When Amazon Meets Google: Product Visualization by Exploring Multiple Web Sources

MENG WANG, Hefei University of Technology
GUANGDA LI, ZHENG LU, YUE GAO, and TAT-SENG CHUA, National University of Singapore

12

Product visualization is able to help users easily get knowledge about the visual appearance of a product. It is useful in many application and commercialization scenarios. However, the existing product image search on e-commerce Web sites or general search engines usually get insufficient search results or return images that are redundant and not relevant enough. In this article, we present a novel product visualization approach that automatically collects a set of diverse and relevant product images by exploring multiple Web sources. Our approach simultaneously leverages Amazon and Google image search engines, which represent domain-specific knowledge resource and general Web information collection, respectively. We propose a conditional clustering approach that is formulated as an affinity propagation problem regarding the Amazon examples as information prior. The ranking information of Google image search results is also explored. In this way, a set of exemplars can be found from the Google search results and they are provided together with the Amazon example images for product visualization. Experiments demonstrate the feasibility and effectiveness of our approach.

Categories and Subject Descriptors: H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Index

General Terms: Algorithms, Experimentation, Human Factors

Additional Key Words and Phrases: Product visualization, Web source, image search

ACM Reference Format:
Wang, M., Li, G., Lu, Z., Gao, Y., and Chua, T.-S. 2013. When amazon meets google: Product visualization by exploring multiple web sources. *ACM Trans. Internet Technol.* 12, 4, Article 12 (July 2013), 17 pages.
DOI:http://dx.doi.org/10.1145/2492690

1. INTRODUCTION

In the past decades, extensive research efforts have been dedicated to building more effective and efficient image search services [Jing and Baluja 2008; Lew 2006; Li et al. 2006; Smeulders et al. 2000]. Product image search, as a specific field of image search, is the most straightforward approach to visualizing products. It can help users get visual knowledge about products and has great commercial potential.

This research has been supported by the National 973 Program of China under grant 2013CB329604, the National Natural Science Foundation of China (NSFC) under grant 6127239, and the Program for New Century Excellent Talents in University under grant NCET-12-0836.
Authors' addresses: M. Wang (corresponding author), School of Computer Science and Information Engineering, Hefei University of Technology, 193 Tunxi Rd., Baohe, Hefei, Anhui, China; email: eric.mengwang@gmail.com; G. Li, Z. Lu, Y. Gao, and T.-S. Chua, School of Computing, National University of Singapore, 25 Heng Mui Keng Terrace, Singapore 119615.

DOI:http://dx.doi.org/10.1145/2492690

In comparison with general image search, product image search has its own properties as there are many resources built with domain knowledge. For example, on many e-commerce Web sites such as Amazon[1] and New Egg[2], typically there exist some high-quality image examples associated with each product. These images are usually created or collected by professional product sellers. Therefore, they are useful for product visualization. However, these images are usually limited and thus may not provide comprehensive visual information. For example, our study shows that for many products Amazon only provides 1 to 5 images.

On the other hand, there are plenty of images that describe a product with different scales, views, and surroundings available on the Internet, which can be easily accessed through image search tools such as Google image search. But simply performing image search with these tools usually returns images that are noisy and redundant as they are indexed by text information.

Therefore, there lacks an approach that can collect a set of relevant and diverse images for product visualization. But it is highly desired as it is useful in a number of applications. First, it can help users easily grasp the visual appearance of a product by observing the collected image set, and it can benefit these users in e-commerce consumption. Second, a lot of research efforts are put on visual product search, such as the work in Chen et al. [2010], He [2011], and Tsai et al. [2010]. These approaches are usually built based on a large-scale database, which needs to contain the descriptive pictures of a lot of products with different scales, views, etc. A method that can collect diverse and relevant product images will be helpful in building the database. Finally, it has already been demonstrated that diversity and relevance are both important for user experience in image search [Deselaers et al. 2009; Hurley and Zhang 2011; Pang et al. 2009; Wang et al. 2010], and it is also the case for product image search.

In this work, we propose an approach that combines Amazon and Google image search services to collect diverse and relevant images for product visualization. Given a product, we can collect a set of example images on Amazon. As previously mentioned, these images are created or collected by professionals and they usually describe the product with different views. We also collect a large set of images by searching the product on Google. However, the images collected in this way are actually noisy and redundant. Our task is to generate a set of image exemplars that satisfy the requirements of both relevance and diversity for describing the product. Since Amazon images are in good quality, we assume that these images can be part of typical exemplars of the product and we only need to identify some other exemplars from Google image search results. Here we propose a conditional clustering approach that is formulated as an affinity propagation problem with regard to the Amazon images and the ranking information of Google image search results as information prior. Finally, the exemplars found in the Google search results together with Amazon example images are presented for visualizing the product.

To our knowledge, this is the first work on collecting diverse and relevant product images. Our work can be understood from different perspectives.

(1) From the Amazon perspective, we can enrich the example images on Amazon with the search results from Google. This can be a service for e-commerce retailers, as these retailers usually provide very few example images. Our approach can automatically enrich them to provide viewers more comprehensive knowledge, such as those images that describe a product with different views, poses, sizes, backgrounds, etc.

[1]http://www.amazon.com
[2]http://www.newegg.com

(2) From the Google perspective, we can refine Google image search results by exploring the example images on Amazon and thus better results in terms of both relevance and diversity can be obtained. That means we build a product image search service that can return results that are both relevant and diverse.
(3) It can also be regarded as a special "metasearch" approach that combines Amazon and Google. In conventional metasearch [Benitez et al. 1998; Kennedy and Chang 2008], the search results from different search engines are usually combined with several rank aggregation algorithms. But in our scheme, the two sources have their own characteristics, and thus we design a special approach to explore their own characteristics.

It is also worth noting that Amazon and Google in our approach actually represent domain-specific knowledge resource and general Web information collection, respectively. Our approach is very flexible. For example, we can replace Amazon and Google with New Egg and Yahoo, respectively, or combine them together. For several products that cannot be found on Amazon, our approach will degenerate to performing the method on merely Google image search results and can still work well. We also would like to emphasize that although our conditional affinity propagation method is designed for product visualization tasks, it is a generalized algorithm and can be applied in many other applications that have a similar information prior.

The rest of the article is organized as follows. Section 2 briefly reviews related work. In Section 3, we introduce the conventional affinity propagation approach. In Section 4, we detail our approach for generating diverse and relevant image exemplars. Our method involves a ranking-based image relevance estimation component, and it is introduced in Section 5. Experiments are provided in Section 6. Finally, we conclude the work in Section 7.

