

A Granular Computing Perspective on Image Organization within an Image Retrieval Context

Orland Hoerber^{1,2} and Minglun Gong¹

¹ Department of Computer Science
Memorial University of Newfoundland
St. John's, NL A1B 3X5
Canada

{hoerber, gong}@mun.ca

² Department of Computer Science
University of Regina
Regina, SK S4S 0A1
Canada

orland.hoerber@uregina.ca

Abstract. The field of granular computing deals with representing, organizing, and processing information based on different levels of abstraction or aggregation. In the domain of image search, an increasingly common approach is to organize and aggregate the retrieved images within multi-level structures. In this paper, we will explore some of the core principles of granular computing within the context of image retrieval, discussing how our hierarchical approach to image clustering supports the searchers' decision-making tasks within the context of image retrieval.

Keywords: granular computing, image retrieval, image organization, multi-resolution SOM, interaction, decision-making

1 Introduction

Granular computing deals with structuring information associated with a problem at different levels of abstraction (granulation). At a high level of abstraction, information is aggregated resulting in a small number of large granules that represent general information. At a low level of abstraction, the degree of aggregation is much lower, resulting in many smaller granules that represent much more specific information. The advantage of problem solving within a granular framework is that it allows irrelevant details to be ignored and the problem to be addressed at the most appropriate level of abstraction. Viewing the problem from a high level of granularity leads to approximate solutions; delving deeper into lower levels leads to successive increases in precision [21].

Bargeila and Pedrycz [2] have suggested that there are three fundamental elements required to engage in granular computing. A granular computing framework must support multiple levels of information granularity, allow for the encoding and decoding of information between these levels of granularity, and

support non-homogeneous computation and analysis methods depending on the level of granularity. Yao [21] reinforced these fundamental elements through the specification of three principles of granular computing: *multilevel granularity*, *granularity conversion*, and *focused effort*.

Granular computing has also been discussed from three complementary perspectives: philosophy, methodology, and computation [21–23]. The philosophical perspective deals with structured thinking (e.g., the human thought process of decomposition and integration). The methodological perspective deals with structured problem solving (e.g., the classical “divide and conquer” approach). The computational perspective deals with structured information processing (e.g., representing the granular information in structures that lend themselves to computation, and the actual process of computation on these structures). The common theme across all of these perspectives is the fundamental notion of multilevel granular decomposition and hierarchical organization.

In this paper, we will discuss how our recent work to develop a visual and interactive interface that supports Web image search tasks [7, 17] follows a granular computing paradigm. We will use Yao’s principles as the basis for not only describing how images are organized, but also how searchers can interact within this interface and the support it provides for decision making in the context of image retrieval. This discussion will primarily follow the methodological perspective on granular computing, focusing on how these systems support the human decision-making processes required in image retrieval.

2 Fundamentals of Image Retrieval

At a conceptual level, image retrieval is not very different from other types of information retrieval (e.g., document retrieval). However, at a practical level, there are significant differences that make it more difficult from an information-centric perspective, but also more effective from a human-centric perspective.

Many image retrieval approaches used on the Web are based on keyword search algorithms [8]. The image index is generated based on terms that are used in relation to the images (e.g., tags used in links, words used on the same page as the image, etc.). Searchers provide a textual description of their image needs, and the underlying search engine matches these terms to the index, returning a ranked set of corresponding images. This approach can work well if the images are accurately and fully described within the index, and if the searcher is able to provide a complete and precise description of what it is they are seeking. Unfortunately, these constraints are seldom met.

The issue of creating a robust index of images on the Web is especially challenging, given the sheer number of images as well as the difficulty with automatically determining accurate index terms. Some have attempted to address this problem from the perspective of content-based image retrieval [13], wherein the searcher provides one or more images or sketches that describe what they are seeking, and matches are made based on visual similarity. While appealing,

such approaches are often not effective due to the differences in how humans and computers evaluate image similarity [4].

