Robust Visual Reranking via Sparsity and Ranking Constraints

Nobuyuki Morioka*
The University of New South Wales & NICTA
Sydney, Australia
nmorioka@cse.unsw.edu.au

Jingdong Wang
Microsoft Research Asia
Beijing, China
jingdw@microsoft.com

ABSTRACT

Visual reranking has become a widely-accepted method to improve traditional text-based image search engines. Its basic principle is that visually similar images should have similar ranking scores. While existing methods are different in specific, almost all of them are based on explicit or implicit pseudo-relevance feedback (PRF). Explicit PRF-based approaches, including classification-based and clustering-based reranking, suffer from the difficulty of selecting reliable positive and negative samples. Implicit PRF-based approaches, such as graph-based and Bayesian visual reranking, deal with such unreliability by making use of the initial ranking in a soft manner, but have limited capability of promoting relevant images and lowering down irrelevant images.

In this paper, we propose ℓ_1 square loss optimization based on sparsity and ranking constraints to detect confident samples which are most likely to be relevant to a query. Based on the discovered confident samples, we present an adaptive kernel-based scheme to rerank the images. The success of our proposed method comes from another important observation that irrelevant images, whether initially positioned at the top or bottom, are usually less-popular and more diverse than relevant images. Therefore, it is robust against outlier images and suitable when relevant images are multimodally distributed. The experimental results demonstrate significant improvement of our method over several existing reranking approaches on both MSRA-MM V1.0 and Web Queries datasets.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Retrieval models

General Terms

Algorithms, Experimentation, Performance

*This work was performed when Nobuyuki Morioka was visiting Microsoft Research Asia as a research intern.

MM '11 Arizona, USA

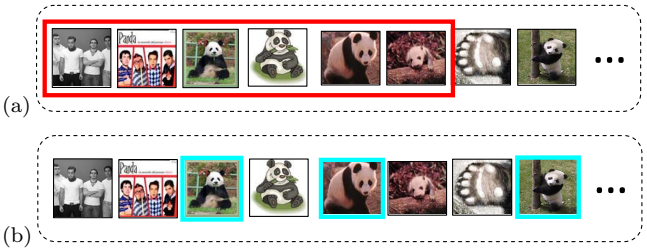

Figure 1: Given a query "panda", (a) pseudo-relevance feedback selects the top images as positive relevant samples indicated in red, and (b) our proposed approach in this paper finds confident image samples indicated in cyan that are thought to be highly relevant to "panda" from the initial rank list via our sparsity and ranking constrained optimization framework.

Keywords

Image Search Reranking, Non-negativity, Sparsity, ℓ_1 Loss

1. INTRODUCTION

Due to ever growing number of online images, image search and retrieval is one of active and challenging areas of multimedia research. Many image search engines such as Bing, Google and Yahoo! have relied on matching textual information of the images, including surrounding texts, titles, captions and URLs, against textual queries given by users. However, because the textual information fails to capture rich image content, their returned results are often inconsistent with the queries [7]. While visual features like color moments [20] and SIFT [15] can be a substitute for the textual information, there is a huge semantic gap between such low-level statistical visual features and the high-level semantic concepts (or queries) [18].

Visual reranking which integrates both the textual and visual information is proposed to return visually satisfying search results. Based on the textual information, it first quickly scans through a whole database to generate a list of images that are potentially relevant to a given query. The images in the list are initially ranked according to their relevance scores calculated from their corresponding textual information. Based on their visual similarities, the images are reordered such that visually similar images are reranked together.

The common idea in many existing reranking algorithms

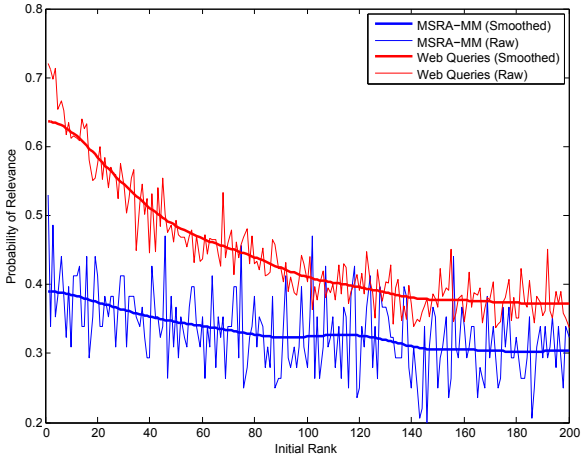

Figure 2: Probability of relevance vs. initial rank position for the MSRA-MM V1.0 and Web Queries datasets. Web Queries have initial ranks highly correlated with relevance. In contrast, MSRA-MM V1.0 contains more noise when the raw data is compared. When smoothed for further analysis, it has low, yet similar relevance probability across all rank positions. Thus, depending on the search engine or dataset used, the pseudo-relevance feedback assumption might not hold.

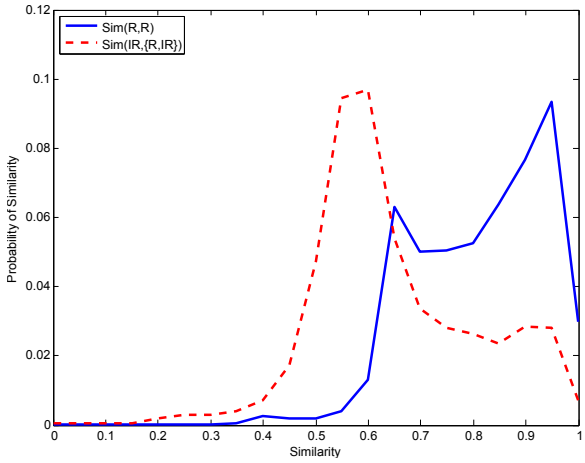

Figure 3: Distributions of visual similarity between images from MSRA-MM V1.0 plotted. The blue curve is the distribution of the similarity between relevant images. The red is the distribution of the similarity from irrelevant images to all images. The relevant images tend to share visual patterns with each other more often than irrelevant images.

is that they implicitly or explicitly utilize pseudo-relevance feedback (PRF) [1]. Explicit PRF-based approaches, including classification-based [29] and clustering-based [4] approaches, take top images as pseudo-positive examples and/or bottom images as pseudo-negative examples to either train a classifier or cluster images to perform reranking. However, the selected pseudo-positive and pseudo-negative examples are not reliable and may potentially be false positives and false negatives respectively which will affect the reranking performance. Figure 1 (a) shows an example of such a scenario where the first two images from the initial ranking will be selected as pseudo-positive, but are in fact false positives. Figure 2 also illustrates that the PRF assumption may not always hold for every dataset used. In contrast, implicit PRF-based approaches, including graph-based [5, 7] and Bayesian visual reranking [22], overcome such unreliability by making use of the initial ranking in a soft manner. It encourages relevant images that are positioned at the bottom to move up by receiving reinforcement from other relevant images positioned at the top. However, its major shortcoming is that it cannot push down irrelevant images that are initially ranked high, as the irrelevant images tend to be more visually diverse than the relevant ones. The diversity of irrelevant images is depicted in Figure 3. Therefore, whether they are explicit or implicit, the PRF-based approaches suffer from unreliability of the initial ranking list where false positives may be selected as pseudo-positive examples.

