STORIES

D-Lib Magazine
November 1999

Volume 5 Number 11

ISSN 1082-9873

Semantics-sensitive Retrieval for Digital Picture Libraries

James Ze Wang
Computer Science and Medical Informatics
Stanford University
wangz@cs.stanford.edu

Jia Li
Computer Science and Electrical Engineering
Stanford University
(currently with Xerox Palo Alto Research Center)
jiali@db.stanford.edu

Desmond Chan
Computer Science
Stanford University
deschan@cs.stanford.edu

Gio Wiederhold
Computer Science and Medical Informatics
Stanford University
gio@cs.stanford.edu

Abstract

We present SIMPLIcity (Semantics-sensitive Integrated Matching for Picture LIbraries), an image database retrieval system, which uses high-level semantics classification and integrated region matching based upon image segmentation. The SIMPLIcity system represents an image by a set of regions, roughly corresponding to objects, which are characterized by color, texture, shape, and location. Based on segmented regions, the system classifies images into categories which are intended to distinguish semantically meaningful differences. These high-level categories, such as textured-nontextured, indoor-outdoor, objectionable-benign, graph-photograph, enhance retrieval by narrowing down the searching range in a database and permitting semantically adaptive searching methods.

1 Introduction

With the steady growth of computer power, rapidly declining cost of storage devices, and ever-increasing access to the Internet, digital acquisition of information has become increasingly popular in recent years. Digital information is preferable because of convenient sharing and distribution. This trend has motivated research in image databases, which were nearly ignored by traditional computer systems because of the large amount of data required to represent images and the difficulty of automatically analyzing images. Currently, storage is less of an issue since huge storage capacity is available at low cost. However, effective indexing and searching of large-scale image databases remains as a challenge for computer systems. The automatic derivation of semantics from the content of an image is the focus of interest for most research on image databases.

In this paper, we present SIMPLIcity (Semantics-sensitive Integrated Matching for Picture LIbraries), an image database retrieval system, which uses high-level semantics classification and integrated region matching based upon image segmentation. The SIMPLIcity system segments an image into a set of regions, which are characterized by color, texture, shape, and location. Based on segmented regions, the system classifies images into categories which are intended to distinguish semantically meaningful differences. The feature extraction scheme is tailored to best suit each semantic class. A measure for the overall similarity between images is defined by a region-matching scheme that integrates properties of all the regions in the images. Armed with this global similarity measure, the system provides users a simple querying interface. The application of SIMPLIcity to a database of about 60,000 general-purpose images shows robustness to cropping, scaling, shifting, and rotation.

2 Related Work in Content-based Image Retrieval

Many content-based image database retrieval systems have been developed, such as the IBM QBIC System [Faloutsos1994, Flickner1995] developed at the IBM Almaden Research Center, the Virage System [Gupta1997] developed by the Virage Incorporation, the Photobook System developed by the MIT Media Lab [Picard1993, Pentland1995], the WBIIS System [Wang1998.1] developed at Stanford University, and the Blobworld System [Carson1999] developed at U.C. Berkeley. The common ground for content-based image retrieval systems is to extract a signature for every image based on its pixel values, and to define a rule for comparing images. This signature serves as an image representation in the "view" of a retrieval system. The components of the signature are called *features*. One advantage of using a signature instead of the original pixel values is the significant simplification of image representation. However, a more important reason for using the signature is the improved correlation with image semantics. Actually, the main task of designing a signature is to bridge the gap between image semantics and the pixel representation. Content-based image database retrieval systems roughly fall into three categories depending on the signature extraction approach used: histogram, color layout, and region-based search. We will briefly review these methods later in this section. There are also systems that combine retrieval results from individual algorithms by a weighted sum matching metric [Gupta1997, Flickner1995], or other merging schemes [Sheikholeslami1998].

After extracting signatures, the next step is to determine a comparison rule, including a querying scheme and the definition of a similarity measure between images. Most image retrieval systems perform a query by having the user specify an image; the system then searches for images similar to the specified one. We refer to this as global search, since similarity is based on the overall properties of images. In contrast to global search, there are also systems that retrieve based on a particular region in an image, such as the NeTra system [Ma1997] and the Blobworld system [Carson1999]. This querying scheme is referred to as partial search.

