Visual Image Browsing and Exploration (Vibe): User Evaluations of Image Search Tasks

Grant Strong, Orland Hoeber, and Minglun Gong

Department of Computer Science, Memorial University
St. John's, NL, Canada A1B 3X5
{strong, hoeber, gong}@cs.mun.ca

Abstract. One of the fundamental challenges in designing an image retrieval system is choosing a method by which the images that match a given query are presented to the searcher. Traditional approaches have used a grid layout that requires a sequential evaluation of the images. Recent advances in image processing and computing power have made similarity-based organization of images feasible. In this paper, we present an approach that places visually similar images near one another, and supports dynamic zooming and panning within the image search results. A user study was conducted on two alternate implementations of our prototype system, the findings from which illustrate the benefit that an interactive similarity-based image organization approach has over the traditional method for displaying image search results.

1 Introduction

Image search tasks can be divided into two fundamentally different categories: discovery and rediscovery. Within a rediscovery task, the searcher knows precisely what image they are looking for and seeks to either find it in the search results collection, or decide that it is not present. In contrast, when a searcher is performing a discovery task, the mental model of the image for which they are searching is often vague and incomplete. Within the search results collection, there may be many images that match the desired image to various degrees. The primary activities for the searcher in such discovery tasks are browsing and exploration.

In this paper, we evaluate how visual image browsing and exploration, as implemented in Vibe, can assist searchers in preforming discovery tasks within the domain of image search. The fundamental premise is that a visual approach to image organization and representation that takes advantage of the similarities between images can enhance a searcher's ability to browse and explore collections of images. Vibe is an example of a web information retrieval support system (WIRSS) [5]; its purpose is to enhance the human decision-making abilities within the context of image retrieval. The primary method of image retrieval used on the Web is based on keyword search. Search engines merely adapt their document retrieval algorithms to the context of images and present the results

in a scrollable list ranked on query relevance. While list interfaces are easy to use there is limited ability to manipulate and explore search results.

To facilitate an exploration of a collection of image search results, Vibe arranges the images by content similarity on a two-dimensional virtual desktop [9, 10]. The user can dynamically browse the image space using pan and zoom operations. As the user navigates, an image collage is dynamically generated from selected images. At the broadest zoom level, the images in the collage are those that best represent the others in their respective neighbourhoods, providing a high-level overview of the image collection. As the searcher zooms in toward an image of interest, more images that are visually similar to the area of focus are dynamically loaded. The benefit of this interaction method is that the user has the ability see as little or as much detail as they wish; a single unified interface provides both a high-level overview and details of a subset of the image collection.

Two different methods for organizing the collection of images in Vibe are discussed and evaluated in this paper. The original design of Vibe displays images in irregular patterns [9], following a messy-desk metaphor. In a preliminary evaluation of the interface, we found that once searchers zoomed into a particular area of interest in the image space, they sometimes experienced difficulties scanning the irregularly placed images within the display. A potential solution to this difficulty is to align the images in the messy-desk arrangement into a more structured neat-desk layout in order to enhance the ability of searchers to linearly scan the images. This method maintains the similarity-based organization of the images, but relaxes the use of distance between pairs of images to represent a measure of their similarity.

Where user productivity and enjoyment are concerned, we feel that the characteristics of Vibe have merit. The results of a user evaluation conducted in a controlled laboratory setting are reported in this paper. The evaluation compares three image search interfaces: messy-desk Vibe, neat-desk Vibe, and a scrollable grid layout similar to that found in Web image search engines.

The remainder of this paper is organized as follows. Section 2 provides an overview of image retrieval and organization. Section 3 outlines the specific features of Vibe and the techniques used to construct the similarity-based image organization. Section 4 describes the user evaluation methods, followed by the results of the study in Section 5. The paper concludes with a summary of the research contributions and an overview of future work in Section 6.

2 Related Work

Techniques for finding specific images in a large image database has been studied for decades [2]. Most current Web-based image search engines rely on some form of metadata, such as captions, keywords, or descriptions; the matching of queries to images is performed using this metadata. Manual image annotation is tedious and time consuming, whereas the results of automatic annotation are still unreliable. Hence, methods for performing retrieval using image content directly,

referred as Content-based Image Retrieval (CBIR) [7, 2], have been extensively studied.

