Visual Structures for Image Browsing

Ricardo S. Torres, Celmar G. Silva, Claudia B. Medeiros and Heloisa V. Rocha
Institute of Computing, University of Campinas
Av. Albert Einstein, 1251, CEP 13084-851
Campinas, SP, Brazil
{rtorres,celmar,cmbm,heloisa}@ic.unicamp.br

ABSTRACT

Content-Based Image Retrieval (CBIR) presents several challenges and has been subject to extensive research from many domains, such as image processing or database systems. Database researchers are concerned with indexing and querying, whereas image processing experts worry about extracting appropriate image descriptors. Comparatively little work has been done on designing user interfaces for CBIR systems. This, in turn, has a profound effect on these systems since the concept of image similarity is strongly influenced by user perception. This paper describes an initial effort to fill this gap, combining recent research in CBIR and Information Visualization, studied from a Human-Computer Interface perspective. It presents two visualization techniques based on Spiral and Concentric Rings implemented in a CBIR system to explore query results. The approach is centered on keeping user focus on both the query image, and the most similar retrieved images. Experiments conducted so far suggest that the proposed visualization strategies improves system usability.

Categories and Subject Descriptors

H.5.2 [**INFORMATION INTERFACES AND PRESENTATION**]: User Interfaces—*Interaction styles*; H.2.8 [**DATABASE MANAGEMENT**]: Database Applications—*Image databases*; H.3.3 [**INFORMATION STORAGE AND RETRIEVAL**]: Information Search and Retrieval—*Query formulation, Relevance feedback, Search process*

General Terms

Design, Experimentation, Human Factors, Management

Keywords

Information Visualization, Visual Structures, focus+context views, Content-based Image Retrieval (CBIR), Image Browsing

CIKM'03, November 3–8, 2003, New Orleans, Louisiana, USA.

1. INTRODUCTION

Advances in data storage and image acquisition technologies have enabled the creation of large image datasets. In order to deal with these data, it is necessary to develop appropriate information systems to efficiently manage these collections. The most common retrieval approach is to attach textual metadata to each image and use traditional database query techniques to retrieve by keyword. An alternative are the so-called *CBIR systems*. Basically, these systems try to retrieve images similar to a user-defined specification or pattern (e.g. shape sketch, image example). Their goal is to support image retrieval based on *content* properties, e.g. shape, color or texture [16]. Research in CBIR systems is multidisciplinary and ranges from finding appropriate indexing and storage schemes for images, to cognitive problems in query specification. From the user's perspective, CBIR systems offer more flexibility in specifying queries than those based on metadata. On the other hand, they present new challenges. The first is how to *interpret* a query – e.g, when a user provides an image as input, what are the similarity criteria to be used. Another problem is *information overload* – how to present the result to the user in a meaningful way. A third issue is that of providing users with tools to *interact* with the system in order to refine their query.

Typically, the result of a query is a set of images, displayed in an Image Browser. Unfortunately, these sets are usually large, so a browsing activity must be performed. The most common result presentation technique is based on showing a two-dimensional grid of thumbnail (miniature) image versions [8, 12]. The grid is organized according to the similarity of each returned image with the query pattern (e.g. from left to right, from top to bottom). It is a $n \times m$ matrix, where position $(1, 1)$ is occupied by a thumbnail of the query pattern, position $(1, 2)$ by the one most similar to it, and so on. This helps browsing, allowing users to simply scan the grid image set as if they were reading a text [13]. This approach, however, displays retrieved images of different similarity degree at the same physical distance from the image query: e.g., images $(1, 2)$ and $(2, 1)$ are displayed at the same physical distance from the query pattern, but the former is more similar to it than the latter. Other display approaches try to consider relative similarity not only between the query pattern and each retrieved image, but also among all retrieved images themselves [14, 17]. These initiatives have the drawback that visually similar images which are placed next to each other can sometimes appear to merge or overlap, making them less eye-catching than if they were separated [13].

