Image Searching Tool Using Category-Based Indexing

Aster Wardhani
School of Software Engineering and Data Communication,
Queensland University of Technology
Brisbane, QLD 4001
a.wardhani@qut.edu.au

Abstract

Searching for an object in a general image collection using current image retrieval systems, is still a problem. The retrieval results contain many unrelated images. In providing an effective and robust image database, objects in an image need to be extracted. Since the number of stored images can be very large, automation is an important aspect. Image indexing is a technique that extracts objects in an image automatically. The aim of this research is to propose a new object based indexing system based on extracting salient region representative from the image and categorising an image into different types.

Different image has different characteristics and often require different image processing techniques. Currently, most content based image retrieval (CBIR) systems operate on all images, without pre-sorting these images into different types. This resulted in limitations on retrieval performance and accuracy. Categories described here are of statistical and syntactical descriptions rather than semantical. By analysing which features are dominant in an image, two outcomes will be obtained: category for that image and salient object. Identifying salient object further reduce the retrieval results into relevant images.

1. Introduction

Tools available for searching for an image within a collections are still far from satisfactory. General images, such as ones found in the internet can be very complex. Currently, standard internet image searching tools such as *Google*, use image filenames as indexing attributes. This result shows that a search keyword “mango” can results in a large number of retrieved images, mostly do not contain the fruit mango. To index an image using its content, there are currently three general approaches: *object recognition*, *statistical analysis* and *image segmentation*. Object recognition techniques are limited to specific domains, e.g., images containing simple geometric objects. This approach has been used to retrieve images of tools and CAD geometric objects [13] and also medical images [10]. For most other types of images, such as images containing people, sceneries, etc, object recognition techniques are infeasible.

To pursuit the complexity of these images, researches then employed statistical indexing based on colour and texture [7], [3]. Images can be retrieved by specifying a combination of RGB colour values, textural measures, and more recently using other features such as shapes and spatial relations between regions in the image [4], [11], [12]. Colour is a low level feature, and by itself cannot adequately describe objects in images. To enable objects to be extracted and indexed within the image, image segmentation technique are used [9]. However, segmentation results of general images are noisy and contain too many regions. Thus, this approach are still limited to simple objects, and thus for general images currently do not provide a meaningful object based representation.

Techniques used by general CBIR systems are generic and aimed to handle all types of images. This is not optimal, since different images have different level of complexity and may require different features and analysis techniques. For example, shape retrieval is not suitable for images containing mostly textures or irregular shapes, such as landscape images. Currently, most content based image retrieval (CBIR) systems operate uniformly on all images, without pre-sorting these images into different types. This has resulted in limitations on retrieval performance and accuracy.

2. Proposed System

Rather than matching the whole image, it is more sensible to firstly *categorise* the image into different types. This is performed by finding the dominant characteristics of the image, such as how much texture, how complex the shapes are, and the presence of a dominant region. This strategy is supported by psychophysical evidence showing that hu-

Category Name	Feature Characteristics
Landscape	Colour = green and blue Spatial relation = vertical layer
People	Colour = human skin Shape = oval
Shape dominant	Number of regions = small Shapes = non complex Figure/background image = yes
Colour dominant	Number of regions = large Colour distribution = smooth
Texture dominant	Number of regions = large Colour distribution = non smooth
Structure dominant	Number of regions = large Shapes = complex

Table 1. Image Categories

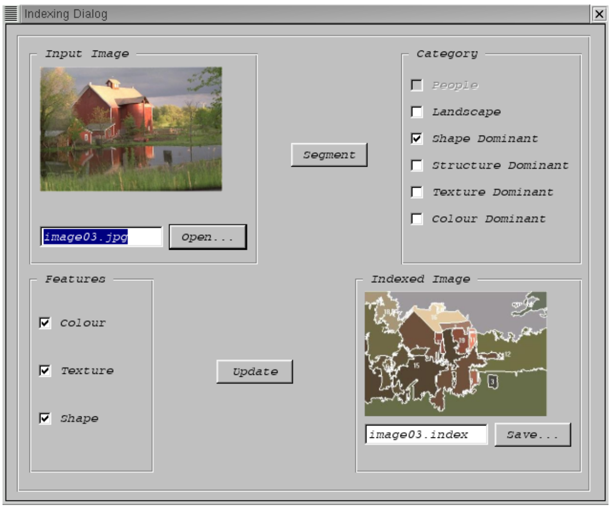


Figure 1. Category Indexing Tool

mans holistically classify visual stimuli before recognising the individual parts [6].

