Semantic Image Browser: Bridging Information Visualization with Automated Intelligent Image Analysis

Jing Yang, Jianping Fan, Daniel Hubball, Yuli Gao, Hangzai Luo and William Ribarsky
Dept of Computer Science
University of North Carolina
at Charlotte
jyang13,jfan,dhubball,ygao,hluo,ribarsky@uncc.edu

Matthew Ward
Dept of Computer Science
Worcester Polytechnic
Institute
matt@cs.wpi.edu

ABSTRACT

Browsing and retrieving images from large image collections are becoming common and important activities. Recent semantic image analysis techniques, which automatically detect high level semantic contents of images for annotation, are promising solutions toward this problem. However, few efforts have been made to convey the annotation results to users in an intuitive manner to enable effective image browsing and retrieval. There is also a lack of methods to monitor and evaluate the automatic image analysis algorithms due to the high dimensional nature of image data, features, and contents.

In this paper, we propose a novel, scalable semantic image browser by applying existing information visualization techniques to semantic image analysis. This browser not only allows users to effectively browse and search in large image databases according to the semantic content of images, but also allows analysts to evaluate their annotation process through interactive visual exploration. The major visualization components of this browser are Multi-Dimensional Scaling (MDS) based image layout, the Value and Relation (VaR) display that allows effective high dimensional visualization without dimension reduction, and a rich set of interaction tools such as search by sample images and content relationship detection. Our preliminary user study showed that the browser was easy to use and understand, and effective in supporting image browsing and retrieval tasks.

Keywords: Image retrieval, image layout, semantic image classification, multi-dimensional visualization, visual analytics.

Index Terms: I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Object recognition; H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—Search process; H.5.2 [Information Interfaces and Presentation]: User Interfaces—Graphical user interfaces

1 INTRODUCTION

Interactive image exploration is important in many grand challenge problems, such as homeland security, satellite image analysis, and weather forecasting. It is also useful in daily life applications such as personal photo management. User studies [3, 13] showed that nowadays the range of exploration of image databases is much wider than just retrieving images based on the presence or absence of objects of simple visual characteristics. For example, the following tasks are common [17]: (a) **target search**. The search for a precise copy of the image in mind, or for another image of the same objects found in an image of interest; (b) **search by association**. The search is an iterative refinement process which at the start has no specific aim other than finding interesting things related to example images. (c) **category search**. The search aims at retrieving arbitrary images representative of a specific class.

The semantic contents of images are more useful for supporting users performing effective interactive image exploration than low level visual features of images. However, in large image collections, such as images available on the Internet, the semantics of images are described only partially or not at all. In order to overcome this problem, semantic image classification techniques, which enable automatic annotation of a large number of images according to their semantic contents [6], have been widely studied in recent years.

However, there is a gap between effective semantic image classification and effective interactive image exploration. There are few efforts at making use of the resulting annotations from automatic semantic image classification processes to support users to effectively browse and retrieve images from large image databases. In addition, since the automatic image classification processes involve high dimensional datasets that are difficult to visually present without special methods from information visualization, they are more or less black-boxes and hard to monitor and evaluate.

To allow effective image database exploration making use of semantic image classification, as well as evaluation and monitoring of automatic image annotation processes, building a bridge between information visualization and semantic image classification seems an urgent task. It can be achieved by providing strong, integrated visual analytic support to automatic image classification systems. We argue that the resulting system will not only facilitate the abilities of users to explore and search large image databases, but also allow expert image analysts to evaluate and monitor automatic image classification processes using their own judgment, innate pattern understanding abilities, and knowledge discovery. In this paper, we call such a system a Semantic Image Browser (SIB). It has the following features:

- It must contain at least one semantic image classification process that automatically annotates large image collections, or it must accept annotating results of large image collections from such processes.

- It must contain one or more coordinated visualization techniques that allow users to interactively explore the image collections and annotations for browsing, searching, or evaluating purposes. It may even allow users to modify the annotations for improving their quality.

- It may contain visualization techniques that allow analysts to evaluate and monitor the automatic annotation process itself through intuitive visual displays and interactions, and allow them to adjust classification algorithms through interactions to improve efficiency and effectiveness.

To support our claims, we have developed a novel SIB containing the above features. It uses a recent semantic image classification process named the concept-sensitive image content representation framework [6] as the annotation engine. It automatically detects semantic image contents (i.e., object classes) and concepts

191

Figure 1: (a) An MDS image overview of the Corel collection (1100 images). (b) The rainfall image view of the collection that shows the correlations between the bottom center image and other images, which are indicated by the vertical distances between them. It can be seen that there is a group of images containing snow and mountains that are very similar to the focus image. A selection by sample image is applied in this view. The sample image is highlighted in pink, and the selected images are highlighted in green.

in large image collections according to perceptual properties, and annotates images using their semantic contents and concepts. Our SIB provides several coordinated views for image browsing and annotation evaluation. The Multi-Dimensional Scaling (MDS) image view maps image miniatures onto the screen based on their content similarities (see Figure 1a) using a fast MDS algorithm [5]: images with similar contents are placed close to each other while images with dissimilar contents are far from each other. The Value and Relation (VaR) [22] content view visually represents the contents of the whole image collection (see Figure 2). The correlations among different contents within the image collection, as well as detailed annotations of each image, are visually revealed in this view. The VaR display is also used to visually convey correlations between the annotations and the low-level features (see Figure 4b). This set of coordinated views provides much more information about image relations and properties than existing image browsers.