This paper presents a new approach to these user interaction problems. This approach is based on adopting recent findings in Information Visualization techniques to provide users with semantically meaningful result presentations, and new kinds of interaction mechanisms. Information Visualization is an important field within the domain of *Human-Computer Interface (HCI)* that aims at studying the use of computer-supported, interactive, and visual representations of abstract data to amplify cognition [3, 5, 15].

The main contributions of this paper are the following:

- presentation of two visualization techniques based on *spiral* and *concentric rings* for exploring query results in an image database. These techniques provide users new means of ranking similar images while at the same time avoid image overlap;

- description of a CBIR prototype developed which incorporates these visualization paradigms. The visualization and interaction properties of the prototype are based on the reference model described in [3].

The rest of this paper is organized as follows. Section 2 characterizes the content-based image retrieval process. Section 3 describes the proposed visualization techniques. Section 4 presents implementation details. Section 5 discusses related work. Section 6 presents conclusions and ongoing work.

2. CONTENT-BASED IMAGE RETRIEVAL SYSTEMS

CBIR is centered on the notion of image *similarity* – given an image database with a large number of images, a user wants to retrieve the set of images which are most "similar" to a query pattern (usually an image). Similarity computation relies on the notion of image *descriptors*. Descriptors are defined as feature vectors whose fields contain values that encode characteristics of an image – e.g. color or texture properties. Similarity between two images is computed by measuring the distance between their feature vectors, using specific distance functions. Usually, the degree of similarity of an image is defined as an inverse function of the distance metric, that is, the larger the distance value, the less similar the image is.

Usually, two kinds of queries are supported by CBIR systems [6]. In a *K-nearest neighbor query (KNNQ)*, the user specifies the number k of images to be retrieved closest to the query pattern. In a *range query (RQ)*, the user defines a search radius r and wants to retrieve all database images whose distance to the query pattern is less than r.

Figure 1 shows an overview of a image database retrieval system. The interface allows a user to specify a query by means of a query pattern and to visualize the retrieved similar images. The query-processing module extracts a feature vector from a query pattern and applies a metric distance (such as the euclidean distance) to evaluate the similarity between the query image and the database images. Next, it ranks the database images according to their similarity and forwards the most similar ones to the interface module. Database images are often indexed according to their feature vectors using structures such as the M-tree [6] to speed up retrieval and distance computation.

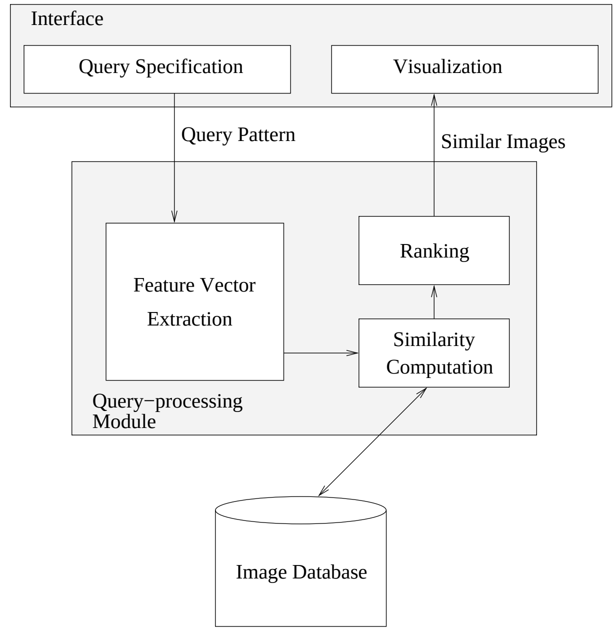


Figure 1: Typical image database retrieval system.

This paper focuses on the interface layer. It uses Information Visualization techniques to enhance similarity comprehension and user interaction in a CBIR system.

3. VISUAL STRUCTURES PROPOSED

Information Visualization is a very important area in HCI. It can amplify cognition in many ways such as: increasing the memory and processing resources available to the users; reducing the search for information, e. g. due to compacting, grouping or visually relating information; enhancing the detection of patterns; enabling perceptual inference operations; using perceptual attention mechanisms for the monitoring of a large number of potential events; and encoding information in a manipulable medium [3].

