# Query Refinement Suggestion in Multimodal Image Retrieval with Relevance Feedback

Luis A. Leiva, Mauricio Villegas, Roberto Paredes
ITI – Institut Tecnològic d'Informàtica
Universitat Politècnica de València
Camí de Vera, s/n - 46022 (Spain)
{luileito,mvillegas,rparedes}@iti.upv.es

## ABSTRACT

In the literature, it has been shown that relevance feedback is a good strategy for the system to interact with the user and provide better results in a content-based image retrieval (CBIR) system. On the other hand, there are many retrieval systems which suggest a refinement of the query as the user types, which effectively helps the user to obtain better results with less effort. Based on these observations, in this work we propose to add a suggested query refinement as a complement in an image retrieval system with relevance feedback. Taking advantage of the nature of the relevance feedback, in which the user selects relevant images, the query suggestions are derived using this relevance information. From the results of an evaluation performed, it can be said that this type of query suggestion is a very good enhancement to the relevance feedback scheme, and can potentially lead to better retrieval performance and less effort from the user.

## Keywords

Image Retrieval, Relevance Feedback, Query Suggestion

## Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Relevance feedback, Query formulation; I.5.5 [**Pattern Recognition**]: Interactive systems

## General Terms

Experimentation, Performance, Human Factors

## 1. INTRODUCTION

Most of the current approaches to retrieve images from online and offline databases are focused on the content-based image retrieval (CBIR) paradigm. However, CBIR systems have not yet gained widespread acceptance, even after more than a decade of research [8]. The main reason for this is what is known as *the semantic gap* [7], which refers to the discrepancy there exists between the low-level content (i.e., the features extracted from the image) and the higher-level concepts (i.e., the interpretation that a person gives to the image in a particular context).

A popular approach for improving the retrieval performance of CBIR systems is by means of relevance feedback. In this approach, the system presents an initial set of images and the user selects which of these are relevant and possibly which are non-relevant. Then, taking into account this new information, the system presents a new set of images, and again the user can select more images. In this way, the user iteratively helps the system to retrieve the desired images. In the literature it has been shown that relevance feedback improves considerably the retrieval performance of a system [3, 4, 5, 9]. Unfortunately, though, if for a given query the image features are unable to represent the relevant images, e.g., when the semantic gap is very wide, then many relevance feedback iterations will be required and few relevant images will be retrieved. A possible way to overcome this problem, is to provide alternative ways of interaction with the user which can give better results when the relevance feedback fails.

Another popular type of user interaction that can improve the retrieval performance is by the system suggesting a refinement of the initial textual query. In contrast to relevance feedback, query suggestion has the benefit that it is optional for the user. If the suggested query is not adequate, then it is unlikely that users will use it, however it is more likely that they will use the good suggestions.

In the literature there is a lot of work on textual query suggestion, although most of them are targeted at document retrieval [10]. Even though most of these can be used in an image retrieval system, they do not consider the way most CBIR with relevance feedback work. Related to image retrieval, Zha and co-authors [10] present what they call the Visual Query Suggestion (VQS) scheme, in which, as the user types the query, the system shows a drop-down list of suggested queries including some examples of images related to that query. In the work of Cui et al. [2], by using a single query image the system suggests a category among a predefined set, such as object, portrait, etc. which is then used to improve the performance of the relevance feedback.

In this work, we propose to integrate a query suggestion scheme into an image retrieval system based on relevance feedback. By providing these two types of user interaction we aim at reducing the effort from the user and obtain better results for queries which are more text driven. The query


*ICMI'11,* November 14–18, 2011, Alicante, Spain.

suggestion is derived by taking advantage of the selection of the relevant images. Each time the user selects an image, the indexed terms associated to the images are analyzed to determine which ones are more characteristic for the selected images in comparison to the rest of the presented images. At the moment, we are not aware of any paper on query suggestion in this manner.

The remainder of the paper is organized as follows. The next section briefly describes our system, the Relevant Image Search Engine (RISE), including the database, how the relevance feedback is performed, and the proposed query suggestion method. Section 3 presents the results of an evaluation performed, followed by a Discussion section. The final section draws the conclusions and directions for future research.