Another approach is to assume that other researchers will be able to continue to improve the information-centric process for matching queries to image indexes, and instead focus on the human-centric aspects of image retrieval. Many of the top search engines provide their image search results in a simple scrollable grid of images. This naïve approach does little to take advantage of the powerful visual capabilities of the human mind [20].

An increasingly popular approach is to support similarity-based image browsing [19, 14], wherein images are organized based on visual similarity and the searcher’s task becomes one of browsing and exploring within this image space. Methods such as these take advantage of the searcher’s ability to easily identify the relevance of images with just a glance. They can be applied to the entire document collection, or to a subset extracted as a result of a user-supplied query. Our research follows this stream, focusing on providing granular support at multiple levels of abstraction for the human-centric decision-making activities that are necessary for effective Web image retrieval.

3 Granular Organization of Images

Due to the ambiguity that is common in Web image search queries [1], the relationships among the set of retrieved images can be rather complicated. Some images may be related to one another because they convey similar meaning, whereas others may be visually similar. These different methods for determining the relationships between images provide different frames of reference for the creation of granular worlds that provide abstract views of the image collection.

Considering the semantic and visual similarity of images, there are three different approaches that can be followed for generating a granular framework for the organization of images: (1) using a semantic frame of reference, images may be organized into granular structures based on the meanings or conceptual features contained within the images; (2) using a visual frame of reference, granules may be defined based on the appearance of the images; and (3) combining the semantic and visual frames of reference, granular structures may be generated that simultaneously group images based both on their meanings and their appearance. Given that searchers are often not only interested in subject matter but also appearance and aesthetics when searching for images, our research follows the last of these options.

3.1 Feature Vector Generation

To achieve our goal of using both semantic and visual information to organize the images, we extract two feature vectors for each image in the collection. The semantic feature vector captures the conceptual meaning of the image, while the visual feature vector describes the appearance and visual characteristics of the image.

To compute the semantic feature vectors, we assume that all images in the collection are associated with one or more tags that describe their contents. This assumption holds for photo collections such as Flickr. For untagged image collections, automatic annotation techniques may be employed such as the one proposed by Li and Wang [9]. Another alternative is to derive tags for the images through a query expansion process, wherein the tags for an image are deduced based on its source expanded query [7, 18].

In order to determine the semantic relationship between any pair of images based on these tags, we must have access to some external knowledge base that understands the semantic relationships between the tags. Wikipedia is well-suited to this task given the fact that it includes many hyperlinked articles on people, places, and things, matching the conceptual subjects of many images. Using Wikipedia, we can compute the relatedness $R(i, j)$ between two images i and j using the average relatedness between their tags:

$$R(i, j) = \frac{\sum_{s \in T(i), t \in T(j)} WLM(s, t)}{\|T(i)\| \times \|T(j)\|}$$

Here, $T(\cdot)$ denotes a set of tags associated with an image, $\|\cdot\|$ represents the number of elements in a set, and $WLM(\cdot, \cdot)$ is the Wikipedia Link-based Measure [10] that computes the semantic relationship between two tags based on the hyperlink structure between their associated articles in Wikipedia.

Given a search results set that contains N images, the above process generates an $N \times N$ table, in which each entry stores the semantic relatedness of two corresponding images. To convert this table into a set of semantic feature vectors, classical multi-dimensional scaling (MDS) is applied [3]. MDS assigns each image i a high-dimensional vector $S(i)$ such that for any pair of images i and j , the distance between $S(i)$ and $S(j)$ is similar to the value of $R(i, j)$.

Determining the visual feature vector is a much simpler operation, given the extensive work that has been devoted to this problem within the domain of computer vision. Many different approaches have been studied that can be applied to this work [15]. One simple yet effective method is to use a colour histogram. The 3D colour space is quantized into a set of bins, and for each image i , we count the number of pixels that belong in each bin. These counts are then normalized, resulting in a visual feature vector $V(i)$.