2.1 Histogram Search

For histogram search [Niblack1993, Flickner1995, Rubner1997], an image is characterized by its color histogram. The drawback of a global histogram representation is that information about object location, shape, and texture is discarded.

2.2 Color Layout Search

The "color layout" approach attempts to deal with the deficiencies of histogram search. For traditional color layout indexing [Niblack1993], images are partitioned into blocks and the average color of each block is stored. Later systems [Wang1998.1] use significant wavelet coefficients instead of averaging. Color layout search is sensitive to shifting, cropping, scaling, and rotation because images are characterized by a set of local properties. The approach taken by the WALRUS system [Natsev1999] to reduce the shifting and scaling sensitivity for color layout search is to exhaustively reproduce many subimages based on an original image. The similarity between images is then determined by comparing the signatures of subimages. An obvious drawback of the system is the sharply increased computational complexity due to exhaustive generation of subimages. Furthermore, texture and shape information is discarded in the signatures because every subimage is partitioned into four blocks and only average colors of the blocks are used as features. This system is also limited to intensity-level image representations.

2.3 Region-based Search

Region-based retrieval systems attempt to overcome the issues with color layout search by representing images at the object-level. A region-based retrieval system applies image segmentation to decompose an image into regions, which correspond to objects if the decomposition is ideal. The object-level representation is intended to be close to the perception of the human visual system (HVS). Since the retrieval system has identified objects in the image, it is easier for the system to recognize similar objects at different locations and with different orientations and sizes. Region-based retrieval systems include the Netra system [Ma1997], the Blobworld system [Carson1999], and the query system with color region templates [Smith1999].

The NeTra system [Ma1997] and the Blobworld system [Carson1999] compare images based on individual regions. Although querying based on a limited number of regions is allowed, the global query is performed by merging single-region query results. The motivation is to shift part of the comparison task to the users. To query an image, a user is provided with the segmented regions of the image, and is required to select the region to be matched and also attributes, e.g., color and texture, of the region to be used for evaluating similarity. Such querying systems provide more control for the users. However, the key pitfall is that the user’s semantic understanding of an image is at a higher level than the region representation. When a user submits a query image of a horse standing on grass, the intent is most likely to retrieve images with horses. But since the concept of horses is not explicitly given in region representations, the user must convert the concept into shape, color, texture, location, or combinations of them. For objects without distinctive attributes, such as special texture, it is not obvious for the user how to select a query from the large variety of choices. Thus, such a querying scheme may add burdens on users without any reward.

Not much attention has been paid to developing similarity measures that combine information from all of the regions. One work in this direction is the querying system developed by Smith and Li [Smith1999]. Their system decomposes an image into regions with characterizations pre-defined in a finite pattern library. The measure defined is sensitive to object shifting. Robustness to scaling and rotation is also not considered by the measure.

3 Related Work in Image Semantic Classification

Although region-based systems attempt to decompose images into constituent objects, a representation composed of pictorial properties of regions is indirectly related to its semantics. There is no clear mapping from a set of pictorial properties to semantics. An approximately round brown region might be a flower, an apple, a face, or a part of sunset sky. Moreover, pictorial properties such as color, shape, and texture of an object vary dramatically in different images. If a system understood the semantics or meaning of images and could determine which features are significant in any object, it would be capable of fast and accurate search. However, due to the great difficulty of recognizing and classifying images, not much success has been achieved in identifying high-level semantics for the purpose of image retrieval. Therefore, most systems are confined to matching images with low-level pictorial properties.

Despite the fact that it is currently impossible to reliably recognize objects in general-purpose images, there are methods to distinguish certain semantic types of images. Any information about semantic types is helpful since a system can constrict the search to images of a particular semantic type. The systems can also improve retrieval by using various matching schemes tuned to the semantic class of the query image. One example of semantics classification is the identification of natural photographs versus artificial graphs generated by computer tools [Li1998, Wang1998.2]. Other examples include the WIPE system to detect objectionable images developed by Wang, et al. [Wang1998.2] and the system to classify indoor and outdoor scenes developed by Szummer and Picard [Szummer1998]. Wang and Fischler [Wang1998.3] have shown that rough but accurate semantic understanding can be very helpful in computer vision tasks such as image stereo matching. Most of these systems use statistical classification methods based on training data.