## 2. RISE SYSTEM OVERVIEW

### 2.1 Web Crawling and Image Indexing

The image database of the RISE system is composed of images found in the Internet. Thus, the RISE system includes a web crawler which is constantly downloading new images and adding them to the database. Currently the database has over 20 million images. In order to guaranty a wide variety of images from scratch, without having to crawl a large portion of the Internet, the crawling has been started using a list of image URLs obtained from popular image search engines (namely Google, Bing and Yahoo) when searching for words in the English dictionary. For each image the crawler stores a thumbnail and a snapshot of the web page that contains that image. Then such web page is analyzed and to each term in the page a relevance score is derived, which depends on: *1)* the number of appearances, *2)* the document object model (DOM) attributes, and *3)* the distance to the image. Finally, the image is indexed using the terms with the highest relevance scores.

### 2.2 Retrieval Engine and Relevance Feedback

The retrieval engine of the RISE system starts by receiving a textual query given by the user. Then the images in the database indexed with the terms in the query are ranked according to the relevance scores derived during the indexing, and the thumbnails of the first $N$ ranked images are presented to the user. At this stage the relevance feedback starts, by letting the user select some of the images as being relevant (the unmarked images are considered as being non-relevant). Using the selected relevant and non-relevant images, the system presents a new set of images which (hopefully) is closer to what the user is searching for. Again, the user can select more relevant images to refine further the results, and this iterative procedure continues until the user is satisfied. The ranking of the images according to the selected relevant images is based on the probabilistic framework proposed by Paredes et al. [5]. For the relevance feedback, we use a multimodal approach, using both the text annotation extracted from web pages, and features extracted from the image [6]. As image features we are using quantized color histograms and as a similarity measure, the Jensen-Shanon divergence.

### 2.3 Query Refinement Suggestion

In order to suggest a new query to the user we propose an approach that essentially accounts for the (weighted) words that appear in the text associated to the relevant images but do not appear in the text associated to the non-relevant images. This selection of words can be accomplished in different ways but here we show the selection that gave us the most promising results.

Let be $\{w_1, \cdots, w_n\}$ the $n$ words of the vocabulary. We denote the set of relevant images as $R$ and the set on non-relevant images as $N$. Let be $W$ the set of words $w_i$ that appear in *all* the relevant images. If this set is empty then no suggestion is produced. If this set is not empty we compute a score for each one of these words:

$$\mathrm{sc}(w_i) = \frac{\sum_{j \in R} t_{ij}}{\sum_{j \in R} t_{ij} + \sum_{k \in N} t_{ik}}, \quad \forall w_i \in W \tag{1}$$

where $t_{ij}$ is the relevance score of the word $w_i$ in the relevant image $j$ and $t_{ik}$ is the relevance score of the word $w_i$ in the non-relevant image $k$. Note that if $\mathrm{sc}(w_i)$ is 1.0 we add the value of $\sum_{j \in R} t_{ij}$ to such score $sc(w_i)$, in order to avoid ties in the image rank.

Then if the scores of the words $w_i \in W$ are higher than a certain threshold and using at most the top 2 of them, these are suggested to the user as a new query. For the evaluation described in the next section, the threshold was not applied, since the objective was to assess the quality of the suggestions using this approach.

A version of the RISE system which includes this proposal for query refinement suggestion is temporarily available at http://risenet.iti.upv.es/qs/.

## 3. EVALUATION

Note that our image database is built from real data gathered from the Internet with completely unsupervised annotations, so we have no ground truth, i.e, labeled samples. Thus, in order to evaluate our approach, we conducted an informal field study. The procedure was simple: to measure the user's subjectivity towards our query suggestion technique.

For the evaluation, we selected 81 of the 99 concepts from the ImageCLEF 2011 dataset[1], and used these as the initial text search query. The reason to remove 18 concepts was because they were related to specific image properties rather than high-level concepts, e.g., "Neutral Illumination", "No Blur", etc.

The task consisted of two stages. First, the user was presented with the first 10 ranked images for the given text query, e.g., "cat". Then the user would select a subset of the images which had a common concept or relation among them, e.g., "all are black cats". If the system was able to derive a query refinement, the user interface (UI) would show it and let the user rate if the suggestion was either good, bad, or neutral. The number of times there was no query (NQ) suggested was also recorded (Table 1). In the second stage of the evaluation, users were presented with the images after following the query suggestion, and they had to mark all of the images considered relevant to the concept they had in mind when selecting the images in the first stage of the evaluation. This two-stage process was repeated for all of the subsets of related images the user could identify. Three people from our department took part in the evaluation.

