Conceptual Query Expansion and Visual Search Results Exploration for Web Image Retrieval

Enamul Hoque, Grant Strong, Orland Hoeber, and Minglun Gong

Abstract Most approaches to image retrieval on the Web have their basis in document search techniques. Images are indexed based on the text that is related to the images. Queries are matched to this text to produce a set of search results, which are organized in paged grids that are reminiscent of lists of documents. Due to ambiguity both with the user-supplied query and with the text used to describe the images within the search index, most image searches contain many irrelevant images distributed throughout the search results. In this paper we present a method for addressing this problem. We perform conceptual query expansion using Wikipedia in order to generate a diverse range of images for each query, and then use a multi-resolution self organizing map to group visually similar images. The resulting interface acts as an intelligent search assistant, automatically diversifying the search results and then allowing the searcher to interactively highlight and filter images based on the concepts, and zoom into an area within the image space to show additional images that are visually similar.

Enamul Hoque
Department of Computer Science, Memorial University, St. John's, NL, A1B 3X5, Canada
e-mail: enamulp@mun.ca

Grant Strong
Department of Computer Science, Memorial University, St. John's, NL, A1B 3X5, Canada
e-mail: strong@cs.mun.ca

Orland Hoeber
Department of Computer Science, Memorial University, St. John's, NL, A1B 3X5, Canada
e-mail: hoeber@cs.mun.ca

Minglun Gong
Department of Computer Science, Memorial University, St. John's, NL, A1B 3X5, Canada
e-mail: gong@cs.mun.ca

1 Introduction

Web image retrieval has traditionally followed an approach that is an extension of Web document search techniques [12]. The textual information surrounding and associated with a particular image is used as the core information that describes the image. Using such textual information allows Web search engines to leverage their existing expertise and infrastructure associated with searching for documents. This approach can work well when images are concisely and accurately described within Web pages, and when searchers provide accurate descriptions of their image needs. Unfortunately, the accuracy of the descriptions given to images on the Web cannot be enforced; nor can we rely upon searchers providing clear textual descriptions of the images they seek. When these conditions are not met, the search results may include many non-relevant images.

Recent studies on user behaviour with respect to image search have found that queries are often very short [1]. The difficulty with short queries is that they can be open to many different interpretations. It is possible that different searchers may enter the same query, but their intentions and needs may vary significantly from each other. In some cases, the searcher may provide a short query because they wish to see a broad range of images associated with their query; in other cases, the short query may include all of the information the searcher can recall at the moment. Simply matching these queries to the information available for the collection of images, and providing a paged grid of image search results may not be an effective approach for image retrieval.

It has been noted that many image search queries are associated with conceptual domains that include proper nouns (i.e., people's names and locations) [1, 9]. Searches of this type will often have a specific focus, but a less specific aim with respect to resolving the image need (i.e., many different images may be viewed as relevant). In situations such as this, it may be beneficial to promote diversity in the image search results, and then allow the searcher to explore within the collection. The goal of our research is to support such an approach to image search.

In this paper we present a conceptual query expansion method combined with a neural network based image organization technique in order to provide a highly diversified collection of images and an interactive interface for assisting searchers to conceptually and visually explore the image search space. Our system automatically extracts a list of concepts from Wikipedia that are related to the query. These concepts are used as the source of the query expansion to retrieve a broad range of images. The images are visually clustered using a similarity-based approach that employs a multi-resolution extension to self organizing maps (SOM). Concepts used for query expansion are also presented to the searcher in a hierarchical manner based on a conceptual ontology.

The benefit of this approach is that it allows searchers to begin with short and ambiguous queries, which are automatically expanded to provide a diverse range of images. The searcher can then browse the hierarchy of concepts that produced the expanded queries, focusing on those that provide a more accurate description of their needs and filtering based on those that are not relevant. In addition, the searcher can visually explore the search results space, zooming into an area that includes images that look like what they are seeking. Used together, these operations empower the searcher to take an active role in the image retrieval process, supporting their ability to refine their image needs as they explore the image search space.