4 The SIMPLIcity Retrieval System

We now present the major contributions of our prototype SIMPLIcity (Semantics-sensitive Integrated Matching for Picture LIbraries) system.

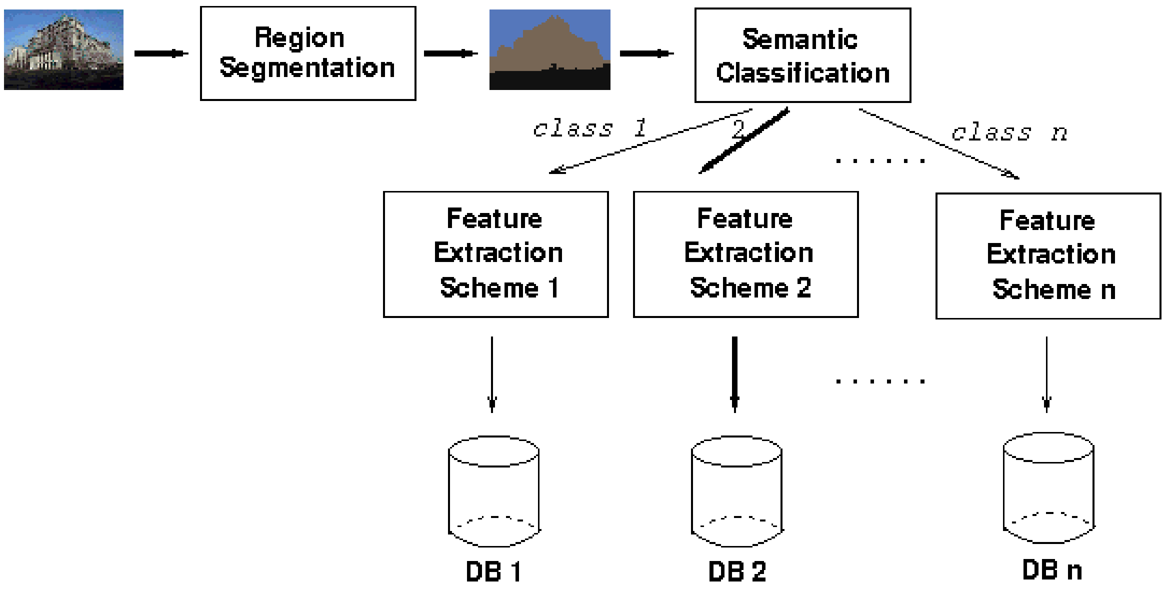


Figure 1. The architecture of the feature indexing module. The heavy lines show a sample indexing path of an image.

The architecture of the SIMPLIcity system is described by Figure 1, the indexing process, and Figure 2, the querying process. Figure 1 shows the indexing process; and Figure 2 shows the querying process. During indexing, the system partitions an image into 4x4 pixel blocks and extracts a feature vector for each block. The k-means clustering approach [Hartigan1979] is then used to segment the image into regions. The segmentation result is fed into a classifier that decides the semantic type of the image. An image is classified as one of the *n* pre-defined mutually exclusive and collectively exhaustive semantic classes. The system can be extended to allow an image being softly classified into multiple classes with probability assignments. As indicated previously, examples of semantic types are indoor-outdoor, objectionable-benign, and graph-photograph images. Features including color, texture, shape, and location information are then extracted for each region in the image. The features selected depend on the semantic type of the image. The signature of an image is the collection of features for all of its regions. Signatures of images with various semantic types are stored in separate databases.