3.2 Similarity-Based Image Organization

To organize the image collection, we use our previously proposed approach [16, 17], which places images on a 2D canvas with the proximity between images indicating the similarities among them. A near-optimal location for each image is obtained by training a Self Organizing Map (SOM) using feature vectors associated with the image set. Once the SOM is trained, an image is placed at the location of its Best Matching Unit (BMU) in the SOM. Since the SOM can effectively map high dimensional space to a 2D canvas in a topology preserving manner, images with similar feature vectors are placed together on the canvas.

As previously noted, our goal in organizing the images is to combine the semantic and visual frames of reference in the production of a granular structure. One option is to use the semantic and visual feature vectors independently within the SOM, determining approximate locations based on semantics first and then more specific locations based on visual similarity. However, such an organization might appear rather disjoint at the transitions between semantic features due to this independence. A more flexible approach is to use these vectors simultaneously within the SOM. That is, a hybrid feature $H(i)$ vector is generated as a weighted concatenation of the semantic and visual feature vectors:

$$H(i) = (\alpha S(i), \beta V(i))$$

When $\alpha \gg \beta$, the approach places semantically close images in the same area of the canvas; at the same time the visual features will also be used to provide a seemingly continuous visual organization. Furthermore, the values of α and β can be controlled by the end user, allowing the searcher to control the degree to which each frame of reference impacts the granularization of the images.

3.3 Hierarchical Structure of Image Space

When the number of retrieved images N is large, it is impractical to display all of these images within the limited resolution of a computer screen at the same time. While traditional search engines address this problem by organizing the image set in a scrollable grid covering multiple pages, this approach is not feasible for our work. Since the images are organized based on semantic and visual features, there is no guarantee that the more relevant images will be placed at the top of the organizational structure. Our solution is to generate a hierarchical granular structure using a multi-resolution SOM. Such a structure provides a high-level overview of the search results set, and allows the searcher to interactively adjust the level of granularity via zoom operations. The details of this interaction will be discussed in the following section.

The process of generating the multi-resolution SOM structure is as follows. After the image set is organized using the SOM, the multi-resolution SOM structure is constructed in a bottom-up approach, with the resolution of the space reduced by half along both dimensions with each step. Hence, a granule at level n corresponds to four granules in level $n - 1$. The image associated with a high-level granule is selected from those associated with the granules it subsumes, where the selection criteria is based on how close the feature vector of an image is to the average feature vector of the group. Hence at a given level of granularity, users only see images that best represent the collection of images contained within the granule, providing searchers with multi-level overviews of the image organization, depending on the level of the SOM that is being shown.

This approach for organizing the search result images in a hierarchy that preserves their conceptual and visual similarity follows the principle of multi-level granularity. Rather than organizing the images once and providing these in a single image space that the searchers must scroll through, our approach generates

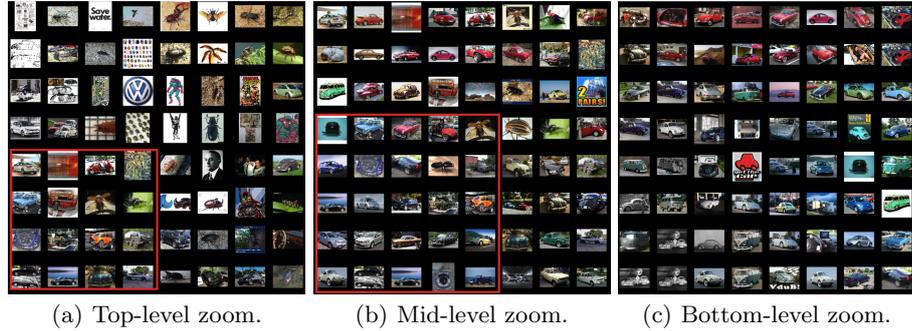


Fig. 1. Zooming from a high level of granularity (a) into lower levels of granularity (b and c) results in a more specific focus on the semantic and visual features that the searcher is seeking. The red box illustrates the region of zoom from the top-level zoom to the mid-level zoom, and from the mid-level zoom to the bottom-level zoom.

a hierarchical granular structure based on both the semantic and visual features of the images. By displaying the images associated with the multi-resolution SOM at a high level of granularity, the searchers are provided with a general overview of the image set. Reducing the level of granularity results in a reduction in the abstraction of the image space, providing images that are more descriptive of the region of the image space that they occupy (see Figure 1).

