COLOUR-TEXTURE FUSION IN IMAGE SEGMENTATION FOR CONTENT-BASED IMAGE RETRIEVAL SYSTEMS

by

OOI WOI SENG

Thesis submitted in fulfilment of the
requirements for the degree of
Master of Science

February 2007

ACKNOWLEDGEMENTS

First of all, I would like to express my heartfelt gratitude to my supervisor, Associate Professor Lim Chee Peng, for his support and motivation throughout this research work. His guidance and incisive advice have inspired me to generate fruitful approaches in achieving the objectives in this research. Without his effort, I would not be able to proceed and bring this research to a completion.

Special thanks, as always, are reserved for my loving parents, who have given their unfailing inspiration, love, and understanding throughout all these years. I am also grateful to my wife, Lim Lay Ee, and all family members, for their support and encouragement.

Last, but not the least, the provision of financial assistance from Universiti Sains Malaysia (Skim Biasiswa Khas USM) is very much appreciated, without which this research work would have never commenced.

TABLE OF CONTENTS

		Page
ACKNOWLEDGEMENTS		ii
TABLE OF CONTENTS		iii
LIST OF TABLES		vi
LIST OF FIGURES		vii
LIST OF ABBREVIATION		xiii
ABSTRAK		xiv
ABSTRACT		xv
CHAPTER 1 : INTRODUCTION		
1.1	Preliminaries	1
1.2	Content-Based Image Retrieval	4
1.3	Problems and Motivation	6
1.4	Research Objectives	8
1.5	Research Methodology and Scopes	9
1.6	Thesis Organisation	11
CHAPTER 2 : LITERATURE REVIEW		
2.1	Introduction	13
2.2	Image Retrieval Systems	13
	2.2.1 Commercial Systems	14
	2.2.2 Prototype System	16
2.3	Image Features	20
	2.3.1 Low-level features	21
	2.3.2 High level-features	24
2.4	Image Segmentation Techniques Used in CBIR	26
	2.4.1 Independent-feature Segmentation	27
	2.4.2 Hybrid Segmentation	31
	2.4.3 Pixel-clustering Segmentation	36

CHAPTER 3 : FEATURE EXTRACTION AND ANALYSIS

3.1	Introduction	41
3.2	Image Segmentation by Pixel Clustering	42
3.3	Colour Space Selection	44
3.4	Colour Space Conversion	50
3.5	Image Smoothing	53
3.6	Colour Feature Extraction	60
3.7	Texture Feature Extraction	63
3.8	Summary	71

CHAPTER 4 : IMAGE SEGMENTATION BY FUSION OF THE COLOUR AND TEXTURE FEATURES

4.1	Introduction	72
4.2	Image Segmentation using the Standard FCM Clustering Algorithm	72
4.3	Modification of the FCM Distance Function	74
4.4	Determining the Optimal Number of Clusters	81
4.5	Region Merging and Labeling	83
4.6	Summary	91

CHAPTER 5 : IMAGE SEGMENTATION EXPERIMENTS

5.1	Introduction	92
5.2	Experiment 1	92
	5.2.1 Experimental Setup	92
	5.2.2 Experimental Results	93
5.3	Experiment 2	95
	5.3.1 Experimental Setup	95
	5.3.2 Experimental Results	96
5.4	Experiment 3	99
	5.4.1 Experimental Setup	99
	5.4.2 Experimental Results	99
5.5	Experiment 4	101
	5.5.1 Experimental Setup	101
	5.5.2 Experimental Results	101

5.6 Summary 104

CHAPTER 6 : APPLICATION: OUTDOOR SCENE AND LANDSAT IMAGERY RETRIEVAL

6.1 Introduction 105
6.2 Data sets 105
6.2.1 The Corel data set 105
6.2.2 The Landsat data set 106
6.3 Extracting Features for Image Querying 107
6.4 Query Interfaces 109
6.5 Image Similarity Measures 114
6.6 Performance Evaluation Criterion 115
6.7 Experiment 1: Outdoor Imagery Retrieval 116
6.7.1 Experimental Setup 116
6.7.2 Experimental Results 117
6.8 Experiment 2: Landsat Imagery Retrieval 123
6.8.1 Experimental Set up 123
6.8.2 Experimental Results 123
6.9 Summary 129

CHAPTER 7 : CONCLUSIONS AND FUTURE WORK

7.1 Conclusions and Contributions of the Research 130
7.2 Suggestions for Future Work 133

REFERENCES 135

LIST OF PUBLICATIONS AND AWARDS 144

LIST OF TABLES

		Page
Table 3.1	The evaluation results for five subjects based on 10 outdoor scene images.	49
Table 3.2	The evaluation results for five subjects based on 10 satellite images.	49
Table 4.1	A comparison of the image segmentation time between the Euclidean distance and the modified CIEDE2000 distance function.	80
Table 4.2	Cluster validity values on the sample data sets for c =2 to 7.	82
Table 5.1	Image segmentation performance using standard FCM as the pixel clustering algorithm.	102
Table 5.2	Image segmentation performance of the proposed modified FCM algorithm.	103
Table 6.1	The feature vector of each segmented region contains eight feature measures. The feature vector represents the content of an image region and it is used for region discrimination.	108

