VisualRank: Applying PageRank to Large-Scale Image Search

Yushi Jing, *Member*, *IEEE*, and Shumeet Baluja, *Member*, *IEEE*

Abstract—Because of the relative ease in understanding and processing text, commercial image-search systems often rely on techniques that are largely indistinguishable from text search. Recently, academic studies have demonstrated the effectiveness of employing image-based features to provide either alternative or additional signals to use in this process. However, it remains uncertain whether such techniques will generalize to a large number of popular Web queries and whether the potential improvement to search quality warrants the additional computational cost. In this work, we cast the image-ranking problem into the task of identifying "authority" nodes on an inferred visual similarity graph and propose *VisualRank* to analyze the visual link structures among images. The images found to be "authorities" are chosen as those that answer the image-queries well. To understand the performance of such an approach in a real system, we conducted a series of large-scale experiments based on the task of retrieving images for 2,000 of the most popular products queries. Our experimental results show significant improvement, in terms of user satisfaction and relevancy, in comparison to the most recent Google Image Search results. Maintaining modest computational cost is vital to ensuring that this procedure can be used in practice; we describe the techniques required to make this system practical for large-scale deployment in commercial search engines.

Index Terms—Image ranking, content-based image retrieval, eigenvector centrality, graph theory.

1 Introduction

Although image search has become a popular feature in many search engines, including Yahoo!, MSN, Google, etc., the majority of image searches use very little, if any, image information. Due to the success of text-based search of Web pages and, in part, to the difficulty and expense of using image-based signals, most search engines return images solely based on the text of the pages from which the images are linked. For example, to find pictures of the Eiffel Tower, rather than examining the visual content of the material, images that occur on pages that contain the term "Eiffel Tower" are returned. No image analysis takes place to determine relevance or quality. This can yield results of inconsistent quality. For example, the query "d80," a popular Nikon camera, returns good results, as shown in Fig. 1a. However, the query for "Coca Cola" returns mixed results; as shown in Fig. 1b, the expected logo or Coca Cola can/bottle is not seen until the fourth result. This is due in part to the difficulty in associating images with keywords and, in part, to the large variations in image quality and user perceived semantic content.

Our approach relies on analyzing the distribution of visual similarities among the images. The premise is simple: An author of a Web page is likely to select images that, from his or her own perspective, are relevant to the topic. Rather than assuming that every user who has a Web page relevant to the query will link to an image that every other user finds relevant, our approach relies on the combined preferences of many Web content creators. For example, in Fig. 1b, many of the images contain the familiar red Coca Cola logo. In some of the images, the logo is the main focus of the image, whereas, in others, it occupies only a small portion. Nonetheless, its repetition in a large fraction of the images returned is an important signal that can be used to infer a common "visual theme" throughout the set. Finding the multiple visual themes and their relative strengths in a large set of images is the basis of the image ranking system proposed in this study.

Our work belongs to the general category of Content-Based Image Retrieval (CBIR), an active research area driven in part by the explosive growth of personal photography and the popularity of search engines. A comprehensive survey on this subject can be found in [1]. Many systems proposed in the past [2], [3], [4], [5] are considered as "pure" CBIR systems—search queries are issued in the form of images and similarity measurements are computed exclusively from content-based signals. On the other hand, "composite" CBIR systems [6], [7] allow flexible query interfaces and a diverse set of signal sources, a characteristic suited for Web image retrieval as most images on the Web are surrounded by text, hyperlinks, and other relevant metadata. For example, Fergus et al. [7] proposed the use of "visual filters" to rerank Google image search results, bridging the gap between "pure" CBIR systems and text-based commercial search engines. These "visual filters" are learned from the top 1,000 search results via Parts-based probabilistic models [8], a form of Probabilistic Graphical Models (PGMs), to capture the higher order relationship among the visual features.

- *Y. Jing is with the Georgia Institute of Technology, Atlanta, and with Google Inc., Research Group, 1600 Amphitheater Parkway, Mountain View, CA 94043. E-mail: jing@google.com.*
- *S. Baluja is with Google Inc., Research Group, 1600 Amphitheatre Parkway, Mountain View, CA 94043. E-mail: shumeet@google.com.*

Manuscript received 23 Sept. 2007; revised 13 Feb. 2008; accepted 21 Apr. 2008; published online 12 May 2008.
Recommended for acceptance by J.Z. Wang, D. Geman, J. Luo, and R.M. Gray. For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMISI-2007-09-0626.
Digital Object Identifier no. 10.1109/TPAMI.2008.121.

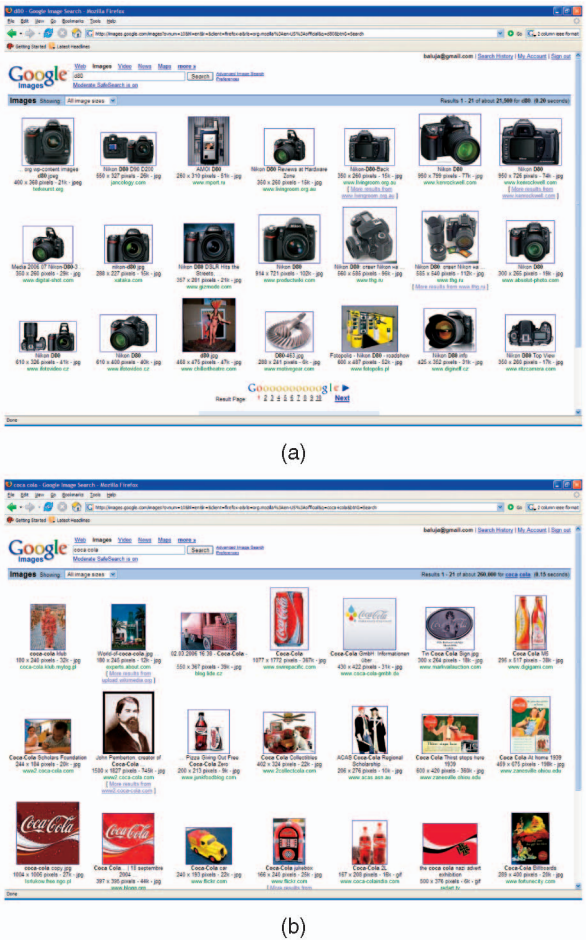


Fig. 1. The query for (a) "d80," a popular Nikon camera, returns good results on Google. However, the query for (b) "Coca Cola" returns mixed results.

However, PGMs have several important limitations for Web image retrieval. First, as generative models are factored according to the structures of the model, a suboptimal model structure can significantly reduce modeling and especially the classification performance [9]. An overly sparse model may neglect important higher order feature dependencies, while learning complex structures and their associated parameters is computationally prohibitive for large-scale Web image retrieval and is prone to data noise, especially given the nature of the Web images and the diverse visual representation of object categories. For example, Web queries with multiple visual concepts such as "Lincoln Memorial" and "Nemo" (shown in Fig. 2) can be particularly challenging [7] for Parts-based probabilistic models. Furthermore, there is an important mismatch between the goal of object category learning and image ranking. Object category learners are designed to model the relationship between features and images, whereas image search engines are designed to model the relationships (order) among images. Although a well-trained object category filter can improve the relevancy of image search results, they offer limited capability to directly control how and why one visual theme, or image, is ranked higher than others.

Fig. 2. Many queries like "Lincoln Memorial" (first two images) and "Nemo" (last three images) contain multiple visual themes.

A simpler and more intuitive alternative is to compute the pairwise visual similarity among images from which a global ordering can be derived and potentially combined with other nonvisual signals. Separating ranking from image representation has practical advantages for Web image retrieval. First, it gives search engine designers the flexibility to customize image similarities through domain engineering. For example, similarity computations that capture higher order feature dependencies[1] and learning techniques can be efficiently employed [11], [12], [13], [14]. Further, even nonvisual information, such as user-generated covisitation [15], [16] statistics, can be easily combined with visual features to make similarity scores more semantically relevant. Second, by separate ranking from the similarity measurement, one can also leverage well-understood ranking algorithms such as PageRank [17] to generate a global ordering given pairwise similarity scores. The simplicity and effectiveness of such approaches were demonstrated by He et al. [18], who first suggested combining PageRank with visual similarity for image retrieval, and which was later extended by Hsu et al. [19] for video retrieval and Joshi et al. [6] in the development of "Story Picturing Engine."

Our work extends [18], [19], [6]; we present *VisualRank*, an end-to-end system, to improve Google image search results with emphasis on robust and efficient computation of image similarities applicable to a large number of queries and images. *VisualRank* addresses new challenges that become apparent when tackling the diversity and magnitude of problems that arise with Web scale image retrieval. A few of the tasks and challenges addressed are described below.

