Content Based Image Retrieval Using Category-Based Indexing

Aster Wardhani and Tod Thomson
Faculty of Information Technology, Queensland University of Technology, Australia.
a.wardhani@qut.edu.au, t.thomson@student.qut.edu.au

Abstract

Currently, most content based image retrieval (CBIR) systems operate on all images, without sorting images into different types or categories. Different images have different characteristics and thus often require different analysis techniques and query types. Additionally, placing an image into a category can help the user to navigate retrieval results more effectively.

To categorise an image, firstly the dominant region needs to be extracted using multi level colour segmentation. Based on the regions' features of colour, texture, shape and relation between regions, the image is then categorised. Users are presented with retrieval results sorted into different categories, where dominant region extraction will allow for object based retrieval to be performed.

1. Introduction

Tools available for searching for an image within an arbitrary image collection, such as in the Internet, are still far from satisfactory. This is because the range of images is wide and the content of the images is complex. Most well-known Internet image searching tools (e.g. Google Image Search - http://images.google.com) use image filename as the primary means of indexing image attributes. This type of image indexing inevitably fails as it is based on the flawed assumption that image content is always reflected correctly by the image filename.

Current CBIR systems typically aim to handle an arbitrary collection of images using the same analysis tool. This is not optimal, as different images have different complexity levels and may require different feature analysis techniques. For example, shape retrieval is not suitable for images containing mostly textures or irregular shapes, such as landscape images. Similarly, query by example is most suitable to images containing single object, where accurate shape analysis and segmentation is required. Currently, most CBIR systems present results without grouping them into categories.

2. Image Category

The proposed CBIR system, rather than matching within the whole image collection, more sensibly first *partitions* the image collection into different categories. This categorisation 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 humans holistically classify visual stimuli before recognising the individual parts [1].

Based on the above intuitions, the following general categories are proposed (Table 1).

- *Semantic*: natural scene, people and geometric object
- *Syntactic*: single, multiple objects
- *Statistic*: smooth, textural

Category Name	Features
Natural Scene	Green & blue spatial relation
People	Human skin hue
Geometric objects	Man made objects
Single object	Figure/ground (F/B)
Multiple objects	Non F/B
Mainly smooth	Smooth colour
Mainly textural	Large variance

Table 1. Image categories

The proposed categories were chosen to provide sufficient grouping of images with similar themes, using general features. The domain assumed is photographic images, but with an unlimited range of themes nature, people, buildings, etc.). Similar work on retrieval of landscape images was performed in [2]. The semantic categories are provided to detect most common types of images, such as those which exist on the Internet. The *single object* category is for images containing only two regions, conforming to the Gestalt figure/ground principle. Historically, the figure/ground separation has

0-7803-8603-5/04/$20.00 ©2004 IEEE

been seen as an important step in object recognition. In CBIR application, this can be used to perform object based queries. The *smooth* and *textural* categories are provided to differentiate the statistical features of the object.

Each image is categorised automatically by analysing the number of regions produced, their colour, texture, shape, location and their relations. Partitioning image results into the proposed categories allows large sets of retrieval results to be organised into groups based on the features of the image's content, making navigation of results faster and easier for the user.

3. Colour Segmentation

Although many image segmentation techniques have been developed and the standard descriptors for visual content have been proposed in MPEG-7, image segmentation is still a challenge. One major drawback of all current segmentation techniques is that they do not produce consistently high quality segmentation results for natural images. Results from existing techniques have the following properties:

- They produce over-segmented results which contain noisy regions at object boundaries and textural areas.
- Demarcation of regions does not always follow perceptual intuition.
- Results are sensitive to thresholds and require manual tuning.

Additionally, for real-time application, many segmentation techniques are computationally expensive. For the purpose of retrieval, the aim of the segmentation stage is to produce a small number of *useable* segments that approximate the users perception of the dominant regions in an image. The second consideration is minimal processing. Thus accurate boundary and details of small objects are not considered.

A relatively simple, effective and fast algorithm for dominant region extraction is the colour segmentation algorithm proposed by Comaniciu and Meer [3]. This technique is based on the mean shift algorithm, "a simple non-parametric procedure for estimating density gradients". The mean shift vector is the vector difference between the local mean and the centre of the window. In [3], it was proven that the mean shift is proportional to the gradient if the probability density of the feature vectors. The mean shift algorithm can be described as follows:

1. Choose a radius r for a search window for initial feature vector density estimation
2. Choose initial location of the window
3. Compute the mean shift vector (MSV)

$$MSV = \frac{r^2}{p+2}\frac{\nabla p(x)}{p(x)} \quad (1)$$

where x is the p-dimensional feature vectors, $p(x)$ is the probability density function of x and $\nabla p(x)$ is the gradient of $p(x)$.