A rich set of interaction tools are provided in this SIB (see Section 4). Users (including image analysts) can interactively browse a large image collection using navigation tools, such as zooming, panning, and distortion, and different image layout strategies. Users can interactively select images based on a sample image and/or desired and undesired contents. Users can interactively compare the contents of one image with all other images, or detect the correlations between one content and all other contents. Users can also manually or semi-automatically modify incorrect annotations according to their observations and domain knowledge. In addition, image analysts can examine the relationships among the low level features and the annotation results, and remove redundant features from the classification process through selection in the feature space. Different views of the SIB are coordinated through selections.

In the following sections, details of the SIB are introduced. A

user study and a case study using the SIB to explore the Corel image collection (1100 images) [1] and analyze its classification process are provided. Interesting patterns were found from the case study, and participants in the user study gave positive feedback on the SIB, which indicates that the SIB serves some key design goals well.

2 RELATED WORK

There exist a few content-based image retrieval approaches. Among them, [16, 14, 15] used MDS to create displays where similar images are displayed close to each other. In these approaches the similarities were calculated based on the lower level visual features of the images. [19] used a 2D grid layout and a spiral layout to present search results of images close to an example image in the low-level feature space. We applied the MDS layout for similarity based image visualization as well as the grid layout for non-overlap image visualization in the SIB. We also support the ability to re-MDS selected images for subsequent queries presented in [16]. However, the capabilities of our approach is beyond the above approaches in that (a) the SIB is based on semantic image classification, which is superior to feature-based image retrieval since there is a gap between the semantics of images from the human point of view and the low-level features; (b) the SIB provides a much richer set of navigation and selection tools than the above approaches. These interactions greatly increase the scalability of the system; (c) the coordinated views provided in the SIB, such as the VaR content view and multiple image layout options, enable more effective and efficient image browsing and search than the above approaches.

Besides the above approaches, there are many other photo browsers, such as PhotoFinder [10] and PhotoMesa [4]. In most of them, limited annotations of images, such as time, location, and events at which the photos were taken, are collected through manual
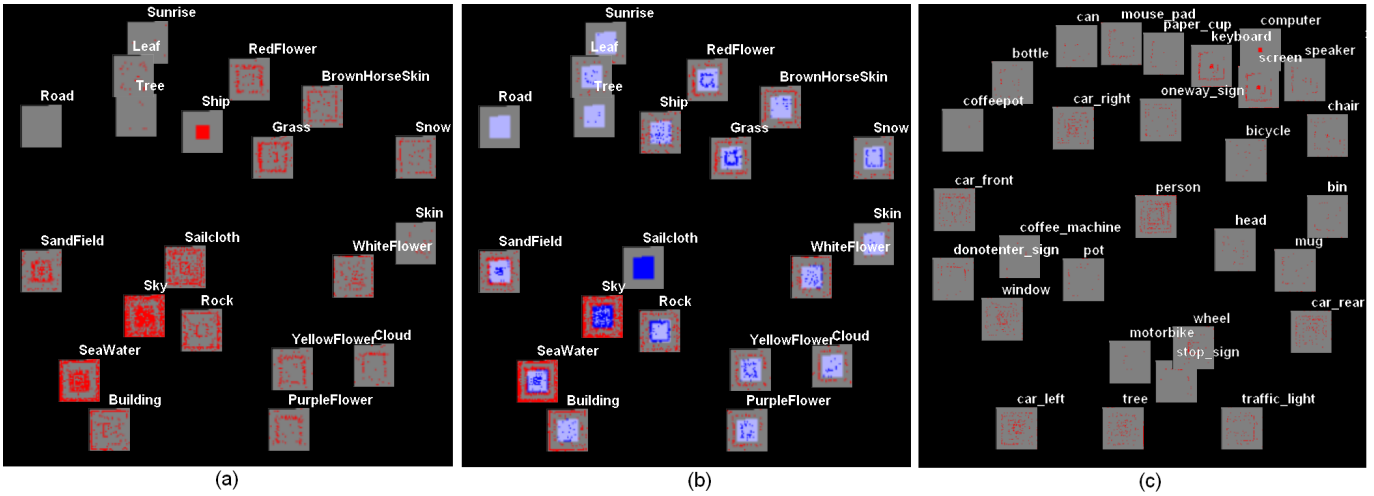
Figure 2: The VaR content view. (a) and (b) show the Corel collection (1100 images and 20 contents) and (c) show the MIT collection (4559 images and 29 contents). In (b), the data items are reordered by their values in the sailcloth dimension. A selection has been performed to highlight images with the sailcloth contents.

annotation and simple semi-automatic annotation techniques such as the Geographic Positioning Systems (GPS) and date recorders built in cameras. The annotations are used to generate meaningful image layout and support visual queries. Although these annotations are useful for certain types of image browsing tasks, such as time-related search, in general they are less detailed than the semantic annotations generated by the automatic image analysis algorithms, and often suffer from absence and incompleteness.

Visual query ability has been provided in many image browsers. For example, PhotoFinder provides visual query widgets and 1-D histograms to facilitate visual queries. 1-D histograms show distributions of the whole image collection and selected subsets in attributes such as time and image ratings. [16] allows users to retrieve images with similar low level features to images in a mouse clicked region. To the best of our knowledge, our prototype is the first system that allows users to directly perform visual queries through a high dimensional visualization where contents of all individual images are explicitly conveyed along with query results.