2 Related Work

Many Web image search approaches are based on traditional keyword based query formulation and matching [12]. Within these, only the text that is associated with the image is considered, without taking advantage of the featuers of the image itself. To address the limitations of this approach, content based image retrieval (CBIR) has been studied as a method for determining image similarity based on low level visual features [4]. However, such approaches can lead to a *semantic gap*: the gap between the way a person finds similarities between images at the conceptual level and the way the system generates similarity based on pixel statistics [5].

One of the recent trends in Web image retrieval employs traditional document search technique combined with CBIR to organize the results with the goal of diversification. In work that was a pre-cursor to Google Swirl, a similarity graph is generated among the results of a textual query using visual features [10]. The PageRank algorithm is applied to the similarity graph to select the authority nodes as the representative images. However it does not analyze semantically related concepts of the textual query; rather it only considers visual features for diversifying the results.

A promising direction for improving the quality of search result in general is the introduction of query expansion based on concepts related to the query [6]. Such an approach is particularly useful for diversifying the search results, and can enable searchers to assist with the query refinement process. However there are a number of challenges associated with conceptual query expansion. The first problem is finding a suitable knowledge base that has sufficient coverage of a realistic conceptual domain. While others have used WordNet to enhance image retrieval [11], it does not contain information pertaining to the proper nouns that are common in image search queries. Using Wikipedia for reformulating queries has shown promise [16], and is the approach we use in our work.

The second challenge is in ranking the extracted concepts for the purposes of selecting the most relevant of these. A useful approach to this problem is to measure the semantic relatedness between the original query and each of the concepts derived from that query. A number of different methods have been devised to use Wikipedia for this purpose, including WikiRelate! [20], Explicit Semantic Analysis (ESA) [7], and Wikipedia Link-based Measure (WLM) [13]. Due to the computationally efficiency and accuracy of WLM, we use this approach in our work.

The last challenge when using conceptual query expansion for the purposes of image search is in the organization of the resulting images. By expanding the query, the number of images within the search results can grow very large. Further, the rank of the search results from a particular query may not be as important as the visual features of the images. As such, a useful approach is to take advantage of the visual features of the images when organizing the search results, and then allow the searcher to browse and explore within the search results space.

Similarity-based image browsing (SBIB) is an approach that takes advantage of the fundamental aspects of CBIR, but eliminates the need for the searcher to identify a set of relevant images *a priori*. Images are organized based solely on their features, allowing searchers to explore the collection even if they do not have a clearly defined information need [8]. The challenge of SBIB is to arrange images based on visual similarities in such a way as to support the browsing and exploration experience. While a number of different approaches have been proposed in the literature [8], we use a method that employs a novel multi-resolution SOM. This approach provides both an organizing structure for the images and a measure of importance that can be used to determine which images to display when space is limited [17, 18]. The interactive features within this approach have been shown to be very useful and easy to use [19].

3 Conceptual Query Expansion for Image Search

One of the main features of this work is the method by which the image search query is automatically expanded. For the short and ambiguous queries that are common in image search, query expansion attempts to capture the various aspects of the query. The objective is to diversify the range of images retrieved, providing a broad view of what is available. The problems of then allowing the searcher to narrow down the image search results and focus on the aspects that match their particular needs are addressed in Section 4.

The process of performing conceptual query expansion of image search queries follows three steps: extracting concepts from Wikipedia, ranking the extracted concepts, and generating the expanded queries. While others have used Wikipedia for query expansion in the context of general Web search [15], the approach we use is novel in that it takes advantage of specific aspects of image search. The details are explained in the remainder of this section.

3.1 Extracting Concepts from Wikipedia

For this work, we use Wikipedia as the core knowledge base for the query expansion process. Wikipedia is an excellent source of information for the purposes of image search since it includes a large number of articles describing people, places, landmarks, animals, and plants. It is densely structured, with hundreds of millions of links between articles within the knowledge base. Most of the articles also contain various representative images and associated textual captions.

