

Lightweight web image reranking

Adrian Popescu, Pierre-Alain Moellic, Ioannis Kanellos, Rémi Landais

▶ **To cite this version:**

Adrian Popescu, Pierre-Alain Moellic, Ioannis Kanellos, Rémi Landais. Lightweight web image reranking. ACM Multimedia 2009 : 17th ACM international conference on Multimedia, Oct 2009, Pékin, China. pp.657 - 660, 10.1145/1631272.1631381 . hal-02345837

HAL Id: hal-02345837
https://hal.science/hal-02345837

Submitted on 4 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Lightweight Web Image Reranking

Adrian Popescu*, Pierre-Alain Moëllic**, Ioannis Kanellos*, Rémi Landais***
*TELECOM Bretagne, France, {adrian.popescu, ioannis.kanellos}@telecom-bretagne.eu
**CEA LIST, France, pierre-alain.moellic@cea.fr
***Exalead, France, remi.landais@exalead.com

ABSTRACT
Web image search is inspired by text search techniques; it mainly relies on indexing textual data that surround the image file. But retrieval results are often noisy and image processing techniques have been proposed to rerank images. Unfortunately, these techniques usually imply a computational overload that makes the reranking process intractable in real time. We introduce here a lightweight reranking method that compares each result not only to the other query results but also to an external, contrastive class of items. The external class contains diversified images; the intuition supporting our approach is that results that are visually similar to other query results but dissimilar to elements of the contrastive class are likely to be good answers. The success of visual reranking depends on the visual coherence of queries; we measure this coherence in order to evaluate the chances of success. Visual reranking tends to emerge near duplicate images and we complement it with a diversification function which ensures that different aspects of a query are presented to the user. Our method is evaluated against a standard search engine using 210 diversified queries. Significant improvements are reported for both quantitative and qualitative tests.

Categories and Subject Descriptors
H.3.1 **Content Analysis and Indexing**

General Terms
Algorithms, Experimentation.

Keywords
Image retrieval, reranking, k-NN.

1. INTRODUCTION
Image retrieval is mainly keyword based. Search engines such as Live or Google only recently introduced content based retrieval, as a complement to textual search, only recently. Results obtained using keyword matching are often irrelevant because the text around images doesn't always describe image content [7], [8].

An important research effort was directed toward developing reranking techniques that exploit image processing; but hard problems are yet to be solved before incorporating image reranking into search engines architectures. Firstly, the topic range: Web image queries address a wide range of subjects and it is impossible to pre-process all possible queries. Consequently, the reranking process should be fast enough to be performed at query time but this is not the case for most existing techniques. Secondly, the discrepancy between query diversification and reranking coherence: queries are conceptually and visually diverse but image reranking performances are good for visually coherent queries; moreover, they are usually tested on narrow domains. For instance, the authors of [8] limit their approach to landmarks while the authors of [11] test it to canine species. To solve this problem, it would be interesting to have a measure which, given any query and corresponding results, may evaluate the chances for image reranking to be successful. Thirdly, a search engine should maximize results precision and cover different aspects of the query in the same time [4] but these two measures are often difficult to maximize simultaneously [1]. We introduce an image reranking technique which tries to cope with the three problems cited above. Central to our approach is the conjunction of a contrast model and a focusing hypothesis, idea borrowed from Tversky's work [11]. The basic idea is that the similarity between two items is defined in contrastive, structuralist terms: it is not only a proximity relationship (the sharing of common features) but also a distance value that quantifies the dissimilarity to an opponent class of items. We translate this principle to images and suppose that an image is relevant for a query if it is visually similar to other query results and dissimilar to an external class which contains diversified images. To determine the visual coherence of a class we consider the best ranked images and compute the average number of neighbors from the external class. Visual reranking tends to favor near duplicate images [4]. We then add a diversification step to our method. We index images associated to 210 diversified queries using a texture-color content descriptor [3]. This descriptor is efficient when indexing heterogeneous datasets and provides a detailed analysis of the performances of our reranking technique.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
MM'09, October 19–24, 2009, Beijing, China.