Based on the above intuition, the following approach to index images using categories is proposed. To provide image retrieval in the internet, this system can be implemented in two steps. Firstly, images are retrieved using filename such as using Google's image searching tool. Google search engine is used as an example as it index a large collections of images in the internet. Results from this search will then be classified into four different general and two semantic categories, shown in Table 1.

Typical indexing system using the proposed category is illustrated in Figure 1. The dialog box shows the features extracted from an image and the measured category for that image. The category is obtained automatically by analysing the composition of colour, texture and structure from the main regions. The regions are produced using the perception-based image segmentation system [14]. By implementing perceptual grouping, the results achieved are clean and only containing significant regions. Some segmentation results using this technique are illustrated in Figure 2 (different colour indicates different region).

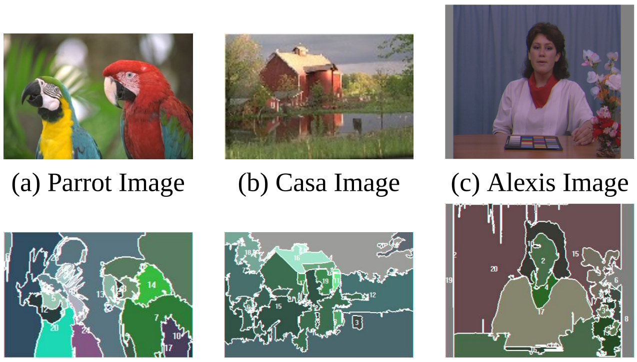
Using this image categorisation, a large retrieval results can be organised into groups that are based on the features of the image content. This would make navigation of results easier. The sheer number of uncorrelated retrieval results makes the searching task difficult and tedious. Using the category, a user could exploit the organisation of images into categories to locate images of interest. The meaning of each category will be explained in Section 3.4. For example, searching for the image "mango" would result in the following categories.

1. *shape dominant* (images likely containing a mango fruit)
2. *colour dominant* (images containing smooth areas, such as mango slices, some mango fruits)
3. *texture dominant* (images containing textural areas, such as mango trees)
4. *structure dominant* (other more complex images, containing structural and geometric regions)

If by "mango" the user interest means a picture of a single mango fruit, she/he would select the shape dominant category. However, if mango refers to a mango tree, the texture dominant category is more relevant. In many cases the categories may coincide with different semantic meanings of the search term. In searching for *people*, colour dominant (of skin colour) may be the most relevant category, whereas in searching for a *house*, structure dominant may be the most relevant.

The features described in Table 1 involves colour histogram, size and location of regions, number of regions and textural descriptions. Using this classification, the image shown in Figure 2(a), can be classified as a *Colour dominant* image, because of the appearance of large areas of smooth colour. Figure 2(b) will be classified as a *Landscape* image with the large tree areas. Figure 2(c) will be classified as a *People* image with the appearance of skin colour region.

Since images are segmented into a set of meaningful regions, retrieval results can be further pruned by performing object based query. In the example above, if the search for a "mango" aimed to retrieve all images that contain (a) mango fruit(s), the user initially is given six different groupings of retrieval results. To retrieve all images that contain a single or a collection of mango fruit(s), further object based query can be performed by firstly select the shape dominant group to choose images that have a single mango object. The user can then select the individual mango region from the image as the query object. The matching will then be performed to



(a) Parrot Image (b) Casa Image (c) Alexis Image

(d) Parrot Result (e) Casa Result (f) Alexis Result

Figure 2. Indexing Results using Perception-Based Segmentation Technique

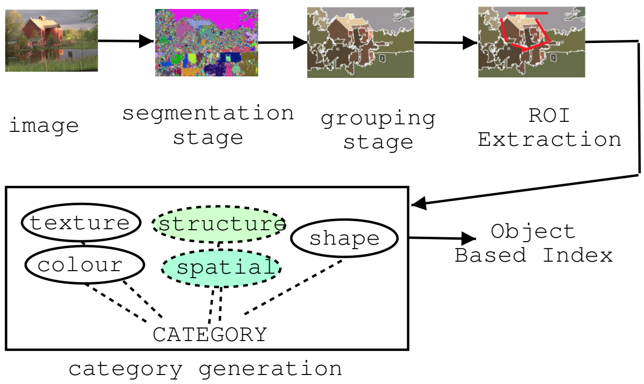


Figure 3. Proposed Image Indexing System

all images that have similar shape and colour combination of the selected region. This will result in retrieval images containing not only a single mango fruit but also images containing (a) mango(s) within other objects. Additionally, within each groupings, further classification can be made, creating a so called "parse-tree" query.

