Organizing and Browsing Image Search Results based on Conceptual and Visual Similarities

Grant Strong Enamul Hoque Minglun Gong Orland Hoeber

Dept. of Computer Sci., Memorial Univ. of Newfoundland, St. John's, Canada

Abstract. This paper presents a novel approach for searching images online using textual queries and presenting the resulting images based on both conceptual and visual similarities. Given a user-specified query, the algorithm first finds the related concepts through conceptual query expansion. Each concept, together with the original query, is then used to search for images using existing image search engines. All the images found under different concepts are presented on a 2D virtual canvas using a self-organizing map. Both conceptual and visual similarities among the images are used to determine the image locations so that images from the same or related concepts are grouped together and visually similar images are placed close to each other. When the user browses the search results, a subset of representative images is selected to compose an image collage. Once having identified images of interest within the collage, the user can find more images that are conceptually or visually similar through pan and zoom operations. Experiments on different image query examples demonstrate the effectiveness of the presented approach.

1 Introduction

The primary method of image retrieval used on the Web is based on keyword search [12]. Search engines have merely adapted their document retrieval algorithms to the metadata (keywords, tags, and/or associated descriptions) of images and present the results in a scrollable list that is ranked based on relevance to the query. While list interfaces are easy to use, there is limited ability to manipulate and explore search results. In addition, keyword search relies on the assumptions that the contents of images are accurately described by the metadata, and that the searcher is able to provide a concise description of what they are seeking; these assumptions are not always valid.

On the other front, content-based image retrieval (CBIR) techniques conduct search using visual features [4, 17]. However, they often 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 [4]. Furthermore, CBIR techniques often require users to draw sketches as visual queries or to rank suggested images, making them somewhat cumbersome to use.

A nice middle ground seems to be searching images using keywords and then organizing the search results using similarity-based image browsing (SBIB) techniques [8]. This allows searchers to use easy-to-construct textual queries, as well as facilitates their locating of desired images through the structured layout of retrieval

results. Google Swirl is such an approach, which uses a visual similarity graph to present the images found through a given textual query [7].

Previous studies have shown that, many image search queries are associated with conceptual domains that include nouns, people's names, and locations [2, 10]. It is advantageous for the search results of such queries to be diverse in nature, both from different conceptual perspectives as well as the visual features. A diversified set of search result images provides a broad scope from which the searcher can seek the images that match their needs. In situations such as this, it may be more beneficial to consider not only visual similarity but also conceptual relatedness between images in organizing the results, and then allow the searcher to focus on a specific area to explore conceptually related images.

Motivated by the above hypothesis, here we propose a novel approach for organizing and browsing textual search results using a combination of conceptual and visual features. Given a user-specified query, our system first performs conceptual query expansion to ensure that a set of conceptually diverse images are retrieved. This is done by automatically extracting a list of concepts from Wikipedia that are relevant to the query, and then perform textual retrieval using both the original query and the related concepts. The relations between the concepts used for query expansion are used to derive a conceptual feature vector for each image, which is used in conjunction with the visual feature vector extracted from the image to form a hybrid feature vector. A self-organizing map based approach is then used to map images onto a 2D canvas, so that the ones with similar concepts and/or visual features are placed close to each other. The searcher can visually explore the search results on the 2D canvas, which initially contains representative images only. Zooming into an area of interests will unveil more conceptually and visually similar images; panning allows the searcher to move within the image space.

1.1 Related Work in Conceptual Query Expansion

A promising direction for improving the quality of search results in general is the introduction of query expansion based on the most related concepts to the query [5]. Such an approach is particularly useful for diversifying the search results covering different concepts and enabling 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 WordNet has been used to improve image retrieval [11], it does not contain information pertaining to the proper nouns that are common in image search queries. As such, using Wikipedia for reformulating queries has shown promise [14], 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! [21], Explicit Semantic Analysis (ESA) [6], and Wikipedia Link-based Measure (WLM) [13]. We use WLM in our work because of its computationally efficiency and accuracy.

1.2 Related Work in Similarity-based Image Browsing

Unlike CBIR, which aims to provide users with the desired images based on a set of input images, SBIB studies how to organize images, either from personal collections or online search results, based on their visual similarities. The challenge of SBIB is to arrange images in such a way as to support the browsing and exploration experience, as well as to facilitate users in locating the image(s) they are looking for.