One of the traditional approaches to present retrieved images in a CBIR system is based on a tabular (grid) disposition. As mentioned in the introduction, this placement affects similarity comprehension, since it displays images with different similarity degrees at the same physical distance to the query pattern.

A solution to overcome this ambiguity is to borrow techniques from the Information Visualization domain. The method proposed here is based on: (1) placing the query pattern at the center of the display, and (2) surrounding it with similar images, with physical distances and sizes proportional to their respective similarity degrees. The less similar an image is, the smaller and the farther apart from the center.

This kind of presentation relies on the fact that the user focus resides on the query pattern and the most similar images. This so-called *focus + context* approach is used to both center the user attention on the result and give the user a contextual notion of the less similar images. Besides, this approach avoids image overlapping, a common problem of some CBIR systems. Two visual structures based on this method place the images along a spiral or concentric rings.

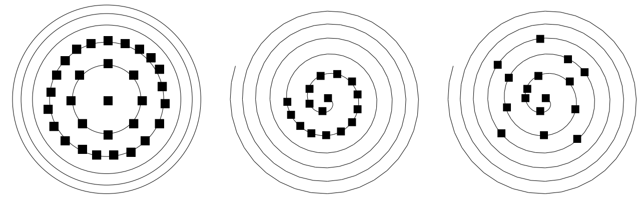
3.1 Concentric Rings Presentation

A ring can be defined as a circle. In polar coordinates a circle can be expressed as $r = k$, where r is the radial distance, and k is a constant. Successive rings are built by changing the k value. Moreover, all rings have the same center and successive rings become closer as k increases. The rings are filled from the innermost ring to the outermost one, according to image ranking. Figure 2a illustrates the concentric ring visual structure implemented, where dots represent image thumbprints.

3.2 Spiral Presentation

The most common planar spirals are the spiral of Archimedes, the hyperbolic spiral and the logarithmic spiral. In order to contemplate the characteristics proposed into our method, a spiral line should become closer to itself as it loops away. This aspect is directly related to the images' size variation along the structure. Since hyperbolic and logarithmic spirals move rapidly away from the origin, they are not appropriate to our goal. Thus, the choice was Archimedes's spiral, expressed in a polar equation as $r = k\theta^a$, where r is the radial distance, θ is the polar angle, k is a constant and a is a constant which determines how tightly the spiral "wraps" around itself. Figure 2b illustrates the spiral adopted, considering $k = 2.5$ and $a = -1.5$. Observe that the greater θ is, the tighter the spiral line becomes, enforcing the focus on the central region.

There are two different ways to display images along a spiral line. The first associates the image ranking to the spiral line, in such a way that the images are disposed successively, at regular distances (Figure 2b). This approach, however, does not present the real similarity degree. A second alternative maps the similarity degree to the spiral line, that is, the image distance to the query pattern is proportional to its similarity degree (Figure 2c).



(a) (b) (c)

Figure 2: Visual Structures. (a) Concentric rings. (b) Spiral mapping image ranking. (c) Spiral encoding image similarity degree. Note that image size reduction along the structures is not shown.

4. IMPLEMENTATION

This section describes the prototype implemented. It is written in Builder C++, running on Windows. It was tested on a database of 11000 fish images and uses two shape descriptors called *Multiscale Fractal Dimension* and *Shape Saliences* [20, 21, 22]. This is part of a biodiversity information system, where users (biologists) explore a database containing images and textual descriptions to find out details about species. Details of this project are outside the scope of the paper [19].

4.1 Formalizing the Visualization Framework

Research in Information Visualization often uses the reference model of [3] as a basis to study the cognitive enhancement provided by visual representations. This model defines a way to analyze successive transformations from raw data to visual representations, taking into account possible human interactions within this process through three transformation stages: *data transformation (DT)*, *visual mapping (VM)* and *view transformation (VT)*. The first stage considers that raw data (data in some idiosyncratic format) are initially transformed into data tables (DT). Tables are next mapped to structures with graphical properties – visual structures, displayed on a screen (VM). Finally, these static structures are transformed into views, which are dynamic, interactive and information-enriched representations (VT).

