Performance measure of the fusion results

Since our objective function is to make the transformed scores of the same Web objects (e.g., duplicate photos) between a non-reference forum and the reference forum as close as possible, it is natural to investigate how close they come to each other and how the scores of the same Web objects change between the two non-reference forums before and after score fusion.

We define the following performance measure—δ measure—to quantify the changes for scores of the same Web objects in different Web forums as

$$\delta_{kl} = Sim(\vec{S}^{lk}, \vec{S}^{kl}) - Sim(\vec{S}_*^{lk}, \vec{S}_*^{kl}) \tag{8}$$

where $\vec{S}^{kl} = (S_1^{kl},\dots,S_{I_{kl}}^{kl})^T, \vec{S}_*^{kl} = (\tilde{S}_1^{lk},\dots,\tilde{S}_{I_{kl}}^{lk})^T$ and

$$Sim(a,b) = \frac{a\cdot b}{|a|\cdot|b|}\,.$$

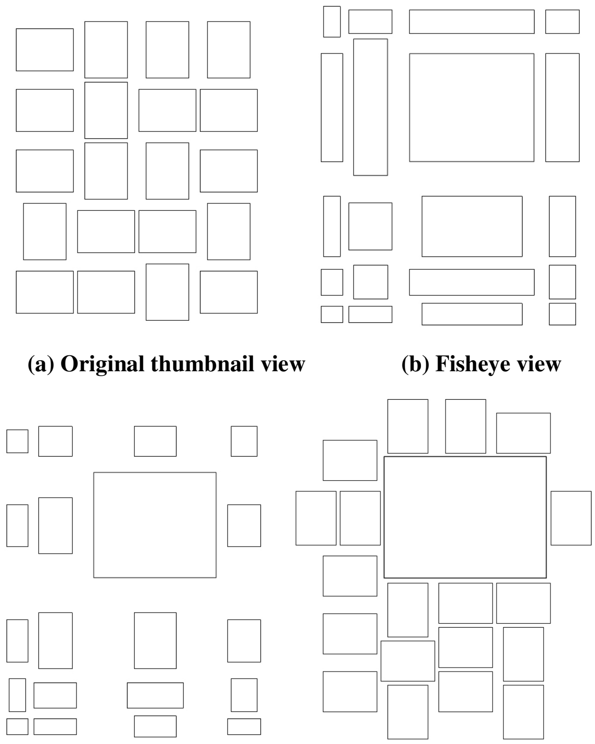
δ_{kl} > 0 means after the score transformation, scores on the same Web objects between the kth and lth Web forum become more consistent, which is what we expect. On the contrary, if δ_{kl} < 0, those scores become more inconsistent.

3.3 Interface Design

3.3.1 Basic search interface

A text-based search user interface is the main user interface in our prototype search engine EnjoyPhoto, as most users are familiar with such UIs in Google and Yahoo. In EnjoyPhoto, users can type any queries and retrieve tens of thousands of images in less than one second, due to the use of efficient indexing structures such as inverted file and cache strategies. One example search result is illustrated in Figure 2.

It is quite common that a user wants to enjoy high-quality photos, but has no idea what he/she wants to search. Therefore, navigation utilities are also useful to provide many starting paths for browsing. For example, in Figure 2, the left panel provides several navigation utilities, including popular queries, category selection, and camera selection. And the right panel provides a window to



(a) Original thumbnail view **(b) Fisheye view**

(c) Modified Fisheye view **(d) Force-Transfer view**

Figure 3. Browsing experiences in different implementations.

randomly show top photos in terms of week/month/year, category, or general queries such as landscape or seasons.

The central part shows search results in a thumbnail list view including 4×5 images. Under each thumbnail, we also display the corresponding image title, which is the most informative summarization provided by photographers. Clicking on a title takes a user to the original web page containing this image.

3.3.2 Browsing interface

The expected user experience in enjoying photos is different from a general-purpose image search. In the latter case, users generally focus on finding or collecting information in the results. A typical UI is Google image search, in which users must frequently open new windows to see high resolution images and switch between the browsing window and detailed image windows. In the former case, though, users generally focus on enjoying rather than searching. Therefore, we propose to adopt an in-place browsing mode in which users can interact with the interested images to dynamically get a detailed high-resolution view without having to switch between multiple windows.