Several SBIB approaches have been proposed, all of which use similar color, structure, and texture features as the basis of similarity measures for their organizations, but differ in how those similarity measures are used to relate images [8]. Heesch and Rüger et al. model relations between images using nearest neighbor networks adapted from document browsing techniques [9]. Search, in their case, is the process of following a path through the series of connections by clicking relevant images. Nguyen and Worring propose a system to meet the three requirements of overview, structure preservation (as it relates to image similarity), and visibility in [15]. They organize using non-linear probabilistic methods and k-means clustering to determine overview images while browsing. In [16], Pečenović et al. use Sammon's projection to map the images onto 2D space for visualization and use a heuristic balanced k-means algorithm for determining representative centroids. Strong and Gong's approach also maps images onto 2D space, but using a multi-resolution self-organizing map algorithm, which can evenly spread images across the available screen space and provide priority information for interactive browsing [18]. User studies have shown that this browsing interface helps to reduce the time users needed to locate the desired images [20].

Strong and Gong's approach is chosen as the basis for organizing image search results in our work. However, a key difference between our approach and all existing SBIB techniques is that we not only organize images based on visual information, but also extract and utilize concept information. Our experiments show that incorporating conceptual information can make the organization results much more meaningful than using visual information alone.

2 The Proposed Approach

2.1 Concept Extraction using Wikipedia

Images retrieved using a user specified query does not carry conceptual information. To address this problem, as well as to obtain a set of conceptually diverse images, we first apply conceptual query expansion to discover different concepts related to the input query. The concepts are then used to diversify the image search results, as well as for generating conceptual feature vectors for the images found.

In 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. Matching a user-supplied query Q to this knowledge base is simply a matter of selecting the best matching article (referred as the home article) using Wikipedia's search feature. In the case where query Q is ambiguous and Wikipedia suggests multiple links, the ones with higher commonness values are used as home articles. Here the commonness value of an article is calculated based on how often it is linked by other articles.

In analyzing Wikipedia, we observed that the in-link articles (ones having links to a home article) and out-link articles (ones to which a home article links) often provide meaningful information that is closely related to one of the home articles, and hence the user-specified query. Therefore, these linked articles are located and their titles are extracted as candidates for related concepts.

For some queries, the total number of linked articles found might be very high and some of them may not be well-related to the query. Thus, a filtering step is necessary to ensure the quality of the concepts that are extracted. The filtering is performed based on the semantic distance between each linked article and its corresponding home article. WLM [13] is used for this purpose, which applies Normalized Google Distance (NGD) on the domain of Wikipedia articles. The NGD between any two articles a and b is calculated using the hyperlink structure of the associated articles to determine how much they share in common. That is:

$$NGD(a,b) = \frac{\log(\max(|A|,|B|)) - \log(|A \cap B|)}{\log(|W|) - \log(\min(|A|,|B|))} \quad (1)$$

where A and B are the sets of all articles that link to the article of a and b, respectively, W is the set of all articles on Wikipedia, and operator $|\cdot|$ computes the number of articles in the set. The distance obtained is between 0 and 1, with a smaller value indicating a higher degree of relatedness.

Once the semantic distance measures for all linked articles are calculated, they are sorted in increasing order. The titles of the top N articles are then selected as related concepts.

2.2 Image Search using Conceptual Query Expansion

Given a user-specified query Q and the corresponding related concepts, $\{R_k | 1 \leq k \leq N\}$, we generate a set of N sub-queries by combining the query with each related concept. Each sub-query $\langle Q, R_k \rangle$ is then used to retrieve a set of images from the Web using the Ajax Goggle API. To avoid duplicate images returned for different sub-queries, a union operation is performed when combining the result sets. All images retrieved are tagged with the corresponding concept R_k. The total number of images retrieved is limited to T, a user tunable parameter.

For example, if a user enters the query "Washington", the system will use the query to perform a search in Wikipedia, which will return multiple articles about "Washington". Articles having higher commonness scores, such as "Washington (state)", "Washington, D.C.", "University of Washington", and "George Washington"

are then selected as home articles. All articles link to or linked by one of the home articles are considered as candidates for related concepts. The top N candidates with smallest semantic distances are used for generating sub-queries (e.g., "Washington Monument" and "White House" for the home article "Washington D.C."; "Martha Washington" and "Benjamin Franklin" for the home article "George Washington").