LIST OF FIGURES

		Page
Figure 1.1	Image retrieval using commercial search engines (front to back: Google, Yahoo, Lycos, and Altavista) tends to return irrelevant images.	2
Figure 1.2	Stone texture images.	3
Figure 1.3	A typical content-based image retrieval system.	5
Figure 2.1	A screenshot of Blobworld system interface.	19
Figure 2.2	A screenshot of Istorama system interface.	20
Figure 2.3	The RGB colour histograms. a) Lena image and its b) red channel, b) green channel, and c) blue channel histogram.	22
Figure 2.4	Examples of textured surfaces. (a) Brick, (b) wood, and (c) gravel.	23
Figure 2.5	Boundary-based representation of shape feature using chain code.	25
Figure 2.6	Region-based representation of shape feature using eccentricity, *E*.	25
Figure 2.7	Image is split into 5 regions.	28
Figure 2.8	Feature independent segmentation.	29
Figure 2.9	Flowchart of the method proposed by Nevatia and Price (1982).	32
Figure 2.10	Integration of colour and texture maps of VisualSEEK segmentation. Joint colour and texture regions are extracted and are represented by blue regions.	35
Figure 2.11	Comparison of segmentation methods. (a) Pixel-clustering segmentation. (b) Hybrid segmentation.	36
Figure 2.12	Natural scene image (obtained from Corel Stock Photo collection) with the object (buildings) situated far from the center of the image.	37
Figure 3.1	Two collections of test images used for experimentation. a) Outdoor images comprise different categories of objects, *i.e.*, living things, night scene, landscape, and building. b) Satellite images acquired by Landsat 1-3.	47

Figure 3.2	The subjective test user interface for the evaluation of the segmentation results in different colour spaces. The center image is the original image and moving from top left in the clockwise direction are the segmentation results applied in RGB, XYZ, $I_1I_2I_3$,YIQ, YCbCr, HIS, HSV and CIELAB colour spaces.	48
Figure 3.3	The evaluation results of the test images being selected at the top 3 position for outdoor scene images and satellite images.	49
Figure 3.4	The overall evaluation results of two collections of test images on different colour spaces.	50
Figure 3.5	The CIELAB colour model.	50
Figure 3.6	RGB to CIELAB conversion.	52
Figure 3.7	Colour distribution of tiger image sample. (a) Colour distribution in RGB colour space. (b) Colour distribution in L*a*b* colour space.	52
Figure 3.8	Median Filter operation. (a) The sliding window of unfiltered values. The median filter would return a value of 4, since the ordered values are 0, 2, 3, 3, 4, 6, 10, 15, 97. (b) Center value (previously 97) is replaced by the median of all nine values (4).	54
Figure 3.9	Images produced after applying colour median filtering using different sizes of window. (a) Original image, (b) 3×3 mask (c) 5×5 mask, (d) 7×7 mask, and (d) 9×9 mask with single iteration.	54
Figure 3.10	Difference segmentation results generated by using various filter sizes. (a) Without image filtering, (b) 3×3 mask, (c) 5×5 mask, (d) 7×7 mask, and (e) 9×9 mask.	56
Figure 3.11	The effect of filtering and window size on the number of segmented regions.	56
Figure 3.12	Median smoothing. (a) One iteration, (b) two iterations, (c) three iterations, (d) four iterations, (e) five iterations, (f) six iterations, (g) seven iterations, and (h) eight iterations.	58
Figure 3.13	The number of regions detected versus the interation curve.	59
Figure 3.14	The recorded filtering time versus the interation curve.	59
Figure 3.15	Colour distribution of the sample image of tiger. (a) Colour distribution in RGB colour space. (b) Colour distribution in CIELAB colour space.	59