The explicit PRF-based approaches can theoretically be improved by selecting different number (N) of top images for each query. According to our simple experiments performed on the publicly available MSRA-MM V1.0 dataset [27], manually selecting the best performing result from different N for each query improves the performance significantly as shown in Figure 4. Although it would be an interesting problem to automatically select N for each query, this is rather difficult to solve and only partially overcomes the unreliability issue. It can exclude irrelevant images that appear around the bottom part of the top images, but it cannot handle the cases similar to the one in Figure 1 where irrelevant images are occupying the first two positions. These irrelevant images will always get selected no matter what N is.

Therefore, we instead redefine the problem as finding a set of confident image samples that are relevant to the query from the top-ranked images. This can be seen as relaxation, because the set of images considered are no longer need to be consecutive (starting from the very top), but are selectively chosen from the top-ranked images. To verify our intuition, another simple experiment is performed where a small number of relevant images from the ground truth are randomly sampled as reference points for reranking. The results are reported in Figure 4. With a small set of relevant samples, the performance improves - even better than when different N is chosen for each query.

Given the new problem of finding confident samples, this paper proposes two optimization methods based on sparsity and ranking constraints to solve the problem. The methods described in this paper utilize the fact that relevant images share visual patterns more often than irrelevant images do. After the confident samples are detected, we use them as reference points to rerank images with an adaptive kernel-based scheme. Our approach may be similar to the exemplar-based classification, but ours is fully unsupervised.

In summary, the main contributions of this paper are as follows:

- We reformulate the new problem of finding confident images from the initial search results, which relaxes the assumption of the psuedo-relevance feedback and

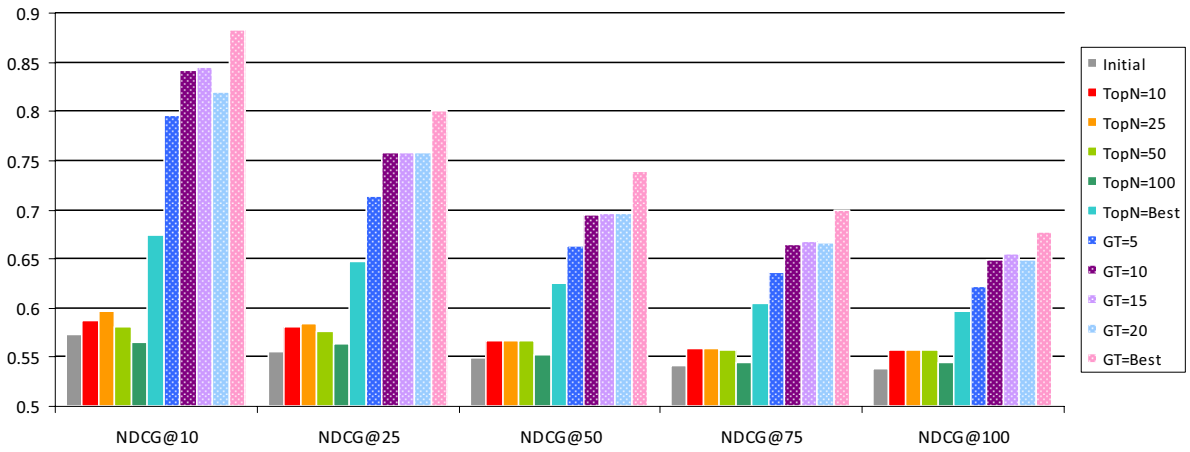

Figure 4: Simple experiments are performed on MSRA-MM V1.0 [27]. TopN=N selects top $N \in \{25, 50, 75, 100\}$ images as reference points for reranking. In contrast, for TopN=Best, we manually inspect different N and choose the best performing one for each query. Hence, its performance improvement can only be achieved theoretically. For GT=N, we manually pick a small number (N) of relevant images from top 100 images using the ground truth labels. While this is ideal, it achieves significantly better performance than PRF-based TopN - motivating our work on automatically detecting confident samples. NDCG is used for performance evaluation (c.f. Section 4.1)

does not assume that all images in the top-N have to be relevant to a query.

- We propose and solve two ℓ_1 loss minimization problems with sparsity and ranking constraints, which result in discovering highly-confident samples and are robust against outlier images that are initially ranked highly.
- We rerank images based on the detected confident samples with a kernel-based approach, which successfully promotes relevant images and lowers down irrelevant images. This greatly improves the reranking performance.

The rest of this paper is organized as follows. Section 2 reviews existing work on visual reranking. Section 3 describes our robust visual reranking approach based on sparsity and ranking constrained optimization framework. Section 4 demonstrates the effectiveness of our approach empirically. Section 5 concludes this paper with the future work.

2. RELATED WORK

This section reviews existing visual reranking methods for both image and video retrieval, as the underlying mechanisms of reranking images and videos are generally similar. Broadly speaking, many existing reranking techniques are either supervised or unsupervised.

The supervised methods generally learn a model from human-labeled data offline and use it during the online reranking process [10, 30]. They tend to focus more on designing good reranking features. Other notable ones include using semantic concepts [17], exploiting the click-through data [6], and analyzing the contextual information [31]. On the other hand, the unsupervised methods are more practical than the supervised ones, as they do not assume any human-labeled data to perform reranking. They are either classification-based [13, 28, 29], graph-based [5, 7, 22] or clustering-based [4].

The classification-based method usually consists of three steps. First, under the pseudo-relevance feedback assumption originally introduced in text retrieval [1], both pseudo-positive and pseudo-negative images are sampled. Then, a classifier is trained using either SVM [29], boosting [28] or ranking SVM [13]. Finally, the trained classifier is used to score each image. The key limitation of this method is that the selected pseudo-positive and pseudo-negative images may not be truly-positive and truly-negative. Hence, it is not always guaranteed to learn a robust classifier.