2.3 Hybrid Feature Vector Generation

The set of images retrieved through conceptual query expansion are highly diverse, both conceptually and visually. While this helps to ensure that the desired images are returned, finding them by going through the whole set can be time consuming. To facilitate searchers in locating the images they seek, a SOM-based image organization technique is applied to the image retrieval results.

SOM is a type of artificial neural network that is trained using unsupervised learning to produce a low-dimensional representation of the input space of the training samples. To use it for image organization, we first need to represent all images using feature vectors, the distances among which indicate the similarities between the corresponding images. Different ways for generating feature vectors from visual information and their performances have been studied previously [19]. While these types of feature vectors can be used to organize images based on color and/or shape similarities, they cannot group conceptually related, but visually different, images together. To address this problem, here we propose a hybrid feature vector.

The hybrid feature vector for an image contains two portions: a conceptual portion determined using the concept tag carried by the image; and a visual portion extracted from pixel intensities and distributions. For the visual portion, we choose the color-gradient correlation approach since it considers both color and general shape information, is efficient to calculate, and offers good organizational performance [19].

To compute the color-gradient correlation of an image I, we first compute the gradient magnitude l_p and gradient orientation θ_p for each pixel p. We then divide the color and gradient orientation spaces into N_C and N_Θ bins, respectively. With functions $C(p)$ and $\Theta(p)$ providing the color and gradient orientation bin indices for pixel p, the sum of gradient magnitudes for all pixels belonging to the k^{th} color and gradient orientation bin can be computed using:

$$m_k = \sum_{C(p)\times N_\Theta + \Theta(p) = k} l_p \tag{2}$$

The visual feature vector $\mathbf{V}(I)$ is then formed using the normalized values of all bins to make the vectors generated from images of different sizes comparable.

Extracting conceptual feature vectors, on the other hand, is not as straightforward. To simplify the problem, we assume that different images retrieved using the same sub-query (i.e., the same related concept) are conceptually the same and, hence have the same conceptual feature vector. Consequently, what we need to do is to derive a feature vector $\mathbf{C}_k$ for each concept R_k used for retrieving images. Even though it is difficult to convert concepts into vectors directly, we can first compute an $N \times N$ semantic distance matrix for the N related concepts using Equation (1):

$$\mathbf{D} = \begin{bmatrix} 0 & NGD(R_1, R_2) & \cdots & NGD(R_1, R_N) \\ NGD(R_2, R_1) & 0 & \cdots & NGD(R_2, R_N) \\ \vdots & \vdots & \ddots & \vdots \\ NGD(R_N, R_1) & NGD(R_N, R_2) & \cdots & 0 \end{bmatrix} \tag{3}$$

where, by definition, we have $NGD(R_k, R_k) = 0$ and $NGD(R_j, R_k) = NGD(R_k, R_j)$. Hence, the above matrix is symmetric.

The above matrix encodes the relatedness information among different concepts, which is then used to generate a set of m-dimensional vectors $\{\mathbf{C}_1, \ldots, \mathbf{C}_N\}$. We need the vectors to model the relatedness information as closely as possible (i.e., the distance between any two vectors is approximate to, if not the same as, the semantic distance between the corresponding concepts). This is the same as minimizing the following least-squares function:

$$\{\mathbf{C}_1, \ldots, \mathbf{C}_N\} = \underset{\mathbf{C}_1, \ldots, \mathbf{C}_N}{\operatorname{argmin}} \sum_{1 \le j,k \le N} \left(\left\| \mathbf{C}_j - \mathbf{C}_k \right\| - \mathbf{D}_{j,k} \right)^2 \tag{4}$$

where $\left\| \mathbf{C}_j - \mathbf{C}_k \right\|$ is the Euclidean distance between the two vectors.

As a result, our task of finding a set of vectors $\mathbf{C}$ based on a given distance matrix $\mathbf{D}$ becomes the classical multi-dimensional scaling problem [3], which can be solved by existing techniques [1].