First, current commercial search engines benefit from a variety of signals (i.e., hyperlinks analysis [17], [20]); many of these signals have proven to be at least as important as the content of the Web documents. This insight from text-based Web documents has motivated our approach of using graph analysis algorithms in addition to visual features. Nonetheless, the use of such systems in large-scale deployment must be carefully weighed with the benefits they provide due to the added complexity and computation. It is important for practical deployment to measure the results on the specific tasks addressed, for example, the number of clicks expected for each image if it is returned as a search result or how closely it matches the expectations of users of the final system. To measure the benefits, we conducted experiments with more

1. For example, geometric constraints [10] can be easily used in conjunction with local descriptors to reduce registration error.

than 2,000 popular commercial queries and 153 human evaluators; it is the largest experiment among the published works for content-based image ranking of which we are aware. Basing our evaluation on the most commonly searched for object categories, we demonstrate that VisualRank can significantly improve image search results for queries that are of the most interest to a large set of people.

Second, we propose a novel extension to previously proposed random-walk models that can take advantage of current progress in image-search and text-based Web search. Instead of combining two sets of rankings (visual and nonvisual) heuristically [7], we demonstrate how the order of placement from Google's image search can be used to bias the computation of the random-walk on the visual-similarity graph, thereby providing a direct integration of nonvisual and visual signals. Intuitively, by treating initial search results as Web documents and their visual similarities as probabilistic *visual hyperlinks*, VisualRank estimates the likelihood of each image being visited by search engine users following these visual hyperlinks, a score dependent on both the initial placement of the images and the collective visual similarities.[2]

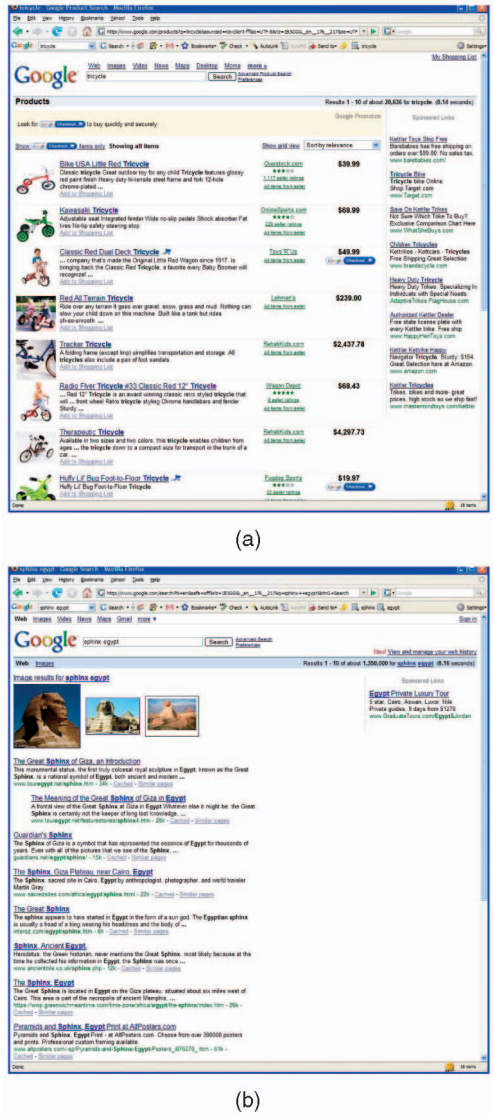
Third, given the scope and diversity of the queries considered, many of the previously commonly used features, such as variants of color histograms, provide unreliable measurements of similarity. We propose and study the use of local descriptor in our ranking framework. Computing local-feature-based pairwise similarity scores for billions of images is computationally infeasible, an important obstacle to overcome before any large-scale deployment of such a method. We use an efficient local descriptor hashing scheme (Locality Sensitive Hashing based on p-stable distributions) to alleviate the computational cost.

Fourth, we examine the performance of the system when adversaries are present. The order in which images are presented from commercial search engines carries great economic importance by driving Web traffic into publishers' Web sites. The economic incentives make commercial search engines targets for manipulative tactics from some publishers (i.e., Web spam and link farms). To address this issue, we also analyze how VisualRank performs in the presence of "spam" or intentionally distracting images.

The remainder of the paper is organized as follows: Section 2 completes the discussion of related work. Section 3 introduces the VisualRank algorithm and describes the construction of the visual similarity graph. Section 4 analyzes VisualRank's performance on queries with homogeneous/heterogeneous visual categories and under adversarial conditions, i.e., the presence of "spam" or distracting images. Section 5 describes an efficient implementation of an end-to-end Web image ranking system. Section 6 presents the experiments conducted and provides an analysis of the findings. Section 7 concludes the paper with a summary and suggestions for future work.

2 Related Work

Differently from pure CBIR systems [3], [5], [4], VisualRank retains the commonly used text query interface and utilizes the visual similarities within the entire set of images for image selection. This approach complements pure CBIR systems in several ways: 1) Text is still the most familiar and, often, the only query medium for commercial search engine users, 2) VisualRank can be effectively used in combination with other CBIR systems by generating a more relevant and diverse set of initial results, which often results in a better starting point for pure CBIR systems, and 3) there are real-world usage scenarios beyond "traditional" image search where image queries are not feasible. In many uses, we need to select a very small set of images to show from potentially millions of images. Unlike ranking, the goal is not to reorder the full set of images but to select only the "best" ones to show. Two concrete usage cases for this are the following: 1) *Google product search*: Only a single image is shown for each product returned in response to a product query, shown in Fig. 3a. 2) *Mixed-Result Search*: To indicate that image results are available when a user performs a Web (Web page) query, a small set of representative images may also be shown to entice the user into trying the image search, as shown in



(a) (b)

Fig. 3. In many uses, we need to select a very small set (1-3) of images to show from potentially millions of images. Unlike ranking, the goal is not to reorder the full set of images but to select only the "best" ones to show. (a) Google product search. (b) Mixed-Result-Type Search.

2. In fact, the navigational hyperlinks in commercial CBIR systems can be considered as such visual hyperlinks.

Fig. 3b. In both of these examples, it is paramount that the user is not shown irrelevant or off-topic, images. Finally, it is worth noting that as good similarity functions are the foundation of CBIR systems, VisualRank can easily incorporate advances in other CBIR systems.

The recently proposed affinity propagation algorithm [21] also attempts to find the most representative vertices in a graph. Instead of identifying a collection of medoids in the graph, VisualRank differs from affinity propagation by explicitly computing the ranking score for all images. Several other studies have explored the use of a similarity-based graph [22], [23] for semisupervised learning. Given an adjacency matrix and a few labeled vertices, unlabeled nodes can be described as a function of the labeled nodes based on the graph manifolds. In this work, our goal is not classification; instead, we model the centrality of a graph as a tool for ranking images. Another related work is by Zhu et al. [23], who propose using a random-walk model on graph manifolds to generate "smoothed" similarity scores that are useful in ranking the rest of the images when one of them is selected as the query image. Our approach differs from that in [23] by generating an a priori ranking given a group of images.

Our work is closely related to [7], as both explore the use of content-based features to improve commercial image search engine. Random-walk-based ranking algorithms were proposed in [18], [19], [6] for multimedia information retrieval; a detailed comparison to these approaches was given in the previous section. Our work is also an extension of that in [24] in which image similarities are used to find a single most representative or "canonical" image from image search results. The "canonical" images are selected as the most densely connected node in the graph. In this work, we use well-understood methods for graph analysis based on PageRank and provide a large-scale study of both the performance and computational cost of such a system.

3 Visual Rank

3.1 Eigenvector Centrality and VisualRank

Given a graph with vertices and a set of weighted edges, we would like to measure the importance of each vertex. The cardinality of the vertices or the sum of geodesic distance to the surrounding nodes are all variations of centrality measurement. Eigenvector Centrality provides a principled method to combine the "importance" of a vertex with those of its neighbors in ranking. For example, other factors being equal, a vertex closer to an "important" vertex should rank higher than others that are further away.

Eigenvector Centrality is defined as the principle eigenvector of a square stochastic adjacency matrix, constructed from the weights of the edges in the graph. It has an intuitive Random Walk explanation: The ranking scores correspond to the likelihood of arriving in each of the vertices by traversing through the graph (with a random starting point), where the decision to take a particular path is defined by the weighted edges.

VisualRank employs the Random Walk intuition to rank images based on the visual hyperlinks among the images. The intuition of using these visual hyperlinks is that if a user is viewing an image, other related (similar) images may also be of interest. In particular, if image u has a visual hyperlink to image v, then there is some probability that the user will jump from u to v. Intuitively, images related to the query will have many other images pointing to them and will therefore be visited often (as long as they are not isolated and in a small clique). The images that are visited often are deemed important. Further, if we find that an image, v, is important and it links to an image w, it is casting its vote for w's importance because v is in itself important; the vote should count more than a "nonimportant" vote.