In contrast to the classification-based method, the graph-based method [5, 7, 22] integrates both the initial ranking and visual consistency between images. It constructs a graph where every pair of visually similar images are connected by an edge and the initial rank information is propagated through the graph until convergence. Unfortunately, the graph-based method has no formal mechanism to explicitly penalize outlier images that are dissimilar from other images since it only reranks visually similar images together. Thus, in theory, if one irrelevant image is placed at the top of the initial search result, it will happily stay there because irrelevant images have often small similarities with other irrelevant images and be pushed down at a lower position only in the rare case of its few similar images having low initial ranks.

The clustering-based method groups images based on their visual similarity as well as the initial ranking information [4, 14]. In [4], the information bottleneck principle is used to derive the best groups that maximize mutual information between clusters and labels. It is shown to perform well on videos where near duplicate shots are often found. Typicality-based reranking [14] is concerned with finding

samples that are typical among all relevant and irrelevant images by exploring the initial ranking too. The two approaches are based on the assumption that both relevant and irrelevant images are visually compact, which may hold in the benchmark TREC video retrieval evaluation [16]. But it does not hold in web image search as shown in Figure 3. Moreover, as concluded in [14], it relies heavily on the initial ranking and suffers from the drawback of graph-based methods where it is limited in lowering down irrelevant images that are initially ranked high.

To overcome this outlier problem, the noise resistant ranking method in [12] computes spectral decomposition of a Laplacian matrix and filters noisy images with sparse eigenbasis fitting. Our reconstruction-based approach with sparsity and ranking constraints also tackles the same problem, but directly seeks a sparse set of confident samples that are relevant to a query. These detected samples are used as reference points to perform reranking such that dissimilar and irrelevant images are ranked low. The sparsity constraint is also adopted in [21] to avoid overfitting of a linear ranking model in the context of directly optimizing ranking measures. In our case, the sparsity constraint is applied on the confident sample detection, not on the ranking model, to only select samples that are relevant with high confidence.

3. ROBUST VISUAL RERANKING

Before describing our approach in detail, we formerly define what visual reranking is. We let an initial ranking list and its corresponding set of visual features of one query be $\bar{\mathbf{r}} = [\bar{r}_1, \bar{r}_2, \cdots, \bar{r}_M]^T \in \mathbb{R}^M$ and $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M]$ respectively where M is the total number of the returned results. We define reranking as:

$$\mathbf{r} = f(\mathbf{X}, \bar{\mathbf{r}}),$$

where f is a function of both $\mathbf{X}$ and $\bar{\mathbf{r}}$ returning another ranking list $\mathbf{r} = [r_1, r_2, \cdots, r_M]^T \in \mathbb{R}^M$ which is considered to be a refined version of $\bar{\mathbf{r}}$.

Our approach consists of two main steps as illustrated in Figure 5. The first step finds confident image samples from the initial search results. We define a function of detecting confident samples as:

$$\mathbf{c} = g(\mathbf{X}, \bar{\mathbf{r}}),$$

where $\mathbf{c}$ is a binary vector. If c_i is 1, then its corresponding image I_i is a confident sample that will be used later for reranking. The second step uses the detected confident samples as reference points to rerank the images and such a reranking step is defined as:

$$\mathbf{r} = h(\mathbf{X}, \mathbf{c}).$$

In the next two sections, we describe these two steps in detail.

3.1 Confident Sample Detection

Our confident sample detection is based on an intuition that relevant images share visual patterns with each other more often than irrelevant images, as the relevant ones tend to be less diverse compared to their counterparts. This is illustrated in Figure 3. Therefore, for each image, we first compute its total similarity degree to all other images in the initial ranking list. For a relevant image, it should be similar to other relevant images making its total similarity degree high. Whereas for an irrelevant image, it should be dissimilar to most images regardless of whether they are relevant or irrelevant. Hence, its total similarity degree should be low.

To measure the total similarity degree, we compute a kernel matrix $\mathbf{K} \in \mathbb{R}^{M \times M}$. Each entry k_{ij} of the matrix stores the similarity between two visual features $\mathbf{x}_i$ and $\mathbf{x}_j$, except that we set all self-similarities to 0. We sum the elements in each row of $\mathbf{K}$ to obtain a column vector $\mathbf{s} = \mathbf{K}\mathbf{e}$ where $\mathbf{e}$ is a vector of all ones. We let each entry s_i of $\mathbf{s}$ be the total similarity degree of an image indexed at i to all other images.

In order to analyze which images *mainly* contribute to the composition of $\mathbf{s}$, we propose a reconstruction-based approach which aims to minimize the error between $\mathbf{s}$ and $\mathbf{K}\mathbf{c}$. If a particular image contributes a lot, then it should be significantly similar to other images and it is most likely to be relevant to the query. As motivated from Figure 4, we only need a few confident samples that are truly relevant to the query to obtain good performance. To achieve this, we add a sparsity constraint on $\mathbf{c}$ to the reconstruction error yielding the following formulation:

$$\begin{aligned} &\min_{\mathbf{c}} \ \|\mathbf{s} - \mathbf{K}\mathbf{c}\|_q^p + \alpha\phi(\bar{\mathbf{r}}, \mathbf{c}) \\ &\text{s.t. } \mathbf{c} \in \{0,1\}^M, \end{aligned} \tag{1}$$

where α is a trade-off parameter that balances the two terms and $\phi(.)$ takes $\bar{\mathbf{r}}$ as one of the input to include the preference of selecting confident samples based on the initial ranking list.

While the above formulation should result in finding a less complex and more robust solution that has as few non-zero entries as possible in $\mathbf{c}$ and achieving a low reconstruction error, it is a binary integer programming problem which is generally NP-hard. Since exhaustive search is computationally expensive, it is not feasible in practice to solve in that way. Hence, we introduce two relaxation schemes, namely non-negative least squares and bounded-variable least squares, such that it can be solved efficiently and effectively.