Figure 3.16	Image segmentation (a) Original images. (b) Luminance and chromatic components are used. (c) Only chromatic component is used. Notice that under segmentation is likely to occur, although using the same number of clusters in segmentation, when the luminance component is absent.	61
Figure 3.17	Image segmentation. (a) Original image. (b) Luminance and chromatic components are used. (c) Only chromatic component is used. Poor segmentation resulting from the absent of luminance component can be clearly seen particularly the result of second test image (bottom right).	62
Figure 3.18	The extraction of colour vectors that represent the colour image. (a) Original RGB image (b) CIELAB image (b) Filtered image.	62
Figure 3.19	Segmentation results when the colour components are combined with each texture descriptors. (a) Contrast, (b) entropy, (c) homogeneity, (d) dissimilarity, (e) energy, and (f) correlation. The original images can be referred to Figure 3.1.	67
Figure 3.20	(a) Original images. The results from (b) to (e) are based on the above mentioned combination schemes.	70
Figure 3.21	Extraction of texture descriptor. (a) CIELAB image. (b) Extraction of contrast and entropy values from the luminance component using co-occurrence matrix method.	70
Figure 4.1	The iteration process of FCM. (a) Original tiger image. Pixel clustering result after (b) 1 iteration, (c) 5 iterations, (d) 10 iterations, (e) 20 iterations, (f) 30 iterations, (g) 40 iterations, (h) 50 iterations, and (i) 53 iterations.	73
Figure 4.2	The iteration process of FCM. (a) Original satellite image. Pixel clustering result after (b) 1 iteration, (c) 5 iterations, (d) 10 iterations, (e) 20 iterations, (f) 30 iterations, (g) 40 iterations, (h) 50 iterations, and (i) 58 iterations.	74
Figure 4.3	Comparison of segmentation results. (a) Original image. (b) Using modified distance function. (c) Using Euclidean distance for both colour and texture components.	78
Figure 4.4	The iteration process of modified FCM. (a) Original tiger image. Pixel clustering result after (b) 1 iteration, (c) 5 iterations, (d) 10 iterations, (e) 20 iterations, (f) 30 iterations, (g) 40 iterations, (h) 50 iterations, (i) 60 iterations, (j) 70 iterations, and (k) 73 iterations.	79
Figure 4.5	The iteration process of modified FCM. (a) Original satellite image. Pixel clustering result after (b) 1 iteration, (c) 5 iterations, (d) 10 iterations, (e) 20 iterations, (f) 30 iterations, (g) 40 iterations, (h) 50 iterations, (i) 60 iterations, (j) 70 iterations, and (k) 80 iterations.	80

Figure 4.6	Region merging and labeling example. (a) Output matrix after FCM. (b) Output matrix after 8-connected component. (c) After region merging. (d) After trimming of small size region.	84
Figure 4.7	The results of segmentation corresponding to difference values of T_a. (a) Original tiger image. (b) Segmentation results without region merging. (c) When T_a = 1%, many non-dominant regions have been merged. (d) When T_a = 2%, the tiger object remained as two regions. (e) Accurate result was obtained when T_a = 3%. Continuing with higher T_a, such as (f) when T_a = 5% and (g) when T_a = 10% do not show improvement. (h) Under-segmentation begins when T_a is set to 17%.	86
Figure 4.8	(a) Original images. (b) Segmentation results without region merging. (c) Segmentation results when T_a = 3%.	87
Figure 4.9	The results of segmentation corresponding to difference values of T_a. (a) Setting T_a = 5% yields four regions. (b) When T_a = 3%, the blue region (red circle) is detected. (c) Reducing T_a = 1% will preserve smaller regions (red circle). (d) Setting T_a = 0.5% yields many details in the form of small regions (red circles).	89
Figure 4.10	(a) Original images. (b) Segmentation results without region merging. (c) When T_a = 1%, non-dominant regions were truncated. Important regions (yellow arrows) can still be detected. (d) If T_a = 2%, some important regions (green arrows) were missing.	90
Figure 5.1	Synthetic colour-texture images with different regions. (a) The original image that comprises a broom texture background and two circular regions with broom (left) and metal sheet (right) texture. (b) The original image that comprises two corns but dissimilar intensity.	93
Figure 5.2	(a) Original image. Segmentation results using colour (b) and texture (c) features alone. (d) Fusion of colour and texture features produces correctly segmented regions.	94
Figure 5.3	(a) Original image. (b) The popcorn regions are correctly separated by colour segmentation. (c) Texture segmentation fails to discriminate the regions due to texture homogeneity. (d) Improvement result is obtained when combine colour and texture features.	95

Figure 5.4	The test images used for the second experiment. (a) Horse image from the outdoor scene collection. (b) Satellite image.	96
Figure 5.5	Segmentation results based on different feature weights. (a) Original image. (b) w_1=0 and w_2=1. (c) w_1=1 and w_2=0. (d) w_1=1 and w_2=0.5. (e) w_1=0.5 and w_2=1. (f) w_1=1 and w_2=1. (g) w_1=0.8 and w_2=1.	97
Figure 5.6	Segmentation results based on different feature weights. (a) Original image. (b) w_1=0 and w_2=1. (c) w_1=1 and w_2=0. (d) w_1=1 and w_2=0.5. (e) w_1=0.5 and w_2=1. (f) w_1=1 and w_2=1. (g) w_1=0.8 and w_2=1.	98
Figure 5.7	Image segmentation comparison: (a) Original images. (b) Results of the proposed system. (c) Results of the Istorama system (obtained from http://uranus.ee.auth.gr/Istorama/).	100
Figure 5.8	Image segmentation comparison: (a) Original images. (b) Results of the proposed system. (c) Results of the Blobworld system (obtained from http://elib.cs.berkeley.edu/photos/ blobworld/start.html).	100
Figure 5.9	The test images used for the forth experiment. (a) Tiger image from the outdoor scene collection. (b) Satellite image from the Landsat image collection.	101
Figure 6.1	Eccentricity.	108
Figure 6.2	Graphical user interface for the image retrieval system.	109
Figure 6.3	The segmentation output panel shows the segmentation result of the query image.	110
Figure 6.4	The properties (visual features) of the selected ROI are displayed beside the segmentation output panel.	110
Figure 6.5	Query control panel.	111
Figure 6.6	(a) to (c) An example of changing the colour of the sky region by selection from a palette.	111
Figure 6.7	Refinement panel.	112
Figure 6.8	Region merging and splitting operation. The sky on the left image consists of three sub-regions. These regions are selected and merged to form a single region (right image). Region splitting is the reverse operation. The splitting process re-segments a region into two or more sub-regions to produce more detailed output.	112
Figure 6.9	Feature weights tuning panel.	113

