THE IMAGE SEARCHING TOOL USING CATEGORY-BASED INDEXING

Tod Nicholas Thomson & Aster Wisnu Wardhani

The Centre for Information Technology Innovation,
Queensland University of Technology.
t.thomson@student.qut.edu.au, a.wardhani@qut.edu.au

ABSTRACT

The aim of this research is to propose a new content based image retrieval (CBIR) system using *categories*. Different images have different characteristics and thus often require different image processing techniques. Most current CBIR systems operate on all images, without pre-sorting images into different categories. This results in limitations on retrieval performance and accuracy. Two *semantic* and four *syntactic* image categories are proposed. The category for an image is generated automatically by analysing the image for the presence of a *dominant object* or for correspondence to an *image 'template'*. Dominant objects are obtained by performing region grouping of segmented thumbnails. The result of this research is a *new* Internet image retrieval and indexing system.

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.

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. Currently, most CBIR systems operate uniformly on all images, without pre-sorting the image collection into different categories. This uniform operation has resulted in limitations on retrieval performance and accuracy.

2. PROPOSED SYSTEM

The proposed CBIR system, rather than matching within the whole image collection, 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 the image has, how complex the shapes are, 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 psychophysical intuitions, the following two step approach for indexing images from the Internet (using categories) is proposed. Firstly, images are retrieved via a filename search, similar to Google Image Search. Secondly, results from the search will then be classified into two semantic categories and four general categories, based on the descriptors given in Table 1.

Category Name	**Feature Characteristics**
Landscape	Colours green & blue Spatial relation in vertical layers
People	Human skin hue
Shape dominant	2 regions foreground/background image Shapes are non-complex
Colour dominant	More than 2 regions Colour distribution smooth (small variance)
Texture dominant	More than 2 regions Colour distribution non-smooth (large variance)
Structure dominant	More than 2 regions Shapes complex Contains geometric objects

Table 1. Image categories

The category for an image is obtained automatically by analysing the composition of colour, texture and structure in the main regions of the image. An example image for each proposed category is shown in Figure 1.

Fig. 1. Image categories (examples)

The regions can be produced using any region-based image segmentation technique (e.g. region growing, split and merge, morphological watersheds). By implementing perceptual grouping [2], the results achieved are clean and only contain significant regions. For the categories described in Table 1 the system considers colour histogram, region size, location of regions, number of regions and textural description to automatically determine the category for the image.

Partitioning image results into the proposed six categories allows large sets of retrieval results to be organised into groups based on the features of the image's content. This grouping makes navigating the results easier for the user. This grouping is done to combat the the sheer number of uncorrelated retrieval results which make searching difficult and tedious in current CBIR systems. The organisation of images into categories could also be used to locate images of interest more quickly. For example, searching for images using the keyword "mango" would result in images being presented to the user in the following way:

- *Shape Dominant* - Images containing a prominent shape (a whole mango fruit).
- *Colour Dominant* - Images containing large smooth areas of similar colour (mango slices, mango fruit pieces).
- *Texture Dominant* - Images containing textural areas (mango trees).
- *Structure Dominant* - Other more complex images, containing structural and geometric regions (not likely for mangoes).

In many cases images from different categories will coincide with different semantic meanings of the search term. The user is provided with some example images from each category for the search term chosen. These example images are used to help the user accurately determine which of the six categories the images they are searching for most likely reside in. Segmentation of the image examples also enables the user to select individual shapes (objects) for object-based queries.

3. METHODOLOGY

Based on the above considerations, the proposed object based image indexing system is designed as shown in figure 2, consisting of the following three stages.

1. The input thumbnail (obtained from an image database such as Google Image Search) is resized to the smallest of the predetermined image sizes and segmented, with the process continuing until a segmented image with the *'appropriate'* number of regions is obtained.
2. The segmentation results are grouped together using the technique proposed in [2]. The low level and high level features of each region in the image are analysed. The dominant region is determined, and descriptors for this region are recorded.
3. The image is categorised into one of six categories. Image category and descriptors about the dominant region are indexed for image retrieval.

The use of thumbnails rather than the original image increases the speed of the system substantially.