3.1.1 Non-Negative Least Squares

In this section, we describe how the formulation given in Eqn. (1) is relaxed into the non-negative least square problem which can theoretically be solved in a polynomial time. We introduce an auxiliary non-negative real vector $\mathbf{z} \in \mathbb{R}^M_{+,0}$ which infers $\mathbf{c}$ as follows:

$$\mathbf{c} = \mathbf{1}_{(\mathbf{z}>0)}, \tag{2}$$

where $\mathbf{1}_{(.)}$ is an element-wise indicator function returning 1 if z_i is positive and 0 otherwise. We replace $\mathbf{c}$ with $\mathbf{z}$ in Eqn. (1) to obtain the following minimization problem:

$$\begin{aligned} &\min_{\mathbf{z}} \ \|\mathbf{s} - \mathbf{K}\mathbf{z}\|_1^2 + \alpha\|\mathbf{D}\mathbf{z}\|_1^2 \\ &\text{s.t. } \mathbf{z} \geq \mathbf{0}. \end{aligned} \tag{3}$$

The ℓ_1 square loss is adopted to measure the reconstruction error between $\mathbf{s}$ and $\mathbf{K}\mathbf{z}$, as the ℓ_1 loss has been shown to tolerate more against outlier images which do not really share visual patterns with other images [3]. This is not the case for the ℓ_2 loss which is known to be sensitive when outliers are present in data. As for the constraints on $\mathbf{z}$, we have the ℓ_1 norm as the sparsity constraint and introduce a diagonal matrix $\mathbf{D}$ to include the initial ranking as prior information. The ℓ_1 norm is preferred over the ℓ_2 norm, as

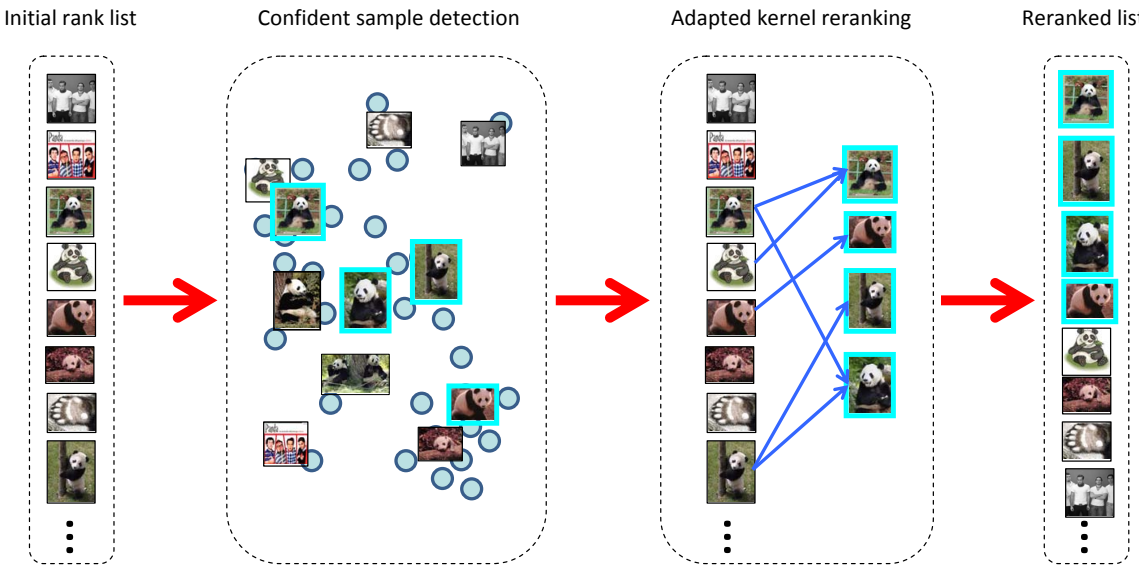

Figure 5: An illustration of robust visual reranking. In the second block, images with cyan bounding boxes correspond to the detected highly-confident samples. Blue circles indicate other images

the latter does not encourage to suppress the entries in $\mathbf{z}$ to zero. Each diagonal entry of $\mathbf{D}$ acts as a weight for the corresponding entry in $\mathbf{c}$. The lower the weight, the higher the chance of the corresponding entry of $\mathbf{c}$ being non-zero and considered as a confident sample. Therefore, $\mathbf{D}$ biases the preference of choosing confident samples more towards the initially-highly-ranked images. We will discuss different solutions in instantiating the matrix $\mathbf{D}$ in Section 3.1.3.

The formulation given in Eqn. (3) may lead to an over-reconstruction of $\mathbf{s}$ by $\mathbf{Kz}$ where some entries of $\mathbf{s}-\mathbf{Kz}$ are negative. As a consequence, among the computed confident samples, some samples may not be truly relevant. The sparsity constraint on $\mathbf{z}$ can handle the problem as the over-reconstruction often results in high ℓ_1 norm of $\mathbf{z}$, but with no gaurantee. To handle this issue properly, we explicitly introduce another constraint of $\mathbf{s}-\mathbf{Kz} \geq \mathbf{0}$ to directly avoid it. This additional constraint in fact helps us to rewrite Eqn. (4) into the form similar to the well-studied non-negative least squares problem [8, 25] given as:

$$\begin{aligned} \min_{\mathbf{z}} \left\| \begin{bmatrix} \mathbf{e}^T\mathbf{s} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{e}^T\mathbf{K} \\ \sqrt{\alpha}\mathbf{e}^T\mathbf{D} \end{bmatrix} \mathbf{z} \right\|_2^2 \\ \text{s.t. } \mathbf{z} \geq \mathbf{0} \\ \mathbf{s} - \mathbf{Kz} \geq \mathbf{0}. \end{aligned} \tag{4}$$

This is essentially a quadratic programming problem which can be solved in polynomial time with any standard quadratic programming solver. The images corresponding to the non-zero entries of the optimal $\mathbf{z}$ are treated as the confident samples of the query.

3.1.2 Bounded-Variable Least Squares

With the formulation given in Eqn. (4), $\mathbf{z}$ can be efficiently solved, but it potentially has one undesired consequence where $\mathbf{z}$ can take values larger than 1. This still results in selecting fewer truly-relevant images and more false-positive images to well fit the reconstruction of the irrelevant images. To resolve this issue, we introduce lower and upper bounds for $\mathbf{z}$ such that its entries will be between 0 and 1. This naturally guarantees the constraint of $\mathbf{s}-\mathbf{Kz} \geq \mathbf{0}$ and turns Eqn. (4) into bounded-variable least squares [2, 19] given as:

$$\begin{aligned} \min_{\mathbf{z}} \left\| \begin{bmatrix} \mathbf{e}^T\mathbf{s} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{e}^T\mathbf{K} \\ \sqrt{\alpha}\mathbf{e}^T\mathbf{D} \end{bmatrix} \mathbf{z} \right\|_2^2 \\ \text{s.t. } \mathbf{0} \leq \mathbf{z} \leq \mathbf{1}. \end{aligned} \tag{5}$$

The above formulation improves the ability of fitting the reconstruction of the relevant images well and excluding the irrelevant images with the help of the ℓ_1 loss. This allows us to further reduce false positives of the confident samples.