While CBIR approaches normally assume that users have clear search goals, Similarity-based Image Browsing (SBIB) approaches cater to users who wish to explore a collection of images, but do not have a clearly defined information need [4]. The challenge of SBIB is to arrange images based on visual similarities in such a way as to support the browsing and exploration experience. This paper investigates whether SBIB techniques, as implemented in Vibe, can improve users' image searching experience and performance.

Several SBIB techniques have been proposed. Torres et al. [11] prescribe ways to enhance CBIR results by browsing them in spiral or concentric ring representations. The position and size of the images vary with their measure of similarity to the query. In Chen et al.'s approach [1], contents of image databases are modelled in pathfinder networks. The result is a branched clustering constructed with reference to the histogram or texture similarity between images. Snavely et al. [8] provide an interesting way to arrange and browse large sets of photos of the same scene by exploiting the common underlying 3D geometry in the scene.

The image browsing technique evaluated in this paper is derived from Strong and Gong's previous work [9, 10]. We adopt their idea of organizing images in 2D space by training a neural network. Two alternative approaches to laying out the images are provided and studied.

3 Vibe

The Vibe technique can arrange images in two alternative ways, which are referred to as messy-desk and neat-desk layouts, respectively. Both layouts place images on a 2D virtual desktop so that visually similar images are close to each other. The difference is that images can be positioned at arbitrary locations in the messy-desk layout, but have to be aligned to a grid in the neat-desk layout. Vibe also supports dynamic pan and zoom operations within the image search results space, allowing the searcher to easily browse and explore the images. The rest of this section discusses the methods for generating these two layouts, and the techniques for supporting interactive exploration and browsing.

3.1 Feature Vector Generation

In order to organize images based on similarity, we need to define a way of measuring the similarity between any two images. Here the similarity is computed using the Euclidean distance between two feature vectors, which are extracted from images to represent the salient information. In this paper, the color-gradient correlation is used since it is easy to calculate and offers good organizational performance [10].

To compute the color-gradient correlation for an input image I, we first compute the gradient magnitude l_p and gradient orientation θ_p for each pixel p. We then divide the colour and gradient orientation spaces into N_c and N_θ

bins, respectively. Assuming that functions $C(p)$ and $\Theta(p)$ give us the colour and gradient orientation bin indices for pixel p, the sum of gradient magnitudes for all pixels belonging to the k^{th} colour and gradient orientation bin can be computed using:

$$m_k = \sum_{p \in I \wedge C(p) \times N_\Theta + \Theta(p) = k} l_p \tag{1}$$

where $N = N_c \times N_\theta$ is the total number of bins. In practice, we set $N_c = 8$ and $N_\theta = 8$, resulting a 64-dimensional feature vector $F(I)$, and then normalize the final vector.

3.2 Messy-Desk Layout

Given a collection of T images, the messy-desk layout tries to position them on a 2D virtual desktop, so that visually similar images are placed together. This layout is generated by training a Self-Organizing Map (SOM), a process similar to the one discussed in [9].

A SOM is a type of artificial neural network that is trained through unsupervised learning. It is used here to map N-dimensional vectors to 2D coordinate space. SOMs consist of $M \times M$ units, where each unit x has its own N-dimensional weight vector $W(x)$. For dimension reduction we ensure that $M \times M \gg T$, making it possible to map distinct vectors to unique locations in the SOM.

The SOM training process requires multiple iterations. During each iteration all images in the collection are shown to the SOM in a random order. When a particular image I is shown, the goal is to find the best match unit B and then update the weight vectors in B's neighbourhood proportionally based on the distance between B and the neighbouring unit in the SOM. After the SOM converges, the coordinates of the best match unit $B(I)$ for each image I gives us the mapping in 2D. The SOM's topology preserving property ensures that images that have similar vectors are mapped to locations that are closer to each other, and vice versa.

3.3 Neat-Desk Layout

The messy-desk layout groups visually similar images together, allowing users to quickly narrow down the search results to a small area of the virtual desktop. However, preliminary evaluations found that users sometimes have difficulty locating the exact image they want because the irregular image layout makes it hard to remember which images have already been inspected. To address this problem, we propose another way to organize images, referred to as the neat-desk layout.