4. Translate the search window by the shift amount
5. Repeat till convergence

An analysis of feature space is performed to detect significant features (regions). Segments in the image correspond to high density regions in the feature space, with the level of segmentation based on the thresholds specified. Three general classes of segmentation resolution are described using this segmentation technique: Under-segmentation, Over-segmentation and Quantization [3]. The over-segmentation predefined threshold is used for segmentation as the algorithm is applied here. Using this setting, experiments show that mean shift algorithm performs better compared to standard segmentation technique such as K-Means, in the sense that we do not need to specify the number of regions produced. Additionally, the formation of regions seems to follow perceptual intuition. Comparisons of segmentation results using this technique is shown below:

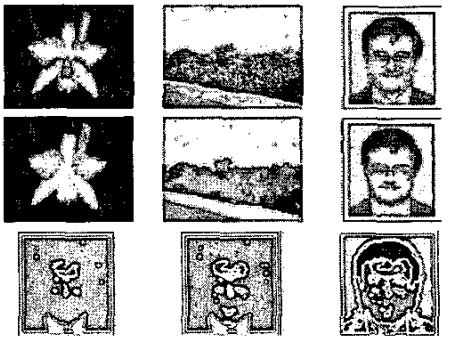


Figure 1. Segmentation comparison: Row 1 / Original, Row 2 / SEGM, Row 3 / K-Means

However, the mean shift algorithm by itself still producing many regions which need further pruning in order to obtain the dominant region. Thus, in addition to this, multi level processing is used to eliminate the need for manual tuning and obtain clean results with a small number of segments. In the past, clean segmentation has been achieved using Gestalt region grouping proposed in [4].

A survey of approximately 100 random images was performed where a number of different image sizes was

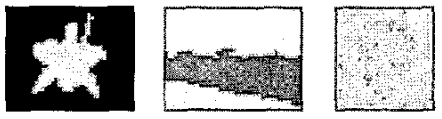
tested. Image thumbnails from Internet were used in this experiment. This allowed processing to be conducted in real-time. The smaller side of the thumbnails was scaled to 16, 32, 48, 64 and 80 pixels. An example of the resulting images from this experiment is given in Figure 2 with the image size and number of segments produced by segmentation (N) are shown.

141x106: orig. 43x32: N = 6 85x64: N = 35

Figure 2. Multi level comparisons

In this experiment 32 pixels was determined to be the size in which the resulting image was most likely to contain a sufficient number of segments (about two to six segments). In most cases 16 pixels caused the segmented image to contain only one segment, and that size is therefore excluded automatically by the system. It can also be seen that as the size of the image increases over 32 pixels the number of segments increases quite rapidly. Thus, the increase in image size be at a rate of eight pixels per resize, not 16 as is the case in this experiment.

An analysis of how segmentation performed at different image sizes, based on the number of segments produced, is performed by the system automatically, for each image. This image resizing begins at the smallest defined resolution, where the length of the smaller side of the image is resized to 32 pixels. The image is continually segmented and the size of the image increased until the result produced has the required number of segments (2 to 6 regions). Additionally, small regions (size less than 100) are merged with larger ones. This application of image segmentation produces results that contain a small number of useful segments that are significant to user perception.



(a)N = 2 (b) N = 5 (c) N = 2

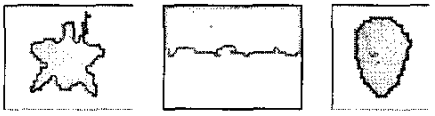
Figure 3. Final segmentation results

4 Dominant Region

One of the most important characteristics of human colour perception is that human eyes cannot simultaneously perceive a large number of regions [5]. Moreover, the number of colours that can be internally represented and identified in cognitive space is about 30 [6]. Based on these findings, the dominant region method is proposed The image segmentation results are used to obtain dominant region based on the assumption that it is not necessary to obtain a complete understanding of a given image. The aim of dominant region extraction is to eliminate background, non-important regions, producing the most prominent region (point of interest). The removal of non-important regions reduces the amount of computationally-expensive segment matching required.

By identifying some key segments such as"trees", "sky", "face" (using skin hue colour), presence of a background region (for identifying F/G image), number of regions, region size and its statistical properties; the presence of each feature will add to the weight for each category in the image. To identify the standard colour for these features, a survey of the average colour found in the dominant segments from sample images was performed. The result of this experiment was that for people images the average hue of dominant region was 15 (out of 255), with tolerance of five percent. For landscape images the average colours were (153,182,224) for sky, (91,110,73) for land and (113,158,194) for sea, in RGB colour, with a tolerance of ten percent. While this method is simple, it currently serves the purpose of this experiment. This can be improved using code-book based methods such as in [7].