4 Granular Interaction Mechanisms

User interaction is a fundamental element of any software system that is designed to support and enhance user-guided tasks or activities [12]. It is also extremely important for systems that seek to take advantage of human visual processing capabilities [20], since such systems are not only designed to visually represent information, but also to allow users to directly interact with this information.

The primary goal for the interaction mechanisms within our system is to allow the searcher to explore within the image space with pan and zoom operations. As the searcher identifies a region of interest within the high-level overview of the image space, they can zoom into this area (using the normal mouse-wheel or two-finger trackpad drag interaction). Since the image sizes remain constant, this zoom operation dynamically creates more space between the images. Once sufficient space is created, the images contained within the next lower level of the multi-resolution SOM are shown. This process continues as the searcher zooms deeper and deeper into the image space, and also works in the reverse direction for zooming out. At the bottom level of the hierarchy, further zooming results in an increase in the image sizes.

This method of interaction follows Shneiderman’s [11] popular Visual Information Seeking Mantra for supporting interaction within visual software: “overview first, zoom and filter, then details-on-demand”. The zoom operation

is an example of Yi et al.'s [24] interaction category of “abstract/elaborate”, where the fundamental goal is to show the user more or less detail depending on the direction of the zoom.

Within the context of granular computing, zooming conforms to the principle of granularity conversion. As the searcher zooms into a region of interest, the granular structure is traversed to a lower-level of granularity, increasing its specificity. As a result, the images that are shown become more specific to the regions of the image space they occupy. Zooming out reverses this process, showing the image collection at a higher-level of granularity and increasing the generality of the information that is shown (i.e., the images are more general representations of the regions they cover). Conceptually, this zoom process represents a movement up or down within the information pyramid [2] produced by the hierarchical granular structure within our system.

5 Support for Intelligent Decision-Making

One of the fundamental features of a Web information retrieval support system is to enhance and promote the searcher's decision-making processes as they evaluate and explore the search results [5]. This holds true regardless of whether the searcher is seeking documents, images, or other types of information.

Our system supports the searcher in making intelligent decisions regarding the images they are seeking at different levels of granularity. When viewing the images at a high level in the multi-resolution SOM (see Figure 1(a)), the searchers are able to make course-grained decisions with respect to which region they should explore in more detail (i.e., zoom into). As the searcher delves deeper into the multi-resolution SOM (see Figure 1(b)), the granularity decreases and the granules become more specific, allowing searchers to make fine-grained decisions regarding the regions in which they feel relevant images may be present. When the searcher arrives at the bottom level in the multi-resolution SOM (see Figure 1(c)), the images are no longer aggregated into granules.

Within each step of traversing the multi-resolution organization of the image space, the number of images that are shown is relatively small (in comparison to the total number of images that are retrieved). This reduction in information supports a more intelligent approach to decision making: rather than considering each image individually for relevance, the searcher considers the representative images at various levels of granularity while traversing the granular structure. As a result, a smaller number of incremental decisions regarding the relevance of images within a given region of the image space are made as the searcher zoom (and perhaps pans) within the image space. At the lowest level of granularity, a small collection of semantically and visually similar images are finally shown, and the searcher can then focus on making a specific relevance decision for each image.