2. RELATED WORK

2.1. Product Image Search and Image Search Diversification

Image search has been deeply investigated in research communities [Wang et al. 2012a, 2012b, 2013], but the literature regarding product-related image search is very sparse. Jing and Baluja [2008] apply a PageRank-like algorithm based on visual links between images to improve the ranking performance for product image search. Xie et al. [2008] propose a client-server architecture for mobile devices to undertake multi-modality search, one of which is content-based product retrieval by queries from a phone camera. For commercial applications, Google and Amazon are able to provide users product search results by simply capturing a view using a mobile phone[3,4].

Regarding image search diversification, there are two typical approaches: one is online ranking that adjusts the order of images to keep the diversity of top search results and the other is clustering that finds a set of representative images. For the first approach, Wang et al. [2010] propose a greedy ordering algorithm that is able to take both relevance and diversity into account by exploring the content of images and their associated tags. Deselaers et al. [2009] present dynamic programming algorithms to jointly optimize the relevance and the diversity of image search results. For the second approach, van Leuken et al. [2009] deploy lightweight clustering techniques in combination with a dynamic weighting function of the visual features to capture the different aspects of image search results. Liu et al. [2009] propose a method to summarize search results by taking relevance and quality as informative prior into an

[3]http://www.google.com/mobile/goggles
[4]http://a9.com

affinity propagation framework. Jia et al. [2008] design a fast sparse affinity propagation algorithm to find exemplars to best represent image search results. However, to our knowledge, there is no research on diverse and relevant product image search, and our work well-complements the existing efforts. In addition, we explore the characteristics of both Amazon and Google to accomplish the task, and this is different from the existing approaches. Our approach is different from Liu et al.'s [2009] method in the following two aspects. First, Liu et al.'s [2009] method simply summarizes Google image search results, whereas our approach simultaneously leverages Amazon and Google image search services. Second, Liu et al.'s [2009] method explores the ranking information of Google search results with a heuristic strategy, whereas our approach estimates a ranking-based relevance measure for Google search results and explores it in the conditional affinity propagation approach. These two differences make our approach superior to Liu et al.'s [2009] method.

2.2. Image Set Summarization

Summarization has been widely investigated to extract representative exemplars from a set of images. Rother et al. [2005] summarize a set of images with a "digital tapestry", that stitches together salient and spatially compatible blocks from the input image set. Jing et al. [2006] propose an efficient method to group Web image search results into clusters and then present them to users. Wang et al. [2006] create a "picture" collage, a 2D spatial arrangement of the images in the input set chosen to maximize the visibility of salient regions. These works usually rely on an explicit or implicit clustering method. While many different clustering methods can be applied, Affinity Propagation (AP), which was first applied in image clustering in Dueck and Frey [2007], attracts great interest for its robustness and ability of automatically determining the number of exemplars. Zha et al. [2009] adopt AP to select a set of representative images associated with a tag. Xu et al. [2011] extend the AP method to deal with heterogeneous data. However, the conventional AP method cannot deal with the case that a set of samples needs to be kept as exemplars. Liu et al. [2011] introduce a strategy to address this issue, but there lacks a deep analysis and the method cannot take the different priorities of samples into consideration. In this work, we extend the conventional AP method to a novel conditional AP approach to accomplish our product visualization task.

3. CONVENTIONAL AFFINITY PROPAGATION

In this section, we first provide a short description of the conventional AP approach. Here we follow the derivation of AP in Givoni and Frey [2009]. Denote by n the number of samples. Let $s(i,j)$ denote the similarity between the i-th and the j-th samples. Let $\{c_{ij}\}_{j=1}^{n}$ be n binary variables associated with the samples, such that $c_{ij} = 1$ if the exemplar for sample i is sample j. All assignments to exemplars and all exemplar choices can be described by the set of n^2 binary variables $\{c_{ij}\}$, $i,j \in \{1,\ldots,n\}$.

To decide exemplars, the algorithm should tend to maximize the similarity between the samples and their exemplars. However, two constraints are needed. The first constraint is that each sample in AP is assigned to a single exemplar, that is, $\sum_{j=1}^{n} c_{ij} = 1$. The second constraint is that the sample i may choose j as its exemplar only if j chooses itself as exemplar. Figure 1 shows a factor graph of the AP approach. The first constraint is introduced via the I function nodes: in every row i of the grid, exactly one c_{ij} must be set to 1. The second constraint is introduced via the E function nodes: for those points to be able to choose sample j as an exemplar, j must choose itself as an exemplar (c_{jj} must be 1). Then, a max-sum formulation can be used, where local functions are

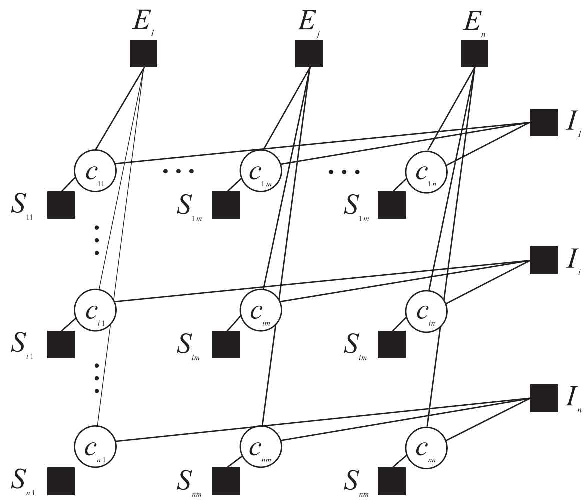

Fig. 1. Factor graph of the conventional AP approach.

added to form the global objective function to maximize, and the I and E functions can be naturally defined as

$$I_i(c_{i1},\ldots,c_{in}) = \begin{cases} -\infty, \text{if } \sum_j c_{ij} \neq 1, \\ 0, \text{otherwise} \end{cases} \tag{1}$$

$$E_j(c_{1j},\ldots,c_{nj}) = \begin{cases} -\infty, \text{ if } c_{jj} = 0 \text{ and } \exists i \neq j, c_{ij} = 1, \\ 0, \text{otherwise.} \end{cases} \tag{2}$$

The S_{ij} function nodes incorporate the similarities between samples and their potential exemplars.