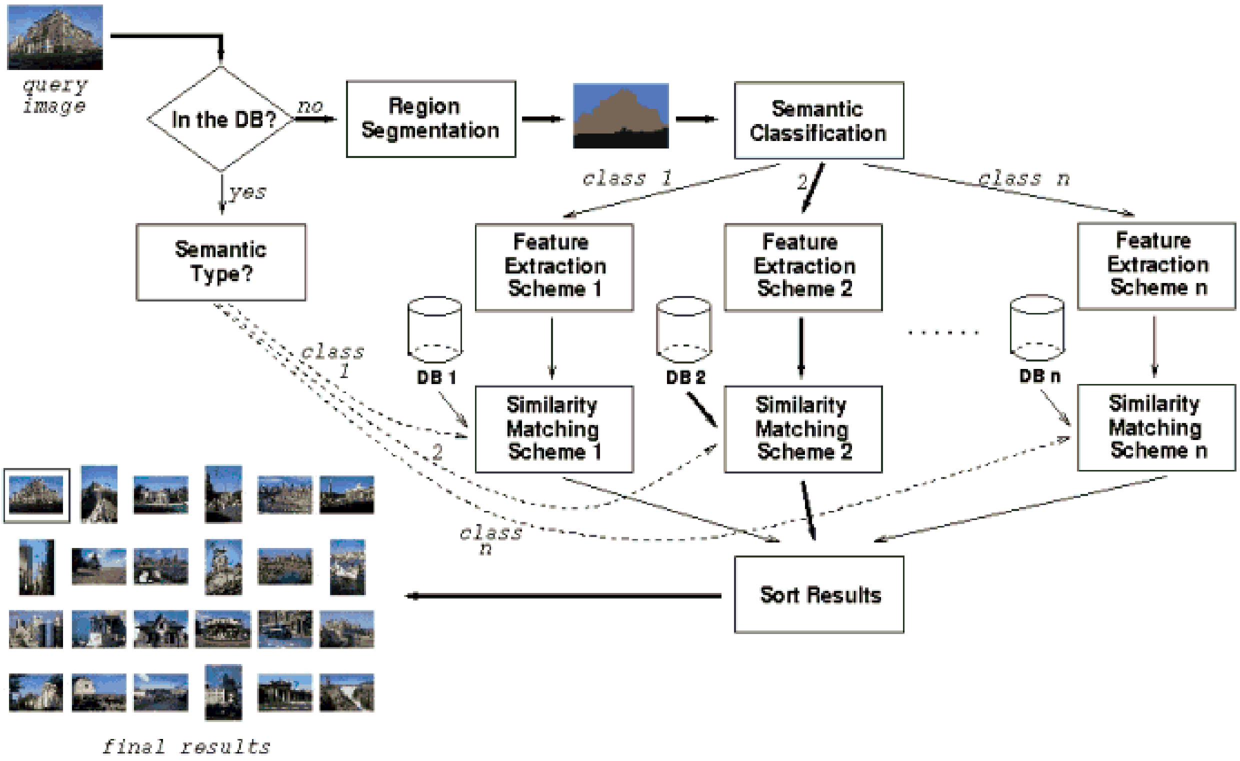


Figure 2. The architecture of the query processing module. The heavy lines show a sample querying path of an image.

In the querying process, if the query image is not in the database as indicated by the user interface, it is first passed through the same feature extraction process as was used during indexing. For an image in the database, its semantic type is checked first and then its signature is extracted from the corresponding database. Once the signature of the query image is obtained, similarity scores between the query image and images in the database with the same semantic type are computed and sorted to provide the list of images that appear to have the closest semantics.

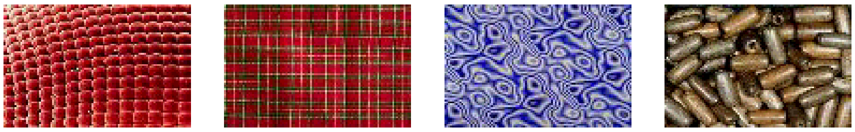


Figure 3. Sample textured images.

For the current implementation of the SIMPLIcity system, we are particularly interested in classifying images into the classes *textured* and *non-textured* [Li1999.2]. By textured images, we refer to images that are composed of repeated patterns and appear like a unique texture surface, as shown in Figure 3. As textured images do not contain clustered objects, the perception of such images focuses on color and texture, but not shape, which is critical for understanding non-textured images. Thus an efficient retrieval system should use different features to depict these two types of images. To our knowledge, the problem of distinguishing textured images and non-textured images has not been explored in the image retrieval literature.

Besides using high-level semantics classification, another strategy of SIMPLIcity to shorten the distance between the region representation of an image and its semantics is to define a similarity measure between images based on the properties of all the segmented regions so that information about an image can be fully used. In many cases, knowing that one object usually appears with another object helps to clarify the semantics of a particular region. For example, flowers often appear with green leaves, and boats usually appear with water.