Beside the similarity based layout and grid layout, there exist many other image layout approaches, such as hierarchical layout [4, 12] for revealing hierarchy structures of the images, graph layout [9] for revealing links among images, map layout [20] to reveal geographic location tags of images, and time quilt and time line layout [8] for time series images. Many of these techniques can be applied to the SIB in the future.

The Value and Relation (VaR) display [22] is a multi-dimensional visualization technique that scales up to hundreds of dimensions. It creates dimension glyphs using pixel-oriented techniques [11] and lays out the glyphs in a 2D display using MDS according to the correlations among the dimensions. Both the texture and closeness of the dimensions indicate the correlations among the dimensions in a VaR display. The VaR display is used the SIB for interactively exploring high dimensional datasets, such as the contents of the images in a large collection.

The VaR display allows users to explore large datasets with real time response for most interactions, since the dimension glyphs are stored as texture objects in OpenGL, thus large datasets do not need to be accessed for interactions such as glyph resizing and relocation. Similar techniques are used in the image views of the SIB. Miniatures of images are stored as texture objects so that users can quickly resize and relocate them. The full-resolution images are only loaded upon user request.

The searching by sample image interaction in the SIB is inspired by the searching by sample document interaction in IN-SPIRE, which is a text document visualization tool [21].In addition, many interactions in the MDS image view of the SIB, such as the rainfall interaction for detecting relationships between one image and other images, distortion, relocation, and resizing are borrowed from the VaR display [22, 23].

3 THE ANNOTATION ENGINE

3.1 Automatic Content-Based Image Annotation

Figure 3: The semantic image classification results for the concept "sea world" with the most relevant objects, such as "sand field" [6].

We use a recently proposed concept-sensitive image content analysis technique [6] as our automatic annotation engine. This technique abstracts image contents by automatically detecting the underlying salient objects (i.e., significant distinguishable regions) in images and associating them with the corresponding semantic objects and concepts according to their perceptual properties. The keywords for interpreting these semantic objects and concepts are used as the keywords for image annotation. For example, the highlighted regions in Figure 3 are the salient objects detected and associated with the semantic object "sand field". Other salient objects are also detected from those images and associated with semantic objects, such as "seawater" and "sky". All the images in Figure

193

3 are associated with the semantic concept "sea world" since the semantic objects "sand field", "seawater" and/or "boat" form the image concept "sea world". With the salient objects for concept-sensitive image content interpretation, the semantic gap between the low-level visual features and the high-level semantic concepts is becoming bridgeable.

In this annotation engine, a set of functions for detecting different types of pre-defined salient objects are used. For example, the detection function for the salient object "sand field" consists of the following components: (a) low-level image segmentation to obtain homogeneous image regions on color or texture by using an automatic image segmentation technique; (b) classifying the relevant image regions as the salient object of "sand field" by using a Support Vector Machine (SVM). By integrating machine learning for salient object detection, this automatic salient object detection technique has achieved very good performance [6]. In the example Corel collection, 20 types of pre-defined salient objects were classified and they were aggregated into 9 different concepts.

3.2 Datasets in the Content-Based Image Annotation Process

In the proposed SIB, it is the data generated by the automated intelligent image analysis process and visualized by the displays that ties analysis and visualization together. It includes:

- Image collections to be annotated and explored. They often contain thousands or more images. An image created by a modern digital camera often contains 5 megapixels or more.

- Low-level visual features of an image or its sub-regions, such as 12-D color histograms, 64-D wavelet texture features, 7-D RGB dominant colors and variances, 7-D LUV dominant colors and variances, 7-D Tamura texture features, and 32-D MPEG-7 color layouts. All these high-dimensional visual features are used to characterize the underlying image contents. Obviously, more visual features can be extracted to characterize various visual properties of images.

- Semantic image contents characterized by the relevant keywords via semantic image classification. One or more salient objects can be detected from an image. Each salient object is associated with the relevant semantic object that can be interpreted by using the corresponding keyword, such as sand field, seawater, or sky, which is called the semantic content (or keyword) of the salient object. An image is thus annotated by the semantic contents of their salient objects. Semantic contents of images (i.e., salient objects) are automatically derived from the relevant visual features by the automatic image analysis process.

- Concepts of images. The annotation process assigns semantic concepts to images based on their semantic contents. For example, an image associated with sand field and seawater will be assigned a "seaworld" concept, and an image containing flowers and trees will be assigned a "garden" concept. Obviously, all these are achieved automatically by the underlying semantic image classification process.

Several useful high dimensional datasets can be derived from the above data:

- Feature dataset - it contains the low level features of the salient objects. In such a dataset, each salient object is a data item and each feature is a dimension. The dimensionality of this dataset can be hundreds or thousands.

- Feature-content dataset - it contains the content annotations of all salient objects (one annotation for each object) in an image collection, as well as the low level features of those salient objects used for generating the content annotations. In such a dataset, each salient object is a data item, the content annotation is a dimension, and the features are other dimensions.

- Content dataset - it contains all content annotations of all images in a collection. It is organized in this way; each image is a data item and each content detected in the collection is a dimension. The value of a data item in a dimension indicates if the image contains that content or not (using value 1 or 0). For example, an image with/without the "sand field" annotation has the value 1/0 in the sand field dimension. We use binary values since it cannot be judged from the number of salient objects of the same content how strongly the image is in that content, since the salient objects are of irregular shapes and areas.