The neat-desk layout constrains images positions to be aligned to a grid. Since a trained SOM cannot guarantee one image per unit, we cannot simply use a SOM has the same number of units as the grid we want to align the images to. Instead, we generate the neat-desk layout from the messy-desk layout. As shown in Figure 1, given the collection of images and their 2D locations in the

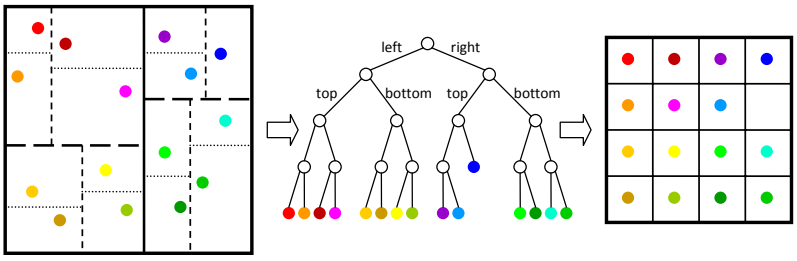


Fig. 1. Converting from a messy-desk to the neat-desk layout using a k-d tree.

messy-desk layout, the k-d tree algorithm is used to arrange the images into a neat-desk layout. The algorithm starts by finding the median value among the horizontal coordinates of all images, and uses this to split the collection into left and right halves. It then computes the median value among the vertical coordinates of images in each half, so that each half is further split into top and bottom quarters. The above two steps are repeated until each node contains at most one image. At the end, all images are contained in the leafs of a balanced binary tree. Based on the position of each leaf, we can assign a unique location to its associated image in the neat-desk layout.

In the messy-desk approach, two images that are very similar to one another will be placed in close proximity. The resulting gaps and irregular placement of images provide a good representation of the visual clustering, but make sequential evaluation of images difficult. The neat-desk layout produces a more regular layout, at the expense of losing the visual encoding of the degree of similarity.

3.4 Determining Display Priority

While the above layouts handle the positioning of the images in a collection, it is impractical to display all images at those positions when the collection is large. To facilitate the selection of images to display at run time, we pre-assign priorities to all images. The priority assignment is based on the criteria that the more representative images should have higher priorities to allow them to be selected first.

For the messy-desk layout, the images' priorities are determined using a multi-resolution SOM [9]. The bottom SOM, the one with the highest resolution, is obtained using SOM training procedure described in Section 3.2. The upper level SOMs are generated from the lower level ones directly without training. This is done by assigning each unit in an upper level SOM the average weight vector of its children in the lower level SOM. The average weight vector is then used to find the best matching image for each unit in the upper level SOMs. The upper level images represent their neighbourhoods below and are given a higher priority for display.

The same principle is applied for the neat-desk layout. The bottom level grid holds all images, each in its assigned location. An upper level grid contains a

Fig. 2. The layout of images using messy-desk Vibe (top row) and neat-desk Vibe (bottom row) for the same collection of images at three different levels of zoom. Note the visual similarity of images that are near to one another.

quarter of the grid points, with each point p linking to four child locations in the lower level grid. To select a single image for the grid point p, we first compute the average vector using images mapped to p's four child locations, and then pick the image that has the vector closest to the average.

3.5 Browsing Interface

Given the images and their mapped locations in either messy-desk or neat-desk layouts, the browsing interface selectively displays images at their mapped locations based on the users' pan and zoom interactions with the interface [9]. The number of images shown depends on the display resolution, the zoom level, and the user specified image display size. If the system is unable to fit all of the available images inside the viewing area, the ones with higher display priorities are shown. Figure 2 shows the three different levels of zoom for both the messy-desk and neat-desk layout methods.

Panning is implemented using a mouse drag operation, which translates the current viewing area. Zooming adjusts the size of the viewing area and is achieved using the normal mouse wheel operations. Zooming out enlarges the viewing area and allows users to inspect the overall image layout on the virtual desktop, whereas zooming in reduces the viewing area, making it possible to show the images in a local region in greater detail. It is worth noting that the zooming operation only changes the image display size when room is available (i.e., the

view is at the lowest level and there are no "deeper" images); otherwise it is provides a filtering operation that pulls and pushes images into and out of the view area.

The browsing interface also provides two ways for adjusting the display size of the images. First, the users can use the combination of the control key and mouse wheel to change the size of all displayed images, which also affects the total number of images that can be shown within the limits of the current view. Secondly, users are able to selectively enlarge an image of interest with a double-click.