Background regions are eliminated by applying the Gestalt figure/ground principle [4]. This is performed by determining the largest region surrounding other objects entirely. After background elimination has been performed, the dominant region will be extracted automatically by analysing the size and location of the remaining regions. If the remainder of this operation is one region, this image can be classified as F/B image. Similar F/B categorisation was performed in [7].



(a) Orchid (b) Scenery (c) Face

Figure 4. Dominant segment

The presence of "trees" or "sky" increases the weight for *natural scene* category. Similarly, regions with skin colour will add weight for the *people* category. All features are combined and ranked. An image can contain more than one category and this can be exploited in the query system example shown in the results in Figure 6.

In addition to providing a rich and compact representation, the categories provide means to capture the general theme of the image. Based on this system, the images in Figure 3 are assigned to the following categories: 'orchid' is *figure-ground*, 'scenery' is *natural scene* and 'face' is *people*.

5 Results

This system is currently in trial on the QIST system (Queensland University of Technology Image Searching Tool), shown in figure 5.

The snapshot shows some results using the search keyword canoe. The first three retrieval results shown are categorised as landscape (natural scene). However, selecting an individual segment will allow further searching on the canoe object to be performed. Other examples of results are shown in Figure 6. Images are retrieved from the Internet via keyword search. The thumbnails produced are then analysed and classified.

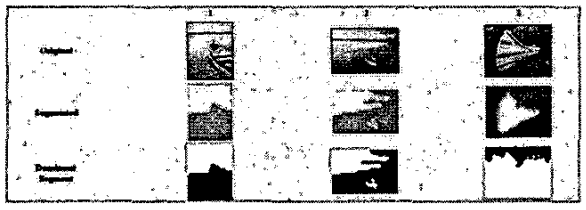


Figure 5. Image retrieval prototype

It is difficult to compare the performance of the QIST system to other CBIR image indexing systems, as QIST retrieves images from the Internet for each search rather than using an image database. However, a time measurement of system performance has been conducted. Multi-resolution colour image segmentation averages 3.281 seconds per image. Dominant region extraction averages 0.575 seconds per image. These timing results where measured for 200 images on a computer with a 450Mhz Pentium III CPU. The initial image size is about 128 pixels width or height. Since thumbnail images are used this initial size varies. The resulting image size is an average of 28 pixels for the image's smaller side. All timing does not take into account any image upload/download time.

6 Conclusion and Future Work

In order to retrieve images from large collections, robust object based CBIR is crucial. This research aims to develop an image retrieval system that extracts the dominant region in an image, placing the image into one or more categories. With dominant region and pre-categorised images, not only objects are extracted,

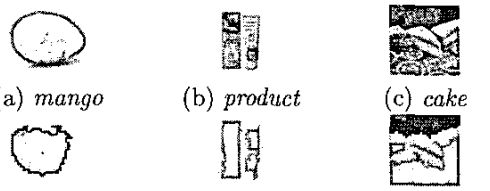


Figure 6. Results with different categories: F/B-figure/ground, SM-smooth, GEO-geometric, MO-multiple object, TEX-textural

but also composition and semantic content for the image. Currently, some results such as people classifier are inaccurate. Many images of people encountered have poor colour quality and noisy. Thus, more robust classifier is required. However, the idea of categorisation has been fulfilled in this paper. Future work includes:

1. Investigate better weighting, priority and hierarchical system within the proposed categories.
2. Investigate a better classifier for different categories.
3. Implement object based searching and queries. With clean and meaningful segmentation produced, this idea can now be realised. Query by example can also be incorporated.

References

[1] 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.

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

[3] D. Comaniciu and P. Meer, "Robust analysis of feature spaces: Color image segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico*, June 1997, pp. 750–755.

[4] A.W. Wardhani, *Application of psychological principal to automatic Object identification for CBIR*, Ph.D. thesis, Information technology, Griffith University, 2001.

[5] A. Mojsilovic "Matching and retrieval based on the vocabulary and grammar of color patterns," *IEEE Trans. of Image Processing)*, vol.1, pp. 38-54, 2000.

[6] J. Derefeldt and T. Swartling "Color concept retrieval by free color naming,", Displays, vol. 16, pp. 67-77, 1995.

[7] B. Leibe and B. Schiele, "Interleaved Object Categorization and Segmentation," *British Machine Vision Conference (BMVC'03)*, 2003.