- Concept dataset - it is a 1-D dataset where each image is a data item, and the concept annotations of the images form a categorical dimension. For example, the values of the images in the dimension can be "seaworld", "garden" or other concepts.

4 IMAGE BROWSING INTERFACE

Providing an intuitive visual interface for users that allows interactive image exploration for large image databases is an important design goal of the SIB. A common strategy toward this goal is to use the layout of images to provide information such as content similarities to users at a glance[12]. Flexible visual queries are also important for effective image retrieval. In the following sections, we introduce our efforts towards effective image browsing and retrieval using meaningful image layout, a unique content dataset visualization, and flexible visual queries.

We first attempted to map the image salient objects to the screen according to their similarities in the feature dataset using MDS. The results were poor; the salient objects looked randomly distributed on the screen and no patterns were detected from the display. The reason is that the feature dataset is in such a high dimensional space that the data items are very sparse in this space. The distances between one item and other items are almost the same. Just as important, the features are non-intuitive to all but image analysts and thus are inappropriate to use by themselves. It seemed to us that we should not base the image layout on such a low-level dataset and that making use of the automatic annotation result was very necessary for effective image browsing.

Our second effort was to sort the images according to their concepts and then show all images in a sequential order using the thumbnail view provided by Microsoft Explorer. Users can then interactively browse the images by moving the scrolling bar and double click a thumbnail to examine it in detail using an image browser such as Microsoft Paint. What surprised us is that this simple solution proved to be very effective for target search in an image collection containing 1100 images in our user study (see Section 6).

However, drawbacks of using this approach were also reported in the user study. For example, it could not provide users a good overview of the image collection; interactions such as visual queries were not supported. Also, as the size of image collections increases, the sequential view could perform worse since users have to move the scrolling bars for long distances and need a good memory to get an overview of the image collection.

In order to overcome these drawbacks, we developed an image browsing interface with the following goals: (1) to provide users an image overview of a large collection so that they can know what kinds of images are in the collection at a glance; (2) to provide users a content overview of the image collection so that they can learn the contents of a collection, their distributions, and their relationships at a glance; (3) to provide users a rich set of interactions for conducting visual queries and analyzing images in detail through simple mouse and keyboard input.
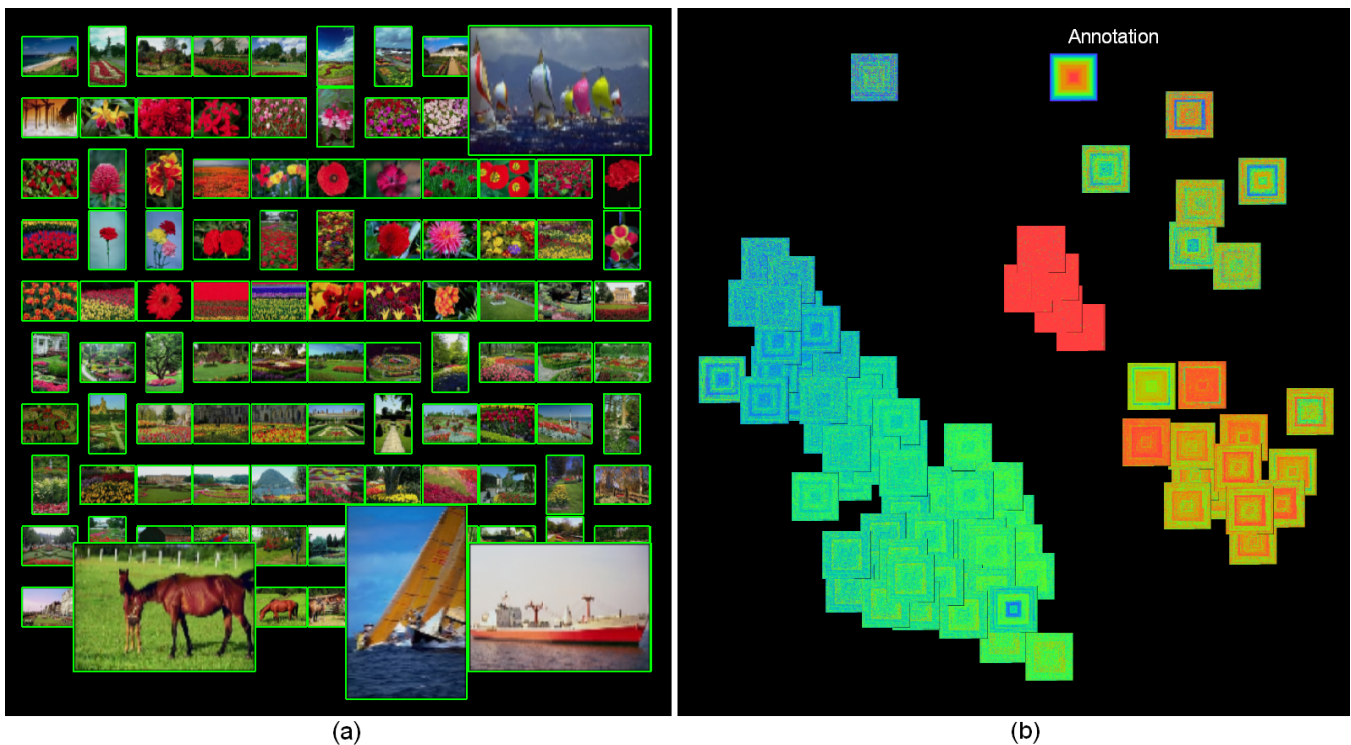
Figure 4: (a) Images with "redflower" annotations in the Corel image collection. Most images do contain red flowers. Several exceptions are enlarged using distortion. (b) The feature-content dataset of the Corel collection (89 dimensions, 10,471 items) in the VaR display.