$$S_{ij}(c_{ij}) = \begin{cases} s(i,j), \text{ if } c_{ij} = 1 \\ 0, \text{otherwise} \end{cases} \tag{3}$$

The max-sum objective function thus can be written as

$$S(c_{11},\ldots,c_{nn}) = \sum_{ij} S_{ij}(c_{ij}) + \sum_i I_i(c_{i1},\ldots,c_{in}) + \sum_j E_j(c_{1j},\ldots,c_{nj}). \tag{4}$$

The AP algorithm solves the preceding optimization problem by propagating two kinds of information between samples: (1) the responsibility $r(i,j)$ transmitted from sample i to sample j, which indicates how well j serves as the exemplar for sample i taking into account other potential exemplars for i; and (2) the availability $a(i,j)$ sent from candidate exemplar j to i, which indicates the appropriateness for sample i to choose sample j as exemplar taking into account the potential samples that may choose j as their exemplar. This information is iteratively updated by

$$r(i,j) = s(i,j) - \max_{j \neq j'}\{a(i,j') + s(i,j')\} \tag{5}$$

$$a(i,j) = \min\left\{0, r(j,j) + \sum_{i' \neq \{i,j\}} \max\{0, r(i',j)\}\right\}. \tag{6}$$

The "self-availability" $a(i,j)$ is updated by

$$a(j,j) = \sum_{i' \neq j} \max\{0, r(i',j)\}. \tag{7}$$

Upon convergence, the exemplar for each image is chosen by maximizing the following criterion.

$$\arg\max_{j}\{r(i,j) + a(i,j)\} \tag{8}$$

AP has a lot of advantages in summarization. First, it only requires pairwise similarities among samples. Second, it simultaneously accomplishes clustering and exemplar selection instead of separating them. Finally, it can automatically determine the number of exemplars. This is advantageous as it is usually difficult for users to set a value for the number of exemplars in advance. Therefore, we also adopt AP to accomplish the product visualization task. But conventional AP cannot handle information priors that certain samples need to be kept as exemplars and several other samples should be selected as exemplars with different probabilities. Thus, we extend it to a novel conditional AP method to address the problem.

4. PRODUCT VISUALIZATION BY CONDITIONAL AFFINITY PROPAGATION

4.1. Task Description and Notations

We now describe our task and notation. Given a product, we denote the set of Google image search results and the image examples on Amazon as $\mathcal{X}_1 = \{x_1, x_2, \ldots, x_n\}$ and $\mathcal{X}_2 = \{x_{n+1}, \ldots, x_{n+m}\}$, respectively. For the images in $\mathcal{X}_1$, we assume they are ordered according to their ranks in the search results (lower number indicates image with higher rank). Usually we have $m \ll n$. The task is to find a representative subset of $\mathcal{X}_1 \cup \mathcal{X}_2$ to visualize the product. As previously analyzed, we would like to preserve all the Amazon examples $\mathcal{X}_2$ in the subset, and thus we only need to further select several representative images from $\mathcal{X}_1$. An intuitive approach is to perform clustering to find exemplars in $\mathcal{X}_1$ and combine them with $\mathcal{X}_2$. However, the exemplars of $\mathcal{X}_1$ generated in this way will be redundant with the images in $\mathcal{X}_2$. In addition, it neglects the ranking information of the Google image search results. Therefore, it is not an optimal way.

4.2. Conditional Affinity Propagation

We now introduce the conditional affinity propagation approach that can tackle the problem. We can see that our task has two characteristics that cannot be handled by the conventional AP approach: (1) the samples in $\mathcal{X}_2$ need to be kept as exemplars; (2) the samples in $\mathcal{X}_1$ should be selected as exemplars with different probabilities according to their ranking orders.

Since the examples in $\mathcal{X}_2$ are kept as exemplars, we only have $n \times (n+m)$ variables, that is, $c_{ij}, 1 \leq i \leq n, 1 \leq j \leq n+m$. For the variables $c_{ij}, j \in \{n+1, n+2, \ldots, n+m\}$, the E function nodes will be removed. That means Eq. (2) turns to

$$E_j(c_{1j}, \ldots, c_{nj}) = \begin{cases} -\infty, \text{if } c_{jj} = 0 \text{ and } \exists i \neq j, c_{ij} = 1, \\ 0, \text{otherwise.} \end{cases} \tag{9}$$

where $j \in \{1, 2, \ldots, n\}$.

In order to take the ranking orders of the samples in $\mathcal{X}_1$ into consideration, we assume that each image has a relevance probability $p_i = Pr(x_i \text{ is relevant})$. Typically, p_i should be high if x_i is highly ranked. The estimation of p_i will be introduced in the

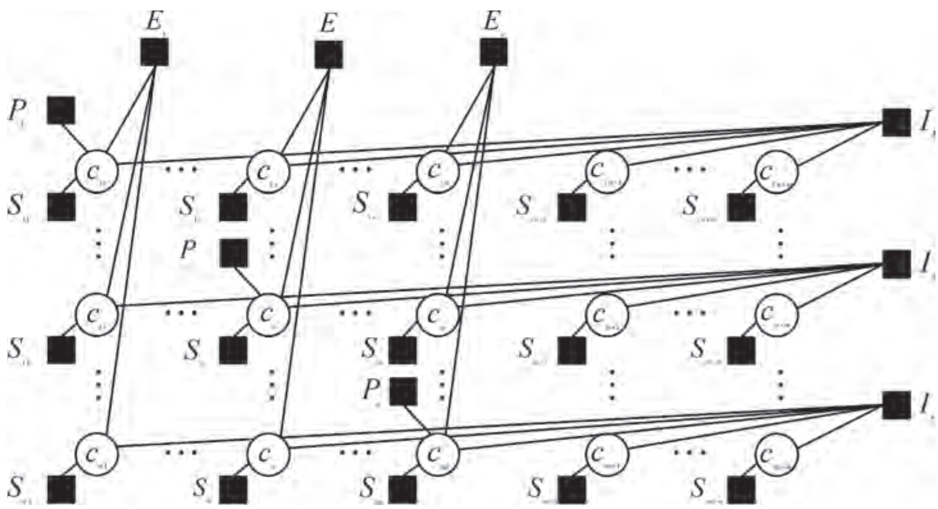

Fig. 2. Factor graph of the conditional AP approach.