We have two implementations for this kind of in-place browsing mode. The first is our Fisheye [17] implementation and the other is Force-Transfer [8]. To obtain a smooth transition, both the Fisheye and Force-Transfer views were implemented in Flash.

3.3.2.1 Fisheye view

Fisheye view, an analogue of a fisheye lens, is a valuable tool to help users see the local detail and global context information simultaneously [17]. Liu et al. [12] employed Fisheye view to display as many as possible (e.g., 200 in their implementation) thumbnail images on one page so that users can efficiently browse more images. However, for the purpose of enjoying photos, such a view looks too crowded and is less attractive for enjoying. To avoid this drawback, in our implementation, we choose to display 4×5 images per page, similar to a traditional thumbnail UI, as shown in Figure 3a. However, once a user clicks on an image, the view will be dynamically and smoothly adjusted by enlarging the selected (clicked) image and meanwhile reducing other images according to the Fisheye algorithm. Immediately after the click, the browsing client starts downloading the high-resolution image, and replaces the enlarged thumbnail with the high-resolution one once it is downloaded.

To avoid exaggerated deformation (see Figure 3b) of reduced images in the same row or column with the clicked image, we modified the Fisheye algorithm to prevent the reduced images not being enlarged in any dimension, as Figure 3c shows.

Without being disturbed by opening new windows and switching between windows, users could be more focused on enjoying photos in the Fisheye view. Moreover, it is still convenient for users to interact with a traditional UI to open a detailed page by clicking the title text under a thumbnail. Apparently, the in-place browsing mode is a valuable complement to the traditional UI.

3.3.2.2 Force-Transfer view

The Fisheye view does provide an efficient way for browsing and enjoying search results. However, the layout in the Fisheye view is still not satisfactory and even looks ugly because of the non-uniform spaces among images, as shown in Figure 3c.

To address this drawback, we employ another layout arrangement algorithm—Force-Transfer [8]—which was originally proposed to avoid overlapping problems in graph layouts. Force-Transfer employs a heuristic method to approximate the global optimal adjustment with local minimal movement. In this work, we adopted it to calculate the final positions of affected images after enlarging an image and made several modifications to further improve user experiences. One of the modifications is to fade out other non-selected images to emphasize the selected one.

Force-Transfer does not need to reduce non-clicked images and thus there is no aspect ratio distortion for each image, as Figure 3d shows. Because Force-Transfer uses the local minimal movement strategy, the use of display space is efficient, and the layout looks more attractive and artistic than the Fisheye view. The transition process of enlarging the clicked image and moving other images looks very smooth. It also greatly improves user experiences.

3.3.3 Photo recommendation

Photo recommendation is a useful feature, as users do not usually have an explicit query in mind, but just desire to enjoy or browse some high-quality photos. Thus, a right panel was designed to randomly show top photos in terms of week/month/year, category, or general queries such as landscape or seasons. Top photos are user independent and could be treated as prior recommendations. Based on a user's browsing history and click log, the system will personalize the recommendation by combining top photos and user's preferences. The detailed algorithm will be described elsewhere. However, it is worth noting that the recommendation algorithm is orthogonal to this UI.

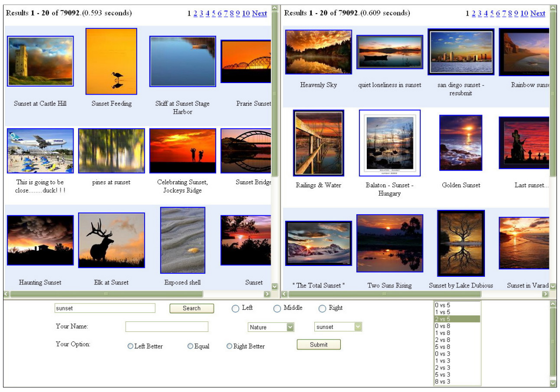


Figure 4. Side-by-side comparison interface.