3. Methodology

Based on the above considerations, the proposed object based image indexing system is described in Figure 3. This system consists of the following stages: segmentation and grouping stage, dominant region extraction and category generation. An input image is firstly segmented. The segmentation results are then grouped together, using the technique proposed in [14]. After this process, both low level and high level features are extracted from each region and by analysing these features, a category and dominant regions will be obtained for that image. To determine which region is dominant, some heuristics will be generated. A relevance feedback such as used in [5], can be used interactively by users to change the chosen object and category within the selected retrieved images.

3.1 Image Segmentation

Although image segmentation techniques have been studied extensively, there are still some drawbacks and issues that need to be solved. One of the drawback is the quality of the segmentation results for natural images. There are various complexities in natural images that need consideration, such as variation of texture and lighting, and complex object shapes. Results from existing techniques have the following properties:

- Over segmented results, containing noisy regions at objects' boundaries and textural areas.
- The demarcation of regions do not always follow perceptual intuitions
- Results are very sensitive to threshold and requiring manual tuning

Selecting the best segmentation technique is an important issue. The difficulty in providing a meaningful segmentation is due to the non-correlation of existing colour distances with human perception of objects. Properties of natural images are too complex, requiring new perceptually intuitive metrics, which can adapt to intensity variations. Additionally, since each object in an image consists of a hierarchical composition, the segmentation process needs to be performed hierarchically. Some perceptual measures, therefore, should be added to improve the segmentation results.

To provide some perceptual foundations to the image segmentation implementation, the use of HVC-based region growing segmentation has been proposed, as reported in [15]. This segmentation approach uses adaptive threshold thus eliminates the need for manual tuning. This is performed by dividing an image into blocks, at each block, the presence of a strong edge indicates pixels requiring high threshold value. This value is then used as a threshold for all pixels in that block. To avoid missing any regions, the threshold is increased by a small amount to produce a slightly over-segmented results. It was perceived that it is better to have an over-segmented image than under-segmented results. Most of the noisy regions will be grouped later in the region grouping stage. Example of results from this segmentation technique is shown in Figure 4. This approach will be used in the proposed image indexing system.

(a) Parrot Image (b) Claire Image (c) Alexis Image

Figure 4. Segmentation Results

(a) Parrot Image (b) Casa Image (c) Alexis Image

Figure 5. Size Grouping Results

3.2 Region Grouping

Gestalt laws state that visual elements that belong to the same object, have the properties of *similarity, proximity, good continuation, closure, common fate, surrounded-ness and relative size, and symmetry* [1]. In the area of psychology, these principles have been accepted as the perceptual grouping laws. For this reason, Gestalt principles are used as the basis for the region grouping stage. The principles of proximity, similarity and good continuation can be used to group segments into regions. The principles of closure, common fate, symmetry and surrounded-ness, however, can be used for higher level grouping. This high level grouping is applied to find relationships between regions or objects' components. This was then translated into a grouping hierarchical formation. The issue would then how should the grouping operation should be performed? What are the grouping algorithm and rules required? To solve this issue, three different features were considered: *texture*, *colour* and *line continuation* [18]. The first grouping performed is the size grouping. The aim of this grouping is to merge noisy areas (region whose size is less than 100) using the similarity of region HVC means. Example of results from this grouping stage is shown in Figure 5.

The next grouping stage is the colour histogram grouping. This is performed by comparing the similarity of two region colour histograms. This is then followed by line continuation grouping. Regions are grouped based on comparing line continuation surrounding regions. Examples of final grouping results is shown in Figure 2. At each grouping step, the number of regions are reduced at each stage of grouping, finally leaving only significant components. These results are more meaningful and provide more data abstraction than the standard image segmentation techniques.

3.3 Dominant Region Extraction

It is difficult to know which object is of interest. An image can contain background objects which may look the same from the computer's point of view with the regions that belong to main objects. There might also be multiple objects. However, generally, there is a process of selection of important from less important objects in an image in human perception. Often the selection is not always based on semantic reasonings. Syntactical relations between regions, such as difference in size and texture can create a point of interests. A small house in the forest as depicted in Figure 3 is an example of region of interest. In Gestalt principles, this is described as figure-background principle (smaller figure against bigger surrounding objects).

The aim of dominant region extraction is to eliminate background, non-important regions, producing the most essential region. The reduction of non useful regions are required to reduced the matching and expensive structural analysis. The background will be eliminated by applying the figure/background principle of Gestalt laws. By analysing that the largest region surrounding other objects entirely can conclude that this region is the background thus eliminated. Similar idea was also used in the form of Region of Interests (ROI), used in a CBIR system proposed by Moghaddam et. al. in [8]. In this system, however, the regions are extracted manually by users.