next section. We assume that the relevances of different images are independent. Denote by $\mathcal{X}_e$ the exemplar set. Thus, we should maximize the relevance probability of the exemplar set, which is equivalent to maximizing $\sum_{x_i \in \mathcal{X}_e} \log p_i$. It is also equivalent to maximizing $\sum_{i=1}^{n} P_i(c_{ii})$, where

$$P_i(c_{ii}) = \begin{cases} \log p_i, \text{if } c_{ii} = 1, \\ 0, \text{otherwise.} \end{cases} \tag{10}$$

We add this term into the max-sum objective function. Therefore, the objective function becomes

$$\begin{aligned} S(c_{11}, \ldots, c_{n,n+m}) = \textstyle\sum_{i=1}^{n} \sum_{j=1}^{n+m} S_{ij}(c_{ij}) + \alpha \sum_{i=1}^{n} P_i(c_{ii}), \\ + \textstyle\sum_{i=1}^{n} I_i(c_{i1}, \ldots, c_{i,n+m}) + \sum_{j=1}^{n} E_j(c_{1j}, \ldots, c_{nj}), \end{aligned} \tag{11}$$

where $\alpha > 0$ is a weighting parameter that modulates the impact of the term that maximizes the relevance probability of the selected exemplars. If it is set to a great value, the algorithm will tend to select samples that are more likely to be relevant. The relevance information will be neglected if it is set to 0. Eq. (11) is exactly the max-sum objective function of the conditional AP approach, and Figure 2 illustrates the binary variable model.

To simplify the formulation, we can define

$$s'(i,j) = \begin{cases} s(i,j) + \alpha \log p_j, \ \text{if } i = j, \\ s(i,j), \ \text{otherwise.} \end{cases} \tag{12}$$

We then define $S'_{ij}(c_{ij})$ as

$$S'_{ij}(c_{ij}) = \begin{cases} s'(i,j), \ \text{if } c_{ij} = 1, \\ 0, \text{otherwise.} \end{cases} \tag{13}$$

The objective function in Eq. (11) then turns to

$$S(c_{11}, \dots, c_{n,n+m}) = \sum_{i=1}^{n} \sum_{j=1}^{n+m} S'_{ij}(c_{ij}) + \sum_{i=1}^{n} I_i(c_{i1}, \dots, c_{i,n+m}) + \sum_{j=1}^{n} E_j(c_{1j}, \dots, c_{nj}). \quad (14)$$

By applying a similar message derivation strategy as the conventional AP, in the conditional AP approach the message rules are updated as follows.

$$r(i,j) = s'(i,j) - \max \left\{ \begin{array}{c} \max\limits_{k \in \{1,\dots,n\} \setminus \{j\}} \{s'(i,k) + \alpha(i,k)\}, \\ \max\limits_{k \in \{n+1,\dots,n+m\} \setminus \{j\}} s'(i,k) \end{array} \right\} \quad (15)$$

$$\alpha(i,j) = \begin{cases} \sum\limits_{k \neq j} \max\{r(k,j), 0\}, \text{if } i = j, \\ \min \left\{ 0, r(j,j) + \sum\limits_{k \notin \{i,j\}} \max(0, r(k,j)) \right\}, \text{otherwise} \end{cases} \quad (16)$$

In the previous two equations, the responsibility $r(i,j)$ indicates how well j serves as the exemplar for sample i taking into account other potential exemplars for i, and the availability $a(i,j)$ indicates how appropriate for sample i to choose sample j as exemplar taking into account the potential samples that may choose j as their exemplar.

After convergence is achieved, the belief that image x_i ($i \in \{1, \dots, n\}$) selects x_j as its exemplar is

$$\arg \max_{j \in \{1,2,\dots,n+m\}} \{r(i,j) + a(i,j)\}. \quad (17)$$

In this way, we can find the exemplar set of $\mathcal{X}_1$, and we can integrate them with $\mathcal{X}_2$ for product visualization.

4.3. Discussion

From Eq. (12) we can see that we actually explore the relevance information of the images in $\mathcal{X}_2$ by adjusting $s(i,i)$. Note that it is a widely used strategy that assigns the prior of exemplar selection by setting $s(i,i)$ to different values [Frey and Dueck 2007]. In several existing works that apply affinity propagation to select exemplars in image search results, the idea of taking the ranking order information into consideration has already been investigated, such as in Liu et al. [2009]. However, conventional approaches usually heuristically convert the ranking orders into confidence or relevance scores and then set $s(i,i)$ accordingly. For example, in Liu et al. [2009], the relevance score at the i-th position is heuristically set to $1 - i/N$, where N is the length of the involved ranking list. Here we have provided an explanation on why we set $s(i,i)$ to take the ranking information of image search results into account and how we should set them. Experiments in Section 6 will demonstrate the superiority of our approach over conventional methods for setting $s(i,i)$.

We can analyze that the computational cost of the conditional approach scales as $O(Tn(n+m))$, where n and m are the numbers of images in $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively, and T is the iteration time (it is set to 100 in our implementation). The value of m is small as the Amazon examples are limited. For Google search results, we only need to use the top results and thus n will not be great. Therefore, the process can be finished rapidly.

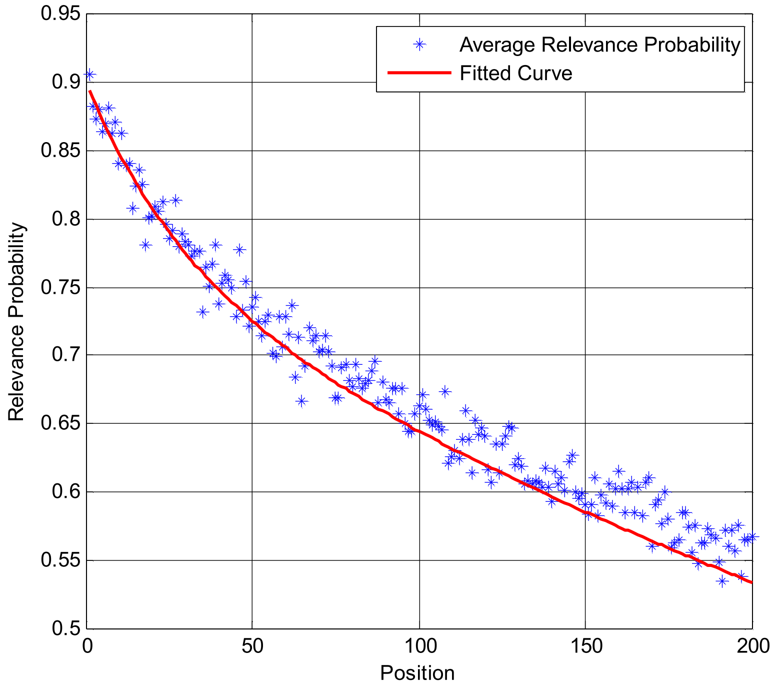

Fig. 3. The average relevance probabilities at different ranking positions and the fitted curve.