3.3.4 Slideshow authoring

As the search results are appealing for most queries, users may want to keep some favorite photos. To support this need, we provide a slideshow-authoring feature in the right panel, as Figure 2 shows. Users can switch from the photo recommendation view to the slideshow-authoring view by clicking the slideshow tab. Users can drag interesting photos to this panel and adjust the sequence of the selected photos. The photos dropped to this panel will be automatically previewed as a slideshow at the top of the slideshow panel. If a user likes the slideshow he/she created, he/she can download the slideshow package. Meanwhile, the system also stores the play list of this slideshow. The stored slideshows can be displayed in the search result thumbnail list, or be recommended to other users in the recommendation view.

This is a very simple authoring tool for slideshows, yet in our user study, most users think it is a "wow" feature and really like it.

4. EXPERIMENTS AND USER STUDIES

In this section, we first present experimental evaluations of the proposed rank fusion algorithm. We then describe the user studies we conducted to investigate the rationale behind building such a vertical search engine and study user interfaces and preferences for the purpose of enjoying high-quality photos.

4.1 Ranking Performance Evaluation

To obtain standardized rating scores that we could use in the ranking function, we first normalized rating scores for each photo forum by using *the Mode and the 90th Percentile* normalization method. We then detected duplicate photos in any two forums. The detection results show that generally 1–3% duplicate photos exist in any two forums even with the strict condition for Dedup. Since a photo forum usually contains more than 100,000 photos, 1–3% still account for 1000–3000 duplicate photos. The number of duplicate photos is large enough for the proposed rank fusion algorithm to find an optimal solution.

Then, we select the largest Web forum as the reference one, and calculate the optimal parameters of the proposed linear transformation model by using the duplicate photo.

Table 1 shows the δ measures on the largest six forums. As δ_{kl} is symmetric and δ_{kk} = 0, we only show the upper triangular part. The NaN value means that no duplicate photos were detected by the Dedup algorithm.

The objective function in a linear model guarantees that the δ measures related to the reference forum A will be no less than 0, as the first row in Table 1 shows. But it does not guarantee that the δ measures on the non-reference forums are no less than 0, because the normalization steps are based on duplicate photos between the reference forum and a non-reference forum. Results shows that all such δ measures are also greater than 0, and indicates that this model is likely to give a good ranking result.

Table 1. The δ measures on the largest six forums.

	A	B	C	D	E	F
A	0	0.0911	0.0672	0.0105	0.0070	0.2220
B		0	0.0659	0.0956	0.0928	NaN
C			0	0.0578	0.0791	0.4618
D				0	0.0566	0.0232
E					0	0.6525
F						0

Note: The underlined number in the first row is related to the reference forum A.

To further evaluate the effectiveness of the proposed rank fusion algorithm, we also conducted a user study for subjective evaluations. Ten subjects were invited to participate in this study. They were recruited from nearby universities. As search engines of both text and image search are familiar to university students, there is no prerequisite criterion for choosing students.

Each participant was required to compare search results generated by two ranking functions with different score transformations. One was transformed by the max–min score normalization, and the other one was transformed by the proposed rank fusion algorithm. We also attempted to tune and compare a simple normalization method, that is, to normalize average scores by linearly transforming them to a fixed interval, e.g 0 to 10. The ranking result is apparently worse than the former two schemes, because an average score inevitably lacks the important information, i.e. the number of reviewers, which reflects the popularity of a photo in a forum. We then discarded this scheme in the user study. We presented to participants a side-by-side comparison interface as Figure 4 shows, in which the system randomly chooses a left or right frame to display two search results. In this way, participants can make a fair comparison. After each search session, a participant can submit his/her feedback (without knowing the ranking methods) by choosing answers among "left better," "equal," and "right better."

The 10 participants provided feedback for 52 queries which were freely chosen by them. Among these queries, the proposed rank fusion method was evaluated to perform better on 29 queries, equal on 13 queries, and worse on 10 queries.

Both the objective δ measure and subjective user study show that the proposed fusion algorithm is more effective in ranking photos from multiple forums than a simple fusion method. Therefore in the following real task user study, we only compare the best ranking function with Google image search engine.