VisualRank (VR) is iteratively defined as the following:

$$VR = S^* \times VR. \tag{1}$$

S^* is the column normalized adjacency matrix S, where $S_{u,v}$ measures the visual similarity between image u and v. Repeatedly multiplying VR by S^* yields the dominant eigenvector of the matrix S^*. Although VR has a fixed-point solution, in practice, it can often be estimated more efficiently through iterative approaches.

VisualRank converges only when matrix S^* is aperiodic and irreducible. The former is generally true for the Web and the latter usually requires a strongly connected graph, a property guaranteed in practice by introducing a damping factor d into (1). In the study presented in this paper, the similarity matrix, S, is symmetric.[3] In cases in which the similarity matrix is symmetric, it should be noted that the use of many forms of damping factors can make the effective modified similarity matrix asymmetric.

Given n images, VR is defined as

$$VR = dS^* \times VR + (1-d)p, \quad \text{where } p = \left[\frac{1}{n}\right]_{n\times 1}. \tag{2}$$

This is analogous to adding a complete set of weighted outgoing edges for all the vertices. Intuitively, this creates a small probability for a random walk to go to some other images in the graph, although it may not have been initially linked to the current image. $d > 0.8$ is often chosen for practice; empirically, we have found the setting of d to have relatively minor impact on the global ordering of the images.

In place of the uniform damping vector p in (2), we can use a nonuniform vector q to bias the computation. For example, we can use it to increase the effect of images ranked high in the initial search engine results since they are selected, albeit through nonvisual features, to be the best match to the query. Vector q can be derived from image quality, anchor page quality, or simply the initial rank from commercial search engines. The intuition is that "random surfers" are more likely to visit and traverse through images that have higher prior expectation of being relevant. For example, if we assume the top m search results from commercial search engines to be of reasonable quality, we can use $q = v_j$, where

$$v_j = \begin{cases} \frac{1}{m}, & j \le m \\ 0, & otherwise. \end{cases} \tag{3}$$

3. Note that the similarity matrix does not necessarily need to be symmetric. For example, consider the case in which one image is an enlarged portion of another image (for example, a close-up of the Mona Lisa); when the scale and/or area of the matching region is considered, the potential advantages of nonsymmetric similarity measures becomes evident.

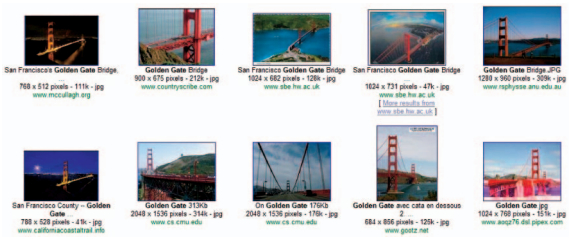


Fig. 4. Similarity measurement must handle potential rotation, scale, and perspective transformations.

As an example of a successful application of Eigenvector Centrality, PageRank [17] precomputes a rank vector to estimate the importance for all of the Web pages on the Web by analyzing the hyperlinks connecting Web documents.[4] Intuitively, pages on Amazon.com are important, with many pages pointing to them. Pages pointed to by Amazon.com may therefore also have high importance. Nonuniform damping vectors were suggested previously by Haveliwala [25] to compute topic-biased PageRank for Web documents.

3.2 Features Generation and Representation

A reliable measurement of image similarity is crucial to the performance of VisualRank since this determines the underlying graph structure. Global features like color histograms and shape analysis, when used alone, are often too restrictive for the breadth of image types that need to be handled. For example, as shown in Fig. 4, the search results for "Golden Gate" often contain images taken from different locations, with different cameras, focal lengths, compositions, etc.

Compared with global features, local descriptors contain a richer set of image information and are relatively stable under different transformations and, to some degree, lighting variations. Examples of local features include Harris corners [26], Scale Invariant Feature Transform (SIFT) [10], Shape Context [27], and Spin Images [28], to name a few. Mikolajczyk and Schmid [29] presented a comparative study of various descriptors; [30], [31] presented work on improving their performance and computational efficiency. In this work, we use the SIFT features, with a Difference of Gaussian (DoG) interest point detector and orientation histogram feature representation as image features. However, any of the local features could have been substituted.

We used a standard implementation of SIFT; for completeness, we give the specifics of our usage here. A DoG interest point detector builds a pyramid of scaled images by iteratively applying Gaussian filters to the original image. Adjacent Gaussian images are subtracted to create DoG images, from which the characteristic scale associated with each of the interest points can be estimated by finding the local extrema over the scale space. Given the DoG image pyramid, interest points located at the local extrema of 2D-image space and scale space are selected. A gradient map is computed for the region around the interest point and then divided into a collection of subregions from which an orientation histogram can be computed. The final descriptor is a 128-dimensional vector by concatenating 4×4 orientation histogram with eight bins. Given two images, we define their similarity as the number of local features shared between them divided by their average number of interest points.

4. The PageRank vector can be precomputed and can be independent of the search query. Then, at query time, PageRank scores can be combined with query-specific retrieval scores to rank the query results. This provides a faster retrieval speed than many query-time methods [20].

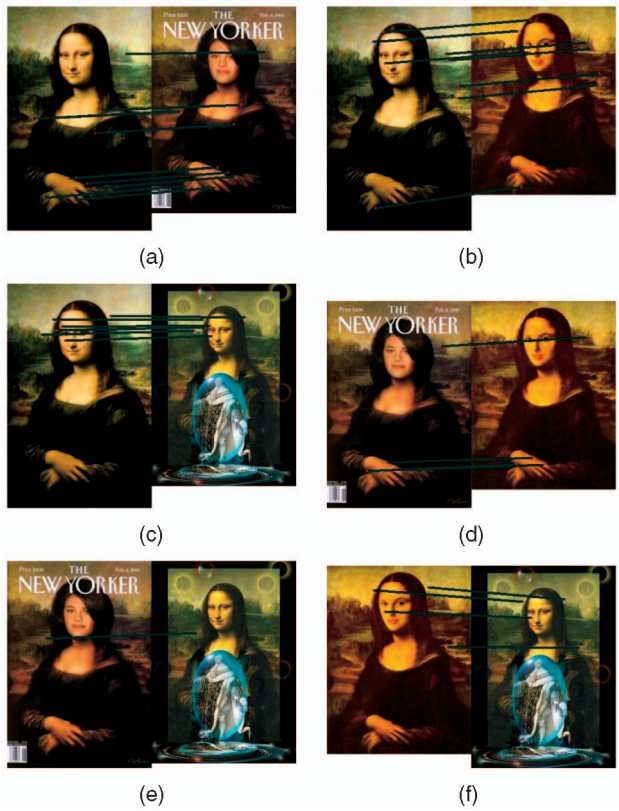


Fig. 5. Since all of the variations (B, C, D) are based on the original painting (A), A contains more matched local features than others. (a) A versus B. (b) A versus C. (c) A versus D. (d) B versus C. (e) B versus D. (f) C versus D.

4 Applying VisualRank

The goal of image-search engines is to retrieve image results that are relevant to the query and diverse enough to cover variations of visual or semantic concepts. Traditional search engines find relevant images largely by matching the text query with image metadata (i.e., anchor text and surrounding text). Since text information is often limited and can be inaccurate, many top ranked images may be irrelevant to the query. Further, without analyzing the content of the images, there is no reliable way to actively promote the diversity of the results. In this section, we will explain the intuition behind how VisualRank can improve the relevancy and diversity of image search results.

4.1 Queries with Homogeneous Visual Concepts

VisualRank improves the relevance of image search results under queries with homogeneous visual concepts. This is achieved by identifying the vertices that are located at the "center" of weighted similarity graph. "Mona-Lisa" is a good example of a search query with a single homogeneous visual concept. Although there are many comical variations (i.e., "Bikini-lisa" and "Monica-Lisa"), they are all based on the original painting. As shown in Fig. 5, the original

Fig. 6. Similarity graph generated from the top 1,000 search results of "Mona-Lisa." The largest two images contain the highest VisualRank.