4.1 Image Overview

Towards the first goal we developed an image overview with MDS or sequential layout. The sequential layout is similar to the view of Windows Explorer; image miniatures are sorted according to their concepts and placed line by line in an array. The MDS layout is generated using the content dataset. First, the distance between each pair of images in the content space is calculated and stored in a distance matrix. Then the matrix is used as input to an MDS process whose output is a position in a 2D space for each image. Images close in the content space will be close to each other in the 2D space, in most cases. The image miniatures are then mapped to their positions in the 2D space. Figure 1a shows the MDS overview of the Corel image collection. It can be seen that this image collection contains boating images, garden images, mountain images, ranching images and so on.

The SIB allows users to switch freely between different views and image layouts and provides many interactions that cause refreshing of the image display. Since image collections often contain thousands or millions of images, the scalability of the image browser is very important. We use two methods to increase the scalability of the SIB. First, since the original images could be very large due to the high resolution of digital cameras, miniatures of the images are displayed rather than the original images unless users explicitly request an original image. Although a simple uniform sampling algorithm is used in the prototype, advanced thumbnail creation techniques, such as [18], can be used in the future to improve the quality of miniatures generated. Second, we load the miniatures as texture objects in OpenGL and map them onto the screen when they are displayed. All interactions in our browser except the first-time loading can be performed in real time, since they only involve different mapping of the constructed texture objects, which can be done very fast in OpenGL. A bottleneck of this approach is that we could use up the texture memory for a very large

image collection. In that case, we can either decrease the resolution of the image miniatures or use other multi-resolution techniques. For example, we can use structure-based sampling to show fewer images while still showing a good overview of the collection.

Besides the sequential and MDS image layout, positioning images according to a hierarchy constructed on the concepts using tree visualization methods might also be a good idea. First, it was revealed in our user study (Section 6) that organizing images by their concepts enables effective browsing and retrieval. Second, showing images using a hierarchical layout has proven to be effective in many image browsers [4, 12]. We plan to build concept hierarchies on the annotation results for large image collections and allow users to interactively explore images stored in the concept tree.

4.2 Interactions in the Image Overview

Similar images will be close to each other in the MDS image overview. On the one hand, this is necessary for creating the overview by showing the distribution of similar images. On the other hand, it prevents users from seeing some images due to the clutter. In order to reduce clutter, the following interactions are provided in the MDS image overview:

- **Reordering**: Users can randomly change the display orders of all images by clicking a *randomizing order* button in the control frame, thus each image can get an equal chance to be "visible". Users can also explicitly bring an image to the front of the display by clicking its visible part or selecting its name from a combo box. Images included in the search result will be automatically brought to the front of the display.

- **Dynamic Scaling**: Users can interactively reduce the sizes of all image miniatures to reduce overlap, or increase their sizes to examine details of individual images.

- **Relocation**: Users can manually change the positions of individual images by mouse dragging and dropping to reduce overlap.

- **Distortion**: Users can enlarge the size of some image miniatures while retaining the size of all other miniatures to examine details within context (see Figure 4a).

- **Showing original image**: users can double click an image to show it at full resolution in a new window. This window can be manually repositioned.

- **Zooming and Panning**: Users can zoom in, zoom out or pan the image display. Used together with dynamic scaling, zooming in allows users to examine local details with less clutter.

Selection is extremely important for the SIB. We allow users to interactively select images according to their similarities to a sample image. What users need to do is simply click the sample image or select the name of it from a combo box, and then interactively change the similarity threshold through a scaling bar. All images whose similarities to the sample image are higher than the threshold will be automatically selected and highlighted (see Figure 1b).

Selected images can be displayed in the image overview in different modes, which are set by the users. In the *display all* mode, all images in the collection are shown and selected images are highlighted and shown on the top of other images. In the *display selected only* mode, only selected images are shown. If the images are laid out in the sequential view (see Figure 4a), the images can either be sorted by their concepts, or by their similarities to the sample image. In the *MDS selected image* mode, a new MDS layout is generated for the selected images according to the correlations among them. This mode provides a good overview for the selected images and is preferred if the selected subset is still large. It also facilitates search by association effectively since selected images are clustered according to their semantic contents in this view; thus users can easily find associated images of different semantic contents.

A *rainfall* mode is provided for the image overview inspired by the rainfall animation in the VaR display [23]. In this mode, the correlations between other images and the image of interest, such as the sample image in a selection, are explicitly revealed through animation. During the animation, the focus image is placed in the middle bottom of the display (the ground) and other images fall to the ground from the top of the display (the sky) at accelerations related to their similarities to the focus images. Images similar to the focus image move toward the ground faster than other images. Figure 1b shows a screenshot of a rainfall animation.