4.2 User Studies

To evaluate the user interfaces, we conducted another series of user studies. Because the 10 subjects were familiar with EnjoyPhoto, another 12 subjects were invited to participate in our study, including 9 males and 3 females. They also were recruited from nearby universities, and we specifically chose "normal" users rather than photography fans, because photography fans might be overly predisposed to prefer the high-quality photos.

We conducted user studies using Internet Explorer 6 on Windows XP with 17-inch LCD monitors set at 1280 × 1024 pixels in 32-bit color. Data was recorded with server logs and paper surveys after each task.

4.2.1 Tasks and results

Through this series of user studies, we were interested in answering the following questions:

1. Do users really enjoy the process in searching and browsing high-quality photos? And could users be motivated to enjoy more photos with such a vertical image search engine?
2. Is a smooth user experience significant in enjoying photos? What is the difference experientially between enjoying high-quality photos and searching general images?

The tasks were designed to probe the answers for the major questions. Each participant was required to complete two tasks in about 60 minutes. Before the user study, we briefly introduced the system to participants and walked them through the features of the experimental interfaces.

Task 1: Searching for materials

This task aims at investigating: 1) whether it is rational to build a vertical image search engine for queries related to location, nature, and daily life categories, 2) if EnjoyPhoto can serve for finding materials for specific purposes, and 3) if users enjoy the process of searching.

We proposed two subtasks for participants. Each subtask is to search two collections of images from EnjoyPhoto and Google image search and then compare and evaluate two image collections in terms of quality and satisfaction. Participants can try any queries to complete the tasks.

1. Search for five more representative and enjoyable images for an unfamiliar location (e.g., London). This scenario is common when someone is going to travel to an unfamiliar place and he or she would like to find some representative and enjoyable images.
2. Search six more images for creating a calendar in one theme (e.g., season, animal, scenery, and portraiture).

Besides for these two mandatory subtasks, three optional subtasks were also provided if a participant wanted to search more. These subtasks included the following:

1. Find five more representative images for a familiar location.
2. Find five more high-quality photos shot by a camera model.
3. Find five portrait pictures for an instant messenger.

After each subtask, the participant was required to answer the following questions:

1. Are you confident that the found images are relevant?
   a) EnjoyPhoto: very confident, somewhat confident, unconfident, very unconfident
   b) Google: very confident, somewhat confident, unconfident, very unconfident
2. Are you satisfied with the search results?
   a) EnjoyPhoto: very satisfied, somewhat satisfied, unsatisfied, very unsatisfied
   b) Google: very satisfied, somewhat satisfied, unsatisfied, very unsatisfied
3. Which collection of images is better in terms of quality?
   EnjoyPhoto (much better), EnjoyPhoto (slightly better), Equally good, Google (slightly better), Google (much better)
4. Did you really enjoy the process of searching?

We realized that the participants' evaluations could be subjective, especially for evaluating images from an unfamiliar location. We therefore invited two "expert" participants as well to further review the found images for each participant and provide feedback to the questions. One expert is familiar with the location in subtask 1 and is responsible for reviewing subtask 1; the other expert is a photography fan and is responsible for reviewing subtask 2. Therefore, for each task, we have evaluations from both an average user and an expert.

For the first three questions, we quantized the answers to scores as follows: For relevance: very confident, somewhat confident, unconfident, and very unconfident are quantized to 4, 3, 2, and 1, respectively. For satisfaction: very satisfied, somewhat satisfied, unsatisfied, and very unsatisfied are quantized to 4, 3, 2, and 1, respectively. For quality: we give EnjoyPhoto scores 5,4,3,2,1 for the five answers, and meanwhile give Google scores 1,2,3,4,5. Based on the quantized scores, we calculate and list the statistical results and ANOVA test result for 12 participants in Table 2 and Table 3 for subtasks 1 and 2.