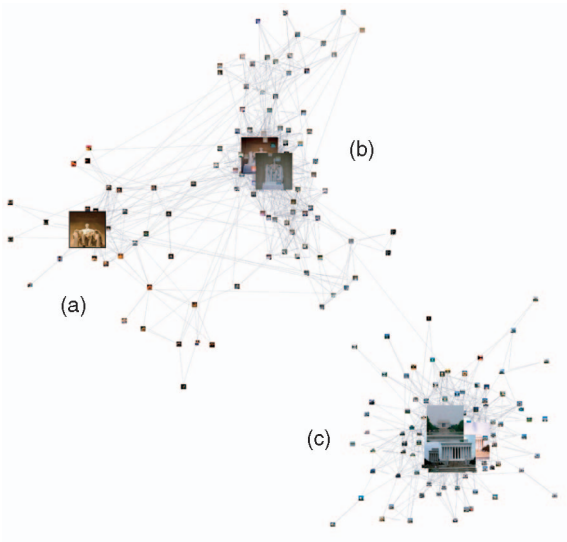


Fig. 7. The top 10 images selected by VisualRank from the 1,000 search results of "Lincoln Memorial." By analyzing the link structure in the graph, VisualRank identifies a highly relevant yet diverse set of images. (a) Night-time photo of the Lincoln statue. (b) Daytime photo of the statue. (c) Lincoln Memorial building.

painting contains more matched local features than others and, thus, has the highest likelihood of visit by an user following these probabilistic visual hyperlinks. Fig. 6 is generated from the top 1,000 search results of "Mona-Lisa." The graph is very densely connected, but, not surprisingly, the centers of the images all correspond to the original version of the painting.

4.2 Queries with Heterogeneous Visual Concepts

VisualRank can improve the relevancy and diversity of queries that contain multiple visual concepts. Examples of such queries that are often given in the information retrieval literature include "Jaguar" (car and animal) and "Apple" (computer and fruit). However, when considering images, many more queries also have multiple canonical answers. For example, the query "Lincoln Memorial," shown in Fig. 7, has multiple good answers (pictures of the Lincoln statue, pictures of the building, etc.). In practice, VisualRank is able to identify a relevant and diverse set of images as top ranking results; there is no a priori bias toward a fixed number of concepts or clusters.

An interesting question that arises is whether simple heuristics could have been employed for analyzing the graph, rather than using a VisualRank/Eigenvector approach. For example, a simple alternative is to select the high degree nodes in the graph, as this implicitly captures the notion of well-connected images. However, this fails to identify the different distinctive visual concepts, as shown in Fig. 8. Since there are more close matches of "Lincoln statue," they reinforce each other to form a strongly connected clique. Further, the random-walk model also accounts for distracting or "spam" images, as will be shown in the next section. Of course, measures can be added to detect these cases; however, VisualRank provides a principled and intuitive method, through a simple fixed-point computation, to capture these insights.

4.3 Performance in the Presence of Distracting Images

Visual similarity among images offers tremendous information about the popularity and relevance of a particular image. However, it can be susceptible to manipulations. For example, in commercially deployed systems, adversary content creators can inflate the rankings of their own images by placing a large quantity of duplicate images on the Web.[5] Although those images may bring additional traffic to their Web site, users may not find them helpful. This practice is analogous to "Link Spam" on the Web, where artificially constructed densely connected Web pages are used to inflate their rankings in regular search engines.

Even in its straightforward implementation, VisualRank is resistant to many forms of similarity link spam by analyzing the global structure of the graph. For example, the top 1,000 images collected with query "Nemo" contained many near/exact-duplicated images of "Nemo Sushi," shown in Fig. 9c. Note that these images reinforce each other. Simpler algorithms, such as selecting nodes with high degree, are easily misled, as shown in Fig. 9b. The use of the damping vector was found to be crucial for the improved performance using VisualRank.

5 Web-Scale Visual Rank System

In this section, we describe our work on scaling VisualRank to work on a large set of real queries.

5.1 Query Dependent Visual Rank

It is computationally infeasible to generate the similarity graph S for the billions of images that are indexed by

5. Note that "adversary" is meant literally; it is common practice for content creators to submit many duplicate or near-duplicate images, Web pages, etc., intentionally designed to bias ranking algorithms to place their content above others.

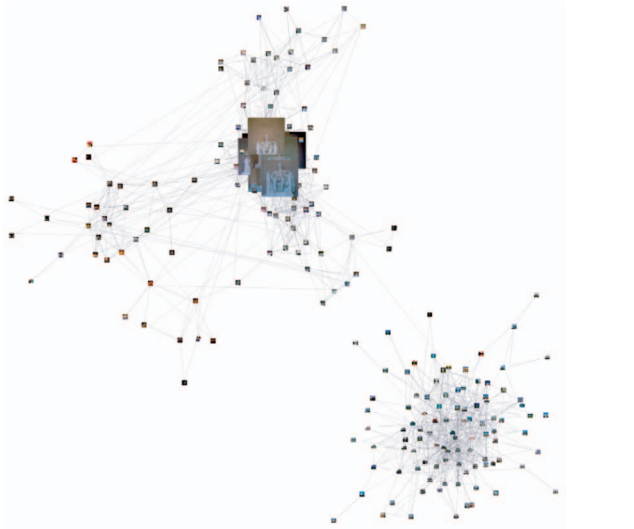


Fig. 8. Alternative method of selecting images with the most "neighbors" tend to generate a relevant but homogeneous set of images.

commercial search engines. One method to reduce the computational cost is to precluster Web images based on using metadata such as text, anchor text, similarity, or connectivity of the Web pages on which they were found. For example, images associated with "Paris," "Eiffel Tower," and "Arc de Triomphe" are more likely to share similar visual features than random images. To make the similarity computations more tractable, a different VisualRank can be computed for each group of such images.

A practical method to obtain the initial set of candidates mentioned in the previous paragraph is to rely on the existing commercial search engine for the initial grouping of semantically similar images. For example, similarly to [7], given the query "Eiffel Tower," we can extract the top-N results returned, create the graph of visual similarity on the N images, and compute VisualRank only on this subset. In this instantiation, VisualRank is query dependent; although the VisualRank of images in the N images is indicative of their importance in answering the query, the same image may have a different score when it is a member of a different set of images that is returned in response to a different query. In the experiment section, we follow this procedure on 2,000 of the most popular queries for Google Product Search.

5.2 Hashing Local Features via p-Stable Distributions

The similarity matrix S^* from (1) can be computed for all unique pairs of images. A more efficient approach is to use a hash table to store all the local descriptors such that similar descriptors fall into the same bin.[6] In the extreme case where only exact duplicates are considered as matches, one can simply use the original descriptor value as the hash key (by converting the 128-dimensional vector into a single long hash key). To match the "similar" non-exact-duplicate local descriptors under different lighting conditions and other variations, a more relaxed distance preserving hashing function can be used.

6. Due to the memory requirement, hashing is practical only for a limited number of local features.

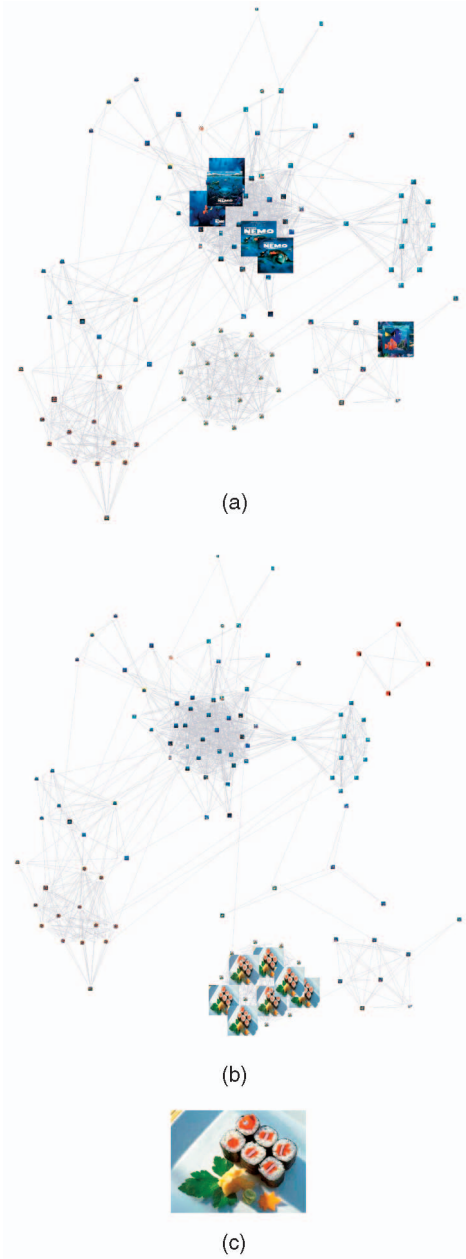


Fig. 9. By analyzing the global structure of the graph, (a) VisualRank avoids selecting images simply because they are close duplicates of each other. The alternative methods of selecting images with (b) high degree is susceptible to this as it finds the (c) spam images repeatedly.