Figure 6.10	The retrieved images panel and retrieval results obtained by first using "sky" as the query. The segmentation outputs and the ROI properties of the retrieved images are produced to help the user understand how the system reasons.	114
Figure 6.11	The 'Next page' and 'Previous page' buttons allows the user to page through the other retrieved images.	114
Figure 6.12	The results of sunset query. The best 8 matches are presented for the same query region selected by the user based on (a) the proposed method, (b) global histogram, and (c) local histogram.	119
Figure 6.13	The results of tiger query. The best 8 matches are presented for the same query region selected by the user based on (a) the proposed method, (b) global histogram, and (c) local histogram.	120
Figure 6.14	The results of sky query. The best 8 matches are presented for the same query region selected by the user based on (a) the proposed method, (b) global histogram, and (c) local histogram.	121
Figure 6.15	The precision versus recall graphs for (a) sunset, (b) tiger, and (d) sky queries. The results show the averaged query performances from a total of five browsing sessions for each sample query.	122
Figure 6.16	The results of vegetation query. The best 8 matches are presented for the same query region selected by the user based on (a) the proposed method, (b) global histogram, and (c) local histogram.	125
Figure 6.17	The results of sea query. The best 8 matches are presented for the same query region selected by the user based on (a) the proposed method, (b) global histogram, and (c) local histogram.	126
Figure 6.18	The results of urban query. The best 8 matches are presented for the same query region selected by the user based on (a) the proposed method, (b) global histogram, and (c) local histogram.	127
Figure 6.19	A precision versus recall graph comparing the three approaches for (a) vegetation, (b) sea, and (c) urban regions queries for Landsat MSS satellite imagery.	128

LIST OF ABBREVIATION

APRP	Adaptive Pattern Recognition Processing
ACA	adaptive clustering algorithm
CBIR	Content-Based Image Retrieval
EM	Expectation Maximization
FCM	Fuzzy *c*-mean
GLCLL	Gray-Level Co-occurrence Linked List
GLCM	Grey-Level Co-occurrence Matrix
HSV	Hue Saturation Value
MSE	Mean Square Error
MSS	Multispectral Scanner
QBIC	Query by Image Content
RGB	Red Green Blue
SRG	Seeded Region Growing

PENGGABUNGAN CIRI-CIRI WARNA-TEKSTUR DALAM SEGMENTASI IMEJ BAGI SISTEM PENCAPAIAN IMEJ BERDASARKAN KANDUNGAN

ABSTRAK

Kemajuan teknologi komputer serta kepopularan World Wide Web telah membawa kepada peningkatan bilangan gambar yang berbentuk digital. Selari dengan perkembangan itu, sistem pencapaian imej berdasarkan kandungan (content-based image retrieval, CBIR) telah menjadi satu topic kajian yang berkembang dengan pesatnya sejak kebelakangan ini. Proses segmentasi merupakan langkah prapemprosesan yang mempunyai pengaruh penting terhadap prestasi sistem CBIR. Oleh itu, dalam penyelidikan ini, satu rangka segmentasi imej yang baru, bersesuaian untuk pertanyaan kawasan (region queries) dalam CBIR, telah dipersembahkan. Teknik yang digunakan merupakan gabungan ciri-ciri warna dan tekstur gambar, dengan bantuan algoritma *fuzzy c-means clustering* (FCM) yang telah diubahsuai. Pada setiap pixel, ciri warna yang diperoleh daripada koordinat warna CIELAB digabungkan dengan ciri tekstur yang diperoleh daripada *Grey-Level Co-occurrence Matrix* (GLCM) untuk membentuk bahagian imej yang mempunyai kesamaan sifat. Kemudian, algoritma gabungan bahagian imej digunakan untuk menggabungkan bahagian-bahagian imej yang tidak dominan. Selepas itu, ciri-ciri yang terkandung pada setiap bahagian imej akan diekstrak dan digunakan dalam proses pencarian. Untuk menguji keberkesanan dan kemampuan sistem yang dicadangkan, beberapa ujikaji yang menggunakan imej luar bilik (outdoor) dan imej satelite telah dilaksanakan. Perbandingan antara sistem yang dicadangkan dengan sistem prototaip, iaitu Istorama dan Blobworld, telah menunjukkan bahawa sistem yang dicadangkan mampu menghasilkan keputusan segmentasi yang lebih baik. Selain itu, melalui keputusan pencapaian imej serta precision-recall analisis, didapati sistem yang dicadangkan dapat memperoleh pencapaian yang lebih baik berbanding dengan histogram tempatan dan global.