Many techniques have been proposed to deal with each transformation step and the underlying data structures [3]. *Details-on-demand* is a method used within the data transformation stage to expand a small set of objects revealing more information about it [1, 3]. *Pan* and *zoom* are commonly used together within view transformations to change the viewer's position and to focus on a specific set of data. The *focus + context* approach is also used within a view transformation stage. It simultaneously combines overview (context) and detail information (focus), using distortion or other specific techniques.

Figure 3 analyzes the architecture of our CBIR prototype under this reference model. The image database represents the raw data. Image processing algorithms automatically extract feature vectors that encode the image content. This extraction phase is the first data transformation, and generates a data table comprised of each image and its feature vector (T1). When a user inputs a query image (QI), a second data transformation occurs: QI's feature vector (FQI) is automatically extracted, and the system executes a matching algorithm to compute the distance from the FQI to feature vectors stored in T1, thus generating a second data table (T2). This table stores the distances from FQI to the feature vector of each database image. A third data transformation occurs when the user specifies a limit to the number of images to retrieve, leading to a third data table (T3) that is a subset of T2. Next, the user chooses the visual structure to be used: traditional (2D grid), spiral or concentric rings. All three visual structures take into account the distance values stored into T3. Finally, the user can interactively manipulate the display using details-on-demand, zoom, pan and focus + context, generating new views of the chosen visual structure and improving user cognition.

4.2 A Sample User Session

Consider the following sample session. Initially, a user specifies a query by providing a query image (the query pattern). Next, the user chooses the descriptor for similarity computation and the visual structure for displaying the query result – 2D grid, rings or spiral.

The 2D grid-based traditional approach just obeys the rank. When it reaches the horizontal end of the screen, it continues the sequence of images on the next line, a typical use of the so-called folding technique [3]. Figure 4a shows a screen copy of this standard visualization approach. The query image is on the left topmost part of the screen. The result of the query, organized in a 2D grid, is on the large part on the right. Since we use shape descriptors, the closest