A dump of the Wikipedia collection was obtained in June 2010, and was preprocessed to support the type of knowledge extraction required for our purposes. In analyzing Wikipedia, we observed that the in-going links (i.e., articles that are linked to the current article) and out-going links (i.e., articles to which the current article links) of an article often provide meaningful information that is closely related to the concept upon which the article is written. Therefore, for each article (concept) within the collection, the in-going and out-going links were detected and extracted as related concepts.

We also found that the captions surrounding the images present within a given article can often provide a valuable perspective on the visual features associated with the article. To ensure the inclusion of all relevant concepts associated with the image captions, we use Wikifier [14] to augment the captions with links to relevant Wikipedia articles that may have been missed by the author of the article, and use these links to extract their associated concepts.

Matching a user-supplied query to this knowledge base is simply a matter of selecting the best matching article (referred to as the home article) from Wikipedia using its search feature. In the case where the query is highly ambiguous and Wikipedia returns multiple articles, the user is prompted to select the home article that is the closest match for their information need. While it is possible for a searcher to enter a query that does not include any matches within Wikipedia, this is highly unlikely given the type of information people commonly look for with image search systems, and how closely this matches the type of information present in Wikipedia (i.e., people, places, and things).

The end result of this process is the selection of the home article, along with a list of all the other articles that are part of the in-going or out-going links for the home article, and the articles that originate from the captions of the images within the home article. These concepts provide the basis for automatic query expansion process.

3.2 Ranking the Extracted Concepts

Due to the richness of Wikipedia, the number of concepts obtained in the process described above may become very large. While it is good that so much information is available for the query expansion process, there is a risk in expanding the query too broadly. In order to address this potential problem, we rank the extracted concepts and use only those that are most similar to the home article.

For each of the candidate articles $\{c_i | 1 \leq i \leq C\}$ extracted from the home article, a semantic relatedness measure is applied between the home article h and the extracted articles. WLM [13] is used for this purpose, taking advantage of the hyperlink structure of the associated articles to find out how much they share in common. In order to give preference to the concepts that have been extracted from the image captions within the home article, we use a re-weighting function to determine the relatedness score:

$$r(c_i, h) = \min(WLM(c_i, h)(1+\alpha), 1)$$

Since WLM provides a value in the [0,1] range, we ensure that the relatedness score remains in this range with the min function. The re-weighting factor α is provided according to the following function:

$$\alpha = \begin{cases} k\frac{C}{N} & \text{if concept } c_i \text{ originates from a caption} \\ 0 & \text{otherwise} \end{cases}$$

Here, C is the number of concepts that have been extracted from the home article, and N is the number of concepts to be selected for query expansion, and k is a system parameter that controls the importance of the concepts derived from the captions. In our prototype implementation $N = 30$, $k = 0.01$, and C commonly ranges from 300 to 600. This results in a 10 - 20% increase in the score for the concepts derived from the captions, with proportionally more importance being given when there are more concepts extracted from the home article.

The outcome of this process is that the top-N concepts are selected from among the candidate articles, such that those which came from the image captions are given preference over those which came from in-going and out-going links of the home article. These concepts are used as the source for the query expansion.

3.3 Generating Expanded Queries

In order to ensure that the expanded queries remain focused on the topic of the query itself, the original query Q is pre-pended to each of the top-N related concepts $\{c_r | 0 \leq r \leq N\}$ as $< Q, c_r >$. We define c_0 to be null, producing the original query plus N expanded queries.