In the proposed system, dominant region will be extracted automatically by analysing the size (largest), location (center) of the regions and appearance of interesting shapes and structures. Interesting shapes will be judged on the region shape regularity and its geometric properties.

3.4 Category Generation

Category generation is then responsible to assign an image type. Image type is aimed to provide sufficient groupings of images with similar characteristics. The issue would then be: "What are the succinct category that can capture different image characteristics?"

To categorise images without performing object recognition, the classification shown in Table 1 is used. It is based on the strength of different features that can be exploited from different image type. For example, images with texture dominant can be handled more effectively with a robust texture matching, whereas images under shape dominant can concentrate on good shape matching.

Shape dominant category is for images containing a small number of regions. These regions also have simple and regular shapes. Images categorised under colour dominant contains large number of smooth (non-textured) regions with less regular shapes. Texture dominant images are images that contain a large number of textural regions. There are many ways in detecting and describing texture. Using results from image segmentation, textural areas can be defined as neighboring regions whose size are small and spurious. These regions were marked as texture map and have been used successfully in segmenting texture for general images as reported in [16]. By measuring how many regions from segmentation are spurious regions, we can estimate whether the image is texture dominant. Using this texture map, texture dominant images can be classified as images containing mostly textural regions.

Images used here is assumed to be general whose semantic are unknown a priori. However, out of these images, many of them contain images which have distinct features, whereby from these features, semantic meaning can be derived. Such example of images are landscape images and images containing people. Landscape images usually contain large regions of sky (certain blue colour) and grass or trees (certain green colour). Using the colour combination, it has been proven successful in retrieving landscape images. Images containing people can be indexed and retrieved by analysing the presence of skin colour in the image. Since the ultimate goal is to perform image classification, using both facts another categories listed in Table 1 are added under classification of people and landscape. It would be possible to add further classification to domain specific images, such as botanical data, museum artefact, etc, whose retrieval has been demonstrated in systems such as [10].

Each of the different category is derived based on the analysis of the features extracted, described above. To allow such classification, each dominant region extracted is analysed for its various features. Both statistical and structural features will be used. This will include colour, texture and shape. For each shape some measures will be generated such as number of corners, degree of curvedness, etc, to distinguish between regular and complex shape.

Another new contribution in this project is the use of object structure [17]. Researches in psychology stated that classification of a scene may remain valid as long as the relative relationships between the image regions remain the same [2]. In the category of *structure dominant*, the existence of certain “interesting” or “prototype” structure will be used to represent an image and used for matching. Rather than matching the whole image or even the whole object (since same object can appear differently in different images), regions and their relations can be used instead. An example of a relational tree extracted is shown in Figure 6.

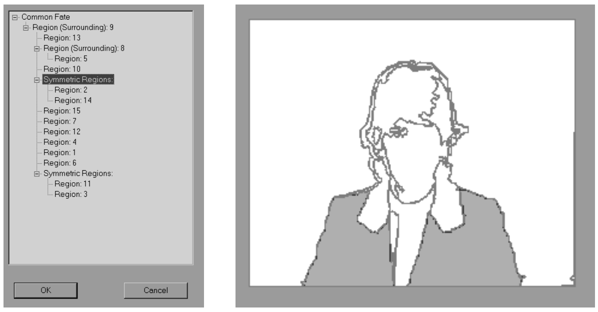


Figure 6. Image structure describes using relational tree

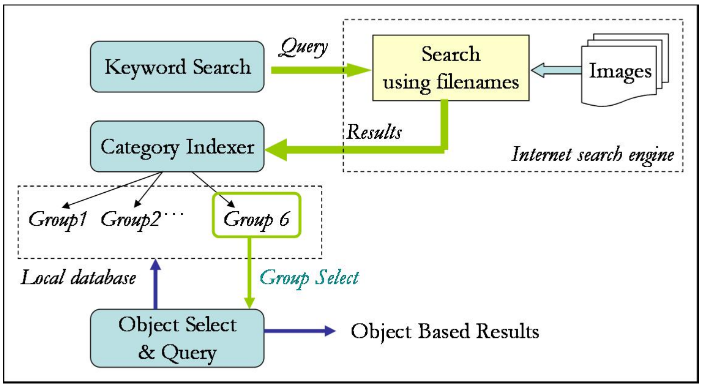


Figure 7. Retrieval System Implementation

3.5 Query Interface

There are many types of queries in CBIR, such as using keyword (semantic), or graphical descriptions (query by example, sketch, etc.). In this system, both queries are combined. Figure 7 shows the design prototype. The system will be implemented to retrieve images in internet. The search starts with query by keyword. The system will then sends this query to standard internet search engine such as *Google* image search tool to retrieve all images based on *image filenames*. The system will download all the links as well as all the thumbnails to be analysed. The analysis will group the retrieval results into 6 different categories described previously. Users can then prune the searching, by navigating the classification and selecting an object from an image. Since each region is segmented in the image, *object based query* can then be performed, reducing further to images that relates to the both *semantic keyword* and *graphical descriptions*.