By defining an overall similarity measure, the SIMPLIcity system provides users with a *simple* querying interface. To complete a query, a user only needs to specify the query image. Compared with retrieval based on individual regions, the overall similarity approach also reduces the influence of inaccurate segmentation. Details of the matching algorithm can be found in [Li1999.2].

5 Experimental Results

The SIMPLIcity system has been implemented with a general-purpose image database (available from COREL) including about 60,000 pictures, which are stored in JPEG format with size 384 x 256 or 256 x 384. These images were segmented and classified into textured and non-textured types. For each image, the features, locations, and areas of all its regions are stored. Textured images and non-textured images are stored in separate databases. An on-line demo is provided at URL: http://WWW-DB.Stanford.EDU/IMAGE/.

5.1 Accuracy

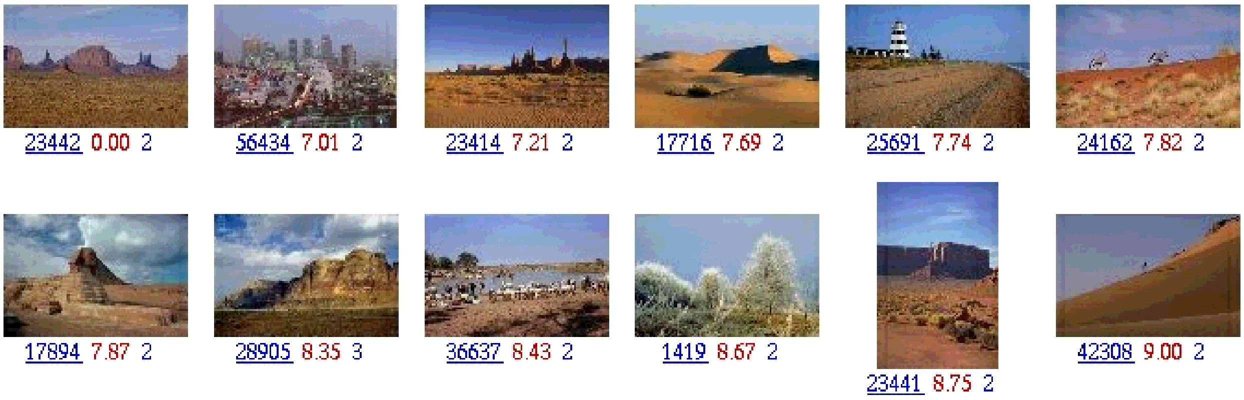


Figure 4. Search Results. The query image is a landscape image on the upper-left corner of the block of images.

For smooth landscape images, which color layout search usually favors, SIMPLIcity performs as well as the color layout approach in general. One example is shown in Figure 4. The query image is the image at the upper-left corner. The underlined numbers below the pictures are the ID numbers of the images in the database. To view the images better or to see more matched images, users can visit the demo web site and use the query image ID to repeat the retrieval.

Figure 5. The query image is a photo of food.

SIMPLIcity performs much better than the color layout approach for images composed of fine details. Retrieval results with a photo of a hamburger as the query are shown in Figure 5. The SIMPLIcity system retrieves 10 images with food out of the first 11 matched images. The top match made by SIMPLIcity is also a photo of hamburger.