Matching local descriptors efficiently has received tremendous research attention in recent years [32], [33], [34], [35], [36], [37]. In particular, Nister and Stewénius [32] proposed the use of "visual vocabularies," a set of distinctive quantitized local descriptors learned via hierarchical k-mean clustering. Raw features are mapped into visual vocabularies by traversing down the vocabulary tree to find the closest leaf. This process can be viewed as constructing a hash function to map raw descriptors into a key, in this case, the visual vocabulary.

For our algorithm, approximation methods [38], [39], [37] to measure similarity are sufficient. Because VisualRank

relies on the global structure of the graph to derive its ranking, it is already quite robust against noise (mismatch or missed matches of local features in images). Intuitively, if the distance measurement captures an overall notion of user perceived similarity, a small difference in the magnitudes of the distance will have negligible effect on the end results. We will use a version of the Locality-Sensitive Hashing (LSH) approximate matching approach.

LSH is an approximate kNN technique introduced by Indyk et al. [39]. LSH addresses the similarity match problem, termed $(r;\epsilon)$-NN, in sublinear time. The goal is formally stated as follows: Given a point q (query) in a d-dimensional feature space, for exact kNN, for any point q, return the point p that minimizes $D(p;q)$. For approximate kNN, if there exists an indexed point p such that $D(p;q) \leq r$, then with high-probability return an indexed point that is of distance at most $(1+\epsilon)r$. If no indexed point lies within $(1+\epsilon)r$ of q, then LSH should return nothing, with high probability. Ke et al. [33] have explored LSH in the task of near-duplicate image detection and retrieval and obtained promising results. The particular hash function in [33] was best suited for the preservation of Hamming distance; for our work, we follow the recent work of Datar et al. [38]. Datar et al. [38] have proposed hash function for l_2 norms, based on p-stable distributions [40]. Here, each hash function is defined as

$$h_{a,b}(V) = \left\lfloor \frac{aV+b}{W} \right\rfloor, \tag{4}$$

where a is a d-dimensional random vector with entries chosen independently from a Gaussian distribution and b is a real number chosen uniformly from the range [0, W]. W defines the quantization of the features and V is the original feature vector. Equation (4) is very simple to implement and efficient.

In practice, the best results are achieved by using L number of hash tables rather than a single one. For each hash table, we reduce the collision probability of nonsimilar objects by concatenating K hash functions. Two features are considered as a match if they were hashed into the same bin in C out of the L hash tables; effectively, this provides a means of setting a minimum match threshold, thereby eliminating coincidental matches that occur in only a few of the tables. We group all of the matched features by their associated image and the similarity matrix, S, is computed by the total number of matches normalized by their average number of local features. The exact parameter settings are given below.

5.3 Summary of the System

The VisualRank system can be summarized as the following four steps. A visual representation of the process is given in Fig. 10.

1. Local features are generated for a group of images, scaled to have a maximum axis size of 500 pixels. From our study, 1,000 Web images usually contain 300,000 to 700,000 feature vectors.
2. A collection of L hash tables $H = H_1, H_2, \ldots, H_L$ is constructed, each with K number of hash functions, as shown in (4). Each of the descriptors is indexed into each of the hash tables. Empirically, we determined that $L = 40$, $W = 100$, and $K = 3$ give good results.
3. For each descriptor, we aggregate objects with identical hash keys across L hash tables. Descriptors that share the same key in more than C hash tables are considered a match $(C = 3)$.
4. We regroup matched features by the images they are associated with. Optionally, for two images and their associated matching feature points, we use a Hough Transform to enforce a loose geometric consistency. A four-dimensional histogram is used to store the "votes" on the pose space (translation, scaling, and rotation). At the end, we select the histogram entry with the most votes as the most consistent interpretation. The surviving matching points are used to compute the similarity score.
5. A pair of images is considered a match if they share more than three matched descriptors. The similarity of two images is computed by the total number of matches normalized by their average number of local features.
6. Given similarity matrix S, we can use the VisualRank algorithm to generate the top N images.

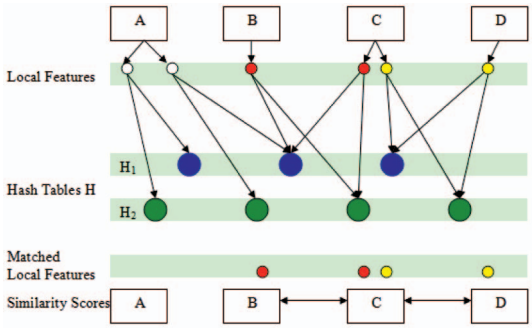


Fig. 10. A visual representation of our hashing scheme. Local features extracted from a collection of images are hashed into a collection of LSH hash tables. Features hashed into the same bin in multiple hash families are considered matches and contribute to the similarity score between their corresponding images.

With the techniques mentioned above and nonoptimized code, it takes approximately 15 minutes to compute and hash the local descriptors for 1,000 images and to compute the full similarity matrix. Although this is a significant computational requirement, it allows us to precompute the results to many popular queries. For example, with 1,000 modest CPUs, the VisualRank for the top 100,000 queries can be computed in less than 30 hours.

6 Experimental Results

To ensure that our algorithm works in practice, we conducted experiments with images collected directly from the Web. In order to ensure that the results would make a significant impact in practice, we concentrated on the 2,000 most popular product queries[7] on Google (product search). Typical queries included "ipod," "Xbox," "Picasso,"

7. The most often queried keywords during a period in August.

TABLE 1
Relevancy Study

"Irrelevant" images per product query	VisualRank	Google
Among top 10 results	0.47	2.82
Among top 5 results	0.30	1.31
Among top 3 results	0.20	0.81

"Fabreze," etc.[8] For each query, we extracted the top 1,000 search results from Google image search in July 2007, with the strict safe search filter. The similarity matrix is constructed by counting the number of matched local features for each pair of images after geometric validation normalized by the number of descriptors generated from each pairs of images.

We expect that Google's results will already be quite good, especially since the queries chosen are the most popular product queries for which many relevant Web pages and images exist. Therefore, we would like to only suggest a refinement to the ranking of the results when we are certain that VisualRank will have enough information to work correctly. A simple threshold was employed if, in the set of 1,000 images returned, fewer than 5 percent of the images had at least one connection, VisualRank was not used. In these cases, we assumed that the graph was too sparse to contain enough information. After this pruning, we concentrated on the approximately 1,000 remaining queries.

It is challenging to quantify the quality of (or difference of performance) of sets of image search results for several reasons. First and foremost, user preference to an image is heavily influenced by a user's personal tastes and biases. Second, asking the user to compare the quality of a *set* of images is a difficult and often time-consuming task. For example, an evaluator may have trouble choosing between group A, containing five relevant but mediocre images, and group B, which is mixed with both great and bad results. Finally, assessing the differences in ranking (when many of the images between two rankings being compared are the same) is error prone and imprecise at best. Perhaps the most principled way to approach this task is to build a global ranking based on pairwise comparisons. However, this process requires a significant amount of user input and is not feasible for large numbers of queries.

To accurately study the performance of VisualRank subject to practical constraints, we devised two evaluation strategies. Together, they offer a comprehensive comparison of two ranking algorithms, especially with respect to how the rankings will be used in practice.

6.1 Experiment 1: User Study on Retrieval Relevancy

This study is designed to study a conservative version of the "relevancy" of our ranking results. For this experiment, we mixed the top 10 VisualRank selected images with the top 10 images from Google, removed the duplicates, and presented them to the user. We asked the user: "Which of the image(s) are least relevant to the query?" For this experiment, more than 150 volunteer participants were chosen and were asked this question on a set of randomly chosen 50 queries selected from the top-query set. There was no requirement on the number of images that they marked.

8. We chose product-related (and travel/landmark) queries for three reasons. First, they are extremely popular in actual usage. Second, they lend themselves well to the type of local feature detectors that we selected in this study (in Section 7, we describe other categories of queries that may benefit from alternative sets of image features). Third, users have strong expectations of what results we should return for these queries; therefore, this provides an important set of examples that we need to address carefully.

TABLE 2
Relevance Comparison per Query

	VisualRank	Google
Outperforming product queries	762	70

There are several interesting points to note about this study. First, it does not ask the user to simply mark relevant images; the reason for this is that we wanted to avoid a heavy bias to a user's own personal expectation (i.e., when querying "Apple," did they want the fruit or the computer?). Second, we did not ask the users to compare two sets since, as mentioned earlier, this is an arduous task. Instead, the user was asked to examine each image individually. Third, the user was given no indication of ranking, thereby alleviating the burden of analyzing image ordering.

It is also worth noting that minimizing the number of irrelevant images is important in real-world usage scenarios beyond "traditional" image search. As mentioned earlier, in many uses, we need to select a very small set of images (1-3) to show from potentially millions of images. Two concrete usage cases that were mentioned earlier include selecting images for Google Product Search and Google Mixed-type search, as shown in Fig. 3.