Given that individual expanded queries have differing degrees of relevance to the original query, we dynamically determine how many images to retrieve for each expanded query based on their relatedness score to the home article:

$$I_r = \frac{r(c_r, h) \times I_t}{\sum_{k=0}^{N} r(c_k, h)}$$

Here, r is the same function used to generate the relatedness score in the concept ranking process, and I_t is the total number of images to be retrieved by all of the queries (we set $I_t = 300$ in the current prototype). Since the null expanded query (c_0) is the original query, we define $r(c_0, h) = 1$ in the above calculation. All of the queries are sent to the Google AJAX Search, and the desired number of images are retrieved. Duplicate images are deleted based on the URL of the source image (as provided by the underlying search engine).

4 Visual and Conceptual Search Results Exploration

The difficulty with retrieving a broad and diversified range of images is how to then present these in a way that allows the searcher to focus on the specific aspect of the query they are interested in. A naïve approach would be to use a traditional paged grid layout of the images, ordered by their rank in the search results list and perhaps the semantic relatedness between their source concept to the original query concept. However, such an approach may not be all that effective in supporting image search tasks since the meaning of the organization of the images may be rather obscure. Instead, we propose a visual and interactive method for exploring the broad range of images retrieved from the expanded queries, taking advantage of both the visual features of the images and the concept from which they came.

4.1 Multi-Resolution SOM-Based Image Organization

In order to organize images based on visual similarity, we must first extract the visual features from the images. While a number of approaches have been studied within the domain of image processing [4], we use color-gradient correlation since it is efficient to calculate and offers good organizational performance [18]. We then train a SOM in a process similar to [17] to organize the images. The topology preserving property of the SOM ensures that images with similar feature vectors are mapped to locations that are closer to each other, and vice versa.

While the trained SOM provides the positioning coordinates for each image, it is impractical to provide and occlusion-free display of all images when the search results collection is large and the display resolution is limited. To facilitate the selection of which images to display under these space constraints, we assign priorities to all images based which are more representative of the images nearby. Of the images in a region of the SOM, a particular image is considered to be more representative if its feature vector is most similar to the average of the feature vectors in that region. This calculation is performed at progressively smaller resolutions, producing in a multi-resolution extension to the SOM [17]. As a result, only images with high display priority are shown when there is insufficient space to display all images. The amount of space available is relative to the screen resolution, as well as two parameters that can be controlled interactively: zoom level and image size.

User evaluations with this framework for visual image browsing and exploration studied the benefits of organizing the images based solely on their similarity (following a messy-desk metaphor) and in a more structured layout (following a neat-desk metaphor) [19]. While this study found that the approach can be very useful in comparison to traditional methods for organizing image search results, the differences between the layouts appeared to be based on personal preferences. For this work, we use the neat-desk layout since it provides less of a departure from what searchers expect in the presentation of image search results, while still maintaining a visual encoding of the degree of similarity.

In order to align the images to a structured grid, we generate a kd-tree [2] using the positions provided by the SOM. At the default zoom level, only the images with the highest display priority are shown. As the user zooms into a particular location, the amount of space available for displaying images grows. To avoid occlusion problems, new images are displayed only when there is sufficient space available between the currently displayed images.

4.2 Concept Hierarchy Focusing and Filtering

In addition to arranging the images from the search results visually, our approach also uses the concepts from which we derived the expanded queries to support focusing and filtering operations. Each of these concepts is mapped to an ontology using DBPedia [3]. This ontology is displayed to the searcher as a hierarchy, with the most semantically similar concepts to the original query placed at the top.