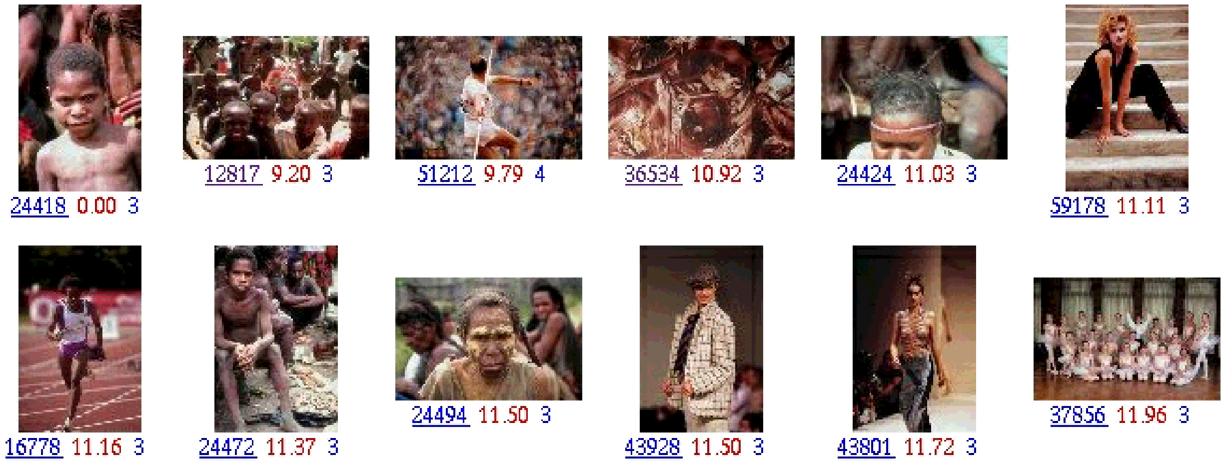


Figure 6. The query image is a portrait image that probably depicts life in Africa.

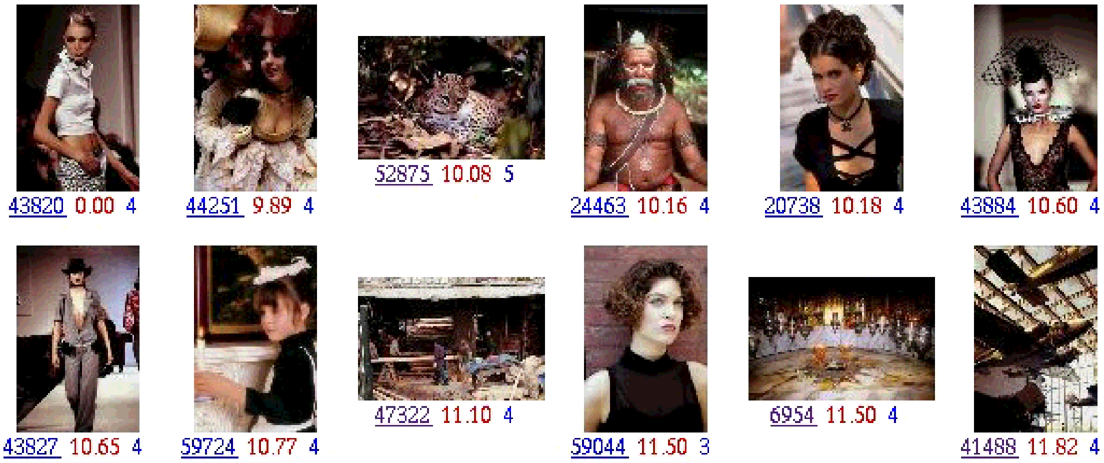


Figure 7. The query image is a portrait image.

Figure 8. The query image is a photo of flowers.

Another three query examples are provided in Figures 6, 7 and 8. The query images in Figures 6 and 7 are difficult to match because objects in the images are not distinctive from the background. Moreover, the color contrast for both images is small. It can be seen that the SIMPLIcity system achieves good retrieval. For the query in Figure 6, only the third matched image is not a picture of a person. A few images, the 1st, 4th, 7th, and 8th matches, depict a similar topic as well, probably about life in Africa. The query in Figure 8 also shows the advantages of SIMPLIcity. The system finds photos of similar flowers with different sizes and orientations. Only the 9th match does not have flowers in it.