These decision-making processes supported by our system conform to the principle of focused effort. That is, decisions are made at the given level of granularity in order to solve the problem of finding relevant images. At a high

level of granularity, the searcher is able to arrive at an approximate solution regarding the region of space that contains the images they are seeking. As they zoom into the image space and are shown the images at a lower level of granularity, they are able to make more precise decisions. When they finally reach the lowest level of granularity, the image space is effectively filtered to show a very focused set of images, upon which low-level relevance decisions can then be made.

6 Conclusion

In this paper, we have outlined how our approach to interactive image organization and exploration follows a model of granular computing, supporting the methodological perspective of structured problem solving that the human mind embraces. The images are organized at multiple levels of abstraction and the searcher is supported in traversing these levels using a zoom operation. At a given level of abstraction, the searcher is able to make decisions of increasing specificity as regions of the image space that contain relevant images are sought. This is in contrast to the common approach in image search of providing a scrollable grid of images, which requires that the searcher make relevance decisions of images individually.

Our future work in this domain includes devising and evaluating additional frames of reference that generate alternative granular worlds for resolving the tasks associated with image retrieval. We also wish to evaluate the benefits of simultaneously showing different granular structures to the searcher, as well as make further enhancements to the interface with the goal of supporting the human-centred aspects of image retrieval on the Web [6].

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. Bargiela, A., Pedrycz, W.: Granular Computing: An Introduction. Springer (2002)
3. Cox, T.F., Cox, M.A.A.: Multidimensional Scaling. Chapman and Hall/CRC, 2nd edn. (2000)
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. 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.: Human-centred Web search. In: Jouis, C., Biskri, I., Ganascia, J.G., Roux, M. (eds.) *Next Generation Search Engines: Advanced Models for Information Retrieval*, pp. 217–238. IGI Global (2012)
7. Hoque, E., Hoeber, O., Strong, G., Gong, M.: Combining conceptual query expansion and visual search results exploration for Web image retrieval. *Journal of Ambient Intelligence and Humanized Computing* (in press)

8. 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)
9. Li, J., Wang, J.Z.: Real-time computerized annotation of pictures. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(6), 985–1002 (2008)
10. 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)
11. Shneiderman, B.: The eyes have it: a task by data type taxonomy for information visualizations. In: *Proceedings of IEEE Symposium on Visual Languages*. pp. 336–343 (1996)
12. Shneiderman, B., Plaisant, C.: *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Addison Wesley, 5th edn. (2009)
13. Smeulders, A.W.M., 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)
14. Snaveley, 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)
15. Strong, G., Gong, M.: Organizing and browsing photos using different feature vectors and their evaluations. In: *Proceedings of the ACM International Conference on Image and Video Retrieval*. pp. 3:1–3:8 (2009)
16. Strong, G., Gong, M.: Similarity-based image organization and browsing using multi-resolution self-organizing map. *Image and Vision Computing* 29(11), 774–786 (2011)
17. Strong, G., Hoerber, 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)
18. Strong, G., Hoque, E., Gong, M., Hoerber, O.: Organizing and browsing image search results based on conceptual and visual similarities. In: *Proceedings of the International Symposium on Visual Computing*. pp. 481–490 (2010)
19. 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)
20. Ware, C.: *Information Visualization: Perception for Design*. Morgan Kaufmann, 2nd edn. (2004)
21. Yao, Y.: The art of granular computing. In: *Proceedings of the International Conference on Rough Sets and Intelligent Systems Paradigms*. pp. 101–112 (2007)
22. Yao, Y.: Granular computing: past, present, and future. In: *Proceedings of the IEEE International Conference on Granular Computing*. pp. 80–85 (2008)
23. Yao, Y.: Human-inspired granular computing. In: Yao, J. (ed.) *Novel Developments in Granular Computing: Applications for Advanced Human Reasoning and Soft Computation*, pp. 1–15. IGI Global (2010)
24. Yi, J.S., Kang, Y., Stasko, J.T., Jacko, J.A.: Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics* 13(6), 1224–1231 (2007)