In the end, the hybrid feature vector $\mathbf{H}(I)$ for a given image I is formed as $\langle \mathbf{C}_{R(I)}, \mathbf{V}(I) \rangle$, where $R(I)$ is the concept used to retrieve image I. Since the conceptual portion has m dimensions and the visual part has $N_c \times N_\theta$ dimensions, the total dimensions of a hybrid feature vector is $m + N_c \times N_\theta$. In this paper, we set $m = 4$ and $N_c = N_\theta = 8$, resulting 68 dimensional hybrid feature vectors.

2.4 Image Organization using Self-Organizing Map

With hybrid feature vectors generated for all images, the next step is to map the vectors onto a 2D virtual canvas. This is achieved through training a SOM, a process similar to the one discussed in [18, 19], with changes applied for handling hybrid vectors. The SOM is capable of organizing non-linear vectors in a topologically meaningful way. Unlike many other dimension reduction and vector embedding techniques, such as Sammon's projection [16], it has an inherent balancing property that seeks to spread the sample vectors over the whole map regardless of the span of the input vectors. Thus, when visualizing the organization of the image search results, images are evenly distributed in the display, making full use of the available screen space.

An SOM consists of 2D network of $M \times M$ interconnected units, where each unit x has, initially, a randomly generated weight vector $W(x)$ associated with it. During each iteration of the training process, all feature vectors affect an area of the map in a random order. The area is chosen by finding the unit with the weight vector that most closely matches the feature vector in terms of minimum distance. Then the best match unit and the neighboring units' weight vectors are interpolated toward the feature vector, where the amount of interpolation varies based on a Gaussian decay. The

overall learning effect dwindles exponentially over time.

Since hybrid feature vectors are used to encode images in our approach, the SOM can be trained based on either visual or conceptual information, resulting images being grouped by either visual or conceptual similarities. More generally, we can organize images by both visual and conceptual information through using a weighted average of both visual and conceptual distances. That is, given two images I and J, whose feature vectors are $\mathbf{H}(I)$ and $\mathbf{H}(J)$, their distance is calculated as:

$$\mathrm{Dist}\big(\mathbf{H}(I), \mathbf{H}(J)\big) = \alpha\big\|\mathbf{C}_{R(I)} - \mathbf{C}_{R(J)}\big\| + (1-\alpha)\|\mathbf{V}(I) - \mathbf{V}(J)\| \quad (5)$$

where the parameter α controls the relative importance of the conceptual distance and visual distance.

When the training is complete, the coordinates of the best match unit for each feature vector give gives us the best position for that feature vector's image in 2D. The SOM's topology preserving property ensures that visually and conceptually similar images are mapped to locations that are closer to each other, and vice versa.