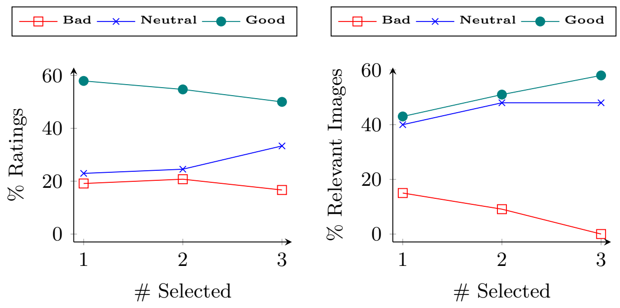
The results of the study are presented in Figure 1 and Tables 1 and 2.

**Table 1:** Results for stage 1 of the evaluation, showing ratings for the suggested queries.

| # selected | # samples | Bad | Neutral | Good | NQ |
|---|---|---|---|---|---|
| 1 | 194 | 35 | 42 | 106 | 11 |
| 2 | 74 | 11 | 13 | 29 | 21 |
| 3 | 24 | 2 | 4 | 6 | 12 |
| >3 | 30 | 0 | 0 | 3 | 27 |
| **Overall** | 322 | 48 | 59 | 144 | 71 |

**Table 2:** Results for stage 2 of the evaluation, showing mean (and SD) values of the number of retrieved relevant images after following the suggested queries.

| # selected | # ratings | Bad | Neutral | Good |
|---|---|---|---|---|
| 1 | 183 | 1.5 (1.5) | 4 (3) | 4.3 (3) |
| 2 | 53 | 0.9 (1.4) | 4.8 (3.2) | 5.1 (3.3) |
| 3 | 12 | 0 (0) | 4.8 (4.8) | 5.8 (2.8) |
| >3 | 3 | 0 (0) | 0 (0) | 9.6 (0.5) |
| **Overall** | 226 | 1.3 (1.5) | 4.3 (3.2) | 4.7 (3.2) |




**(a)** **(b)**

**Figure 1:** Evaluation results. 1a: Average rating of the suggested queries in relation to the number of initially selected images. 1b: Percentage of images considered as relevant after following the query suggestions in relation to the number of initially selected images.

# 4. DISCUSSION

Regarding the first stage of the evaluation, the first thing to note is that as more images are selected, it is less probable that the system will suggest a query (see Table 1). This is understandable since it is less likely that there will be common terms to all selected images. Moreover, the terms associated to each image depends totally on the Web pages in which the image appears, thus not all of the images will be well annotated. Nonetheless, most of the suggested queries were rated as being good, which indicates that this approach of deriving the suggestions based on the selected relevants can be quite useful.

Regarding the second stage of the evaluation, as expected, the query suggestions which were rated as good or neutral, retrieved more relevant images than the bad query suggestions (see Figure 1b). This is convenient since it is unlikely that a user will use a suggestion considered to be bad. A peculiar behavior that is also observed is that the performance tends to be better for suggestions that were derived using more selected relevant images. Then, overall, as more images are selected, it is less likely that the system will suggest a query, however if there is a suggestion it tends to be a better one.

Another observation from the evaluation was that the quality of the query suggestions depends highly on the particular query. There are some queries where the images presented to the user clearly belong to different subgroups, which, if selected, most of the time a query will be suggested that relates to that subgroup. An example of a query that gives good suggestions is shown in Figure 2. In such figure it can be observed that when only one image is selected, the suggested queries tend to be very specific to that image. This is another possible factor why when fewer images were selected, the percentage of retrieved relevant images was lower: a very specific query might be very good, however this also means that there will be less images available of this type.

# 5. CONCLUSION AND FUTURE WORK

In this paper we have presented a simple method to derive a query refinement suggestion by taking advantage of the selection of relevant images commonly used in image retrieval with relevance feedback. This proposal has been implemented in a real-world image search prototype which can be accessed at http://risenet.iti.upv.es/qs/. An evaluation was performed and from which it has been observed that the proposed method most of the time is effective and may reduce user workload while improving user experience. The perceived (subjective) satisfaction towards the system was positive, although several enhancements remain to be implemented in order to consolidate our approach.