In order to quantify the effectiveness of visual features, VisualRank was computed with a uniform bias vector, ignoring order of placement in the original search results. We measured the results for Google and VisualRank for three settings: the number of *irrelevant* images in the top 10, top 5, and top 3 images returned by each of the algorithms. Table 1 contains the comparison results. Among the top 10 images, VisualRank produced an average of 0.47 irrelevant results; this is compared with the 2.82 by Google; this represents an 83 percent drop in irrelevant images. When looking at the top 3 images, the number of irrelevant images for VisualRank dropped to 0.20, while Google dropped to 0.81.

In terms of overall performance on queries, as shown in Table 2, VisualRank contains fewer irrelevant images than Google for 762 queries. In only 70 queries did VisualRank produce worse results than Google. In the remaining 202 queries, VisualRank and Google tied (in the majority, of these, there were no irrelevant images). Fig. 11 provides a query-by-query analysis between VisualRank and existing Google image search. The Y axis contains the number of "irrelevant" images and the X axis lists the type of queries. The order of queries are sorted by the number of "irrelevant" images retrieved by the Google image search engine for better visualization.

To present a complete analysis of VisualRank, we describe two cases where VisualRank did not perform as expected. VisualRank sometimes fails to retrieve relevant images, as shown in Fig. 12. The first three images are the logos of the company that manufactured the product being searched for. Although the logo is somewhat related to the

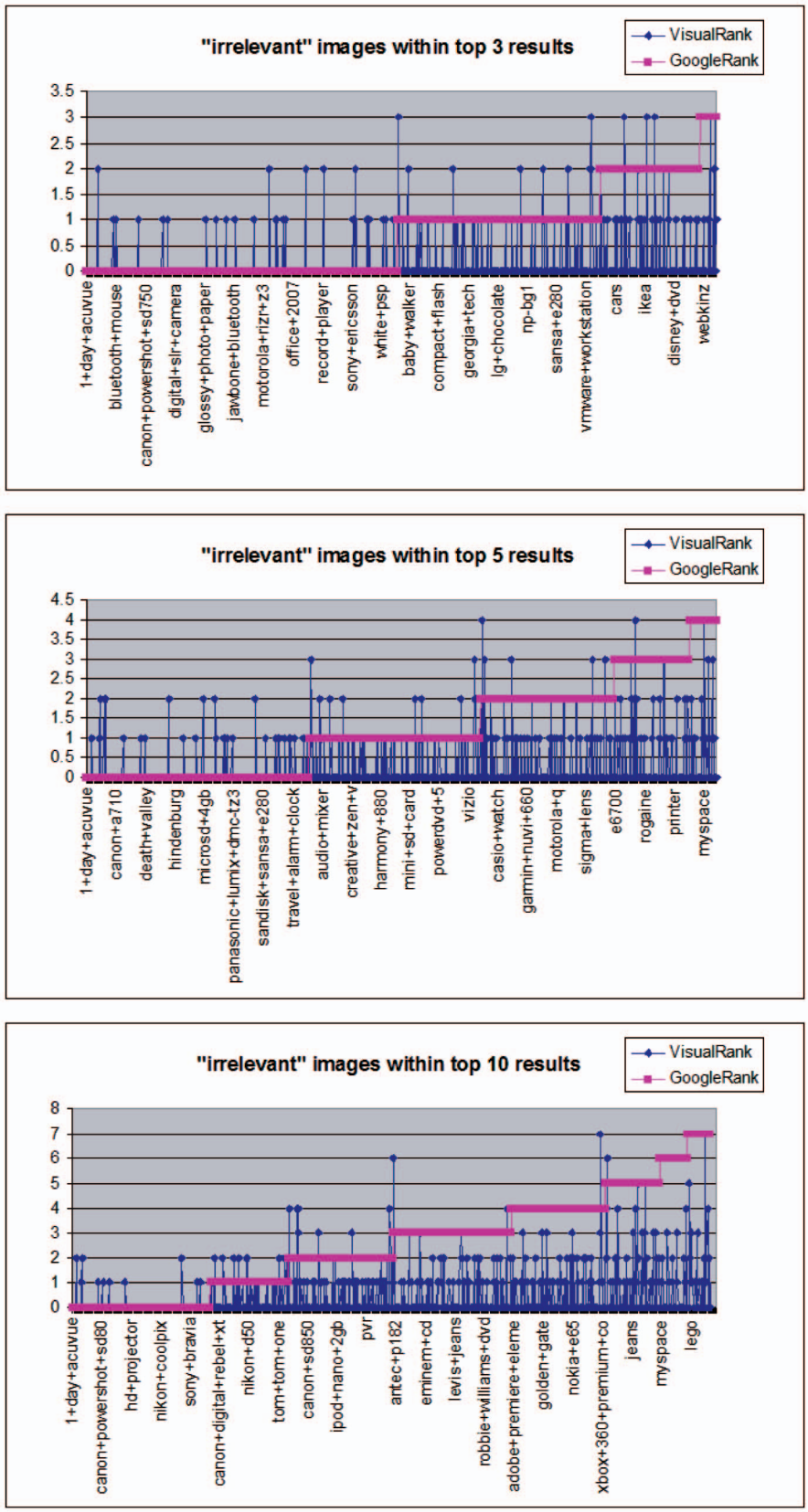


Fig. 11. Number of irrelevant images (per query) retrieved by competing algorithms. For visual clarity, a subsample of corresponding queries (cars, etc.) is shown under the x-axis. The queries are sorted by the number of irrelevant results retrieved by the Google image search.

query, the evaluators did not regard them as relevant to the specific product for which they were searching. The inflated logo score occurs for two reasons. First, many product images contain the company logos, either within the product itself or in addition to the product. In fact, extra care is often taken to make sure that the logos are clearly

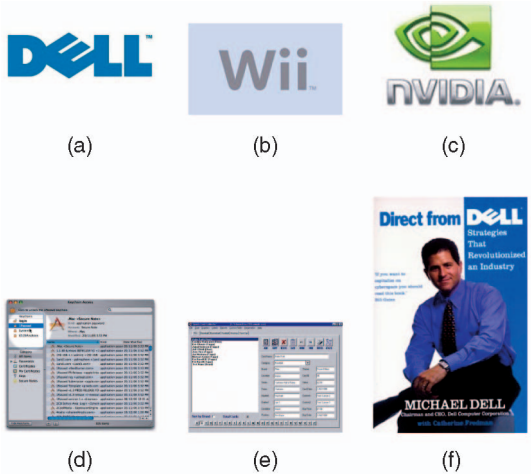


Fig. 12. The particular local descriptors used provided a bias to the types of patterns found. These VisualRank selected images received the most "irrelevant" votes from the users for the queries shown. (a) Dell computer. (b) Nintendo Wii system. (c) 8800 Ultra. (d) Keychain. (e) 2PS2 network adapter. (f) Dell computer.

visible, prominent, and uniform in appearance. Second, logos often contain distinctive patterns that provide a rich set of local descriptors that are particularly well suited to SIFT-like feature extraction.

A second, but less common, failure case is when screen shots of Web pages are saved as images. Many of these images include browser panels or Microsoft Windows control panels that are consistent across many images. It is suspected that these mismatches can easily be filtered by combining VisualRank with other sources of quality scores or measuring the distinctiveness of the features not only within queries but also across queries, in a manner similar to using TF-IDF [41] weighting in textual relevancy. In fact, as shown in the next sections, some of the mismatches can be easily filtered by biasing the computation of VisualRank with the initial order of placement from Google image search results.

6.2 Experiment 2: Additional Comparison with Alternative Methods

As mentioned earlier, in Section 3.1, it is possible to extend the VisualRank algorithm to take into account the information conveyed in the order of results returned from Google's image search. Google presents the most relevant images in the order of the most relevant to the least relevant, based on its own internal metrics. The next set of experiments provides a small example of how this ordering information can be incorporated into the random-walk model. Instead of the uniform damping factor that is commonly used in PageRank systems, we use a nonuniform damping vector in (3) to increase the effect of images highly ranked in the search engine results as these images were selected (by Google, with nonvisual features) as the best match to the query. We call the resulting algorithm $\mathrm{VisualRank}_{bias}$. Table 3 contains the set of manually computed comparison results.[9] Among the top 10 images, $\mathrm{VisualRank}_{bias}$ produces an average of 0.17 irrelevant results, compared with 0.47 by VisualRank with a uniform damping vector; this represents a 64 percent drop in irrelevant images. When looking at the top 3 images, the number of irrelevant images for $\mathrm{VisualRank}_{bias}$ dropped to 0.04, while VisualRank dropped to 0.20. Through a deeper analysis of the results, it was determined that $\mathrm{VisualRank}_{bias}$ does not contain the type of "irrelevant" images shown in Figs. 12d and 12e.