3.1.3 Sparsity and Ranking Constraints

The two proposed formulations are both driven by sparsity and ranking constraints on $\mathbf{z}$. The sparsity constraint aims to discover a few images that are indeed relevant to the query such that they become reference points to perform the kernel-based reranking described later. The ranking constraint utilizes the initial ranking $\bar{\mathbf{r}}$ as an additional constraint on $\mathbf{z}$ such that initially highly-ranked images have a higher chance of being selected as the confident samples. This is motivated by Figure 2 indicating top-ranked images often have a higher probability to be relevant. The ranking constraint is achieved by introducing a diagonal matrix $\mathbf{D}$ and linearly combining it with $\mathbf{z}$ in the ℓ_1 norm. The important question is how to set each diagonal entry d_{ii} of $\mathbf{D}$ in order to achieve the preference of the confident samples according to the initial ranking. We give three possible solutions, namely linear, step-wise and shrinkage, which are defined as follows:

$$d_{ii} = \eta(i+\nu), \tag{6}$$
$$d_{ii} = \eta\lceil (i+\nu)/\epsilon \rceil, \tag{7}$$
$$d_{ii} = \eta(\max(i-\epsilon, 1)+\nu)), \tag{8}$$

where η acts like a normalization constant to make the en-

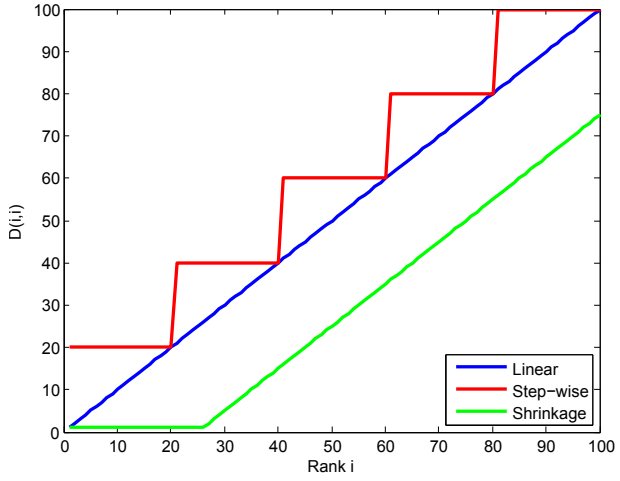

Figure 6: Three possible penalty matrices for the ranking constraint, namely linear in blue, step-wise in red and shrinkage in green.

tries of $\mathbf{D}$ sum up to 1. The diagonal entries of the three different matrices are plotted in Figure 6. In the case of $\mathbf{D}$ being set to an identity matrix, we have no confident sample preference towards the initial ranking, i.e. all images are equally likely to be a confident sample.

Our sparsity and ranking constraints differ from the adaptive Lasso [32], as we have either non-negativity or bounds to further constrain $\mathbf{z}$ and the weights defined in $\mathbf{D}$ are more natural to combine the additional useful initial ranking information. To make our problem computationally more feasible, we use $\|.\|_1^2$ instead of $\|.\|_1$ norm. Other penalty matrices, such as fused Lasso [23], ℓ_1 trend filtering [9] and generalized Lasso [24] where off-diagonal entries of the matrices are also non-zero, have been explored by the statistical community before. It is not yet clear and still open how we can utilize them in the reranking problem.

3.2 Adaptive Kernel-Based Reranking

After discovering the confident samples using one of the two optimization methods described, we use the kernel-based approach to rerank the images in the initial ranking list. We choose the kernel-based reranking over the graph-based reranking, due to its simplicity and computational efficiency. The justification of the kernel-based approach is also supported by Figure 3 presented earlier in this paper. The similarities between relevant images are generally higher than those between relevant and irrelevant images and between irrelevant images. Therefore, the relevant images receive higher support from the discovered confident samples while the irrelevant images receive lower support. Even when false-positive images are selected as confident samples, they will not promote the ranks of other irrelevant images to go up because the similarities to those irrelevant images are low.

To formalize, for each image, we compute its rank by summing the responses of the Gaussian kernels with the discovered confident samples which is given as:

$$r_i = \sum_{m=1}^{M} \mathbf{1}_{(z_m>0)} \exp\left(\frac{\|\mathbf{x}_i - \mathbf{x}_m\|^2}{2h^2}\right). \quad (9)$$

Instead of the Gaussian kernels used above, we may generalize the formula as:

$$r_i = \sum_{m=1}^{M} \mathbf{1}_{(z_m>0)} K(\mathbf{x}_i, \mathbf{x}_m). \quad (10)$$

Such a generalization is useful when the only available information is a kernel matrix (not raw visual features). This implies that the whole reranking algorithm can work directly from the kernel matrix, as the confident sample detection is also based on the kernel matrix.

In summary, we outline the whole reranking procedure as follows:

1. Retrieve visual features $\mathbf{X}$ of images in the initial rank list of a query.
2. Obtain a kernel matrix by computing cosine similarity between visual features (i.e. $\mathbf{K} = \mathbf{X}^T\mathbf{X} - \mathbf{I}$). We suppress self-similarity.
3. Compute the total similarity vector $\mathbf{s}$ by summing the columns of $\mathbf{K}$, i.e. $\mathbf{s} = \mathbf{K}\mathbf{e}$.
4. Find the optimal $\mathbf{z}$ according to Eqn. (4) or Eqn. (5) to find confident samples.
5. Infer the final rank $\mathbf{r}$ with Eqn. (9).

3.3 Time Complexity

We analyze the time complexity of our reranking algorithm. Given the initial ranking list of a query, we first compute a kernel matrix $\mathbf{K}$ which takes $O(M^2)$ where M is the length of the ranking list. To run the optimization to detect confident samples from the top N images, it runs ordinary least squares on an active set of variables over N iterations where each iteration takes $O(N^3)$. In practice, it often converges in significantly less than N iterations and the number of active variables at each iteration is also smaller than N. Therefore, it can be solved very efficiently. After the confident samples are found, the adaptive kernel-based reranking takes $O(MN)$ to compute reranked scores. Since the purpose of our algorithm is to rerank the initial search results, its speed is independent to the number of images in a database. On a desktop with 2.4GHz Intel Core Duo and 2GB memory, our MATLAB implementation takes about 0.3 seconds only to rerank each query for M and N being 850 and 100 respectively. We believe this is reasonable for a real-time application.