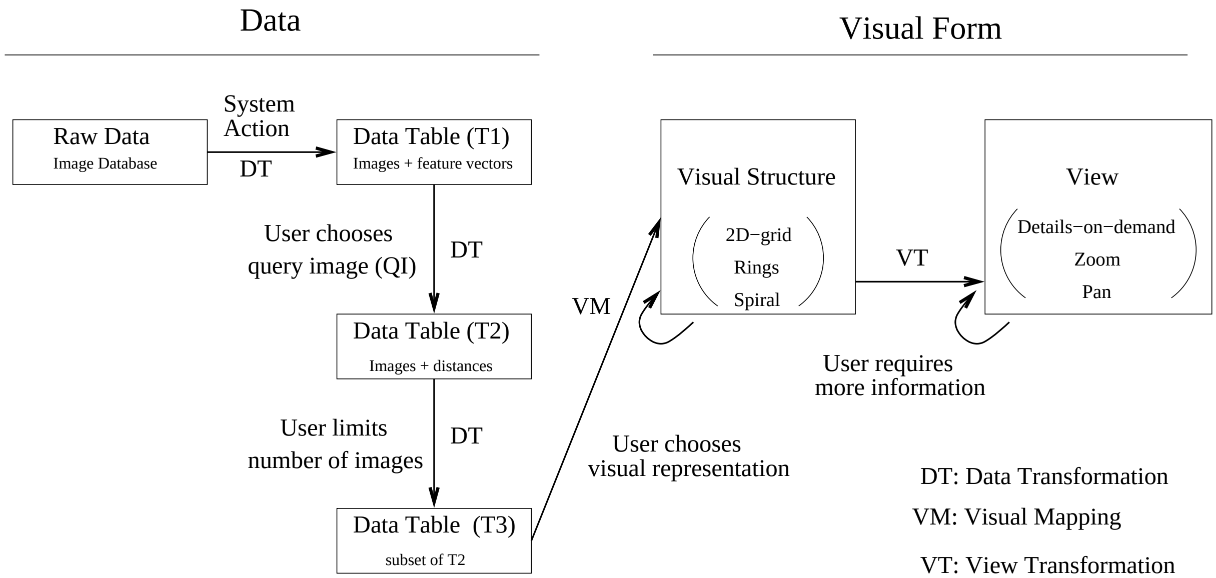
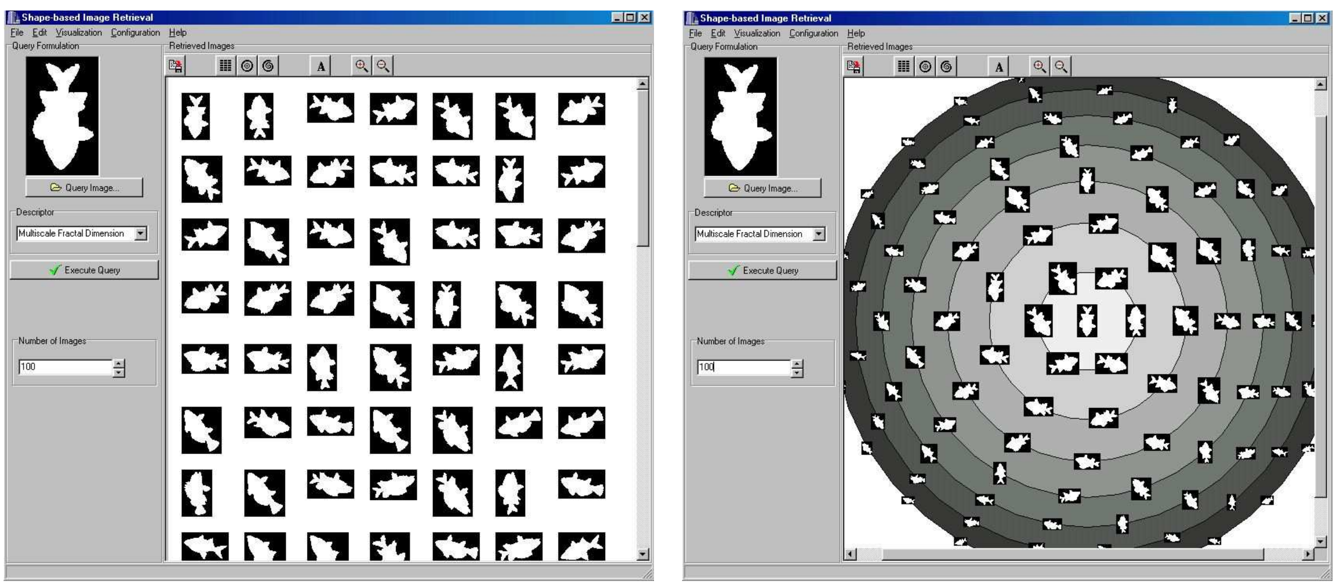


Figure 3: Information Visualization phases on a CBIR system.



(a) (b)