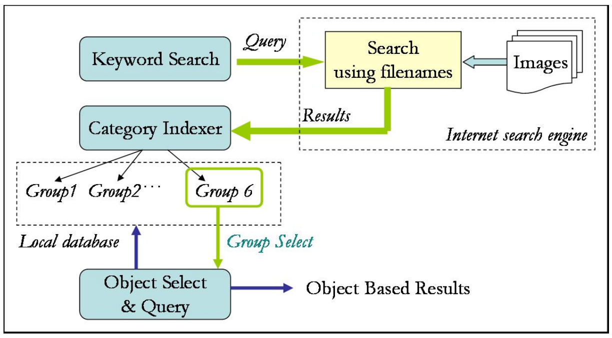


Fig. 2. Design of the system

3.1. Image Segmentation & Size Reduction

Although many image segmentation techniques have been developed and the topic of optimal segmentation algorithms has been studied extensively, there are still many issues related to image segmentation that need to be resolved. 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:

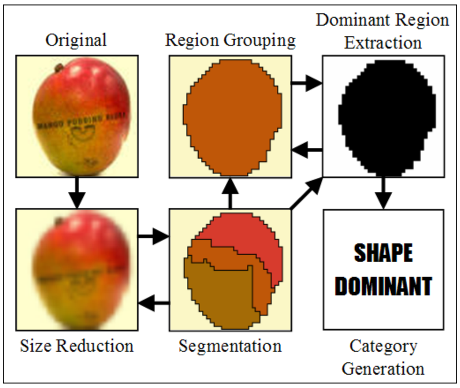


Fig. 3. Image processing stages

Orchid	Scenery	Face
132x106	105x79	92x98

Fig. 4. The original test images

- 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.

Selecting the best segmentation technique is an important issue. In this system a new segmentation algorithm is not suggested. Image segmentation is provided via the implementation of a known *good* algorithm which is applied intuitively based on some developed heuristics. The aim of this is to produce segmentation results that contain a small number of *useable* segments that coincide more closely with the users perception of the dominant segments in an image. In addition to this we use image resizing and segment testing to eliminate the need for manual tuning. To avoid missing any regions, the threshold is increased by image resizing small amounts to produce a slightly over-segmented result. Though the number of segments produced by the segmentation technique is small, it is perceived that it is better to have slightly over-segmented rather than under-segmented results. The problem of segmentation noise will be solved at the region grouping stage. A description of the segmentation algorithm chosen, and how it is applied in the system is given following.

The segmentation technique used is SEGM [3]. This technique is based on the mean shift algorithm, "a simple non-parametric procedure for estimating density gradients". 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 thresholding specified. The correct level of thresholding producing an image with the required level of segmentation. 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.

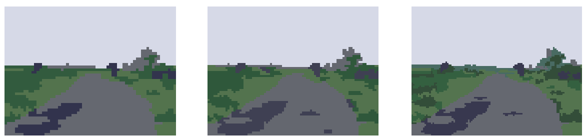
Initial segmentation results (without adaptive thresholding via image resizing) are shown in Figure 5. The number of segments produced in each case is shown. Even though the input images are only thumbnails, the number of segments produced (n) is too high to be useful for identifying a focus (dominant feature) in the image. Hence image resizing is required before more useful regions (regions for indexing the image) can be identified.

n = 152	n = 73	n = 42

Fig. 5. Original thumbnail segmentation (SEGM)

A survey of approximately 100 random images was performed where a number of different image sizes was tested. These image sizes were where the smaller side of the image was scaled to 16, 32, 48, 64 and 80 pixels. An example of the resulting images from this experiment is shown in Figure 6 with the number of segments produced by segmentation (n) shown.

141x106	21x16	43x32
Original	n = 2	n = 6

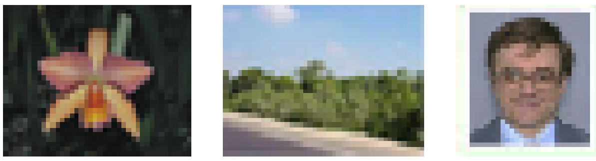


64x48	85x64	80x24
n = 24	n = 35	n = 56

Fig. 6. Resize experiment (example)

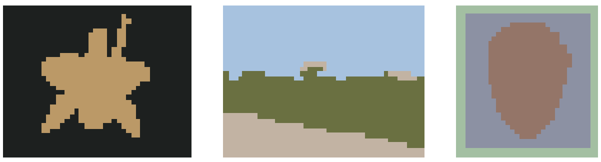
In this experiment 32 pixels was determined to be the size in which the resulting image was most likely to contain a sufficient number of segments (about two to six segments). In most cases 16 pixels caused the segmented image to contain only one segment, and that size is therefore excluded automatically 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 it is suggested that the increase in image size be at a rate of eight pixels per resize, not 16 as is the case in this experiment. An example of images resized to 32 pixels and segmented using SEGM, including the number of segments produced, is shown in Figures 7 and 8.