4. EXPERIMENTS

4.1 Experimental Setup

Experiments are conducted on the MSRA-MM V1.0 [27] and Web Queries [10] datasets both of which are publicly available for download. The MSRA-MM dataset consists of 68 diverse yet representative queries collected from the query log of Bing. In total, it contains 60,257 images and there are roughly 900 images for each query. As for the ground truth labels, it has three relevant levels from 0 to 2: (0) irrelevant, (1) fair, and (2) relevant. The Web Queries dataset, on the other hand, consists of 354 representative queries collected from Google. In total, it contains 71,478 images and there are about 200 images for each query. In contrast to MSRA-MM, while the number of queries is much larger for Web Queries, the images returned per query are much low (some

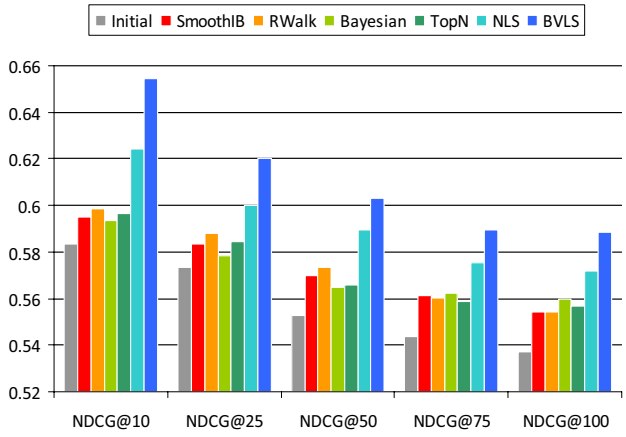

Figure 7: Experimental results on MSRA-MM V1.0.

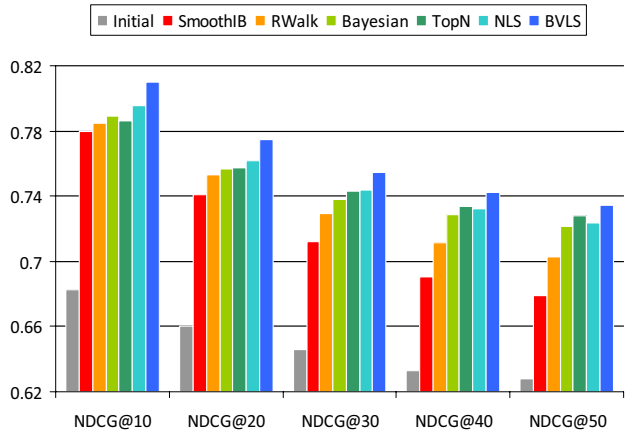

Figure 8: Experimental results on Web Queries.

queries have less than 20 images). The ground truth labels are in binary: (0) irrelevant and (1) relevant.

As for the visual features, we use the thumbnail images instead of the original images to densely sample 128 dimensional SIFT descriptors from 16×16, 24×24 and 32×32 sized patches at every 8 pixels step from each image in both datasets with VLFeat [26]. Then, a random collection of 200,000 descriptors is used to learn 1024 visual words with k-means clustering and the bag-of-words descriptor is computed for each image. We include the spatial location information of each local feature as well [11]. To detect the confident samples, the top 100 images are considered as candidates for each query. Accordingly the dimension of the kernel matrix $\mathbf{K}$ becomes $M \times 100$ and the total similarity vector $\mathbf{s}$ is calculated from this matrix. The linear weight is adopted for $\mathbf{D}$ to induce the ranking constraint. For each experiment, the bandwidth for the adaptive kernel-based reranking is fixed as 0.8 and 1.5 for MSRA-MM and Web Queries respectively. To perform the actual optimization of non-negative least squares and bounded-variable least squares, we use MATLAB's lsqlin(.) which solves least squares with linear constraints.

In order to measure the reranking performance for each query, we adopt the Normalized Discounted Cumulated Gain (NDCG) which is a standard evaluation method in information retrieval when relevance levels are more than two. For a given query, the NDCG at position p can be calculated as

$$\text{NDCG@}p = \frac{1}{Z} \sum_{j=1}^{p} \frac{(2^{t_j} - 1)}{\log(1 + j)},$$

where t_j is the relevance level of the jth image in the refined ranking list. Z is a normalization constant which makes NDCG@p be 1 for a perfect ranking. For both MSRA-MM V1.0 and Web Queries datasets, Z is calculated based on the ground truth labels provided. To compute the overall performance, NDCGs are averaged over all queries for each dataset.

4.2 Performance Comparison

We compare the following reranking methods to demonstrate the effectiveness of our approach.

- SmoothIB [4] - A clustering based approach based on pseudo-relevance feedback and information bottleneck clustering over visual features with the help of the smoothed initial ranking.
- Random walk reranking (RWalk) [5, 7] - A popular graph-based method that integrates initial ranking and visual similarity between images.
- Bayesian reranking (Bayesian) [22] - Another graph-based method that exploits initial ranking by considering their pairwise difference.
- TopN - The top N images for each query are assumed to be confident samples and the adaptive kernel-based reranking (Eqn. (10)) is performed. N is tuned (as 25) in a brute-force manner to get the best performance.
- NLS - Confident samples are detected based on non-negative least squares (Eqn. (4)) and then the adaptive kernel-based reranking is performed.
- BVLS - Confident samples are detected based on bounded-variable least square (Eqn. (5)) and then the adaptive kernel-based reranking is performed.

Choice of the parameters for each existing method is critical for optimal performance. We have carefully evaluated different parameters and the best performing result is reported in the performance comparison for each method.

4.2.1 *MSRA-MM V1.0*

The results for the MSRA-MM dataset are shown in Figure 7. It is clear from the results that our reranking methods based on the confident sample detection outperform other existing reranking methods across all evaluated NDCGs. When compared against the improvement of both NLS and BVLS from the initial ranking, other reranking methods do not improve much especially for NDCGs at earlier positions. The main reason for this is that the PRF assumption, which is used explicitly in SmoothIB and TopN and implicitly in Graph and Bayesian, does not hold well for this dataset. As previously shown in Figure 2, the probability of relevance given a rank position stays relatively low and similar across all ranks. Hence, the results on this dataset demonstrate

Method	MAP
Initial	0.570
Smooth [4]	0.616
RWalk [5, 7]	0.634
Bayesian [22]	0.658
TopN	0.666
NLS	0.664
BVLS	**0.670**
Query-Rel. [10]	0.649

Table 1: Mean average precision evaluated for different reranking methods on the Web Queries dataset.

the robustness of our confident sampled-based rereanking methods against the unreliability of PRF.