5. RANKING-BASED RELEVANCE ESTIMATION FOR GOOGLE IMAGE SEARCH RESULTS

In the conditional AP method introduced earlier, the prior relevance probabilities of the Google image search results are involved (see Eq. (10)). Now we introduce the estimation of the prior relevance probabilities. Let p_i denote the relevance probability of the i-th image in a ranking list obtained by the Google image search engine. We investigate the relationship between p_i and the position i with a large number of queries. Actually, we can define

$$p_i = E_{q \in \mathcal{Q}}[\hat{p}(q,i)]\,, \tag{18}$$

where $\mathcal{Q}$ denotes the whole query space, $E_{q \in \mathcal{Q}}$ means the expectation over the query set $\mathcal{Q}$, and $\hat{p}(q,i)$ indicates the relevance ground truth of the i-th search result for query q. Therefore, the most intuitive approach is to estimate p_i by averaging $\hat{p}(q,i)$ over a large query set.

Here we use 100 product queries to estimate p_i. The relevance score of each search result is manually labeled to be 0 or 1 if the image is irrelevant or relevant to the query. Figure 3 shows the averaged relevance score curve for the 100 queries with respect to the ranking position. However, although we can see that the curve tends to decrease when ranking position increases, it is not smooth enough. There are fluctuations, and this is not consistent with the prior knowledge that the expected relevance scores should be decreasing with respect to ranking positions. This can be attributed to the fact that the queries used to estimate the initial relevance queries are still insufficient. Here we smooth the curves with a parametric model. We assume $p_i = ae^{-i/b} + ce^{-i/d}$, where a, b, c, and d are the four parameters in an exponential function. We then fit this function with the points. We can estimate the parameters with the mean squared loss criterion. Here we adopt a gradient descent method, and the values of these four parameters are estimated to be 0.133, 30, 0.767, 609, respectively. Figure 3 also shows

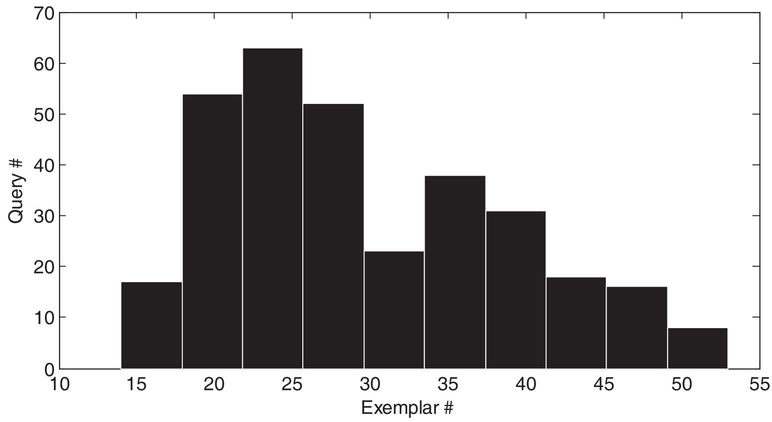

Fig. 4. Distribution of the exemplar numbers obtained by the proposed approach. We can see that for most products the number of exemplars is between 20 and 40.

the fitted curves from which we can see that they reasonably preserve the original information.

6. EVALUATION

6.1. Experimental Settings

We conduct our experiments with 300 products selected from the electronics and furniture domains from Amazon. Note that these 300 products have no intersection with the 100 product queries used to estimate the prior relevance probabilities of Google image search results (see Section 5). They are mainly rigid products and we do not consider nonrigid product categories, such as *clothes* and *hats*. It is due to the following two reasons. First, it does not make much sense for the visualization of many nonrigid objects. Second, if we really need to visualize a product category, we can generate its visual example set by assembling the examples of the rigid products that belong to the category.

For each product, we collect images from Amazon and Google respectively. For Amazon images, we perform a simple duplicate removal step. The process is as follows. We check the Amazon image list from top to bottom, and if an image is close to a sample that appears above it, we remove it. More specifically, we remove the i-th image if there exists $j < i$ that satisfies $s(i,j) > 0.95$, where $s(i,j)$ is the cosine similarity of the i-th and j-th images. After removing duplicates, the number of images for a product on Amazon mainly varies from 1 to 7. We collect the top 200 images from the Google search results. For feature representation, we employ a Difference-of-Gaussian (DoG) method to detect keypoints and from each keypoint we extract 128-dimensional SIFT features [Lowe 2004]. The SIFT features are grouped into 160,000 clusters with hierarchical k-means [Nister and Stewenius 2006]. Therefore, each image is represented by a 160,000-dimensional Bag-of-Visual-Word (BoVW) histogram. Image similarity is estimated with the cosine similarity of the histograms. Note that DoG + SIFT may not be optimal for image similarity estimation, and several other methods such as dense SIFT plus different pooling strategies can be applied. But a comprehensive investigation of image representation is beyond the scope of this work, and thus we leave it to our future work.

Note that our approach involves a parameter α (see Eq. (11)). To tune this parameter, we randomly select 30 queries and generate the results with different values of α,

Table I. The 20 Most Popular Products for which We Provide Detailed Results

Index	Product
1	Kindle
2	Apple iPod touch(4th Generation)
3	Iomega eGo 2 TB Desktop Hard Drive
4	Roku XD Streaming Player 1080p
5	TomTom XL 340TM 4.3-Inch GPS
6	Apple iPod shuffle 2 GB Silver (4th Generation)
7	TomTom XXL 540TM 5-Inch Widescreen Portable GPS Navigator
8	Motorola SB6120 Modem
9	Apple TV MC572LL/A (2010)
10	Garmin Forerunner 305 GPS Receiver With Heart Rate Monitor
11	Nikon D3100 DSLR
12	Canon EOS 5D
13	Logitech Webcam Pro C910
14	Western Digital My Passport Essential SE 1 TB
15	SanDisk 16GB Cruzer Micro USB Drive
16	Acme Sectional Sofa
17	Retro Style Chairs Set Of 4
18	Milano Espresso and Mocha Sofa
19	Safco Compact Mobile File Cart
20	Walnut 5-tier Leaning Ladder Book Shelf

including 0.001, 0.01, 0.1, 1, 10, 100, and 1000. Five users observe the results and pick up the best image set in terms of product visualization and the value of α by a voting.