2. RELATED WORK
Image reranking can be performed using textual information associated to images, visual description or a combination of the two. In [7], the authors adapt the PageRank algorithm to image retrieval in order to find "authority nodes" in a visual similarity graph. Both homogeneous and heterogeneous visual concepts are discussed but the approach is only tested on product images and it largely outperforms the Google standard search. Van Leuken et al. [12] propose techniques for diversifying image search results based on visual clustering. Clustering is applied to both ambiguous and non-ambiguous queries and it is evaluated against manually clustered search results. Tests show that the approach tends to reproduce manual clustering in a majority of cases. Deselaers et al. [4] discuss the joint optimization of search precision and diversity, with a focus on diversity. They implement a dynamic programming algorithm applied on top of a greedy selection and test their approach on a heterogeneous test database (ImageCLEF 2008 photo retrieval task [1]). An improvement of diversity, accompanied by a small precision loss is reported when comparing results to ImageCLEF runs.

Cai et al. [2] propose a hierarchical clustering approach in order to discover semantic clusters within Web search results. Their method uses textual, visual and link analysis and is mainly designed for ambiguous queries. In [8], the authors introduce a multimodal clustering technique (based on k-Means) to produce relevant and diversified results for Flickr landmarks images. Tags, user related information, geotags and temporal information are combined to propose highly accurate results. This technique surfaces images that are well linked to items uploaded by a large number of users, giving thus a social relevance to best ranked results. Whereas the technique in [8] is tuned for landmarks, [9] implements a shared nearest neighbors algorithm (s-NN) which clusters both tags and visual content for any given query. Unfortunately, the technique in [9] is not fast enough to be performed at query time. Compared to approaches like [8] or [11], our technique is domain independent. Moreover, it is quite fast because the underlying algorithm, *k nearest neighbors* (k-NN), is simpler to compute than most other classification methods. The computational complexity has a linear variation with the number of considered images.

In [6], reranking is applied on video search results. Initial text search results are reranked using multimodal pair-wise similarity. The reranking problem is formulated using as a random walk by building a context graph. More recently, in [10], reranking is seen as a global optimization problem within a Bayesian framework by maximising the ranking score using visual similarity features (global color descriptor) between video shots and minimizing the ranking distance based on the initial text-based ranking. The paper is mainly focused on the likelihood optimisation by proposing two distances between two ranked lists.

Though efficient, techniques like s-NN [9] or dynamic programming [4] are computationally expensive and are hard to apply under real time constraints. The use of an external class which helps surfacing relevant images is central to our method. To the best of our knowledge, such an approach was not used for image reranking. Another particularity of our method is the introduction of a measure that tries to evaluate if the visual reranking will be efficient for a given query or not. A large number of features can be used to describe visual content. Global image descriptors are computed in [11], local features are extracted in [7] or [9] and a combination of the two types of features is used in [4]. Choosing the correct descriptors or combining them are indisputably complex problems, but they fall outside the scope of this paper.

3. IMAGE RERANKING

The introduction of content based image processing techniques in Live Search and Google Image proves the feasibility of applying such techniques to large volume of images. However, in order for the search process to be computationally efficient, the indexing process needs to be performed offline. We pre-index our images using a global texture-color descriptor presented in [3]. Local based approaches provide more robust information but are clearly more expensive due to the high dimensionality of classical local features and usually need nearest neighbors approximation to perform points matching, like in [7] with an LSH approach used to speed up the construction of the connectivity graph: for 1000 images (about 500,000 local features), 15 minutes were necessary to compute the full similarity matrix. At query time, we select only images associated to the textual query from the index and calculate the similarity matrix dynamically. Such a process takes 0.8 s on average on a 3.0 GHz Intel processor.

Our reranking technique is based on the visual similarity between image search results and on their dissimilarity to an external class. The external class was created by launching a query with "test" in Flickr and recuperating 300 images from different users. A more judicious choice would be to manually build the external class so as to maximize the diversity of its elements. If search results contain an important number of irrelevant images [4], we presume that i) noisy results are weakly related to relevant results and ii) relevant results are visually related to other answers to the same query. Images in the external class are added to query results in order to find out which elements are close to the class itself and far from the external class. To express the relatedness of each image (noted img_i) to its class, we compare it to other query results and to the external class using content description and finally count the number of extraneous items that are found among the k nearest neighbors of the image (ext_{imgi}). A small number of neighbors from the external class indicates that the image is closely related to other query results.