4.3 Content Overview

The content overview is generated by visualizing the content dataset in the VaR display. Figure 2a is the content overview of the Corel image collection. In this view, each content is represented by a block. Each image is mapped to a pixel (which might be enlarged to a small block in interactive exploration) in each block whose color represents if the image contains that content. In Figure 2a the red/grey color indicates that the image contains/does not contain the content. The pixels representing the same image are in the same position in all blocks so that users can associate them together. The VaR content overview conveys the following information to users:

- Contents of all images in the collection detected by the automatic annotation process. Each block in the display is labeled by the content it represents, thus contents of the image collection can be observed by scanning the labels.

- Distribution of images containing each content. It can be observed by the distribution of red pixels in each block. It can be seen from Figure 2a that sky appears in most images of the Corel image collection.

- Correlations among the contents. They can be observed by examining the positions of the blocks and their textures. For example, Figure 2a shows that sand field often appears together with seawater in the Corel image collection.

- Contents of selected images. Figure 2b highlights the selected images in the VaR content view. The images are sorted by their values in the sailcloth dimension. Red/grey is turned to blue/light grey for selected images. It can be seen in Figure 2b that all images with the sailcloth content are selected and that many of them do not contain seawater.

4.4 Interactions in the VaR Content Overview

The VaR display provides a rich set of interaction tools to users, such as clutter reduction, reordering, and detection of correlations between a dimension of focus and other dimensions [22]. All those interactions are provided in the VaR content overview.

According to the binary nature of the content dataset and the need for searching by contents in image retrieval, a special set of interactions is added to the VaR content view. Users can start a new search for images with (without) certain contents, reduce a selected subset by requiring that the search results must (or must not) contain certain contents, or increase a selected subset by adding images with/without certain contents. All those interactions are performed by clicking related content blocks while holding certain function keys. Searching by sample image and searching by content can be combined together through these interactions since the search results are shared by different views.

4.5 Coordination between the Image Overview and the Content Overview

The image overview and the content overview provide different views for the same image collections and are closely coordinated. In particular, each image has its visual representations in both the image overview and the content overview. Selected images are highlighted in both overviews.

Users may prefer different views in different exploration stages. For example, if users want to select images based on their relationships to a sample image, they may use the image overview. If they want to select images by their contents, they may start from the content display. In addition, the VaR content view is more scalable than the MDS image view. Thus for very large image collections, starting from the VaR content overview and switching to the image view after reducing the number of images to be shown through selections might be an effective strategy.

5 VISUAL IMAGE ANALYSIS

Not only expert image analysts, but also common users, can be interested in the accuracy of the annotation. In addition, analysts are interested in the correlations among the content annotations of salient objects and the low level features (i.e., the feature-content dataset) since this information is useful for improving the annotation process.

5.1 Annotation Evaluation and Improvement

The image view and content view and their interactions provide a powerful tool for users to evaluate the annotation results by comparing annotations of images with their real contents. Figures 4a and 2b show two examples of using the SIB to examine annotations of the Corel image collection. In the first example, images with "red-flower" annotations are selected and shown in the sequential image view in Figure 4a. It can be seen that the annotation for red flowers

is pretty good since most selected images do contain red flowers. There are only a small number of exceptions, some of which are enlarged in the display. In the second example, images with "sailcloth" annotations are selected. From the VaR content view (see Figure 2b), it is seen that many images containing the "sailcloth" annotation do not contain the "seawater" annotation! To investigate further, we selected all images with "sailcloth" and without "seawater". From the image view of the selected images, we found that most selected images did not contain sailcloth. Through this simple visual exploration, we learned that the automatic annotation technique we used is good at detecting "redflower", but needs to be improved in detecting "sailcloth".

How to improve the annotation through visualization is important in our visual analytic approach. Currently, a simple manual annotation interaction is provided to fix the wrong annotation produced by the automatic process. By double clicking an image, a dialog with all contents in the collection will be popped up, with a check box beside each content. If the clicked image is annotated as containing a content, the check box beside it is checked. Users can manually change the status of the check boxes to change the annotation. The underlying feature-content, content and concept datasets will be modified as a consequence. In the future, we will explore how to pass the feedback to the automatic classification process to improve its accuracy when annotating new images. This will be necessary as the size of the image collection grows.

Since the automatic annotation process is not 100% accurate and its accuracy varies depending on content types, we need an indicator of reliability for the users. We show the reliability of annotating by surrounding the content blocks in the VaR display with a frame. The frame of an unreliably annotated content will be yellowish to give users a warning against depending on it, while a green frame indicates that the users can use the content safely. The reliability information can be provided by the automatic annotation process or manually set up by the analysts.

5.2 Annotation Process Monitoring and Improvement

The VaR display scales effectively to hundreds of dimensions. Our preliminary research showed that it provides analysts a great opportunity to visualize the low-level feature dataset and feature-content dataset to obtain an intuitive impression of the relationships among the features and the annotations. Figure 4b shows the feature-content dataset of the Corel image collection in the VaR display. It contains 89 dimensions and 10,471 data items. It can be seen that there are some features more closely related to the annotation than the others. It can also be seen that many features used for classification are nearly identical to other features and could be removed from the annotation process. The automatic selection for distinct dimensions provided by the VaR display (see [22] for more detail) provides analysts a powerful tool to remove those redundant features. We will work on integrating these visualization tools with the classification process for making the latter more transparent and improving its efficiency.