Table 2 shows that EnjoyPhoto significantly outperforms Google in terms of satisfaction (P-value < 0.05) and quality (P-value << 0.05). However, there is no apparent difference for relevance; with the mean value, Google received slightly higher marks than EnjoyPhoto (3.25 vs. 3.0). As discussed in the introduction, relevance is pretty good for most search engines and it is reasonable that the responses for relevance are similar. By means of satisfaction, we mean that the found images are both representative (relevant) and enjoyable. Based on these two criteria, EnjoyPhoto performs better than Google (P-value < 0.05). However, if we focus on quality, EnjoyPhoto shows significantly better performance (P-value << 0.05) as the found images are really artistic and attractive.

Table 2. Statistical results for subtask 1.

	Stat.	Relevance	Satisfaction	Quality
Google	Mean	3.25000	3.12500	2.00000
	Variance	0.63043	0.46196	0.95652
Enjoy Photo	Mean	3.00000	3.41667	4.00000
	Variance	0.52174	0.34058	0.95652
ANOVA	F(1,47)	4.05882	4.28517	25.09091
	P-value	0.05578	0.04986	0.00005

Table 3 shows that EnjoyPhoto significantly outperforms Google in terms of all three factors (the P-values for three factors are very supportive to this conclusion), as the task of searching images for creating a calendar explicitly emphasizes image quality. The reason that participants were not satisfied with Google for relevance is that they usually felt there were too many personal photos or graphics in the search result pages. Therefore, they had to spend more time to find the desired images. In addition, some participants didn't like that Google mixes high- and low-quality images together in the search results, which look annoying and are unattractive to continue searching.

Table 3. Statistical results for subtask 2.

	Stat.	Relevance	Satisfaction	Quality
Google	Mean	3.16667	3.04167	1.83333
	Variance	0.66667	0.56341	1.01449
Enjoy Photo	Mean	3.66667	3.70833	4.16667
	Variance	0.31884	0.30254	1.01449
ANOVA	F(1,47)	8.62500	10.51429	32.20000
	P-value	0.00741	0.00359	0.00001

After two mandatory subtasks, 8 out of 12 participants showed more interest in searching images and took one more optional task (three for optional task 1, one for optional task 2, and four for optional task 3). Overall, 10 out of 12 participants said that their searching experience in EnjoyPhoto is really interesting and pleasant.

This study shows that EnjoyPhoto is capable of finding relevant image materials, especially when quality is an important issue.

Task 2: Evaluating browsing interfaces

This task focuses on investigating and evaluating browsing interfaces for enjoyable or recreational purposes. We asked participants to evaluate three aspects in our UI design.

First, each participant was required to compare four image results browsing UIs after 10 minutes of practice. The four UIs are the Fisheye view (Figure 3c), the Force-Transfer view (Figure 3d), opening an image in a new window, and opening an image in the same window. The last two are actually very familiar to participants, as these are the traditional ways for browsing search results. Then the participants were required to answer the following questions:

1. Which UI do you prefer/like?
   Fisheye, Force-Transfer, New window, Same window
2. Why do you prefer such a UI? (multiple options)
   Convenient, Smooth, Cool, Speed

Second, for the Force-Transfer view, participants were required to evaluate different settings including the zoom-in ratio for the clicked image and the fade-out effect for non-clicked images. They then answer the following questions:

1. Which zoom-in ratio do you prefer? 2×, 3×, 4×
2. Do you prefer the fade out (50%) effect? Yes, No

Finally, we showed each participant the slideshow-authoring feature and asked if he/she liked this feature for enjoying high-quality photos and his/her comments to this feature.

Table 4. Votes for reasons of choosing an in-place browsing mode.

Reason	Convenient	Smooth	Cool	Speed
Votes	11	5	4	7

The participants' feedback was really encouraging. Ten participants chose the Force-Transfer view, and two participants chose the modified Fisheye view. Nobody chose the traditional image-result browsing UI. We summarize the reasons that they chose the in-place browsing mode in Table 4. It is noteworthy that almost all participants (11 out of 12) chose "convenient" as a reason. It indicates that there is much room for improvement in traditional image search and browsing UIs.

Although most image-result browsing UIs simply follow text-result browsing UIs, an image UI has its own unique and advantage properties. First, an image thumbnail is more informative than a text snippet. Second, it's very easy to generate image thumbnails in different resolutions or to enlarge images. The in-place browsing mode actually leverages these properties, and users feel this is more convenient. The transition effect of dynamically enlarging the selected image and moving other images presents users with smooth and cool experiences. Moreover, because right after the click the browsing client starts rendering the transition effect and meanwhile downloading the high-resolution image, the waiting time seems shorter than with a traditional UI. This technique is being widely used in more and more applications.