Finally, to compare against a pure CBIR system, we formulate an alternative heuristic approach, named HeuristicRank. HeuristicRank retains the top 10-15 Google search results and uses their closest visual neighbors (with the metric defined in Section 3.2) as the next set of best images. Those images that do not have a visually similar neighbor are considered noise and are removed from the ranking. The top 20 resulting images are compared with VisualRank, using the rank given by Google's search results. In this experiment, we also manually evaluated the images generated by the two competing algorithms. The results are reported in Table 3.

Table 4 demonstrates that VisualRank outperforms HeuristicRank in 664 queries, while it was outperformed by HeuristicRank in 182 queries. Upon further analysis of the results, we find that the accuracy of HeuristicRank highly depends on the top search results. Although HeuristicRank can remove some outliers by eliminating images without a similar neighbor, there are many "irrelevant" images, like the "Nemo Sushi" shown in Fig. 9c, with near duplicates among the top 1,000 search results.

6.3 Experiment 3: Satisfaction and Click Measurement

Results from Experiment 1 show that VisualRank can effectively decrease the number of irrelevant images in the search results. However, user satisfaction is not purely a function of relevance; for example, numerous other factors, such as the diversity of the selected images, must also be considered. Assuming the users usually click on the images they are interested in, an effective way to measure search quality is to analyze the total number of "clicks" each image receives.

We collected clicks for the top 40 images (first two pages) presented by the Google search results on 130 common product queries. The VisualRank for the top 1,000 images for each of the 130 queries is computed and the top 40 images are reranked using VisualRank. To determine if the ranking would improve performance, we examine the number of clicks each method received from only the top 20 images (these are the images that would be displayed in the first page of results (on http://images.google.com)). The hope is that, by reordering the top 40 results, the best images will move to the top and will be displayed on the first page of results. If we are successful, then the number of clicks for the top 20 results under reordering will exceed the number of clicks for the top 20 under the default ordering.

It is important to note that this evaluation contains an *extremely severe bias that favors the default ordering*. The ground truth of clicks an image receives is a function not only of the relevance to a query and quality of the image *but also of the position in which it is displayed*. For example, it is often the case that a mediocre image from the top of the

9. For the following experiment, another large-scale user study was beyond the resources available. The results are manually computed to the best of our abilities. Nonetheless, a complete user study is left for future research.

TABLE 3
Relevancy Study

"Irrelevant" images per product query	VisualRank$_{bias}$	VisualRank	HeuristicRank
Among top 20 results	0.23	0.83	1.93
Among top 10 results	0.17	0.47	1.42
Among top 5 results	0.12	0.30	0.86
Among top 3 results	0.04	0.20	0.65

first page will receive more clicks than a high-quality image from the second page (default ranking 21-40). If VisualRank outperforms the existing Google Image search in this experiment, we can expect a much greater improvement in deployment.

When examined over the set of 130 product queries, the images selected by VisualRank to be in the top 20 would have received approximately 17.5 percent more clicks than those in the default ranking. This improvement was achieved despite the positional bias that strongly favored the default rankings.

6.4 An Alternate Query Set: Landmarks

To this point, we have examined the performance of VisualRank on queries related to products. It is also interesting to examine the performance on an alternate query set. Here, we present the results of an analogous study to the product-based one presented to this point; this study is conducted with common landmark related queries.

For this study, we gathered 80 common landmark related queries. Typical queries included: "Eiffel Tower," "Big Ben," "Coliseum," and "Lincoln Memorial." Similarly to product queries, these queries have rigid canonical objects that are central to the answer. Table 5 shows the performance of VisualRank when minimizing the number of irrelevant queries in the top 10, top 5, and top 3 results. As was seen in the experiments with product images, VisualRank significantly outperforms the default rankings at all of the measured settings. Table 6 shows the number of queries for which VisualRank outperformed Google and vice versa. Note that the default Google rankings rarely outperformed VisualRank; however, there were a large number of ties (32) in which Google and VisualRank had an equal number of irrelevant images.

For the last measurement, we examine the clicks that would have been received under VisualRank-based reordering and under default settings. In 50 of the queries, VisualRank would have received more clicks, while, in 27 of the queries, the default ranking would have. The remaining three queries tied.

7 Conclusions and Future Work

The VisualRank algorithm presents a simple mechanism to incorporate the advances made in using link and network analysis for Web document search into image search. Although no links explicitly exist in the image search graph, we demonstrated an effective method to infer a graph in which the images could be embedded. The result was an approach that was able to outperform the default Google ranking on the vast majority of queries tried, while maintaining reasonable computational efficiency for large-scale deployment. Importantly, the ability to reduce the number of irrelevant images shown is extremely important not only for the task of image ranking for image retrieval applications but also for applications in which only a tiny set of images must be selected from a very large set of candidates.

Interestingly, by replacing user-created hyperlinks with automatically inferred "visual hyperlinks," VisualRank seems to deviate from a crucial source of information that makes PageRank successful: the large number of *manually* created links on a diverse set of pages. However, a significant amount of the human-coded information is recaptured through two mechanisms. First, by making VisualRank query dependent (by selecting the initial set of images from search engine answers), human knowledge, in terms of linking relevant images to Web pages, is directly introduced into the system. Second, we implicitly rely on the intelligence of crowds: The image similarity graph is generated based on the common features between images. Those images that capture the common themes from many of the other images are those that will have higher relevancy.

The categories of queries addressed, products, and landmarks lend themselves well to the type of local feature detectors that we employed to generate the underlying graph. One of the strengths of the approach described in this paper is the ability to customize the similarity function based on the expected distribution of queries. Unlike many classifier-based methods [7], [42] that construct a single mapping from image features to ranking, VisualRank relies only on the inferred similarities, not the features themselves. Similarity measurements can be constructed through

TABLE 5
Relevancy Study

"Irrelevant" images per landmark query	VisualRank	Google
Among top 10 results	0.35	3.64
Among top 5 results	0.18	1.73
Among top 3 results	0.03	0.94

TABLE 4
Relevance Comparison per Query

	VisualRank	HeuristicRank
Outperforming product queries	664	182

TABLE 6
Relevance Comparison per Query

	VisualRank	Google
Outperforming landmark queries	46	2

numerous techniques and their construction is independent of the image relevance assessment. For example, images related to people and celebrities may rely on face recognition/similarity, images related to landmarks and products (i.e., Eiffel Tower) may use local descriptors, other images, such as landscapes, may more heavily rely on color information, etc. Additionally, within this framework, context-free signals, like user-generated covisitation [15], can be used in combination with image features to approximate the visual similarity of images.

There are many avenues open for future exploration; they range from the domains to which these techniques can be deployed to refinements of the algorithms presented for accuracy and for speed. Three directions for future study that we are pursuing are given here. First, we are working on both labeling and ranking unlabeled images from personal image collections. The underlying graph structures explored in this paper are relevant to this task. Further, the ability to use unlabeled images to create linkages between images that are labeled and those that need labels is an active area of research.

Second, we would like to study the relationship between image similarity and "likelihood for transition" more extensively. For example, although a user is likely to transit between two related images, the exact behavior may be more complicated. For example, a user may not want to transit to images that are too similar. Therefore, learning the weighting function for creating the similarity weights that may *decrease* the weight of images that are too similar (perhaps near duplicates) may further improve the diversity of the top ranked images.

Third, we are extending this work to other domains that are not amenable to traditional text-based analysis. These domains include video and audio analysis.

ACKNOWLEDGMENTS

The authors would like to thank D. Sivakumar, James M. Rehg, Henry Rowley, Michele Covell, and Dennis Strelow for the fruitful discussions on local features and graph analysis algorithms and Mithun Gupta for his help in implementing LSH.

REFERENCES

[1] R. Datta, D. Joshi, J. Li, and J. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age," *ACM Computing Surveys,* vol. 40, no. 2, 2008.

[2] R.P.A. Pentland and S. Sclaroff, "Content-Based Manipulation of Image Databases," *Int'l J. Computer Vision,* vol. 18, no. 3, pp. 233-254, 1996.

[3] A.W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-Based Image Retrieval at the End of the Early Years," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 12, pp. 1349-1380, Dec. 2000.

[4] W.-Y. Ma and B.S. Manjunath, "A Toolbox for Navigating Large Image Databases," *Multimedia System,* vol. 3, no. 7, pp. 184-198, 1999.

[5] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 8, pp. 1026-1038, Aug. 2002.

[6] D. Joshi, J.Z. Wang, and J. Li, "The Story Picturing Engine—A System for Automatic Text Illustration," *ACM Trans. Multimedia, Computing, Comm. and Applications,* vol. 2, no. 1, pp. 68-89, 2006.