Considering our approach generates r exemplars for a given product, we compare our approach with the following four methods.

(1) We directly use the top r Google images.
(2) We use the combination of m Amazon examples and the top $r - m$ Google images.
(3) We use the combination of m Amazon examples and the exemplars of Google images generated by the conventional AP algorithm.
(4) We also use the conditional AP method to select exemplars, but we set the relevance probabilities of Google image search results as p_i to $1-i/200$, where i means ranking order. The comparison with this method is in order to verify the effectiveness of our ranking-based relevance estimation component.

We denote our approach as "Amazon + Google_CAP". The aforesaid four to-be-compared methods are denoted as "Google_Top", "Amazon + Google_Top", "Amazon + Google_AP", and "heuristic Relevance Setting", respectively.

Figure 4 illustrates the distribution of the exemplar numbers obtained by our approach. We compare the preceding methods in terms of both relevance and diversity of the generated exemplars. For each collected image, its relevance is labeled by the voting of three volunteers. We also pick up the 20 most popular products according to their customer reviewers on Amazon and provide detailed experimental results for them. These products are illustrated in Table I. We use 1 to 20 to denote them for simplicity.

6.2. Experimental Results

We first evaluate the relevance of the exemplar set generated by different methods. Figure 5 illustrates the average precision of the exemplars generated by different methods over the 300 products (here the precision means the ratio of relevant images

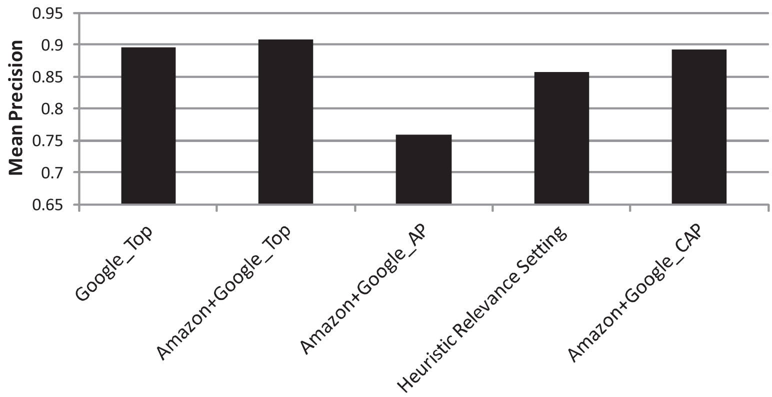

Fig. 5. The average of the precision measurements of the exemplar set generated by different methods over 300 products.

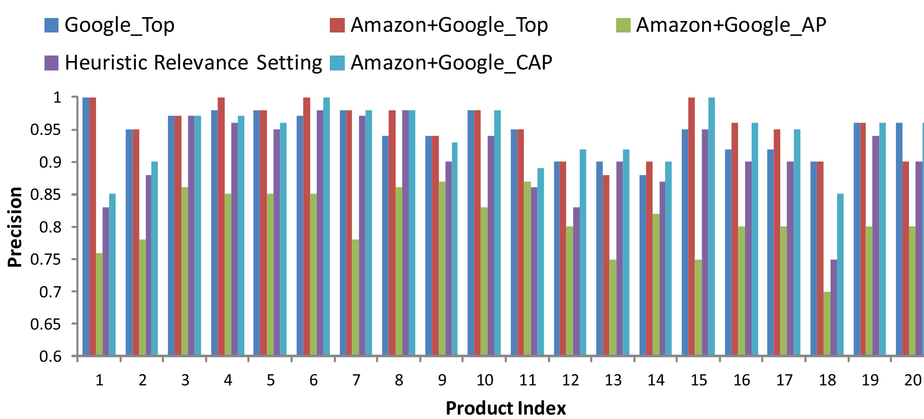

Fig. 6. The precision comparison of different methods for the 20 most popular products.

in an exemplar set). From the figure we have the following observations. The "Amazon + Google_Top" method achieves the best result. The proposed "Amazon + Google_CAP" approach performs slightly worse than the "Amazon + Google_Top" and "Google_Top" methods in terms of relevance. This is because the "Amazon + Google_CAP" has a diversification process, whereas "Amazon + Google_Top" and "Google_Top" do not have. However, we can see that the proposed approach can still achieve very high relevance and its difference with the "Amazon + Google_Top" and "Google_Top" methods is actually very small. The average precision measurements of these three methods are all close to 0.9. The "Heuristic Relevance Setting" and the "Amazon + Google_AP" methods are worse than the proposed "Amazon + Google_CAP" approach. This is due to the fact that: (1) the heuristic estimation of the relevance probability in "Heuristic Relevance Estimation" is significantly biased towards their real values and the AP method in "Amazon + Google_AP" does not take the ranking order information of Google image search results into consideration. Therefore, several nonrelevant images have been selected as exemplars in these two methods. Figure 6 presents the detailed precision measurements for the 20 popular products in Table I. We can see that the results for most of them are consistent with the observations.

We then evaluate the diversity of the generated exemplars. This is because usually more diverse exemplars are able to cover more visual appearances of a product. Here

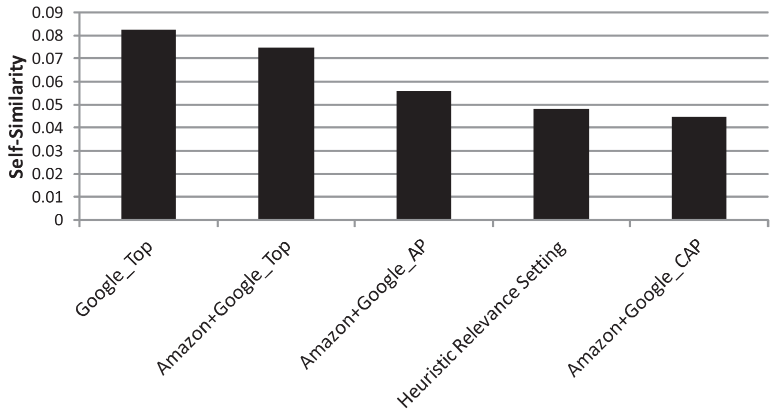

Fig. 7. The average self-similarity measurements of the exemplar set generated by different methods over 300 products.