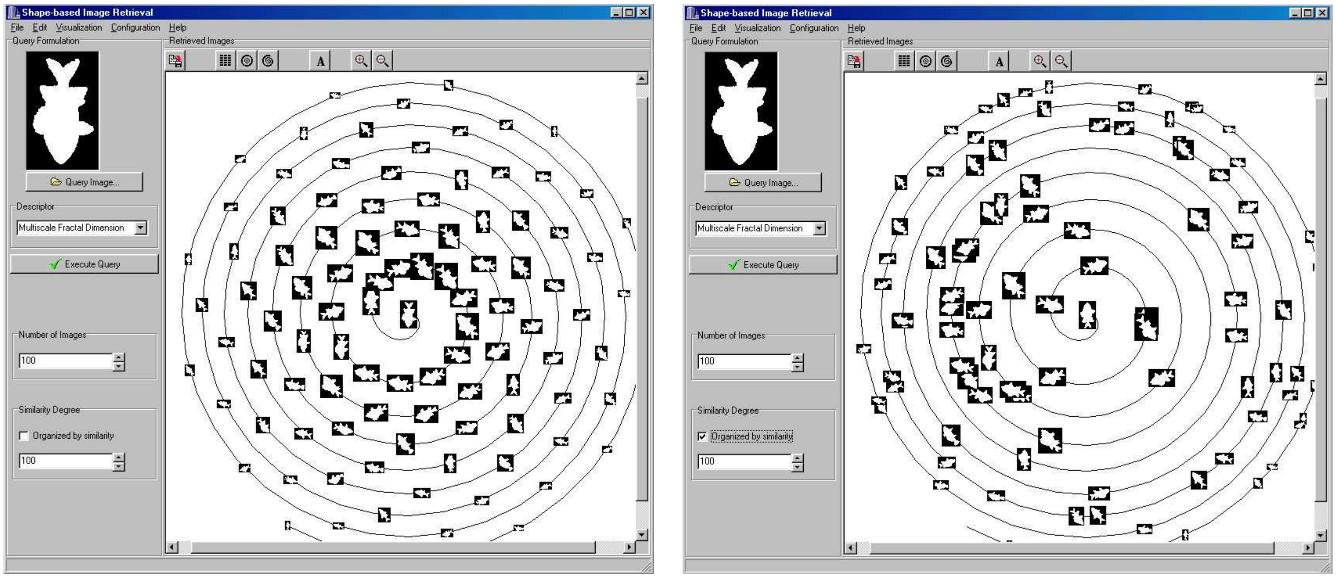
Figure 4: Prototype screen shots. (a) 2D grid approach. (b) Concentric rings approach.

results are fish images whose shapes are most similar to the query pattern's shape. Thus, image rotation or scaling are not taken into consideration in similarity computation. This is a nice property of the shape descriptors used – see [20, 21, 22] for more details.

Figure 4b shows the result of the same query using the *concentric rings* visualization approach. This screen shot enhances the ring structure with increasing levels of gray to help focus user attention on the center and provide a better separation among rings (another technique in visualization theory). The query image is at the center of the rings. Images at rings farther from the center are less similar than those along closer rings.

In a similar fashion, the *spiral* approach also places the query image in the center, and fills the spiral with the retrieved images. Figure 5 presents the two available spiral variants. Figure 5a shows a spiral in which images are placed successively, at regular distances. Figure 5b, in turn, places the ranked images within the spiral considering their similarity degree. This latter approach, however, can overlap images with similar distance to the query pattern.

Users can interact with the result in many ways. Besides zoom and pan operations, they can select a specific image as a new query, or obtain a detail-on-demand box with a real-sized image and its filename. The user can also control the number of images displayed (simulating a KNN query) for all three visual representations. In the case of spiral representation, the user can threshold the retrieved images



(a) (b)

Figure 5: Prototype screen shots. Spiral placement based on (a) image ranking and (b) similarity.

by their degree of similarity. This corresponds to a range query, where the search radius is controlled by the user.

The user can specify a new query either by selecting an image from the retrieved image set or by providing a new image file name. Besides, the user can provide new query parameters (e.g. kind of descriptor or number of retrieved images) via the textual controls at the left part of the screen.

4.3 Relevance Feedback

Relevance feedback is a commonly accepted method to improve interactive retrieval effectiveness [11]. Basically, it is composed of three steps: (a) an initial search is made by the system with a user-supplied query pattern, returning a small number of images; (b) the user then indicates which of the retrieved images are useful (relevant); (c) finally, the system automatically reformulates the original query based upon user relevance judgments. This process can continue to iterate until the user is satisfied.

The proposed visual structures can also be used to improve user interaction in the relevance feedback process. Two kinds of interaction based on *direct manipulation* are foreseen. First, a user can move images along the spiral line. By taking an image away from the spiral center, the user informs the system that this image is not relevant. The opposite situation is also true: moving an image closer to the spiral center increases its relevance for future queries. A similar interaction can occur on the concentric ring visual structure. In this case, users can move an image across rings – relevance increases as the image is moved to a position closer to the center.