The image resizing begins at size 32 pixels (width or height, whichever is smaller). This starting point is based on the experiment which resulted in an indication that 32 pixels gave the best segmentation result. If the number of segments produced (n) is too large (n ¿ 6), the image size is reduced. However, if the number of segments produced is too small (n ¡ 2) the image size is increased. In many cases the image does not require further resizing from the initial size of 32. A number of segments between two and six is sufficient for the purpose of extracting the *dominant* region. In this case, the details of small regions are not considered important, and thus a small number of regions is all that is required. This application of image segmentation produces results that contain a small number of useful segments, whilst producing segments that are significant to user perception. The pseudocode for the segmentation and resizing performed by the system is shown in Figure 9.



39 x 32 43 x 32 32 x 34

Fig. 7. Resize images 32 pixels (maintain ratio)



n = 2 n = 5 n = 3

Fig. 8. Segmentation of 32 pixel images (SEGM)

3.2. Region Grouping & Dominant Region Extraction

In the area of psychology, Gestalt principles have been accepted as the perceptual grouping laws and thus are used as the basis of the region grouping algorithm [4]. Gestalt principles are applied to address the issue of how region grouping should be performed, specifying what region grouping rules are required. From region grouping the number of regions is reduced, leaving only significant regions. These significant regions are more useful for image indexing and category generation than are the regions produced by image segmentation alone. In the region grouping algorithm presented here, size, colour and line continuation grouping methods are considered.

```
n = 0
image size = 32 // 16..24..32..40..48..56..64
WHILE(n < 2 || n > 6)
{
  IF(n < 2)
    increase image size
  ELSE IF(n > 6)
    decrease image size
  ELSE
    BREAK
}
```

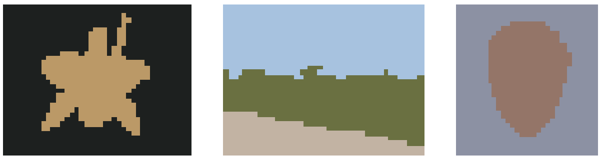
Fig. 9. Segmentation & image resizing pseudocode

The first grouping performed is size grouping. The aim of this grouping stage is to merge noisy segments. This size grouping is performed in conjunction with the next grouping stage, colour histogram grouping. This is performed by comparing the similarity of two region's colour histograms. A combination of these two types of regions grouping is initially considered by the system in removing segmentation noise from the system. Each segment with a size less than 100 pixels is merged with the neighbouring segment with the most similar colour. The pseudocode for merging segments based on size and colour is given as follows):

```
WHILE(there are more segments in the image)
  IF( !(current segment >= 100 pixels) )
  {
    WHILE(there are more neighbours)
    {
      test neighbours colour
      IF(colour closest found so far)
        save segment number of closest colour segment
    }
    merge current segment with closest colour segment
  }
DONE
```

Fig. 10. Segment merging (size/colour) pseudocode

The last grouping performed is line continuation grouping. Regions are grouped based on comparing line continuation into surrounding regions. Figure 11 shows the number of segments remaining (n) after the three types of region grouping have been performed.



$n = 2$ $n = 3$ $n = 2$

Fig. 11. After segment grouping

After region grouping has been performed the dominant region in the image must be determined (and any possible further grouping based on dominant region performed). The aim of dominant region extraction is to eliminate background, non-important regions, producing the most prominent region (region of interest). The removal of non-important regions reduces the amount of computationally-expensive segment matching required. Background regions are eliminated by applying the Gestalt figure/background principle [2]. This is performed by determining the largest region surrounding other objects entirely. It can be concluded that this region is the background and thus it can be eliminated. After the background elimination has been performed, the dominant region will be extracted automatically by analysing the size and location of the remaining regions. The largest and most centred region is usually the one chosen as the dominant region in the image.

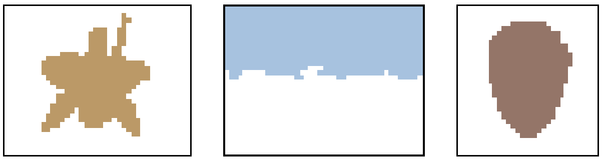


Fig. 12. Dominant segment (with average colour)

A five-step process is performed in the segment merging and dominant region extraction stage:

1. The *image map* from the output of the image segmentation stage is loaded. The image map is a two dimensional array, with the same proportions as the image. It contains the value representing the segment each pixel belongs to, at the corresponding point in the array.