COLOUR-TEXTURE FUSION IN IMAGE SEGMENTATION FOR CONTENT- BASED IMAGE RETRIVAL SYSTEMS

ABSTRACT

With the advances in computer technologies and the popularity of the World Wide Web, the volume of digital images has grown rapidly. In parallel with this growth, content-based image retrieval (CBIR) is becoming a fast growing research area in recent years. Image segmentation is an important pre-processing step which has a great influence on the performance of CBIR systems. In this research, a novel image segmentation framework, dedicated to region queries in CBIR, is presented. The underlying technique is based on the fusion of colour and texture features by a modified *fuzzy c-means clustering* (FCM) algorithm. For each image pixel, the colour components of the CIELAB colour space are combined with texture features, computed from the *Grey-Level Co-occurrence Matrix* (GLCM), to form regions that exhibit homogeneous properties. A region merging algorithm is applied to recursively merge non-dominant regions. Then, the visual properties of each region are indexed and used in a query. To evaluate the effectiveness and applicability of the proposed method, a series of experiments using outdoor and satellite scene images has been performed. The proposed method shows superior segmentation performances when compared with those from existing CBIR prototype systems, i.e., Istorama and Blobworld. Through the retrieval results and the precision-recall analysis, it is demonstrated that the proposed system is effective, and compares favorably with global and local histogram methods.

CHAPTER 1

INTRODUCTION

1.1 Preliminaries

With the steady growth of computer technology, rapidly declining cost of storage devices, and the ever-increasing imaging on the World-Wide Web, the volume of image data continues to grow exponentially over the year. It is estimated that 80 billion of new images are created each year, and 7 million new images are added to the web daily (eVision, 2001). However, a full use of these huge image collections is impossible unless the images are effectively organised, and fast and accurate searching and retrieval methods are deployed.

Text-based retrieval is a common technique to provide information relating to the content of the database. Current web-based search engines such as Yahoo and Google rely largely on keywords, tag, subject heading, and creation date to handle image management, retrieval, and image database functions. These text descriptors are not derived automatically from an image, but are manually annotated by humans. To retrieve images from an image database, a user has to supply keywords, and the images returned are associated with these keywords.

However, use of text in retrieving visual data has been less successful. A common problem encountered when retrieving images by text is the high irrelevant returned results. For example, assume that one wants to retrieve cancer images from the Web using a commercial image search engine. The search is started by supplying "cancer" as the keyword. As shown in Figure 1.1, the returned images, however, comprise skin cancer images, medical books related to skin cancer therapy, cancer scanning machines, and even cancer horoscope just to name a few. Numerous results

returned are beyond one's expectation. One has to keep on navigating the results for the desired images which may probably end up viewing several tens of irrelevant web pages.

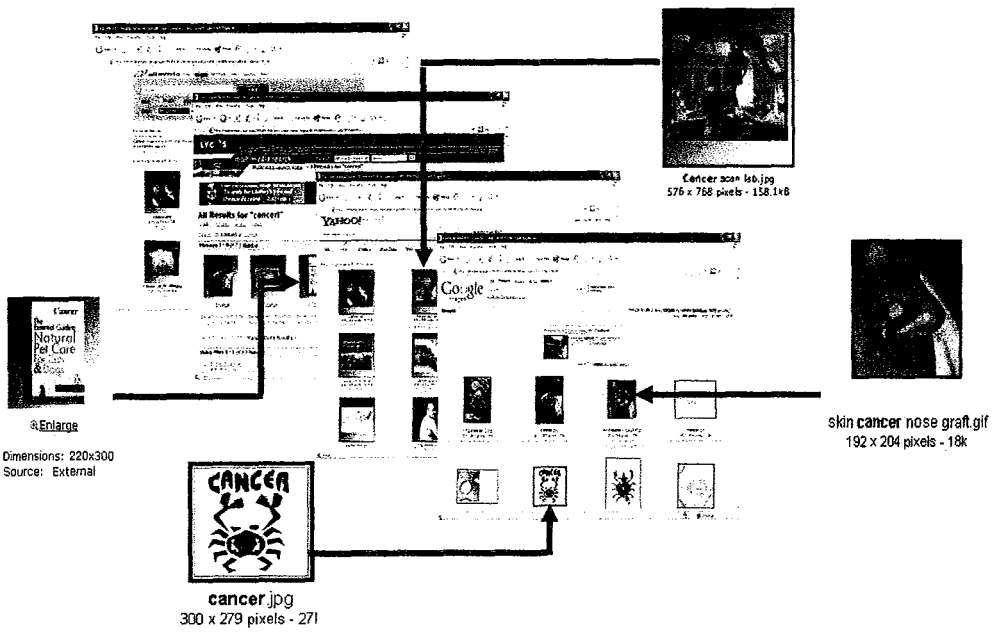


Figure 1.1: Image retrieval using commercial search engines (front to back: Google, Yahoo, Lycos, and Altavista) tends to return irrelevant images.