2.5 Image Browsing Interface

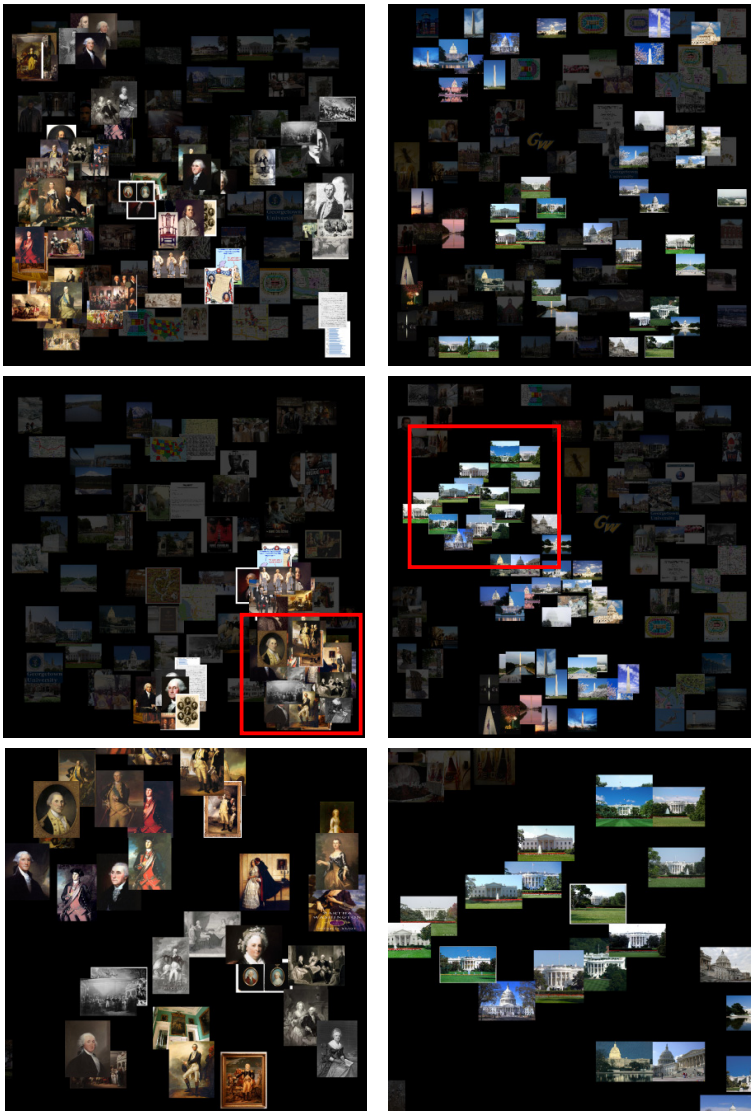
Taking the images and their mapped locations as input, the browsing interface dynamically generates an image collage based on which portion of the 2D virtual canvas is currently in view. The users can adjust the viewing area though intuitive panning and zooming operations. Once the viewing area is set, all of the images inside the area are candidates for generating the image collage, but only a number of representative images are actually used. An image is chosen to represent nearby images if its feature vector is close to the average of feature vectors in the neighborhood.

The number of images actually used depends on the screen size of the browsing window, as well as the user specified image display size. The zooming operation does not affect the display size of each individual image, but reduces the portion of the canvas that is viewable. As a result, when users zoom in, they can observe more images in the region rather than see the same set of images at higher resolution.

Under a typical browsing scenario, the user initially sees only a couple representative images. These may come from specific expanded queries that produce groups of images, or from other visually similar images that are obtained from multiple expanded queries. Once having identified images of interest, the searcher can find more images with similar concepts or visual features by zooming into the area.

3 Experiment Results

Thus far in this paper, we describe how to infuse concept information into the image search process and how to utilize the concept information in organizing the resulting images. Next we examine the merit of doing this for the purposes of browsing and exploring image search results.



Washington Washington, D.C.

Figure 1: Comparison between image organization results generated using the visual feature vector alone (top) and the hybrid feature vector (middle), as well as the zoomed-in views of the areas marked by the red squares (bottom). For illustration purposes, images of several persons related to George Washington are highlighted in the left dataset, and the ones related to buildings in Washington, D.C. are highlighted in the right dataset. One can readily see that the grouping of the highlighted images is much better with the hybrid feature vector (middle) than the visual feature vector (top).

The proposed algorithm is tested using a variety of datasets, all are retrieved using Google image search with the above described conceptual query expansion procedure. The results for two queries are shown here to demonstrate the performance of the algorithm under different levels of query ambiguity. The first one, "Washington", is highly ambiguous and the related images range from landmarks (Washington Monument and Space Needle), to persons (George Washington and Denzel Washington), and maps (Washington, D.C. and Washington state). It is also one of the examples used by Google Swirl, the result of which can be found at [7]. The second query simply appends "D.C." to the original query, which increases the specificity of the results significantly. Yet, the query expansion process still finds multiple related concepts, such as "Washington Monument", "White House", "United States Capitol", etc.

The organization of the results (shown in Figure 1) suggest that, since conceptually related images may be visually different, organizing images using visual features only does not place them together. Using concept information, on the other hand, not only can images retrieved using the same concept be grouped together, but also, images from related concepts can be placed at nearby locations. Hence, once users identify the images of interest, they can easily zoom into the area to find more conceptually related images.

4 Conclusions

In this paper, we present a novel approach of organizing and browsing image search results based on visual and conceptual similarities. The main contributions of this work are the infusing of related concepts into image search results through query expansion, the encoding of both visual and conceptual information in feature vector generation, the adopting of hybrid feature vectors in the training of the SOM, and finally the organizing of the image search results based on both conceptual and visual features. The benefit of this organization is that it groups images from the same or related concepts together, while simultaneously grouping visually similar objects, allowing users to explore their own areas of interests based on both conceptual and visual features of the images. Moreover, our experimental result shows that, this visualization technique can be particularly helpful for dealing with an ambiguous query, by separating images of ambiguous concepts from each other into their own area and placing visually and conceptually similar concepts near one another.