4 Evaluation

In order to explore the differences between the traditional grid layout of image search results and the interactive content-based approach implemented in Vibe, a user evaluation was conducted in a controlled laboratory setting. In this study, the messy-desk Vibe (Vibe-m) and the neat-desk Vibe (Vibe-n) are compared to a grid layout (Grid). In order to reduce the interaction differences between the systems being studied, the Grid was implemented as a single scrollable grid (rather than the more common multi-page approach).

4.1 Methods

Although a number of options are available for studying search interfaces [6], we conducted a user evaluation in a laboratory setting in order to obtain empirical evidence regarding the value of the similarity-based approach for image search results representation. The controlled environment of the study allowed us to manage and manipulate the factors that we believed would have an impact on a participants performance and subjective reactions. At the same time, we were also able to ensure that the search tasks each participant performed were the same.

The study was designed as a 3×3 (interface $\times$ search task) between-subjects design. Each participant used each interface only once, and conducted each search task only once. To further alleviate potential learning effects, a Graeco-Latin square was used to vary the order of exposure to the interface and the order of the task assignment. Prior to performing any of the tasks, participants were given a brief introduction to the features of each of the three interfaces.

A set of three situated search tasks were provided to the participants, for which they used either Vibe-m, Vibe-n, or the Grid. For each task, participants were given a scenario in which they were asked to find five images that were relevant to the described information need (see Table 1). The tasks were chosen to be somewhat ambiguous, requiring the participants to explore the search results in some detail. The images used for all three datasets were obtained from Google Image Search by searching with the corresponding keywords. In addition, the order of images displayed in the Grid follow the order returned by Google search.

Table 1. Tasks assigned to participants in the user evaluation.

Query	*Information Need*
"Eiffel Tower"	Find five images of sketches of the Eiffel Tower.
"Notre Dame"	Find five images of the stained glass windows of the Notre Dame Cathedral.
"Washington"	Find five images of Denzel Washington.

For each task, measurements of time to task completion, accuracy, and subjective measures were made. Pre-study questionnaires were administered to determine prior experience with image search, educational background, and computer use characteristics. In-task questionnaires measured perceptions of quality of the search results and ease of completing the task. Post-study questionnaires followed the guidelines of the Technology Acceptance Model (TAM) [3], measuring perceived usefulness and ease-of-use, along with an indication of preference for an image search interface.

4.2 Participant Demographics

Twelve individuals were recruited from the student population within our department to participate in this study. They reported using a wide range of systems for the purposes of searching for images. These included the top search engines (e.g., Google, Bing, and Yahoo), other online services (e.g., Flickr, Picasa, and Facebook), and desktop software (e.g., iPhoto, Windows Photo Gallery, and file browsers). As a result, we can conclude that all of the participants in the study were very familiar with the traditional grid-based approach to image layout.

5 Results

5.1 Time to Task Completion

The average time required to complete the three tasks with the three interfaces are illustrated in Figure 3. Clearly, these results are somewhat varied. For the

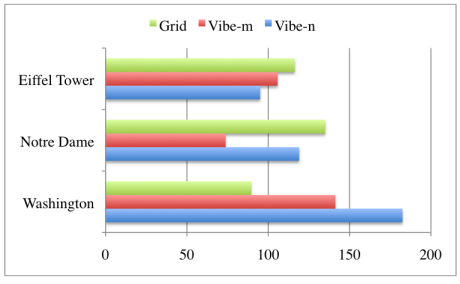


Fig. 3. Average time to task completion measurements from the user evaluation.

"Eiffel Tower" and "Notre Dame" tasks, participants performed better using both versions of Vibe than the Grid. However, which version of Vibe performed better was different between the two tasks. For the "Washington" task, participants performed better using the Grid than either version of Vibe.

ANOVA tests were preformed on these measurements to determine whether their differences were statistical significant. Among these results, only three were significant. For the "Notre Dame" task, the time taken to complete the task using Vibe-m was faster than both the Grid ($\mathbf{F(1,7) = 12.4, p < 0.05}$) and Vibe-n ($\mathbf{F(1,7) = 8.15, p < 0.05}$). For the "Washington" task, the time to completion using the Grid was faster than Vibe-m ($\mathbf{F(1,7) = 6.49, p < 0.05}$). For the rest of the pair-wise comparisons, the differences were not statistically significant. For most combinations of tasks and interfaces, there was a high degree of variance in the time to task completion measurement, indicating that the ability to complete the tasks is more a function of the skill and interest of the participant than the interface used to browse, explore, and evaluate the image search results.