Firstly, even though it is unlikely that a user will use a query suggested which is not considered adequate, it would be ideal to find a way to avoid showing these low-quality suggestions. A simple threshold can partially deal with this, however a more advanced method would be desired. For example, we could analyze what would happen when the images are re-ranked after following the suggested query. If the new top-ranked images are not clearly similar to selected relevant images and dissimilar to the non-relevants, it is an indication that the suggested query is not adequate.

Another problem of the proposed approach is that it is limited by the annotations that were obtained from the Web pages containing the images. A better performance would be expected if, for instance, we would use a thesaurus such as Wordnet[2] to somewhat clean up the annotations and use hypernyms [1] to find the common terms to the selected relevant images. Also there are images on the Web with few text around it, text that does not give any clue regarding the content of the image, or even with no text at all. For these cases it would be required to have a technique which automatically assigns them proper annotations.

Finally, we are considering to incorporate natural language processing techniques, not only to build query suggestions, but also to constrain initial user queries. While the latter falls under the relatively well-covered natural language understanding domain (that is, convert text into easy-to-manipulate structures for computers), the former case is related to natural language generation (convert information from computers into readable human language), and we feel it is more challenging. In this way, for instance, following Figure 2d, instead of suggesting "marathon people", a more appropriate query would be "marathon with people"; and


**(a)** Initial set of images for the query “marathon”.


**(b)** Number of selected images: 1. Suggested query: “marathon brandenburg gate”.


**(c)** Number of selected images: 1. Suggested query: “marathon belfast city”.


**(d)** Number of selected images: 4. Suggested query: “marathon people”.

**Figure 2:** Example of query refinement suggestions when searching for “marathon”. The suggestions tend to be very specific when only one image is selected (2b,2c). Conversely, when several images are selected the suggestions tend to be a common characteristic among them (2d).

searching for “marathon with people” and “marathon without people” indeed should produce slightly different results — although many image search engines today would generally treat both queries the same way.

## Notes

[1] http://imageclef.org/system/files/concepts_2011.txt
[2] http://wordnet.princeton.edu/

## Acknowledgements

This work is supported by the Spanish MICINN under the MIPRCV Consolider Ingenio 2010 (CSD2007-00018) and RISE (TIN2008-04571) projects, and by the Generalitat Valenciana under grant Prometeo/2009/014.

## REFERENCES


[1] K. Chandramouli, T. Kliegr, J. Nemrava, V. Svatek, and E. Izquierdo. Query refinement and user relevance feedback for contextualized image retrieval. In *Proc. of VIE*, pp. 453–458, 2008.

[2] J. Cui, F. Wen, and X. Tang. User intention modeling for interactive image retrieval. In *Proc. ICME*, pp. 1517–1522, 2010.

[3] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40(2):1–60, 2008.

[4] E. Hörster, R. Lienhart, and M. Slaney. Image retrieval on large-scale image databases. In *Proc. CIVR*, pp. 17–24, 2007.

[5] R. Paredes, T. Deselaers, and E. Vidal. A probabilistic model for user relevance feedback on image retrieval. In *Proc. MLMI*, pp. 260–271, 2008.

[6] F. M. Segarra, L. A. Leiva, and R. Paredes. A relevant image search engine with late fusion: Mixing the roles of textual and visual descriptors. In *Proc. IUI*, pp. 455–456, 2011.

[7] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Trans. PAMI*, 22(12):1349–1380, 2000.

[8] J. Urban and J. M. Jose. Can a workspace help to overcome the query formulation problem in image retrieval? In *Proc. ECIR*, pp. 385–396, 2006.

[9] J. Urban, J. M. Jose, and C. J. van Rijsbergen. An adaptive technique for content-based image retrieval. *Multimedia Tools Appl.*, 31(1):1–28, 2006.

[10] Z.-J. Zha, L. Yang, T. Mei, M. Wang, Z. Wang, T.-S. Chua, and X.-S. Hua. Visual query suggestion: Towards capturing user intent in internet image search. *ACM Trans. Multimedia Comput. Commun. Appl.*, 6(3):13:1–13:19, 2010.