The future work includes plans to allow the user to provide further input on the process of query refinement and image search results organization. We are also interested in enhancing the quality of feature vectors through incorporating other types of visual features. Finally, we want evaluate the benefit of this approach through user evaluations.

References

1. Algorithmics Group: MDSJ: Java Library for Multidimensional Scaling (Version 0.2),

http://www.inf.uni-konstanz.de/algo/software/mdsj/ (2009)
2. P. André, E. Cutrell, D. S. Tan and G. Smith: Designing novel image search interfaces by understanding unique characteristics and usage. In: Proc. IFIP Conference on Human-Computer Interaction (2009) 340–353
3. I. Borg and P. Groenen: Modern Multidimensional Scaling: Theory and Applications. 2nd ed., Springer (2005)
4. 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) (2008) 1-60
5. B. M. Fonseca, P. Golgher, B. Pôssas, B. Ribeiro-Neto and N. Ziviani: Concept-based interactive query expansion. In: Proc. ACM International Conference on Information and Knowledge Management (2005) 696–703
6. E. Gabrilovich and S. Markovitch: Computing semantic relatedness using wikipedia-based explicit semantic analysis. In: Proc. International Joint Conference on Artificial Intelligence (2007) 1606-1611
7. Google: Google Image Swirl, http://image-swirl.googlelabs.com/ (2009)
8. D. Heesch: A survey of browsing models for content based image retrieval. Multimedia Tools and Applications 40 (2) (2008) 261-284
9. D. Heesch and S. Rüger: Image Browsing: A semantic analysis of NNk networks. In: Proc. International Conference Image and Video Retrieval (2005) 609-618
10. B. J. Jansen, A. Spink and J. Pedersen: An analysis of multimedia searching on AltaVista. In: Proc. ACM SIGMM International Workshop on Multimedia Information Retrieval (2003) 186–192
11. D. Joshi, R. Datta, Z. Zhuang, W. P. Weiss, M. Friedenberg, J. Li and J. Z. Wang: PARAgrab: A comprehensive architecture for web image management and multimodal querying. In: Proc. International Conference on Very Large Databases (2006) 1163-1166
12. M. L. Kherfi, D. Ziou and A. Bernardi: Image Retrieval from the World Wide Web: Issues, Techniques, and Systems. ACM Computer Survey 36 (1) (2004) 35-67
13. D. Milne and I. H. Witten: An effective, low-cost measure of semantic relatedness obtained from wikipedia links. In: Proc. AAAI Workshop on Wikipedia and Artificial Intelligence (2008) 25-30
14. D. Myoupo, A. Popescu, H. L. Borgne and P. A. Moëllic: Multimodal image retrieval over a large database. Lecture Notes in Computer Science (2010) 1-8
15. G. P. Nguyen and M. Worring: Interactive access to large image collections using similarity-based visualization. J. Vis. Lang. Comput. 19 (2) (2008) 203-224
16. Z. Pečenović, M. Do, M. Vetterli and P. Pu: Integrated Browsing and Searching of Large Image Collections. In: Proc. International Conference on Advances in Visual Information Systems (2000) 279-289
17. 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 Transactions on Pattern Analysis and Machine Intelligence 22 (12) (2000) 1349-1380
18. G. Strong and M. Gong: Browsing a large collection of community photos based on similarity on GPU. Advances in Visual Computing. Springer Berlin, Las Vegas, NV, USA, (2008) 390-399
19. G. Strong and M. Gong: Organizing and Browsing Photos Using Different Feature Vectors and Their Evaluations. In: Proc. International Conference on Image and Video Retrieval (2009) 1-8
20. G. Strong, O. Hoeber and M. Gong: Visual image browsing and exploration (vibe): user evaluations of image search tasks (2010)
21. M. Strube and S. P. Ponzetto: WikiRelate! computing semantic relatedness using wikipedia. In: Proc. AAAI Conference on Artificial Intelligence (2006) 1419-1424