The responses to the zoom-in ratio preference seem rather varied and subjective. Four chose 2×, two chose 3×, four chose 4×, one chose 2× to 3×, and one didn't care about the ratio. One participant suggested that it would be perfect if the zoom-in ratio is customizable by a slider. We do think this is a great idea, as the slider adjusting could bring even cooler transition effects by dynamically adjusting the layout.

For the fade-out effect, 8 out of 12 participants provided positive feedback by choosing yes, while the other 4 didn't find it necessary because the enlarged image is already attractive. We could probably just set the fade-out effect as a configurable feature.

The feedback on the slideshow feature was actually better than what we expected. Although it is a simple idea, 11 participants really liked this feature. Several participants evaluated this as a "wow" feature and could not wait to use it. Because of the high photo quality, participants were excited by the idea of picking their favorite photos and playing them locally.

4.2.2 Overall comments

Besides collecting answers to the participants' specific questions, we also asked for participants' overall comments to the system. Many of them commented that the coverage is still limited compared to commercial search engines like Google's image search. Nevertheless, they all acknowledged that the photo qualities in EnjoyPhoto are generally much better than Google.

Because EnjoyPhoto is built as a vertical image search, it is not going to replace or compete with general image search engines, but it is designed to serve specific user needs. Furthermore, EnjoyPhoto can be combined with a general image search engine to avoid its shortcomings in coverage and improve the general image search engine's overall performance. For example, if a user submits a query to the general image search engine, the query will be first directed to EnjoyPhoto. If there are a sufficient number of returned images (e.g., $> 5,000$ images), this means that EnjoyPhoto has good coverage for this query, and the general image search engine will directly return EnjoyPhoto's results to the user. Otherwise, the general image search will return its own search results.

Another valuable comment is about browsing UIs. Although the in-place browsing mode is convenient for browsing, the traditional UI is still useful when a user needs to see the context of an image. We are investigating how to design the interface and operation for displaying the context of an image in Force-Transfer view. Simply put, we can directly display the original Web page containing this image but remove critiques and other non-related elements in that page. Artistically, as each image has rich metadata about the title, category, photographer, camera setting, and a short description, we can design a unified artistic view for photos from various photo forums.

5. CONCLUSIONS

We have built a vertical image search engine, EnjoyPhoto, by leveraging rich metadata from various photo forum web sites to meet the user's need to enjoy high-quality photos, which is impossible in traditional image search engines. To effectively rank photos coming from multiple forums, we proposed a novel rank fusion algorithm that elegantly utilizes duplicate photos to normalize rating scores. To further improve user experiences in enjoying photos, we designed an in-place browsing interface, as well as other interface features such as photo recommendations and slideshow authoring, enabled by rich metadata and rating information. Experimental results and user studies show that the proposed rank fusion algorithm is effective. Additionally, the database of 2.5-million high-quality images in EnjoyPhoto demonstrates a comparable performance with the database of 1 billion images in Google for queries related to location, nature, and daily life categories. Also, the in-place browsing interface—e.g., the Force-Transfer view—is evaluated to be much more convenient than traditional interfaces. We believe such a vertical search engine will make it possible not only for users to browse and enjoy high-quality photos, but also for general image search engines to improve their search performance.

6. REFERENCE

[1] Burges C., Shaked T., Renshaw E., *et al*. Learning to Rank Using Gradient Descent. In *Proc. of the 22nd International Conference on Machine Learning*, 2005, 89-96.

[2] Chen, Z., Liu, W.Y., Hu, C.H., Li, M.J., and Zhang H.J. A Web Image Search Engine. In *Proc. ACM SIGIR*, 2001.

[3] Dwork C., Kumar R., Naor M., and Sivakumar D. Rank Aggregation Methods for the Web. In *Proc. of the 10th International WWW Conference*, Hong-Kong, 2001.