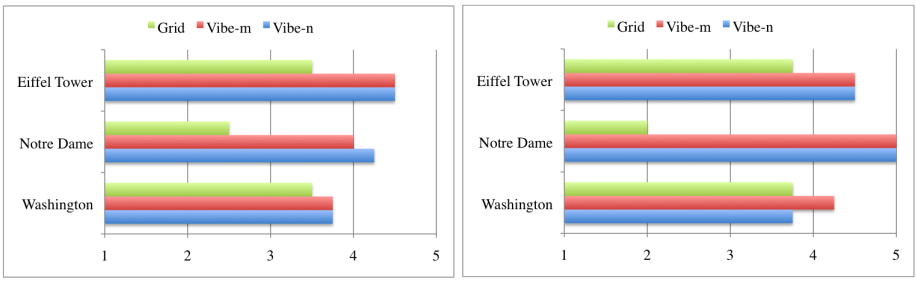
One aspect of particular note is the situation where the Grid allowed the participants to complete the "Washington" task faster than with either version of Vibe. Within Vibe, the system was effective in grouping images with similar global features, but not very effective in putting together images with similar local features. Since the images that contain people are strongly influenced by the background, these images are not necessarily placed together in Vibe. While participants were able to navigate to a location of interest easily, if they were unable to find enough relevant images in that location (e.g., images of Denzel Washington), they were hesitant to zoom out and continue exploring. As a result, it took them longer to find the images than sequentially searching the image space. Nevertheless, this suggests that the users were able to use the spatial layout information presented in the Vibe interface effectively. As the methods for grouping images based on local features improve, issues such as this will be eliminated.

5.2 Accuracy

After the participants completed the tasks, the five selected images were carefully inspected to verify their relevance to the information need. ANOVA tests across all three tasks indicate that there are no statistically significant differences in the accuracy when using the different interfaces ("Eiffel Tower": $F(2,11) = 1.29, p = 0.32$; "Notre Dame": $F(2,11) = 1.00, p = 0.41$; "Washington": $F(2,11) = 0.346, p = 0.72$). The average number of errors ranged from zero to 0.75. This result indicates that the exploratory nature of Vibe neither helped nor hindered the participants in deciding the relevance of individual images to the search task.

5.3 Subjective Reactions

After each task was complete, participants were asked to indicate their degree of agreement to statements related to the quality of the search results and the



(a) quality of search results (b) ease of the search task

Fig. 4. Average response to statements related to the search tasks.

ease at which they were able to complete the task (using a five-point Likert scale where high values indicated agreement). The average responses to these questions are reported in Figure 4. For the "Eiffel Tower" and "Notre Dame" tasks, one can readily see that participants perceived the search results to be of higher quality and the tasks to be easier to perform when using either version of Vibe compared to the Grid. For the "Washington" task, it appears that since there was some difficulty with Vibe being able to organize the local features of people in the images properly, the participants provided similar responses for all three interfaces.

The statistical significance of these results were evaluated using pair-wise Wilcoxon-Mann-Whitney tests. Significance was found only for certain comparisons in the "Notre Dame" query. For the quality of search results measure, only the Grid vs. Vibe-n ($\mathbf{Z = -2,055, p < 0.05}$) comparison was statistically significant. For the ease of search task measure, only the Grid vs. Vibe-m ($\mathbf{Z = -2.494, p < 0.05}$) and Grid vs. Vibe-n ($\mathbf{Z = -2.494, p < 0.05}$) comparisons were statistically significant.

Since the data from these in-task questionnaires was rather sparse, questions related to the overall perception of the usefulness and ease of use of the interface were collected in the post-study questionnaire, using the TAM instrument. Since this data was not collected in the context of a particular task, aggregate results of all participants and all TAM statements are shown in Figure 5. Wilcoxon-

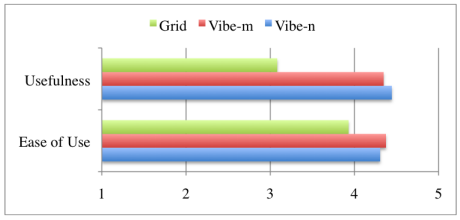


Fig. 5. Average response to statements regarding to the usefulness and ease of use of the interface.