Currently, segmentation and feature extraction stages have been performed and the overall retrieval system is currently being developed.

4. Conclusions

In order to retrieve images from large collections, a robust object based CBIR is crucial. This research aims to develop an image retrieval system that is based on extracting the dominant figure / region in the image, which subsequently placing an image into one of the proposed generic categories. The indexing information consists of not only low level but high level features. Images will be classified into a set of types. An indexing template will be generated automatically for each type, based on visual observation of which combination of features occurs for that type of image. Such template will be used to match images against the query information. We need to investigate the suitable and succinct set of features for each type of template.

This research will provide a new image retrieval system that provides users with the ability to further classify the content of an image. The impact from this method is more accurate retrieval results. An image will be represented by rich descriptions that relate directly to the content of the image.

References

[1] V. Bruce and P. Green. *Visual Perception: Physiology, Psychology and Ecology*. Lawrence Erlbaum Assoc. Hove and London, 2nd ed edition, 1990.

[2] C. Cave and S. Kosslyn. The role of parts and spatial relations in object identification. *Perception*, 22:229–248, 1993.

[3] T. S. Chua, K.-L. Tan, and B. C. Ooi. Fast signature-based color-spatial image retrieval. In *Proceedings of IEEE conference on multimedia computing and systems*, 1997.

[4] M. Flickner, H. Sawhney, H. Niblack, J. Ashley, Q. Huang, 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, No.9:23–31, 1995.

[5] F. Jing, B. Zhang, F. Lin, W.-Y. Ma, and H. jiang Zhang. A novel region-based image retrieval method using relevance feedback. In *ACM Workshop on Multimedia*, pages 28–31, 2001.

[6] P. Lipson, E. Grimson, and P. Sinha. Configuration based scene classification and image indexing. In *Proceedings of IEEE conference on computer vision and pattern recognition*, 1997.

[7] H. Lu, B. Ooi, and K. Tan. Efficient image retrieval by color contents. In *Proceedings of 1994 international conference on applications of databases*, 1994.

[8] B. Moghaddam, H. Biermann, and D. Margaritis. Region-of-interest and spatial layout for content-based image retrieval. Technical Report TR-2000-35, MERL - mitsubishi electric research laboratory, November 2000.

[9] A. Mohan, C. Papageorgiou, and T. Poggio. Example-based object detection in images by components. *IEEE transactions on pattern analysis and machine intelligence*, 23(4):349–361, April 2001.

[10] A. Mojsilovic and J. Gomes. Semantic based categorization, browsing and retrieval in medical image databases. In *Proceedings International Conference Image Processing*, 2002.

[11] S. Sclaroff, L. Taycher, and M. La Cascia. Imagerover: A content based browser for the world wide web. In *IEEE Workshop on Content-based Access of Image and Video Libraries*, pages 2–9, 1997.

[12] J. Smith and S. Chang. Visualseek: a fully automated content-based image query system. *ACM Multimedia '96*, 1996.

[13] G. Srinivas, E. Fasse, and M. Marefat. Retrieval of similarly shaped parts from a cad database. *Systems, Man, and Cybernetics*, 3, 1998.

[14] A. Wardhani. *Application of psychological principal to automatic Object identification for CBIR*. PhD thesis, Information technology, Griffith University, 2001.

[15] A. Wardhani and R. Gonzalez. Automatic object extraction for content based retrieval. In *Proc. of International Conference on Telecommunication*, volume 2, pages 1005–10, 1997.

[16] A. Wardhani and R. Gonzalez. Using high level information for region grouping. In *Proc. of the IEEE Region 10 Conference (Tencon'97): Speech and Image Technologies for Computing and Telecommunications*, pages 339–342, 1997.

[17] A. Wardhani and R. Gonzalez. Automatic image structure analysis. In *Proc. of the IEEE International Conference of Multimedia Systems (ICMCS'98)*, pages 180–188, 1998.

[18] A. Wardhani and R. Gonzalez. Automatic object identification for image indexing. In *Proc. of the International Wireless Telecommunications Symposium (IWTS'98)*, pages 330–334, 1998.