The value of k is an important parameter of the reranking procedure and we empirically fixed it at *10*. In [7], the authors make a distinction between visually heterogeneous and visually homogeneous queries (*Apple* and *Mona Lisa*). Image reranking is particularly interesting for queries with a large number of results (hundreds or more); a value of k which is significantly smaller than the number of results facilitates the discovery of different aspects of the query. For instance, images of *Apple* as *fruit* and *Apple* as a *device* are visually dissimilar and will tend to be classified with images that correspond to the same sense of the term. The reranked list of results will propose images with small ext_{imgi} firstly, because they are well linked to the class and are likely to be relevant. Clearly, a value of $k = 10$ will determine a lot of equal ext_{imgi} scores; thus, in order to differentiate between images with such scores, we introduce a second score int_{imgi}, which represents the cumulated sum of visual distances between the image and the 5 nearest neighbors from the class. At equal ext_{imgi}, images with small int_{imgi} will be presented firstly. The authors of [4] and [12] note that visual reranking techniques, such as ours, tend to generate results with rather reduced diversity. Since the similarity matrix between the images associated to a query is already computed, we may use it in order to diversify results. In order to ensure that diversified results will be chosen among well linked images, we retain only the best 30% reranked elements (which are more likely to be relevant than other images) and try to find diversified items among them. Once we fixed the number of images the system will finally present to the user, the diversification process is iterated until enough images are retained. We build a list of diversified results by adding new elements to the list whenever these new images are different enough from images that were already selected. To express difference, we count the number of nearest neighbors of the new image that are not nearest neighbors of selected images and use it as threshold. The value of the threshold varies from 1 to 11 and this variation defines an acceptability criterion that is more and more relaxed. The process stops when there are enough elements in the list of diversified results. This list includes 20 elements, a number which roughly corresponds to the number of images on a Web search engine results page. The complexity of the diversification is equal to the product between the size of the list and the number that represents the 30% best ranked results. For a results set containing 300 images, the diversification takes around one second on a 1.6GHz processor and this without any focus on algorithmic optimization.

To characterize the visual coherence (*viscor*) of a query, we average ext_{imgi} for the N best ranked images associated to a query:

$$viscor = \frac{1}{N} \sum_{i=1}^{n} ext_{imgi} \qquad (1)$$

Small values of *viscor* indicate that the query is visually coherent and that the visual reranking is likely to be successful. Our notion of visual coherence is different from the binary separation of queries in visually homogeneous and heterogeneous proposed in [7]. For instance, a query with *Europe* has a low visual coherence and corresponds to a heterogeneous query as defined in [7]. A query with *Monet paintings* is heterogeneous according to [7] but has a good visual coherence because our k-NN algorithm stimulates the discovery of local regularities (here individual paintings).

4. EVALUATION

Our reranking technique is evaluated on a diversified test dataset comprising 210 concepts which were illustrated with Exalead images [5]. Queries were selected by Exalead according to the following criteria: i) *frequency*–queries should be chosen among the most frequent queries; ii) *diversity*–queries should treat a large range of the target domain (geographic entities, celebrities' names, artefacts…); iii) *visual coverage*–this criterion is related to visual coherence. Examples of concepts in the database include: *airplane, Eiffel Tower, crowd* or *Björk*. Up to 300 Exalead images were retained for each query.

The effects of the reranking on results precision and diversity were analyzed at a query level in a user study with 22 participants. Then, we performed a smaller scale precision evaluation where three assessors evaluated the P@10 for the original, the reranked and the diversified results. Finally, we reused results of the precision test to assess the utility of *viscor* (i.e. the visual coherence measure), by means of a threshold on reranking results.

4.1 Pertinence vs. Diversity

In our user study the participants were asked to compare the accuracy and the diversity of the results for Exalead images, for the visual reranking technique and for the visual reranking plus diversification. Participants were contacted via e-mail; the participation was voluntary. In order not to overload participants, we asked them to evaluate at most 30 queries. The test dataset was split in seven equal parts; participants had to deal with different parts of the dataset. Since the queries in the dataset were diversified, it was possible that some of them were unknown to participants and these last were instructed to assess only queries they knew well enough. Each query was presented on a distinct page; the top 12 results for each method were displayed on separated columns. To avoid the formation of evaluation patterns, the results columns on different pages were presented in a different order. Participants were asked to evaluate global accuracy and diversity on a scale ranging from 0 (bad quality) to 4 (very good quality) for both query and each retrieval method. The accuracy of results for visually reranked results is significantly higher compared to "Exalead" accuracy (2.94 vs. 2.57); but the results diversity is smaller (2.02 vs. 2.84). To test statistical significance of results difference between "Exalead" and reranked results, we performed a paired T-test (with $p < 0.05$) and the result (0.00257) shows that the two distributions are statistically different. Values for "Exalead" and "Rerank" in table 1 confirm that visual reranking is efficient in surfacing relevant elements but hurts results diversity. As for results after diversification, the average accuracy is 2.74, compared to 2.57 for "Exalead". The result of the T-test for accuracy in this case (0.1342) is clearly less convincing but the accuracy gain is obtained with little diversity loss (2.76 vs. 2.84). The diversification function has its acceptability parameter set up to a limit case and with a relaxation of this parameter, it is easy to obtain performances ranging from "Rerank+Diversification" to "Rerank" (table 1).