The main reason of this problem is owing to the search that relies entirely on keywords which are tagged on the images. If the image content is not properly described during annotation, a misconnection between image documents and irrelevant images are likely to be returned. One may claim that if the image content is thoroughly described and associated with proper keywords, the irrelevant issue can be overcome. However, describing an image to a satisfactory degree of details requires a long list of descriptions. Moreover, accurate descriptions of image visual properties such as texture and shape by keywords are difficult to accomplish in practice even by experienced users. For instance, try to associate and compare the stone texture images in Figure 1.2 by words. Even with a long list of descriptions, it is still difficult to describe the images accurately.

(a) (b)

Figure 1.2: Stone texture images

In addition to the difficulty to describe image content with text, text-based retrieval suffers from two other drawbacks. First, manual annotation of keywords to every image is inherently labour intensive. The time taken to annotate an image ranges from 7 minutes to more than 40 minutes (Eakins and Graham, 1999). The task becomes nearly impossible for a large volume of databases consisting thousands of images. Referring to the stone texture example in Figure 1.2, how long would it take to describe those images? Imagine how many hours would be lost if one had to describe thousands of those images? This exhaustive process will be a source of considerable frustration. As a result, the description for these images might probably end up with a single "stone texture" keyword. Second, image annotation is subjective. A review, conducted by Markey (1984), on inter-indexer consistency revealed that there were wide disparities in the keywords that different individuals assigned to the same picture.

To overcome the deficiencies of text-based retrieval, the research of indexing image features for representation has risen since the early 90s. This technology is generally referred to as Content-Based Image Retrieval (CBIR), which is designed to complement text-based retrieval systems with an image visual search capability. With this technology, a user can retrieve desired images from the database by supplying sample images or sketches. Hence, more accurate results can be obtained since the similarity of the input and stored images is determined based on image content rather

than the texts or keywords tagged to the images. After a decade of research, the CBIR technology is now beginning to move out of the laboratory and into the marketplace. However, there are still many issues that remain unsolved.

In the following sections, a brief introduction to CBIR is provided. The problem of inaccurate segmentation faced by the CBIR research community, and motivation for the novel approaches developed in this thesis are presented. Then, the research objectives are defined. This is followed by a discussion on the methodology and scope of this research. An overview of the organisation of this thesis is included at the end of this chapter.

1.2 Content-Based Image Retrieval

CBIR has been an active research area owing to the need for effective and efficient ways of searching large collection of digital images. Generally, this technique employs primitive image features, such as colour, texture, and shape, to characterise image content for both stored and query images. The similarity between images is computed using some similarity measure. The main advantage of this technology is its ability to automatically extract image content for retrieval without human intervention. Moreover, it allows users to compose query using sample images. Keywords are avoided to initiate the search operation. Thus, the difficulty of associating correct keywords to images is diminished. A typical CBIR system is depicted in Figure 1.3.

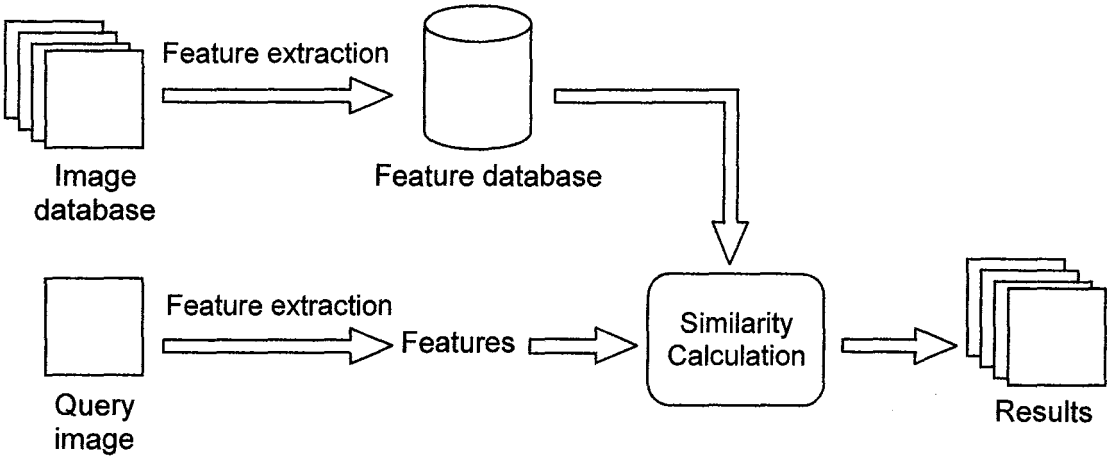


Figure 1.3: A typical content-based image retrieval system

In general, all images in the database are processed in advance to extract the features which represent the contents of the images. This step is very time consuming since it has to process sequentially all the images in the database. Hence, it is usually done off-line to reduce the computational overhead for query processing.