4.2.2 Web Queries

Figure 8 shows that for the Web Queries dataset, the performance of all methods are generally better against the initial rank. This is because unlike the previous dataset, the pseudo-positive examples are more reliable for this dataset as shown in Figure 2. While the existing methods perform well on this dataset, BVLS achieves better performance. NLS achieves better results than TopN and Bayesian at earlier positions and comparable results at later positions. As consistently observed from these two datasets, for earlier NDCGs, BVLS achieves high performance - implying it is effective in pulling more relevant images to appear in the first 10 to 20 positions of the reranked list. This is useful in practice, as users tend to see the first 10 to 20 images more often than the images at later positions.

Since the relevance labels provided by this dataset are in binary, we also report mean average precision (MAP). BVLS and NLS obtain MAP of 0.670 and 0.664 respectively. The results for other reranking methods are shown in Table 1. For a fair comparison, we report the result for the query-relative feature reranking [10] based only on the visual features instead of the one utilizing both the visual and textual features, as both BVLS and NLS exploit the visual features only. It is worth noting that TopN achieve marginally better than NLS. This may be due to the fact that the PRF assumption holds well for Web Queries and the number of top images to consider is tuned extensively. Since MAP takes into account of the whole image list unlike NDCG, it suggests that while BVLS is effective in pulling the relevant images to high positions, it tends to not care too much about the ordering of the bottom images.

4.3 Empirical Analysis

This section analyzes our reranking algorithm in detail. In particular, we check the correctness of the detected confident samples and evaluate the parameters associated to the sparsity and ranking constraints as well as the kernel-based reranking.

4.3.1 The Correctness of Confident Samples

We evaluate how many of the confident samples detected with our approach are indeed correct according to the ground truth. We use precision to measure the performance of the confident sample detection. The results are reported in Table 2. When compared against the method that simply picks the top N number of images based on the PRF assumption, both of our optimization methods, in particular BVLS, achieve better precision. Not only our approach picks the confident samples more accurately, it also determines the number of confident samples for each query automatically.

Method	MSRA-MM	Web Queries
TopN=25	0.384/0.890	0.621
TopN=50	0.370/0.882	0.569
TopN=100	0.349/0.862	0.509
NLS	0.393/0.892	0.638
BVLS	**0.423/0.916**	**0.682**

Table 2: Precision of confident sample detection based on non-negative least squares (NLS) and bounded-variable least squares (BVLS). For MSRA-MM V1.0, since there are three different relevance levels, we evaluate the precision in two ways. The first is to only consider relevance scale of 2 being positive and the rest being negative. The second is to treat relevance scales of 2 and 1 as positive.

For qualitative analysis, Figure 9 shows the confident samples detected from images associated with the query of "panda" from the MSRA-MM dataset. While there are several irrelevant images in the initial ranking list (especially, the first two images are irrelevant), our confident sample detection has successfully discovered the confident samples that are relevant. The kernel-based reranking based on these confident samples achieves visually much better results.

4.3.2 The Sparsity and Ranking Constraints

In this section, we study the effect of changing both the sparsity parameter α and the ranking constraint for BVLS on MSRA-MM. We evaluate α in a range between 20 and 160 with a 20 step increment and use the three ranking constraints described in Section 3.1.3, i.e. linear, step-wise and shrinkage. We also compare against when no initial ranking preference is used, i.e. the ranking constraint matrix set to an identity. For all three ranking constraints, ν is set to 50, such that the weight assigned at each rank position does not decrease sharply. This is based on the observation of Figure 2 for MSRA-MM that the relevance probability is relatively uniform compared against Web Queries. For both step-wise and shrinkage constraints, ϵ is set to 10. η for all three constraints normalizes the sum of the diagonal entries of $\mathbf{D}$ to be 1.

The results are shown in Figure 10. For both NDCG@10 and NDCG@50, the peaks of the four different constraints (including no ranking constraint) are observed when the sparsity is 120. While the performance of no ranking constraint is still better than other existing reranking methods on this dataset (c.f. Figure 7), we can see the clear advantage of including the initial ranking as the confident sample selection preference. All three ranking constraints are better than the one with no ranking preference. For NDCG@10, the linear constraint clearly performs the best amongst the three ranking constraints, and for NDCG@50, all constraints achieve about the same performance. This suggests that unless we confidently know what kind of ranking constraint is suited for the dataset we use, it is better to use the simple linear constraint, as both step-wise and shrinkage require an additional parameter to be tuned.

Figure 9: A qualitative analysis of our reranking approach for the query of "panda" from the MSRA-MM V1.0 dataset. (a) The first 50 images from the initial rank list shown. Each image has a color-coded border. Green, blue and red indicate relevance scales 2, 1 and 0 respectively. (b) Confident samples selected from the initial rank list are shown and unselected ones are suppressed for clarity. (c) The first 50 images of the reranked list shown.

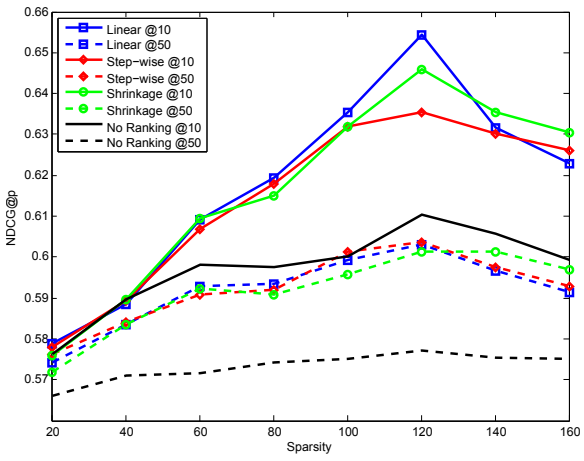

Figure 10: Different sparsity parameters and ranking constraints are evaluated. The ranking constraint improves from just having the sparsity constraint of z.

4.3.3 The Bandwidth of Gaussian Kernel

The adaptive kernel-based reranking use a Gaussian kernel which requires the bandwidth parameter h to be tuned. It may be desirable to select a different bandwidth for each query. However, this is beyond the scope of this paper. Also, Figure 11 indicates that our approach is not very sensitive to the selection of the bandwidth parameter. We evaluate different bandwidths selected between 0.5 and 1.0 with a 0.1 step increment. While the appropriate bandwidth selection seems slightly critical to obtain the optimal performance for NDCG@10, other NDCGs is not so sensitive to the bandwidth selection.