6 USER STUDY

A user study has been conducted to evaluate the prototype by comparing it with the sequential thumbnail view of Microsoft Explorer on a high resolution 4 megapixel desktop display. For Explorer, two modes are used: the images are randomly sorted (Random Explorer); the images are sorted by their semantic concepts generated by our classification process (Sorted Explorer). A sequential layout similar to Sorted Explorer is actually available in the SIB, but it was hidden from the subjects. Subjects could only use the sequential layout in the SIB to view search results. The Corel image collection was used for the user study.

Ten subjects participated in the user study. One subject was a psychology major graduate student, five subjects were gradu-

ate students in the field of visualization, two subjects were researchers/post doctors in visualization, and two were senior Ph.D. students who were involved in the development of the annotation engine. The subjects did the user study one by one on the same desktop with the same instructor. Each subject used both Sorted Explorer and the SIB. Half of the subjects used Sorted Explorer first and another half used the SIB first. Since Random Explorer was expected to have significantly worse performance, it was only tested on 3 subjects after they used Sorted Explorer and the SIB.

In experiments on the SIB, there was a 20 minute training session and a further 10 minutes free exploration time preceding a set of tasks. Experiments on Explorer were almost the same, except that there was no training since subjects were all familiar with Explorer. A post-test survey for user preference and discussion were conducted immediately after a user finished all experiments.

Two sets of equivalent (in difficulty) tasks were used. Half of subjects used task set 1 for Sorted Explorer and set 2 for the SIB, and another half used task set 2 for Sorted Explorer and set 1 for the SIB. For Random Explorer task set 1 (1 subject) or 2 (2 subjects) were used. There were three tasks in each task set. For the first task, there are three trials. In each trial, users were presented with an image from the collection and were asked to search for the image from the 1100 images. The time taken to complete each trial was recorded. A time out of 180 seconds was used for each trial. In the second task, users were asked to find images containing particular features (e.g. images containing sand field and water). The final task required users to approximate what proportion of the images in the collection contained particular contents (e.g. the percentage of images in the collection containing mountains).

For the first task, we calculated the percentage of failed trials, average time and standard deviation of succeeded trials. With Random Explorer, 22% (2 of 9) trials failed. The average time and standard deviation were 81 and 29 seconds. For Sorted Explorer, 6.6% (2 of 30) trials failed. The average time and standard deviation were 29 and 20 seconds. With the SIB, 20% (6 of 30) trials failed. The average time and standard deviation were 45 and 26 seconds. Subjects found the target images significantly faster using Sorted Explorer and the SIB than using Random Explorer. To conduct this task using Explorer, subjects scanned the image collections. For the SIB, subjects usually searched by contents in the VaR content view or searched by sample images in the MDS view, and then switched to the sequential image views to search for the target image from the selected image subset.

All failed trials with the SIB were due to inaccurate annotated contents of the target images or sample images used in the search. It is expected that the fail rate of the SIB will decrease when the accuracy of the automatic classification increases. However, it seems that annotation at the concept level is more "error tolerant" than in the content level since there were less failed trials with Sorted Explorer. The reason is that an image with a wrong annotation in a salient object might still be classified into the correct concept, since its salient objects playing important roles in concept classification were properly annotated. According to this observation, we will use concepts more often in the future development of the SIB.

According to our observation, the SIB was slightly less efficient in task 1 than Sorted Explorer due to the following reasons: the names of some contents were confusing; subjects spent time to adjust the search to bypass wrong annotations; there was only one view in Explorer while users needed to switch among multiple views in the SIB; and subjects were not familiar with the menu and quick buttons for search and selection in the SIB. However, we expect that the performance of sorted Explorer will decrease much faster than the SIB as the size of the image collection increases, since no overview is provided in Explorer. Task 2 showed similar patterns as task 1.

The SIB showed significant benefits in performing task 3, with a

much greater accuracy in the results compared with those obtained using Explorer. This demonstrates the power of the SIB in providing a good overview of the entire database.

In the post-test survey, a 1 to 10 scale was used. The average rating for the question "prefer using the SIB or Explorer for image searching was 7.7 (1 for Explorer, 10 for the SIB). The average rating for usefulness of the MDS overview, search by example, search by content, changing glyph size, and randomizing orders were 6.2, 6.4, 9.2, 8.8, and 6.3 respectively (1 for useless, 10 for very useful). Subjects commented that the interactions in the SIB made the experience more enjoyable than Explorer. Users were particularly interested in the MDS display for its ability to display the entire database on a single screen. They complained that Explorer could not provide such a good overview. We also collected many valuable suggestions for improving the SIB, such as putting the image view and the VaR content view side by side so that both can be seen at the same time, allowing selection of multiple images using dragging and dropping in the MDS image views so that adjacent images can be selected at the same time, and providing a sample image beside each content block in the VaR view to make it easier to understand and find a content.

7 CONCLUSION AND FUTURE WORK

The major contributions of this paper are: (a) a novel semantic image browser that is among the first attempts to bridge information visualization with automated intelligent image analysis in the semantic level; (b) an MDS image layout based on semantic similarities that provides promising image overviews for large image collections; (c) an VaR content display that visually represents the content of a large image collection; (d) a rich set of interaction tools such as combinable searching by sample and searching by contents; and (e) visualizations and interactions that allow image analysts to visually monitor, evaluate, and improve their annotation processes.