[4] Fagin R., Kumar R., and Sivakumar D. Comparing Top K Lists. *SIAM Journal on Discrete Mathematics*, 17(1), 2003.

[5] Frankel, C., Swain, M.J., and Athitsos, V. WebSeer: An Image Search Engine for the World Wide Web. *Technical Report TR-94-14,* Computer Science Department, University of Chicago, 1996

[6] Freund Y., Iyer R., Schapire R. E., and Singer Y. An Efficient Boosting Algorithm for Combining Preferences. *Journal of Machine Learning Research*, 4(1), 2004, 933-969.

[7] Google image search; http://images.google.com.

[8] Huang, X., and Lai. W. Force-Transfer: A New Approach to Removing Overlapping Nodes in Graph Layout. In *Proc. The 25th Australian Computer Science Conference*, 2003.

[9] IMDB, Formula for calculating the top-rated 250 titles in imdb; http://www.imdb.com/chart/top.

[10] Joachims, T. Optimizing Search Engines using Clickthrough Data. In *Proc. 8th ACM SIGKDD*, 2002, 133-142.

[11] Li, Z.W., Xu, G., Li, M., Ma, W.Y., and Zhang H.J. Grouping WWW Image Search Results by Novel Inhomogeneous Clustering Method. In *Proc. 7th International Multimedia Modeling Conference*, 2005.

[12] Liu, H., Xie, X., Tang, X.O., Li, Z.W., and Ma, W.Y. Effective Browsing of Web Image Search Results. In *Proc. 6th ACM SIGMM International Workshop on Multimedia Information Retrieval* (*MIR 2004*), 2004.

[13] Nie Z., Zhang Y., Wen J.R., and Ma W.Y. Object-Level Ranking: Bringing Order to Web Objects. In *Proc. 14th International Conference on World Wide Web*, 2005.

[14] Page, L., Brin, S., Motwani, R. and Winograd, T. The PageRank Citation Ranking: Bringing Order to the Web. *Technical report*, Stanford Digital Libraries, 1998.

[15] PicSearch; http://www.picsearch.com.

[16] Robertson E., Walker S., and Jones S., Hancock-Beaulieu M.M., Gatford M. Okapi at TREC-3, In *Proc. Second Text Retrieval Conf.* (*TREC-3*), 1995.

[17] Sarkar, M., and Brown, M.H. Graphical Fisheye Views. *Communications of the ACM*, 1993.

[18] Savakis, A.E., Etz, S.P. and Loui, A.C. Evaluation of Image Appeal in Consumer Photography. In *Proc. SPIE Human Vision and Electronic Imaging V,* SPIE, 2000, 111-120.

[19] Sclaroff, S., Cascia, M.L. and Sethi, S. Unifying Textual and Visual Cues for Content-Based Image Retrieval on the World Wide Web. *Computer Vision and Image Understanding*, 75, 1/2 (1999):86-98.

[20] Smith, J.R. and Chang, S.F. WebSeek: Visually Searching the Web for Content. *IEEE MultiMedia*, 4, 3 (1997):12-20.

[21] Sullivan, D. Hitwise Search Engine Ratings. *Search Engine Watch Articles*, August 23, 2005; http://searchenginewatch.com/reports/article.php/3099931.

[22] Süsstrunk, S. and Winkler, S. Color Image Quality on the internet. In *Proc. IS&T/SPIE Electronic Imaging 2004: Internet Imaging V*, IS&T/SPIE, vol. 5304 (2004):118-131.

[23] Tong, H.H., Li, M.J., Zhang H.J., He, J.H., and Zhang C.S. Classification of Digital Photos Taken by Photographers or Home Users. In *Proc. Pacific-Rim Conference on Multimedia* (*PCM*), 2004, 198-205.

[24] Wang B., Li Z.W., Li M.J., and Ma W.Y. Large-Scale Duplicate Detection For Web Image Search, in *Proc. ICME*, 2006.

[25] Yahoo image search; http://images.yahoo.com.

[26] Yee, K.P., Swearingen, K., Li, K., and Hearst, M. Faceted Metadata for Image Search and Browsing. In *Proc. ACM SIGCHI*, 2003