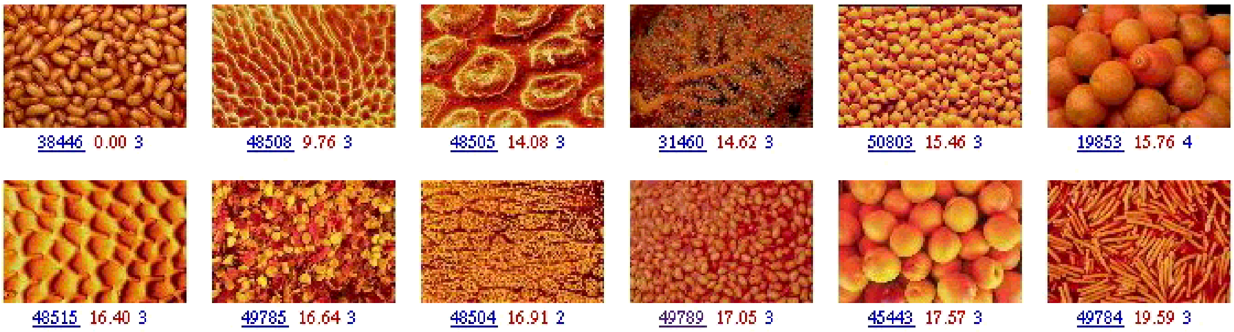
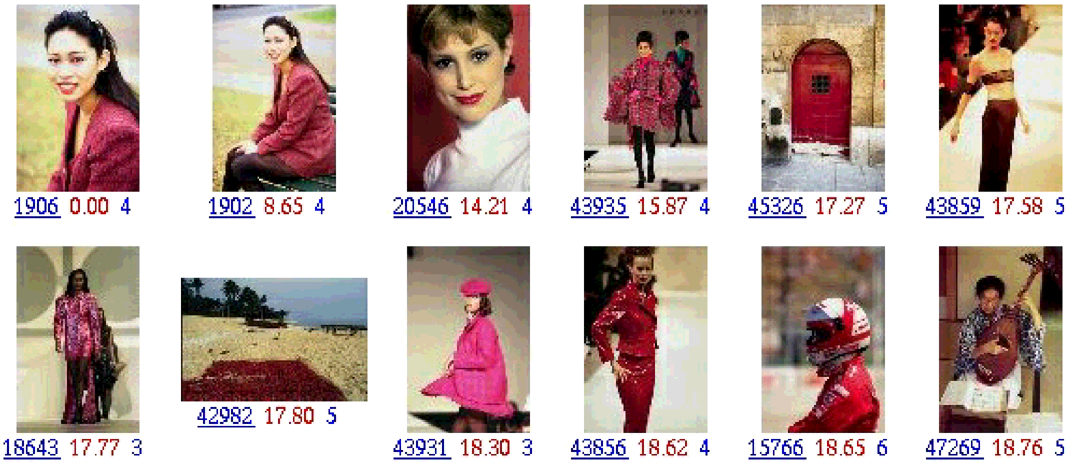


Figure 9. The query image is a textured image.

An example of textured image search is shown in Figure 9. The granular surface in the query image is matched accurately by the SIMPLIcity system.

5.2 Robustness to Scaling, Shifting, and Rotation

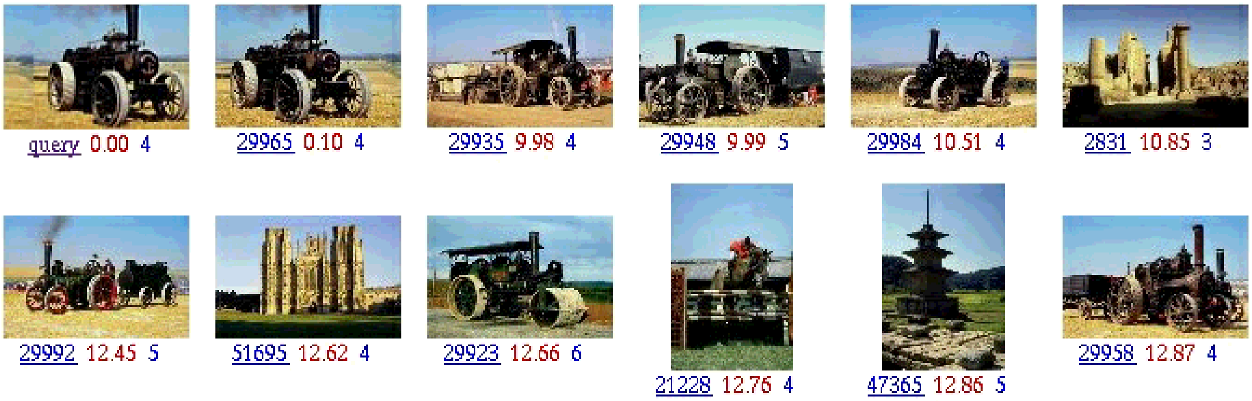


Test 1

Test 2

Figure 10. The robustness of the SIMPLIcity system to image cropping and scaling.

To show the robustness of the SIMPLIcity system to cropping and scaling, querying examples are provided in Figure 10. As we can see, one query image is a cropped and scaled version of the other. Using either of them as query, SIMPLIcity retrieves the other one as the top match. Retrieval results based on both of the queries are good.



Test 1