Table 1. Accuracy and diversity for the three tested techniques averaged on a panel of 22 participants. The scale is from 0 (bad) to 4 (good quality results).

	Method		
	Exalead	**Rerank**	**Rerank+Diversification**
Accuracy	2.57	2.94	2.74
Diversity	2.84	2.02	2.76

The evaluation of image search results is a subjective and context dependant task. In our test, we also noted important variations between the participants. Accuracy varies between 1.87 and 3.21 for "Exalead", between 2.31 and 3.68 for "Rerank" and between 2.125 and 3.43 for "Rerank+Diversification". When considering accuracy, all users preferred Rerank to Exalead and only 5 users out of 22 preferred "Exalead" to "Rerank+Diversification". Visual reranking seems to be preferred to a classical keyword-based approach by a large majority of the users. A results example is presented in figure 1.

4.2 Precision Evaluation

The user study focused on a global characterization of answers sets; but we also wanted to assess the precision at 10 (P@10) for all the three methods. To do this, we selected 60 queries from the test dataset and computed the P@10 for each query.

Table 2. P@10 for a sample of 60 queries and three users.

	P@10		
	Exalead	**Rerank**	**Rerank+Diversification**
User 1	0.628	0.693	0.615
User 2	0.671	0.735	0.676
User 3	0.713	0.807	0.747

In table 2, precision at 10 for "Exalead" varies between 0.628 and 0.713 and between 0.693 and 0.807 for "Rerank". The performed T-tests show statistically significant differences between "Rerank" and "Exalead" for the three users. All participants ranked the three methods in the order they have also done for the global evaluation: "Rerank" scores best followed by "Rerank+Diversified" for User 2 and User 3 and by "Exalead" for User 1. These concordant results indicate that there is a correlation between the global assessment of results quality and the detailed assessment using P@10. Globally, P@10 results confirm the global quality evaluation and show that the highest accuracy is obtained for "Rerank", followed by "Rerank+Diversified" and "Exalead".

4.3 Role of Visual Coherence

We hypothesize that it is worth reranking the results for a given query only if its associated visual coherence is sufficiently big (small value of *viscor* defined in equation 1). We use *viscor* as a

threshold to decide if a query should be reformulated or not and present accuracy results for viscor varying from 0.1 to 7, with a step of 0.1. The use of *viscor* to decide which queries should be reranked introduces a slight improvement of results (0.1 for all three participants with a threshold value around 2). To confirm the results reported here, the utility of *viscor* should be evaluated on larger scale query samples; and, of course, with more participants.

5. CONCLUSION

We introduced an image reranking method that relies on the use of an external class in order to surface relevant images. The method improves results accuracy but hurts diversity and a diversification function was introduced as a compromise. We also defined a visual coherence measure and used it to evaluate if reranking is likely to improve results for a particular query or not. Preliminary tests show that this measure improves results over the use of visual reranking for all queries. Our reranking method is generic, fast and easy to integrate in existing Web image search architectures.

We currently investigate the introduction of other content descriptors in the reranking framework, focusing on the use of the visual coherence measure for automatically selecting the best descriptor (or combination of descriptors) for a query. We will also compare our approach with techniques such as VisualRank. Finally, we will investigate the effect of constructing the external class manually and the performances of the method when retaining more than 300 images per query.

6. ACKNOWLEDGMENTS

This research is part of Georama, a French research project funded by ANR.

7. REFERENCES

[1] T. Arni, P. Clough, M. Sanderson, and M. Grubinger. "Overview of the ImageCLEFphoto 2008 photographic retrieval task." In *CLEF 2008 Workshop Working Notes*.