[7] R. Fergus, P. Perona, and A. Zisserman, "A Visual Category Filter for Google Images," *Proc. Eighth European Conf. Computer Vision,* pp. 242-256, 2004.

[8] R. Fergus, P. Perona, and A. Zisserman, "Object Class Recognition by Unsupervised Scale-Invariant Learning," *Proc. Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 264-271, 2003.

[9] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian Network Classifiers," *Machine Learning,* vol. 29, pp. 131-163, 1997.

[10] D.G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int'l J. Computer Vision,* vol. 60, no. 2, pp. 91-110, 2004.

[11] E. Xing, A. Ng, M. Jordan, and S. Russel, "Distance Metric Learning, with Applications to Clustering with Side-Information," *Proc. 15th Conf. Advances in Neural Information Processing Systems,* vol. 15, pp. 450-459, 2002.

[12] K. Weinberger, J. Blitzer, and L. Saul, "Distance Metric Learning for Large Margin Nearest Neighbor Classification," *Proc. 18th Conf. Advances in Neural Information Processing Systems,* vol. 18, pp. 1437-1480, 2006.

[13] A. Frome, Y. Singer, F. Sha, and J. Malik, "Learning Globally-Consistent Local Distance Functions for Shape-Based Image Retrieval and Classification," *Proc. 11th IEEE Int'l Conf. Computer Vision,* pp. 1-8, 2007.

[14] I. Simon, N. Snavely, and S.M. Seitz, "Scene Summarization for Online Image Collections," *Proc. 12th Int'l Conf. Computer Vision,* 2007.

[15] S. Uchihashi and T. Kanade, "Content-Free Image Retrieval by Combinations of Keywords and User Feedbacks," *Proc. Fifth Int'l Conf. Image and Video Retrieval,* pp. 650-659, 2005.

[16] S. Baluja, R. Seth, D. Siva, Y. Jing, J. Yagnik, S. Kumar, D. Ravichandran, and M. Aly, "Video Suggestion and Discovery for YouTube: Taking Random Walks through the View Graph," *Proc. 17th Int'l World Wide Web Conf.,* 2008.

[17] S. Brin and L. Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Computer Networks and ISDN Systems,* vol. 30, nos. 1-7, pp. 107-117, 1998.

[18] X. He, W.-Y. Ma, and H. Zhang, "Imagerank: Spectral Techniques for Structural Analysis of Image Database," *Proc. Int'l Conf. Multimedia and Expo,* vol. 1, pp. 25-28, 2002.

[19] W.H. Hsu, L. Kennedy, and S. Chang, "Video Search Reranking through Random Walk over Document-Level Context Graph," *Proc. 15th Int'l Conf. Multimedia,* pp. 971-980, 2007.

[20] J.M. Kleinberg, "Authoritative Sources in a Hyperlinked Environment," *J. ACM,* vol. 46, no. 5, pp. 604-632, 1999.

[21] B.J. Frey and D. Dueck, "Clustering by Passing Messages between Data Points," *Science,* vol. 315, pp. 972-976, 2007.

[22] R.I. Kondor and J. Lafferty, "Diffusion Kernels on Graphs and Other Discrete Structures," *Proc. 19th Int'l Conf. Machine Learning,* pp. 315-322, 2002.

[23] X. Zhu, Z. Ghahramani, and J.D. Lafferty, "Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions," *Proc. 20th Int'l Conf. Machine Learning,* pp. 912-919, 2003.

[24] Y. Jing, S. Baluja, and H. Rowley, "Canonical Image Selection from the Web," *Proc. Sixth Int'l Conf. Image and Video Retrieval,* pp. 280-287, 2007.

[25] T. Haveliwala, "Topic-Sensitive Pagerank: A Context-Sensitive Ranking Algorithm for Web Search," *IEEE Trans. Knowledge and Data Eng.,* vol. 15, no. 4, pp. 784-796, July/Aug. 2003.

[26] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," *Proc. Fourth Alvey Vision Conf.,* pp. 147-151, 1988.

[27] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching and Object Recognition Using Shape Contexts," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 4, pp. 509-522, Apr. 2002.

[28] S. Lazebnik, C. Schmid, and J. Ponce, "A Sparse Texture Representation Using Affine-Invariant Regions," *Proc. Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 319-324, 2003.

[29] K. Mikolajczyk and C. Schmid, "A Performance Evaluation of Local Descriptors," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 27, no. 10, pp. 1615-1630, Oct. 2005.

[30] S. Winder and M. Brown, "Learning Local Image Descriptors," *Prof. Conf. Computer Vision and Pattern Recognition,* 2007.

[31] H. Bay, T. Tuytelaars, and L.V. Gool, "Surf: Speeded Up Robust Features," *Proc. Ninth European Conf. Computer Vision,* pp. 404-417, 2006.

[32] D. Nistér and H. Stewénius, "Scalable Recognition with a Vocabulary Tree," *Proc. Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 2161-2168, 2006.

[33] Y. Ke, R. Sukthankar, and L. Huston, "Efficient Near-Duplicate Detection and Sub-Image Retrieval," *Proc. ACM Int'l Conf. Multimedia,* pp. 869-876, 2004.
[34] Y. Ke and R. Sukthankar, "PCA-SIFT: A More Distinctive Representation for Local Image Descriptors," *Proc. Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 506-513, 2004.
[35] G. Schindler, M. Brown, and R. Szeliski, "City-Scale Location Recognition," *Proc. Conf. Computer Vision and Pattern Recognition,* 2007.
[36] E. Nowak and F. Jurie, "Learning Visual Similarity Measures for Comparing Never Seen Objects," *Proc. Conf. Computer Vision and Pattern Recognition,* 2007.
[37] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object Retrieval with Large Vocabularies and Fast Spatial Matching," *Proc. Conf. Computer Vision and Pattern Recognition),* 2007.
[38] M. Datar, N. Immorlica, P. Indyk, and V.S. Mirrokni, "Locality-Sensitive Hashing Scheme Based on p-Stable Distributions," *Proc. 20th Symp. Computational Geometry,* pp. 253-262, 2004.
[39] P. Indyk, R. Motwani, P. Raghavan, and S. Vempala, "Approximate Nearest Neighbor—Towards Removing the Curse of Dimensionality," *Proc. 30th ACM Symp. Computational Theory,* pp. 604-613, 1998.
[40] P. Indyk, "Stable Distributions, Pseudorandom Generators, Embeddings, and Data Stream Computation," *Proc. 41st IEEE Symp. Foundations of Computer Science,* pp. 189-197, 2000.
[41] G. Salton and M.J. McGill, *Introduction to Modern Information Retrieval.* McGraw-Hill, 1983.
[42] G. Park, Y. Baek, and H. Lee, "Majority Based Ranking Approach in Web Image Retrieval," *Lecture Notes in Computer Science,* vols. 27-28, pp. 499-504, 2003.

Yushi Jing received the BS and MS degrees (with distinction) from the Georgia Institute of Technology (Georgia Tech), Atlanta, and is currently a PhD candidate there, focusing on probabilistic graphical models and boosting. Currently, he is also a full-time member of the Google Research group, working on computer vision, social networks and machine learning. His research and engineering efforts have shaped various Google products, with press coverage from *Scientific American*, the BBC, and the *New York Times*. He won the Distinguished Student Paper Award at the International Conference on Machine Learning (ICML) and the Computational Perception Fellowship given by Georgia Tech. He also has a strong interest in social science and business analytics and previously consulted for the United Nations as well as for a preeminent Japanese advertising agency on social network analytics. He is also pursuing MS candidacy in International Affairs from Georgia Tech. He is a member of the IEEE and the IEEE Computer Society.

Shumeet Baluja received the BS degree (high distinction) from the University of Virginia in 1991 and the PhD degree in computer science from Carnegie Mellon University in 1996. He is currently a senior staff research scientist at Google, where he works on a broad set of topics, ranging from image processing and machine learning to wireless application development and user interaction measurement. He was formerly the chief technology officer of JAMDAT Mobile, Inc., where he oversaw all aspects of technology initiation, development, and deployment. Previously, he was a chief scientist at Lycos Inc., where he led the Research and Applied Technology Group in the quantitative and qualitative analysis of user behavior, including data mining and trend analysis, advertisement optimization, and statistical modeling of traffic and site interactions. As the senior vice president of R&D at eCompanies LLC, he spearheaded the creation of their wireless practice and was responsible for finding and evaluating new technologies and technology entrepreneurs. He has filed numerous patents and has published scientific papers in fields including computer vision and facial image processing, advertisement display and optimization, automated vehicle control, statistical machine learning, and high-dimensional optimization. He is a member of the IEEE and the IEEE Computer Society.