4.4 Experimentation

Experiments conducted so far did not gather enough data to prove the superiority of spiral or concentric rings over 2D grids. So far, our experiments have been conducted with a limited number of users, that are not experts on research on fish. This has been a limitation factor on interface evaluation, since we would have to consider many kinds of user profile. Nevertheless our experiments allow the following preliminary conclusions:

- alternative (multiple) views of a result are much more useful than just the usual 2D grid, offering users distinctive perceptions of relative distances and similarities;

- when the query for k nearest neighbor images involves large values of k, the result clutters the screen. In this case, spiral and ring presentations help zooming into the desired result. For small values of k (typically when results can be seen in one horizontal line) users see no advantage in using alternatives to 2D grid.

- users were not aware that extended visualization presentations were possible. Faced with alternatives, they began demanding further extensions. The prototype presents the results ranking all images w.r.t. the query pattern. An extension would be to allow clustering images according to the relative distance to each other. Another request is for 3D presentation, though recent results seem to indicate that for this kind of query 2D presentations are better cognition-wise [7].

5. RELATED WORK

Information Visualization is attracting considerable attention in several domains, such as data mining, data exploration and knowledge discovery (e.g., [10, 23]). In particular, classification in data mining is often visualized in terms of data clusters, where each class instance is presented as a point in a 2D or 3D space. Each cluster represents a data class, and the distances among clusters allow users to deduce the relative similarity among classes and their instances.

2D grid presentation can be found in several image database systems [8, 12, 16]. [2] and [7] try to improve this visual structure by studying zoom properties to enhance image browsing. Rodden *et al.* [13], in turn, investigates whether it benefits users to have sets of thumbnails arranged according to their similarity, so images that are alike are placed together. They describe experiments to examine whether similarity-based arrangements of the candidate images help in picture selection.

Stan *et al.* [17] describe an exploration system for an image database, which deals with a tool for visualization of the database at different levels of details based on a multidimensional scaling technique. This visualization technique groups together perceptual similar images in a hierarchy of image clusters. Retrieved images can overlap. The overlap problem is also found in El Niño image database [14]. In this context, Tian *et al.* [18] propose a PCA (Principal Component Analysis)-based image browser which looks into an optimization strategy to adjust the position and size of images in order to minimize overlap (maximize visibility) while maintaining fidelity to the original positions which are indicative of mutual similarities.

Spirals and rings are used to visualize information in different domains [4, 9, 24, 25]. [4] and [24] investigate the use of spirals to visualize time-series. They display data along a spiral to highlight serial attributes along the spiral axis and periodic ones along the radii. Mackinlay *et. al.* [9], in turn, use a spiral for calendar visualization, building iconic representations of past daily calendar entries, positioned on a spiral. A radial layout is used in [25] to visualize graphs. In this approach, graph nodes are arranged on concentric rings around the focus node. Each node lies on the ring corresponding to its shortest network distance from the focus.

6. CONCLUSION

This paper presented a new approach to improve user interaction in CBIR systems based on applying Information Visualization research to construct CBIR interfaces. It discusses two visualization techniques based on Spiral and Concentric Rings to explore query results. These visual structures are centered on keeping user focus on the query image and on the most similar retrieved images. These strategies improve traditional 2D grid presentation and avoid image overlaps, commonly found in CBIR systems.

Ongoing work includes the finalization of user experiments, and the definition of a new visualization strategy. This strategy extends the proposed methods by considering the mutual similarities among retrieved images. At the same time, relevance feedback principles are being incorporated to the prototype.

7. ACKNOWLEDGMENTS

This work was supported by FAPESP (Proc. 01/02788-7), the PRONEX/SAI Project, CNPq and CAPES.

8. REFERENCES

[1] C. Ahlberg and B. Snheiderman. Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays. In *ACM Conference on Human Factors in Computing Systems*, pages 313–317, 479–480, 1994.

[2] B. B. Bederson. Photomesa: A Zoomable Image Browser Using Quantum Treemaps and Bubblemaps. In *ACM Symposium on User Interface Software and Technology*, pages 71–80, 2001.

[3] S. K. Card, J. D. Mackinlay, and B. Shneiderman. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann Publishers, 1999.