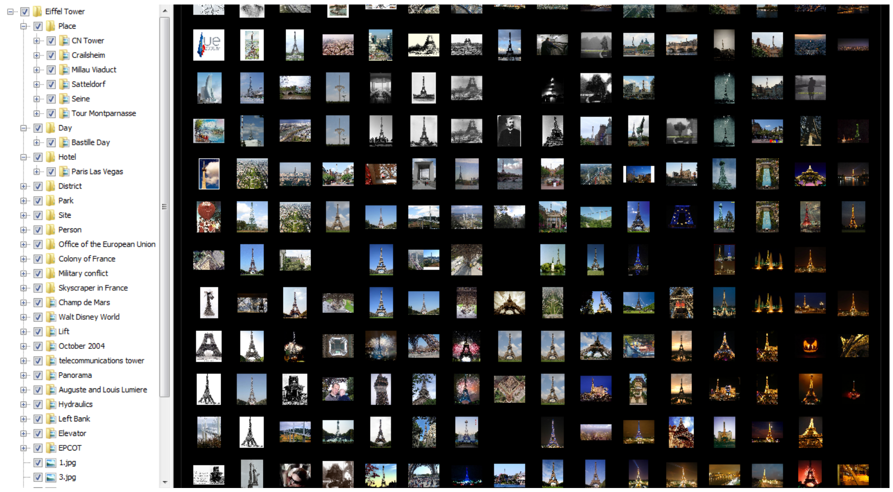
The searcher can use this hierarchy of concepts for both focusing and filtering operations. By clicking on any of the concepts, all the images that were retrieved as a result of that concept are pulled to the front of the display (temporarily increasing their display priority within the image organization process described above); the remaining images are dimmed giving the focused images more visual prominence.

The searcher can use checkboxes associated with each node in the concept hierarchy to filter the search results, removing the associated images from the display. At any time during the use of these conceptual filtering and focusing operations, the searcher can perform additional visual exploration, zooming into an area of interest to show additional visually similar images. Screenshots of the search results exploration interface are provided in Figure 1.

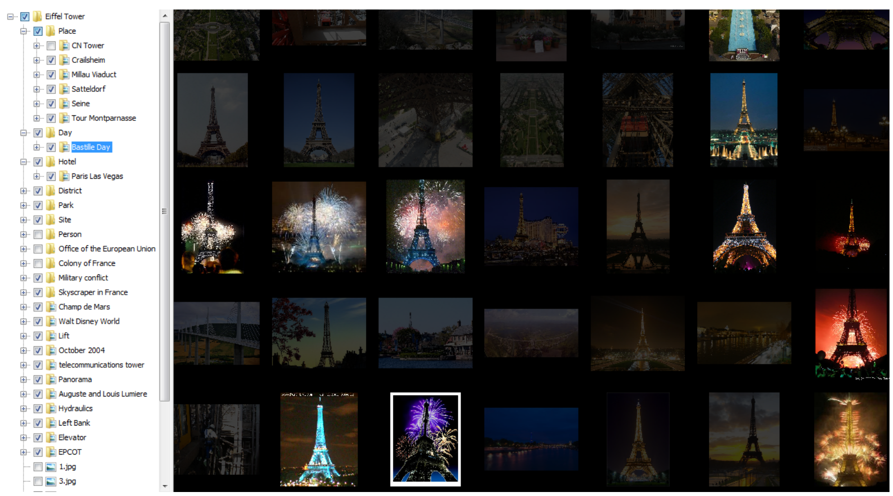
5 Conclusions & Future Work

In this paper, we describe an approach for performing conceptual query expansion, producing a diversified set of image search results which are then organized based on their visual features, and presented within an interactive interface. The primary contributions of this work are the novel use of Wikipedia for image retrieval, and the interactive support provided for conceptual and visual exploration within the image search space.

Future work includes adding features to support complex multi-concept queries, adding additional features that support interactive query refinement loops and query-by-example, and evaluating the approach through user studies. We are also examining the benefit of including conceptual information within the image organization process.



(a) The search results include images from the expanded queries, organized based on their visual similarity. Due to space constraints, only images with a high display priority are shown.



(b) After filtering the search results to remove irrelevant concepts, the search results are focused on the concept of "Bastille Day" which brings those images to the foreground, and zoomed to show the images in greater detail.

Fig. 1 Search results from the query "Eiffel Tower". The images from the expanded query are provided in (a); the search results are filtered, focused, and zoomed in (b).