Besides the Corel dataset, the MIT dataset with 4559 images and 29 classified contents [7], which is a subset of the LabelMe dataset [2], has also been tested in the SIB without any problems. Figure 2c showed the content view of this image collection. We plan to test the SIB with larger image collections in the future. For larger datasets, the MDS algorithm can still be fast by trading efficiency with accuracy. In addition, the MDS layout is only calculated once when the dataset is first-time loaded and the image positions are then recorded in a file for future use. For the VaR display, the largest dataset tested so far contained 838 dimensions and 11,413 data items (see [23]). It is expected that the VaR display can effectively visualize image collections with hundreds of contents with its rich set of interaction tools. Thus the major bottleneck of the SIB for scalability is the available texture memory of OpenGL for the thumbnails, and the available screen space for distinct images in the MDS overview.

In the future, we plan to improve the scalability of the SIB using techniques such as concept hierarchies and sampling techniques. We also plan to integrate the time and location information provided by digital photos with the automatic annotation results to generate a more powerful semantic image browser with time and spatial information. Moreover, we want to strengthen the role of visualization in improving the semantic image classification process. Finally, we will conduct further user studies.

REFERENCES

[1] The corel image collection. http://www.corel.com.

[2] The labelme dataset. http://people.csail.mit.edu/brussell/research/ LabelMe.

[3] L. Armitage and P. Enser. Analysis of user need in image archives. *Journal of Information Science*, 23(4):287–299, 1997.

[4] B. Bederson. Photomesa: a zoomable image browser using quantum treemaps and bubblemaps. *UIST 2001*, pages 71–80, 2001.

[5] C. Bentley and M. Ward. Animating multidimensional scaling to visualize n- dimensional data sets. *Proc. IEEE Symposium on Information Visualization*, pages 72–73, 1996.

[6] J. Fan, Y. Gao, H. Luo, and G. Xu. Automatic image annotation by using concept-sensitive salient objects for image content representation. *International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 361–368, 2004.

[7] Y. Gao, J. Fan, H. Luo, X. Xue, and R. Jain. Automatic image annotation by incorporating feature hierarchy and boosting to scale up svm classifiers. *ACM Multimedia*, 2006.

[8] D. Huynh, S. Drucker, P. Baudisch, and C. Wong. Time quilt: scaling up zoomable photo browsers for large, unstructured photo collections. *CHI Extended Abstracts 2005*, pages 1937–1940, 2005.

[9] T. Jankun-Kelly and K. Ma. Moiregraphs: Radial focus+context visualization and interaction for graphs with visual nodes. *Proc. IEEE Symposium on Information Visualization*, pages 59–66, 2003.

[10] H. Kang and B. Shneiderman. Visualization methods for personal photo collections: Browsing and searching in the photofinder. *IEEE International Conference on Multimedia and Expo (III) 2000*, pages 1539–1542, 2000.

[11] D. Keim, H.-P. Kriegel, and M. Ankerst. Recursive pattern: a technique for visualizing very large amounts of data. *Proc. IEEE Visualization '95*, pages 279–286, 1995.

[12] J. Kustanowitz and B. Shneiderman. Meaningful presentations of photo libraries: rationale and applications of bi-level radial quantum layouts. *ACM/IEEE Joint Conference on Digital Libraries*, pages 188–196, 2005.

[13] S. Ornager. Image retrieval: Theoretical and empirical user studies on accessing information in images. *ASIS 97: Proceedings of the 60 th ASIS Annual Meeting*, 34:202–211, 1997.

[14] K. Rodden, W. Basalaj, D. Sinclair, and K. Wood. Evaluating a visualization of image similarity as a tool for image browsing. *Proc. IEEE Symposium on Information Visualization*, pages 36–43, 1999.

[15] K. Rodden, W. Basalaj, D. Sinclair, and K. Wood. Does organisation by similarity assist image browsing? *Proc. ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 190–197, 2001.

[16] Y. Rubner, C. Tomasi, and L. Guibas. A metric for distributions with applications to image databases. *IEEE International Conference on Computer Vision*, pages 59–66, 1998.

[17] A. 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 (TPAMI)*, 22(12):1349–1380, 2000.

[18] B. Suh, H. Ling, B. Bederson, and D. Jacobs. Automatic thumbnail cropping and its effectiveness. *UIST 2003*, pages 95–104, 2003.

[19] R. Torres, C. Silva, C. Medeiros, and H. Rocha. Visual structures for image browsing. *CIKM 2003*, pages 49–55, 2003.

[20] K. Toyama, R. Logan, and A. Roseway. Geographic location tags on digital images. *ACM Multimedia 2003*, pages 156–166, 2003.

[21] J. Wise, J. Thomas, K. Pennock, D. Lantrip, M. Pottier, A. Schur, and V. Crow. Visualizing the non-visual: Spatial analysis and interaction with information from text documents. *Proc. IEEE Symposium on Information Visualization*, pages 51–58, 1995.

[22] J. Yang, A. Patro, S. Huang, N. Mehta, M. Ward, and E. Rundensteiner. Value and relation display for interactive exploration of high dimensional datasets. *Proc. IEEE Symposium on Information Visualization*, pages 73–80, 2004.

[23] J. Yang, M. Ward, E. Rundensteiner, and W. Ribarsky. Value and relation display: Interactive visual exploration of large datasets with hundreds of dimensions. Submitted to IEEE Transactions on Visualization and Computer Graphics, Visual Analytics Special Issue.