5. CONCLUSION AND FUTURE WORK

In this paper, we have formulated a problem of finding confident image samples that are relevant to a given query. Two optimization solutions that exploit sparsity and ranking constraints are proposed, namely non-negative least squares and bounded-variable least squares. The detected confident samples are used as reference points for the kernel-based approach to rerank images. The experimental results have shown that the performance of our approach is significantly better than other existing reranking approaches. It can also be computed efficiently. For our future work, we would like to explore other optimization techniques to further reduce false-positive confident samples.

6. REFERENCES

[1] J. Carbonell, Y. Yang, R. Frederking, R. Brown,

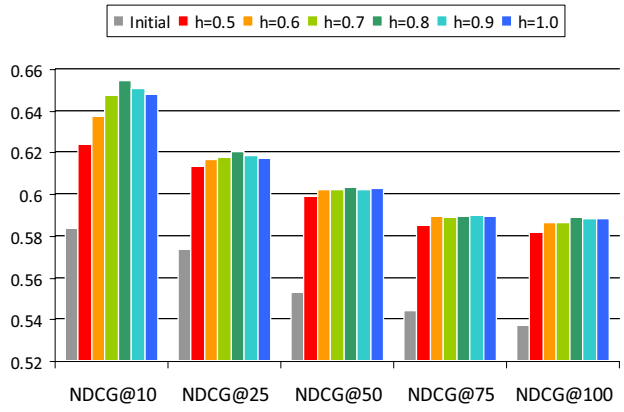

Figure 11: Different bandwidths of the Gaussian kernel used for reranking are evaluated on the MSRA-MM V1.0 dataset.

Y. Geng, and D. Lee. Translingual information retrieval. In *IJCAI*, 1997.

[2] T. Coleman and Y. Li. A Reflective Newton Method for Minimizing a Quadratic Function Subject to Bounds on Some of the Variables. *SIAM Journal on Optimization*, 1996.

[3] A. Eriksson and A. van Den Hengel. Efficient computation of robust low-rank matrix approximations in the presence of missing data using the L1 norm. In *CVPR*, 2010.

[4] W. Hsu, L. Kennedy, and S.-F. Chang. Video search reranking via information bottleneck principle. In *ACM Multimedia*, 2006.

[5] W. Hsu, L. Kennedy, and S.-F. Chang. Video search reranking through random walk over document-level context graph. In *ACM Multimedia*, 2007.

[6] V. Jain and M. Varma. Learning to re-rank: query-dependent image re-ranking using click data. In *WWW*, 2011.

[7] Y. Jing and S. Baluja. VisualRank: Applying PageRank to Large-Scale Image Search. *PAMI*, 2008.

[8] H. Kim and H. Park. Sparse non-negative matrix factorization via alternating non-negativity-constrained least squares for microarray analysis. *Bioinformatics*, 2007.

[9] S.-J. Kim, K. Koh, S. Boyd, and D. Gorinevsky. L1 trend filtering. *SIAM Review*, 2009.

[10] J. Krapac, M. Allan, J. Verbeek, and F. Jurie. Improving web-image search results using query-relative classifiers. In *CVPR*, 2010.

[11] S. Lazebnik, C. Schmid, and J. Ponce. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In *CVPR*, 2006.

[12] W. Liu, Y.-G. Jiang, J. Luo, and S.-F. Chang. Noise Resistant Graph Ranking for Improved Web Image Search. In *CVPR*, 2011.

[13] Y. Liu, T. Mei, J. Tang, X. Wu, and X.-S. Hua. Learning to video search rerank via pseudo preference feedback. In *ICME*, 2008.

[14] Y. Liu, T. Mei, M. Wang, X. Wu, and X.-S. Hua. Typicality-Based Visual Search Reranking. *IEEE Transactions on Circuits and Systems for Video Technology*, 2010.

[15] D. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *IJCV*, 2004.

[16] M. Naphade, L. Kennedy, J. Kender, S.-F. Chang, J. Smith, P. Over, and A. Hauptmann. A Light Scale Concept Ontology for Multimedia Understanding for TRECVID 2005, 2005.

[17] A. Natsev, A. Haubold, J. Tesic, L. Xie, and R. Yan. Semantic concept-based query expansion and re-reranking for multimedia retrieval. In *ACM Multimedia*, 2007.

[18] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-Based Image Retrieval at the End of the Early Years. *PAMI*, 2000.

[19] P. Stark and R. Parker. Bounded-variable least-squares: an algorithm and applications. *Computational Statistics*, 1995.

[20] M. Stricker and M. Orengo. Similarity of color images. In *SPIE Storage and Retrieval for Image and Video Datasets III*, 1995.

[21] Z. Sun, T. Qin, Q. Tao, and J. Wang. Robust Sparse Rank Learning for Non-Smooth Ranking Measures. In *SIGIR*, 2009.

[22] X. Tian, L. Yang, J. Wang, X. Wu, and X.-S. Hua. Bayesian Visual Reranking. *IEEE Transactions on Multimedia*, 2010.

[23] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistics Society Series*, 2005.

[24] R. Tibshirani and J. Taylor. The Solution Path of the Generalized Lasso. *The Annals of Statistics*, 2010.

[25] M. van Benthem and M. Keenan. Fast algorithm for the solution of large-scale non-negativity constrained least squares problem. *Journal of Chemometrics*, 2004.

[26] A. Vedaldi and B. Fulkerson. VLFeat: An Open and Portable Library of Computer Vision Algorithms, 2008.

[27] M. Wang, L. Yang, and X.-S. Hua. MSRA-MM: Bridging Research and Industrial Societies for Multimedia Information Retrieval, 2009.

[28] R. Yan and A. Hauptmann. Co-retrieval: A boosted reranking approach for video retrieval. In *CIVR*, 2004.

[29] R. Yan, A. Hauptmann, and R. Jin. Multimedia Search with Pseudo-Relevance Feedback. In *CIVR*, 2003.

[30] L. Yang and A. Hanjalic. Supervised reranking for web image search. In *ACM Multimedia*, 2010.

[31] W. Zhou, Q. Tian, L. Yang, and H. Li. Latent visual context analysis for image re-ranking. In *CIVR*, 2010.

[32] H. Zou. The Adaptive Lasso and Its Oracle Properties. *Journal of the Americal Statistical Association*, 2006.