References

1. André, P., Cutrell, E., Tan, D.S., Smith, G.: Designing novel image search interfaces by understanding unique characteristics and usage. In: Proceedings of the IFIP Conference on Human-Computer Interaction, pp. 340–353 (2009)

2. Bentley, J.L.: Multidimensional binary search trees used for associative searching. Communications of the ACM **18**(9), 509–517 (1975)
3. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S.: DBpedia - a crystallization point for the web of data. Journal of Web Semantics: Science, Services and Agents on the World Wide Web **7**(3), 154–165 (2009)
4. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Image retrieval: Ideas, influences, and trends of the new age. ACM Computing Surveys **40**(2), 1–60 (2008)
5. Enser, P., Sandom, C.: Towards a comprehensive survey of the semantic gap in visual image retrieval. In: Proceedings of the International Conference on Image and Video Retrieval, pp. 291–299 (2003)
6. Fonseca, B.M., Golgher, P., Pôssas, B., Ribeiro-Neto, B., Ziviani, N.: Concept-based interactive query expansion. In: Proceedings of the ACM International Conference on Information and Knowledge Management, pp. 696–703 (2005)
7. Gabrilovich, E., Markovitch, S.: Computing semantic relatedness using Wikipedia-based explicit semantic analysis. In: Proceedings of the International Joint Conference on Artificial Intelligence, pp. 1606–1611 (2007)
8. Heesch, D.: A survey of browsing models for content based image retrieval. Multimedia Tools and Applications **42**(2), 261–284 (2008)
9. Jansen, B.J., Spink, A., Pedersen, J.: An analysis of multimedia searching on AltaVista. In: Proceedings of the ACM SIGMM International Workshop on Multimedia Information Retrieval, pp. 186–192 (2003)
10. Jing, Y., Baluja, S.: VisualRank: Applying PageRank to large-scale image search. IEEE Transactions on Pattern Analysis and Machine Intelligence **30**(11), 1877–1890 (2008)
11. Joshi, D., Datta, R., Zhuang, Z., Weiss, W.P., Friedenberg, M., Li, J., Wang, J.Z.: PARAgrab: A comprehensive architecture for web image management and multimodal querying. In: Proceedings of the International Conference on Very Large Databases, pp. 1163–1166 (2006)
12. Kherfi, M.L., Ziou, D., Bernardi, A.: Image retrieval from the world wide web: Issues, techniques, and systems. ACM Computing Surveys **36**(1), 35–67 (2004)
13. Milne, D., Witten, I.H.: An effective, low-cost measure of semantic relatedness obtained from Wikipedia links. In: Proceedings of the AAAI Workshop on Wikipedia and Artificial Intelligence, pp. 25–30 (2008)
14. Milne, D., Witten, I.H.: Learning to link with Wikipedia. In: Proceedings of the ACM Conference on Information and Knowledge Management, pp. 509–518 (2008)
15. Milne, D., Witten, I.H., Nichols, D.M.: A knowledge-based search engine powered by Wikipedia. In: Proceedings of the ACM Conference on Information and Knowledge Management, pp. 445–454 (2007)
16. Myoupo, D., Popescu, A., Borgne, H.L., Moëllic, P.A.: Multimodal image retrieval over a large database. To Appear in Lecture Notes in Computer Science (2010)
17. Strong, G., Gong, M.: Browsing a large collection of community photos based on similarity on GPU. In: Proceedings of the International Symposium on Advances in Visual Computing, pp. 390–399 (2008)
18. Strong, G., Gong, M.: Organizing and browsing photos using different feature vectors and their evaluations. In: Proceedings of the International Conference on Image and Video Retrieval, pp. 1–8 (2009)
19. Strong, G., Hoeber, O., Gong, M.: Visual image browsing and exploration (Vibe): User evaluations of image search tasks. In: Proceedings of the International Conference on Active Media Technology, pp. 424–435 (2010)
20. Strube, M., Ponzetto, S.P.: WikiRelate! computing semantic relatedness using Wikipedia. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 1419–1424 (2006)