2. *Segment information is calculated* from the image map. This includes information about the average colour of the segment and its size. This information is required before any segment merging can begin.

3. A *list of neighbouring segments* is recorded for each segment in the image. This list is used to determine the possible segments that the segment can be merged with.

4. *Size merging* is performed (Figure 10).

5. The *dominant region* is extracted from the image (as above).

3.3. Category Generation

Category generation is responsible for assigning the image to one of the six aforementioned categories, with the category aiming to provide sufficient grouping of images with similar characteristics. The question is then, “What are the succinct number of categories that can capture the range of image characteristics?” To categorise images without performing object recognition, the descriptors shown in Table 1 are used. These descriptors are based on a number of features that can be easily detected in images, but are often best found via the application of a non-generic image analysis technique (i.e. texture dominant images are handled most effectively with robust texture matching, whereas shape dominant images are compared with good shape matching).

The *shape dominant* category is for images containing only two regions, with simple, regular shapes and for those conforming to the Gestalt figure/background principle. The *colour dominant* category is for images containing regions with a smooth colour distribution and less regular shapes. The *texture dominant* category is for images that contain highly textural regions (using the procedure proposed in [5]). The *structure dominant* category is for images that include straight lines, geometric shapes or conform with a structural template. The *people* category is for images with prominent regions with a hue in the human skin range. The *landscape* category is for pictures of landscapes that obey the landscape template and have regions with colours in the ‘sky’, ‘land’ and ‘sea’ colour ranges.

Figure/background detection is used by the category generation system in conjunction with a count of the number of segments in an image, in order to determine if an image is shape dominant. Firstly, all images with only two regions are declared shape dominant. Secondly, images have a background that surrounds all other segments in the image are declared shape dominant, inline with the figure/background principle [2]. Images are tested to determine if all corners of the image are the same colour or are in the same image segment. If this property is true then they are known as ‘figure/background’ images, and are declared shape dominant.

In categorising people and landscape images an analysis of the average colour found in the dominant segments of a sample of those image types was performed. The result of this experiment was the development of an average colour reading for the sky, land and sea parts of landscape images and the average hue for people images. These values are used by the category generation system, when determining image category. For people images the average hue of dominant region was 15, 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. In landscape image categorisation, the image must also conform to the landscape image template.

Another new contribution in this research is the use of object structure [6]. 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 [7]. 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 structure description is the region relations in the landscape template. In the landscape category a template pertaining to the layout of sky, land and sea (they must be in layers that appear one above the other in the image) will be used as part of the matching criteria, for images that fall into that category. This layering template, combined with matching the colours of the regions to the template colour averages, will create a robust system for categorising landscape images.

The pseudo code for the current version of the category generation code is given in Figure 13. Based on this system, the example images (Figure 4) are assigned to the following categories: 'orchid' is *shape dominant*, 'scenery' is *landscape* and 'face' is *people*.

```
//Landscape
IF(any BIG segment like SKY
|| any BIG segment like LAND)
  category = "landscape"

//People
ELSE IF(any BIG segment like SKIN)
  category = "people"

//Shape Dominant
ELSE IF( number of segments == 2 )
  category = "shape dominant"

//Shape Dominant (continued)
ELSE IF(four corners have same segment label)
  //image is Figure/Background
  merge all remaining segments
  category = "shape dominant"

//Colour, Texture & Structure Dominant
ELSE
  WHILE(all segments have not been considered)
    WHILE(this segment has more neighbours)
      WHILE(current segment != BIG)
        merge adjacent segment with current segment
DONE
```

Fig. 13. Category generation pseudocode

4. CONCLUSION

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 from an image, placing the image into one of six generic categories. These categories are developed from an understanding of how psychological principles apply to computer vision, and thus include low level and high level image features. With pre-categorised images we can more intelligently apply image processing techniques and thus increase the quality of the results provided to the end-user. A prototype of the image searching tool using category-based indexing is in production, with completion scheduled for November 2003.

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

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

[5] A.W. 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*, 1997, pp. 339–342.

[6] Aster Wardhani and Ruben Gonzalez, "Automatic image structure analysis," in *Proc. of the IEEE International Conference of Multimedia Systems (ICMCS'98)*, 1998, pp. 180–188.

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