[2] D. Cai, X. He, Z. Li, W.-Y. Ma, and J.-R. Wen. "Hierarchical clustering of www image search results using visual, textual and link information." *Proc. of ACM MM'04*.

[3] Y.-C. Cheng, S.-Y. Chen. "Image classification using color, texture and regions." *Image Vision Computing*, 21(9), 2003.

[4] T. Deselaers, T. Gass, P. Dreuw, H. Ney. "Jointly Optimising Relevance and Diversity in Image Retrieval." *Proc of CIVR '09*.

[5] Exalead – http://exalead.com

[6] W.H. Hsu, L. Kennedy, S-F. Chang, "Video Search Reranking through Random Walk over Document-Level Context Graph". *Proc. of ACM MM'07*.

[7] Y. Jing, S. Baluja. "VisualRank: Applying PageRank to Large-Scale Image Search" Transactions On Pattern Analysis and Machine Intelligence, Vol 30, No 11, Novembre 2008.

[8] L. Kennedy, M. Naaman. "Generating Diverse and Representative Image Search Results for Landmarks." *Proc. of WWW 2008*, April 2008, Beijing, China.

[9] Pierre-Alain Moëllic, Jean-Emmanuel Haugeard, Guillaume Pitel. "Image clustering based on a shared nearest neighbors approach for tagged collections." *Proc. of CIVR '08*.

[10] X. Tian, L. Yang, J. Wang, Y. Yang, X. Wu, X-S. Hua, "Bayesian Video Search Reranking". *Proc of. ACM MM'08*.

[11] A. Tversky. "Features of similarity." *Psychological Review*, 84 (4), 1977, pp. 327-352.

[12] R. H. van Leuken, L. Garcia, X. Olivares, R. van Zwol. "Visual Diversification of Image Search Results." *Proc. of WWW 2009*, April 2009, Madrid, Spain.

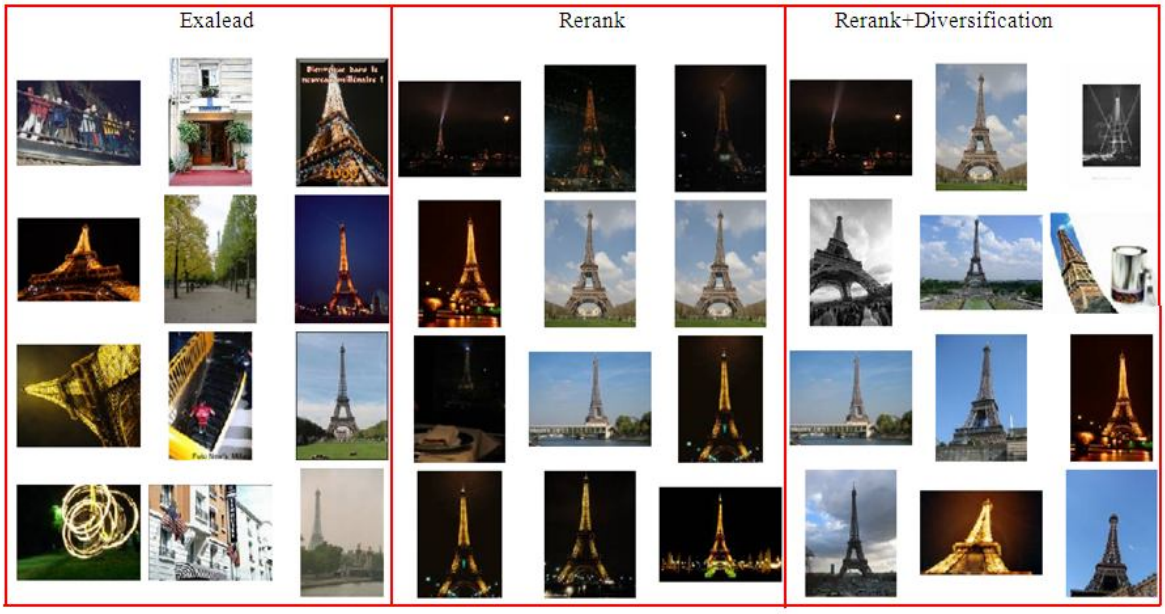


Figure 1. **Results for Eiffel Tower using Exalead, the Reranking procedure and the Reranking+diversification procedure.**