The extraction process assigns a set of identifying descriptors to each image which will be used by the system later in the matching phase to retrieve relevant images. The descriptors are stored in the database as depicted in Figure 1.3. To reduce the processing time, effective feature extraction algorithms are needed. For instance, the algorithm should only process new images added to the database, leaving the pre-extracted images intact.

During the query operation, the query image undergoes the same extraction procedure. Image retrieval is then performed by a matching algorithm, which compares the features or the descriptors of the query image with those of the images in the database using some similarity metric. The images in the database are then ranked according to their similarity with the query, and the highest ranking images are retrieved.

CBIR is a leading edge technology and substantially revolutionises the classical way of image retrieval. Research in CBIR has found promising results in many areas, and they have been used as a part of database management systems in various applications. Since the advent of CBIR, many special issues of leading journals (Tamura and Yokaya, 1984; Bimbo *et al.*, 1999; Basu *et al.*, 2002; Guan *et al.*, 2002) and books (Faloutsos, 1996; Gong, 1998; Bimbo, 1999; Lew, 2001; Castelliand Bergman, 2002) have been dedicated to this topic. In the following, a list which represents only a fraction of areas in which CBIR has been successfully implemented is presented:

i) Intellectual property such as trademark, image copy detection (Kato, 1992; Eakins *et al.*, 1999)
ii) Biomedical applications (Muller *et al.*, 2004)
iii) Web-related applications (eVision, 2001; Kherfi *et al.*, 2004)
iv) Architectural and engineering design (Eakins, 1993; Yang *et al.*, 1994)
v) Journalism and advertising (Gupta, 1997)
vi) Remote sensing (Kitamoto, 1993)
vii) Cultural (museums, art galleries, Hirata and Kato 1992; Holt and Hartwick, 1994)
viii) Image security filtering (Wang *et al.*, 1998)
ix) Crime prevention (Eakins *et al.*, 1998)

1.3 Problems and Motivation

Constructing an efficient CBIR system requires expertise from different disciplines to cover several major aspects; such as system design, feature extraction, feature vector quantization, similarity measures, perception analysis, semantic analysis. Besides, other aspects including high-level abstract description of the scene from the

extracted low-level features *e.g.* using automated feedback mechanisms such as relevance feedback that further refines the search process (Harman, 1992) are needed.

Despite the intensive investigation and development of CBIR systems, there are still many issues that require further research. One of the difficulties inherent in the design of a CBIR system is the construction of an efficient image segmentation algorithm.

Image segmentation is the basis of CBIR. The goal of image segmentation is to divide an image into a set of disjoint regions uniquely corresponding to objects in the input image. Useful information which describes the image objects, such as colour, texture, and shape features can then be indexed for retrieval. Hence, the retrieval performance depends critically on the accuracy of image segmentation. Inaccurate segmentation leads to inferior performance in both primitive and semantic retrieval. Thus, a good segmentation algorithm is essential before high level concepts can be embedded into the retrieval system.

A number of image segmentation techniques have been developed to characterize the image contents. However, most techniques consider only single image feature, *e.g.* colour (Smith and Chang, 1996; Tremeau and Borel, 1997) or texture (Manjunath *et al.*, 2001; Roula *et al.*, 2001) during segmentation. Such techniques do not yield satisfactory results when it comes to complex scene images where both colour and texture features are important for content discrimination. To solve this problem, hybrid approaches have been proposed. For example, Nevetia and Price (1982) presented a hybrid method by combining edge-based and region-based segmentation in aerial image segmentation. Other similar methods of integrating the results of edge detection and region growing have also been proposed (Pavlidis and Liow, 1990; Chu and Aggarwal, 1993; David *et al.*, 2001). The main aim of these

hybrid approaches is to improve the segmentation results based on the combined outcome of different segmentation methods. However, hybrid segmentation is a complicated process. Different segmentation processes are carried out independently, and the results of these processes are then combined for further analysis and interpretation. Hence, hybrid segmentation is time consuming and demands a higher computational load.

Recently, pixel clustering techniques that combine image features for segmentation have received much attention (Chang and Wang, 1996; Stefania *et al.*, 1999; Ma and Manjunath, 2000; Carson *et al.*, 2002). Image features such as colour and texture are combined during segmentation to improve the results. Pixel clustering techniques are useful in solving complex scene segmentation. Complex scene images are particularly hard to treat owing to varying lighting conditions and highly overlapping of homogeneous regions which belong to different objects. Fusion of the colour and texture features has proven to be effective and efficient in partitioning complex scene images into meaningful regions. However, the problem of selecting the right features and the extraction techniques is still an open issue. More rigorous researches to good segmentation is desired.