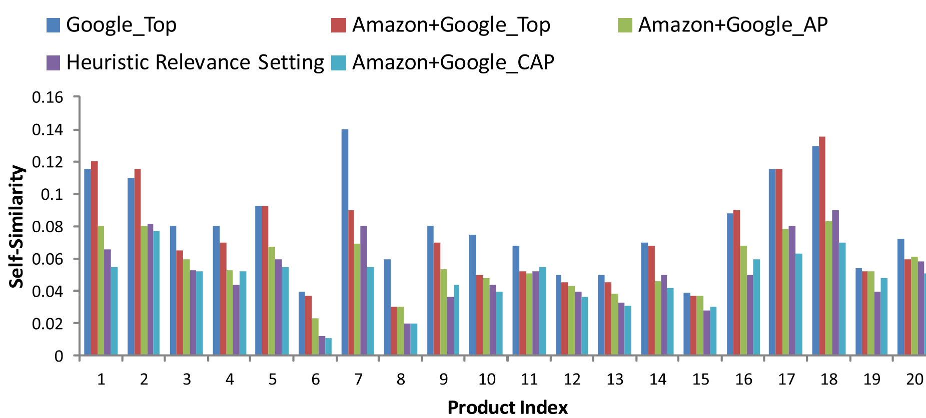

Fig. 8. The self-similarity comparison of the exemplar sets generated by different methods for the 20 most popular products.

we define a self-similarity measure to evaluate the diversity of an exemplar set, which is measured by calculating the average of the pairwise similarities of all exemplar pairs. Therefore, a low self-similarity indicates a high visual diversity. Figure 7 compares the average self-similarity measures of the five methods over the 300 products. From the results we can see that the "Amazon + Google_AP", "Heuristic Relevance Setting" and "Amazon + Google_CAP" achieve lower self-similarity measures than the other two methods. This is because these three methods all adopt an AP or conditional AP process. Among these three methods, the proposed "Amazon + Google_CAP" approach achieves the lowest self-similarity measure. Figure 8 presents the detailed self-similarity measurements for the 20 popular products in Table I.

From the preceding quantitative study, we can see that the proposed "Amazon + Google_CAP" approach can achieve a good trade-off between relevance and diversity. It is only slightly worse than "Google_Top" and "Amazon + Google_Top" in terms of relevance, but it is much better in terms of diversity. Figure 9 demonstrates the images for an example product *Roku XD Streaming Player 1080p* obtained from Amazon, Google, and our approach. We only illustrate the top 15 images for Google's search results here. From the results we can see that the images obtained by our method cover more

Fig. 9. Comparison of the images obtained by different methods for the product “Roku XD Streaming Player 1080p”. (a) Images on Amazon; (b) top results obtained by Google; (c) exemplars obtained by our approach.

widely than the Amazon examples and Google image search results. Although there is great deal of redundancy in the Google image search results, the results obtained by our approach do not have the redundancy problem.

Finally, we conduct a user study to compare our approach with the four methods for product visualization. There are 15 evaluators that are familiar with image search and e-commerce participating in the study. Each user is asked to freely pose a query among the 300 products and then compare the results of our approach, that is, “Amazon + Google_CAP”, with the other four methods. They are asked to evaluate which exemplar set can better visualize a product and in this process they consider both the relevance and coverage of the images. The two sets of results are shuffled and blended to keep a fair comparison. We adopt a 3-level scale to quantify the results. The score 0 is assigned to the worse method and 1, 2, and 0 are assigned to the other method if it is better, much better, or comparable to this one, respectively. Since there are disagreements among the evaluators, we perform a two-way analysis of variance (ANOVA) test [Minium et al. 1970] to statistically analyze the comparison. Tables I, II, III, and IV demonstrate the four comparisons. As can be observed from the tables, the average rating scores of our method are higher than the other four methods. Therefore, users confirm the superiority of our approach. From the p-values of the ANOVA test, we can

Table II. Mean and Standard Deviation Values of the Rating Scores Converted from the User Study of the Comparison of "Amazon + Google_CAP" and "Google_Top" and ANOVA Test Results

Amazon + Google_CAP vs. Google_Top		The factor of method		The factor of user	
Amazon_Google_CAP	Google_Top	F-statistic	p-value	F-statistic	p-value
1.67 ± 0.72	0.07 ± 0.26	46.34	8.49×10^{-6}	0.43	0.94

The p-values show that the difference of the two methods is significant and the difference of users is insignificant.

Table III. Mean and Standard Deviation Values of the Rating Scores Converted from the User Study of the Comparison of "Amazon + Google_CAP" and "Amazon + Google_Top" and ANOVA Test Results

Amazon + Google_CAP vs. Amazon + Google_Top		The factor of method		The factor of user	
Amazon + Google_CAP	Amazon + Google_Top	F-statistic	p-value	F-statistic	p-value
1.47 ± 0.74	0.07 ± 0.26	35.48	3.51×10^{-5}	0.49	0.90

The p-values show that the difference of the two methods is significant and the difference of users is insignificant.

Table IV. Mean and Standard Deviation Values of the Rating Scores Converted from the User Study of the Comparison of "Amazon + Google_CAP" and "Amazon + Google_AP" and ANOVA Test Results

Amazon + Google_CAP vs. Amazon + Google_AP		The factor of method		The factor of user	
Amazon + Google_CAP	Amazon + Google_AP	F-statistic	p-value	F-statistic	p-value
1.4 ± 0.74	0.07 ± 0.26	16.29	0.0012	1.12	0.4205

The p-values show that the difference of the two methods is significant and the difference of users is insignificant.

Table V. Mean and Standard Deviation Values of the Rating Scores Converted from the User Study of the Comparison of "Amazon + Google_CAP" and "Heuristic Relevance Setting" and ANOVA Test Results

Amazon + Google_CAP vs. Heuristic Relevance Setting		The factor of method		The factor of user	
Amazon + Google_CAP	Heuristic Relevance Setting	F-statistic	p-value	F-statistic	p-value
0.73 ± 0.59	0.067 ± 0.26	28.63	0.0001	0.51	0.91

The p-values show that the difference of the two methods is significant and the difference of users is insignificant.

see that the superiority of our approach is statistically significant and the difference of the evaluators is insignificant.