[4] J. V. Carlis and J. A. Konstan. Interactive Visualization of Serial Periodic Data. In *11th Annual Symposium on User Interface Software and Technology*, Nov. 1998.

[5] E. H. Chi. A Taxonomy of Visualization Techniques Using the Data State Reference Model. In *IEEE Symposium on Information Visualization*, pages 69–76, 2000.

[6] P. Ciaccia, M. Patella, and P. Zezula. M-tree: An Efficient Access Method for Similarity Search in Metric Spaces. *VLBD Journal*, pages 426–435, 1997.

[7] T. T. A. Combs and B. B. Bederson. Does Zooming Improve Image Browsing? In *Proceedings of the Fourth ACM International Conference on Digital Libraries*, pages 130–137, 1999.

[8] M. Flickner, H. Sawhney, W. Niblack, Q. H. J. Ashley, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by Image and Video Content: the QBIC System. *IEEE Computer*, 28(9):23–32, Sep 1995.

[9] G. G. R. J. D. Mackinlay and R. DeLine. Developing Calendar Visualizers for the Information Visualizer. In *ACM Symposium on User Interface Software and Technology*, pages 109–118, 1994.

[10] D. A. Keim. Information Visualization and Visual Data Mining. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):1–8, Jan/March 2002.

[11] S. D. MacArthur, C. E. Brodley, A. C. Kak, and L. S. Broderick. Interactive Content-Based Image Retrieval Using Relevance Feedback. *Computer Vision and Image Understanding*, 88(2):55–75, 2002.

[12] V. E. Ogle and M. Stonebraker. Chabot: Retrieval from Relational Database of Images. *IEEE Computer*, 28(9):40–48, Sep 1995.

[13] K. Rodden, W. Basalaj, D. Sinclair, and K. Wood. Does Organization by Similarity Assist Image Browsing? In *ACM Conference on Human Factors in Computing Systems*, volume 3, pages 190–197, 2001.

[14] S. Santini, A. Gupta, and R. Jain. Emergent Semantics through Interaction in Image Databases. *IEEE Transactions on Knowledge and Data Engineering*, 13(3):337–351, May/June 2001.

[15] B. Shneiderman. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *IEEE Symposium on Visual Languages*, pages 336–343, Sep 1996.

[16] 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):1349–1380, December 2000.

[17] D. Stan and I. K. Sethi. eID: a System for Exploration of Image Databases. *Information Processing nad Management*, 39:335–365, 2003.

[18] Q. Tian, B. Moghaddam, and T. S. Huang. Display Optimization for Image Browsing. In *2nd International Workshop on Multimedia Databases and Image Communications*, Sep 2001.

[19] R. S. Torres. *Integrating Image and Spatial Data for Biodiversity Information Management*. PhD thesis, Institute of Computing, University of Campinas, 2004. Under development.

[20] R. S. Torres, A. X. Falcão, and L. F. Costa. Shape Description by Image Foresting Transform. In *14th International Conference on Digital Signal Processing*, Santorini, Greece, July 2002.

[21] R. S. Torres, A. X. Falcão, and L. F. Costa. A Graph-based Approach for Multiscale Shape Analysis. Technical Report IC-0303, Institute of Computing, University of Campinas, January 2003.

[22] R. S. Torres, E. M. Picado, A. X. Falcão, and L. F. Costa. Effective Image Retrieval by Shape Saliences. In *Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI03)*, São Carlos, São Paulo, Brazil, 12-15 October 2003.

[23] S. Wang and H. Wang. Knowledge Discovery Through Self-Organizing Maps: Data Visualization and Query Processing. *Knowledge and Information Systems*, 4(1):31–45, 2002.

[24] M. Weber, M. Alexa, and W. Muller. Visualizing Time-Series on Spirals. In *IEEE InfoVis Symposium*, pages 7–14, 2001.

[25] K.-P. Yee, D. Fisher, R. Dhamija, and M. A. Hearst. Animated Exploration of Dynamic Graphs with Radial Layout. In *IEEE Symposium on Information Visualization*, pages 43–50, 2001.