Table 2. Statistical analysis (Wilcoxon-Mann-Whitney tests) of the responses to the TAM questions.

	Grid vs. Vibe-m	*Grid vs. Vibe-n*	*Vibe-m vs. Vibe-n*
Usefulness	$\mathbf{Z = -7.578, p < 0.001}$	$\mathbf{Z = -7.966, p < 0.001}$	$Z = -0.967, p = 0.334$
Ease of Use	$\mathbf{Z = -2.775, p < 0.05}$	$\mathbf{Z = -2.206, p < 0.05}$	$Z = -0.785, p = 0.432$

Mann-Whitney tests were performed on the responses using a pair-wise grouping of the interfaces. The results from this statistical measure are reported in Table 2, showing that participants found either version of Vibe more useful and easy to use than the Grid. The differences between Vibe-m and Vibe-n were not found to be statistically significant.

5.4 Preference

At the end of the study, participants were asked to indicate their preference for an image search interface. Four participants indicated a preference for Vibe-m (33%), six for Vibe-n (50%), and two for the Grid (17%). This clearly indicates a high degree of preference for the dynamic layout and interactive features of Vibe. A Wilcoxon signed rank sum test found statistical significance ($\mathbf{Z = -2.309, p < 0.05}$) in the preference of Vibe over the Grid. The preference between the messy-desk and neat-desk layouts was not statistically significant ($Z = -0.632, p = 0.53$).

6 Conclusions & Future Work

In this paper, we present an interactive visual interface that supports the browsing and exploration of image search results (Vibe). Two different versions of Vibe were created and studied in comparison to the commonly used grid layout. The messy-desk layout version of Vibe places images on a 2D virtual desktop, using the distance between images to represent their similarity. The neat-desk layout adds structure to the image arrangement. Both versions of Vibe provide dynamically generated collages of images, which can be interactively panned and zoomed. As the searcher zooms into an area of interest and more space is created in the view, more images from the search space are dynamically displayed. This interaction results in a filtering and focusing of the search space, supporting the searcher in discovering relevant images.

As a result of the user evaluation, we conclude that Vibe can improve the time it takes to find relevant images from a collection of search results. However, there are situations where the overhead of browsing and exploring outweighs the time saved in finding relevant images. Further study is required to examine the boundary conditions for increasing or decreasing searcher performance.

During the study, the perception of search results quality and ease of completing the tasks was higher for Vibe than for the grid layout. However, the

degree and significance of this result was dependent on the task. By the end of the study (after each participant was exposed to each of the three interfaces), measurements of usefulness and ease of use showed a clear and statistically significant preference for Vibe. These results indicate that the the participants were able to see the value in using Vibe for their image search tasks, even though the time taken to find relevant images was not necessarily improved. Further validation of this outcome was provided by the fact that 83% of the participants preferred to use Vibe over a grid layout.

In terms of the differences between the messy-desk and neat-desk layout, no clear conditions were found in this study indicating when one layout method was superior to the other. Whether a participant found one or the other easier to use may simply be a matter of personal preference. However, further study to identify the value of one layout method over the other will be of value.

References

1. Chen, C., Gagaudakis, G., Rosin, P.: Similarity-based image browsing. In: Proceedings of the IFIP International Conference on Intelligent Information Processing. pp. 206–213. Beijing, China (2000)
2. 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)
3. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. Management Information Systems Quarterly 13(3), 319–340 (1989)
4. Heesch, D.: A survey of browsing models for content based image retrieval. Multimedia Tools and Applications 42(2), 261–284 (2008)
5. Hoeber, O.: Web information retrieval support systems: The future of Web search. In: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence - Workshops (International Workshop on Web Information Retrieval Support Systems). pp. 29–32 (2008)
6. Hoeber, O.: User evaluation methods for visual Web search interfaces. In: Proceedings of the International Conference on Information Visualization. pp. 139–145. IEEE Press, Los Alamitos, CA, USA (2009)
7. Smeulders, A., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(12), 1349–1380 (2000)
8. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: Exploring photo collections in 3d. In: Proceedings of the ACM International Conference on Computer Graphics and Interactive Techniques. pp. 835–846 (2006)
9. 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)
10. 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)
11. Torres, R.S., Silva, C.G., Medeiros, C.B., Rocha, H.V.: Visual structures for image browsing. In: Proceedings of the International Conference on Information and Knowledge Management. pp. 49–55 (2003)