7. CONCLUSION AND FUTURE WORK

In this article, we present a novel solution for product visualization by simultaneously leveraging Amazon and Google image search engines. We propose a conditional clustering approach that is formulated as an affinity propagation problem incorporating the Amazon images and the ranking order of Google image search results as information prior. Several images from Google's search results are selected by the clustering method as a complementation of Amazon images. In this way, the finally obtained results maintain both diversity and relevance and can be used for product visualization. Experiments are conducted on a set of products and the results demonstrate the feasibility and effectiveness of our approach. In future work, we plan to improve the scheme by integrating more Web sources, including both more domain-specific knowledge rescouses (such as eBay and New Egg) and more image search engines (such as Yahoo and Bing).

REFERENCES

Benitez, A., Beigi, M., Beigi, I., and Chang, S. F. 1998. A content-based image meta-search engine using relevance feedback. *IEEE Internet Comput. 2*, 4, 59–69.

Chen, Y., Yu, N., Luo, B., and Chen, X. W. 2010. Ilike: Integrating visual and textual features for vertical search. In *Proceedings of the ACM International Conference on Multimedia*.

Deselaers, T., Gass, T., Dreuw, P., and Ney, H. 2009. Jointly optimising relevance and diversity in image retrieval. In *Proceeding of the ACM International Conference on Image and Video Retrieval (CIVR'09)*. 39:1–39:8.

Dueck, D. and Frey, B. J. 2007. Non-metric affinity propagation for unsupervised image categorization. In *Proceedings of the IEEE International Conference on Computer Vision*.

Frey, B. J. and Dueck, D. 2007. Clustering by passing messages between data points. *Sci. 315*, 5814, 972–976.

Givoni, I. E. and Frey, B. J. 2009. A binary variable model for affinity propagation. *Neural Comput. 21*, 6, 1589–1600.

He, J., Lin, T.-H., Feng, J., and Chang, S.-F. 2011. Mobile product search with bag of hash bits. In *Proceedings of the ACM International Conference on Multimedia*.

Hurley, N. and Zhang, M. 2011. Novelty and diversity in top-n recommendation - Analysis and evaluation. *ACM Trans. Internet Technol. 10*, 4.

Jia, Y., Wang, J., Zhang, C., and Hua, X. S. 2008. Finding image exemplars using fast sparse affinity propagation. In *Proceeding of the 16th ACM International Conference on Multimedia (MM'08)*. 639–642.

Jing, F., Wang, C., Yao, Y., Deng, K., Zhang, L., and Ma, W. Y. 2006. Igroup: Web image search results clustering. In *Proceedings of the ACM International Conference on Multimedia*.

Jing, Y. and Baluja, S. 2008. Pagerank for product image search. In *Proceeding of the International World Wide Web Conference*.

Kennedy, L. and Chang, S. F. 2008. Query-adaptive fusion for multimodal search. *Proc. IEEE 96*, 4, 567–588.

Lew, M. S., Sebe, N., Djeraba, C., and Jain, R. 2006. Content-based multimedia information retrieval: State of the art and challenges. *ACM Trans. Multimedia Comput. Comm. Appl. 2*, 1, 1–19.

Li, J., Allinson, N., Tao, D., and Li, X. 2006. Multi-training support vector machine for image retrieval. *IEEE Trans. Image Process. 15*, 11, 3597–3601.

Liu, D., Wang, M., Hua, X. S., and Zhang, H. J. 2011. Semi-automatic tagging of photo albums via exemplar selection and tag inference. *IEEE Trans. Multimedia 13,* 1, 82–91.

Liu, R., Yang, L., and Hua, X. S. 2009. Image search result summarization with informative priors. In *Proceeding of the 9th Asian Conference on Computer Vision (ACCV'09)*. 485–495.

Lowe, D. 2004. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis. 60,* 2, 91–110.

Minium, E. W., King, B. M., and Bear, G. 1970. *Statistical Reasoning in Psychology and Education*. Wiley, New York.

Nister, D. and Stewenius, H. 2006. Scalable recognition with a vocabulary tree. In *Proceedngs of the IEEE International Conference on Computer Vision and Pattern Recognition*.

Pang, H., Shen, J., and Krishnan, R. 2009. Privacy-preserving, similarity-based text retrieval. *ACM Trans. Internet Technol. 10*, 1.

Rother, C., Kumar, S., Kolmogorov, V., and Blake, A. 2005. Digital tapestry. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*.

Tsai, S. S., Chen, D., Chandrasekhar, V., Takacs, G., Cheung, N. M., et al. 2010. Mobile product recognition. In *Proceedings of the ACM International Conference on Multimedia*.

Smeulders, A. W. M., Member, S., Worring, M., Santini, S., Gupta, A., and Jain, R. 2000. Content-based image retrieval at the end of the early years. *IEEE Trans. Patt. Anal. Mach. Intell. 22*, 12, 1349–1380.

van Leuken, R. H., Garcia, L., Olivares, X., and Van Zwol, R. 2009. Visual diversification of image search results. In *Proceedings of the 18th International Conference on World Wide Web (WWW'09)*. 341–350.

Wang, J., Sun, J., Quan, L., Tang, X., and Shum, H. 2006. Picture collage. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*.

Wang, M., Yang, K., Hua, X. S., and Zhang, H. J. 2010. Towards a relevant and diverse search of social images. *IEEE Trans. Multimedia 12,* 8, 829–842.

Wang, M., Li, H., Tao, D., Lu, K., and Wu, X. 2012a. Multimodal graph-based reranking for web image search. *IEEE Trans. Image Process*. *21*, 11, 4649–4661.

Wang, M., Ni, B., and X. S. Hua, T. S. C. 2012b. Assistive tagging: A survey of multimedia tagging with human-computer joint exploration. *ACM Comput. Surv. 44*, 4.

Wang, M., Gao, Y., Lu, K., and Rui, Y. 2013. View-based discriminative probabilistic modeling for 3d object retrieval and recognition. *IEEE Trans. Image Process. 22*, 4, 1395–1407.

Xie, X., Lu, L., Jia, M., Li, H., Seide, F., and Ma, W. Y. 2008. Mobile search with multimodal queries. *Proc. IEEE 96,* 4, 589–601.

Xu, H., Wang, J., Hua, X. S., and Li, S. 2011. Hybrid image summarization. In *Proceedings of the ACM International Conference on Multimedia*.

Zha, Z. J., Yang, L., Mei, T., Wang, M., and Wang, Z. 2009. Visual query suggestion. In *Proceedings of the ACM International Conference on Multimedia*.

Received April 2012; revised October 2012; accepted April 2013