1.4 Research Objectives

A CBIR system comprises many parts, from system design, segmentation, feature extraction, to image retrieval. Each part of the system can be a research area on its own. Owing to the wide scope of research in CBIR, this thesis does not attempt to cover all aspects of the problem in building a CBIR system. Instead, this research focuses on the development of a framework for image segmentation and feature extraction, which is considered as the main mechanism of a CBIR system.

This research work is geared towards achieving the following objectives:

(1) To devise an effective method to select the best image features and extraction techniques used for image segmentation.

(2) To develop a robust pixel clustering algorithm based on the fusion of image features.

(3) To validate the effectiveness of the proposed algorithm applied to complex scene segmentation.

(4) To demonstrate the applicability of the proposed algorithm in CBIR applications

During the course of achieving these objectives, extensive empirical analyses have been conducted to compare the effectiveness of various indexing approaches prior to selection. Hence, the resulting algorithm is based on the combination and refinement of various approaches, which lead to an improvement in quality of image segmentation.

1.5 Research Methodology and Scopes

The goal of this research is to develop a novel algorithm based on the fusion of multiple features to address the problem of image segmentation in CBIR systems. The research methodology consists of four parts. The first part examines the suitable image features and the extraction techniques used in image segmentation. A thorough experimental analysis is conducted to select the best combination technique to integrate multiple image features for segmentation. The experimental analysis employs a qualitative measurement method based on human's perceptual judgements.

The second part studies the construction of a pixel-clustering algorithm to combine the extracted colour and texture features. To effectively exploit the usefulness of these features, the fuzzy clustering algorithm used in pixel classification is modified

to take into account the heterogonous properties of colour and texture features during integration. The outcome is a set of segmented regions representing the image visual content.

The third part addresses the problem of over segmentation. A robust post-processing method is crucial for the success of a useful segmentation algorithm. A region merging algorithm is constructed to combine disconnected regions based on mutual similarity of colour and texture information.

Lastly, the fourth part concerns the extraction of compact attributes from the segmented image regions. These attributes form the basis of image discrimination. The proposed algorithm is performed entirely in the spatial domain in order to reduce the complexity of mapping the image into frequency domains.

To evaluate the effectiveness and applicability of the proposed algorithm, a series of experiments using outdoor and satellite scene images have been performed. The outdoor scene collection, which consists of 2000 images, is obtained from the Corel Stock Photo collection (Corel Corp., 1990). The satellite images are Landsat images obtained from the "LANDSAT Images of the USA" website (Landsat MSS Imagery, 1998). The Landsat image collection (200 images) was acquired by a four-band Multispectral Scanner (MSS) remote sensing instrument aboard LANDSAT 1, 2 and 3.

An interactive region-based image retrieval system is developed to demonstrate the applicability of the proposed approach to image retrieval. Image visual features such as colour, texture, and shape features are used as the criteria for images matching. The image retrieval performances are compared with two traditional image

retrieval techniques, *i.e.*, global histogram and local histogram intersection (Swain and Ballard, 1991) by means of recall and precision measurements (Jones, 1981).

The uncertainty information from the segmentation results is not considered in this study. Other topics like system design, image compression, multi-dimensional indexing techniques, MPEG-7 standardization, and relevance feedback are also beyond the scope of this research work.

1.6 Thesis Organisation

This thesis is organised in accordance with the objectives mentioned above, as follows.

Chapter 2 presents the background information of CBIR and a literature review on image segmentation techniques used in practical CBIR systems. The existing CBIR systems are first presented. This is followed by a discussion on researches related to image segmentation. Image segmentations techniques based on pixel clustering are examined in detail. Several examples of hybrid systems are also reviewed.

Chapter 3 concentrates on the selection of image features and the extraction techniques used for image segmentation. Colour model selection is the first step in colour image segmentation. Then, texture descriptors and the texture extraction method are determined based on an extensive empirical analysis. The extracted colour and texture features for individual pixels are subjected to a pixel clustering algorithm presented in the next chapter.

Chapter 4 describes the application of a fuzzy clustering algorithm (Bezdek, 1981) to cluster individual pixels into groups that exhibit homogeneous properties. The

fuzzy c-means distance function used in pixel classification is modified to combine the advantages of both colour and texture information extracted from an image. The modification leads to superior segmentation results particularly in segmenting complex scene images. Then, post-processing operations are introduced. Operations which include merging and truncation of non-dominant regions are examined in detail.

Chapter 5 demonstrates the effectiveness of the algorithm applied to complex scene segmentation. Four experiments are devised. The first experiment demonstrates the usefulness of integrating colour and texture information in image segmentation. The second experiment shows the weight tuning process of the proposed FCM distance function in assisting the segmentation algorithm to discriminate image objects. The third experiment compares the proposed segmentation results with two existing colour-texture integrated systems. The forth experiment measures the image segmentation time of the proposed algorithm, and compares the results with those from the standard fuzzy c-means clustering algorithm.

Chapter 6 demonstrates the applicability of the proposed approach in CBIR applications. Two databases containing outdoor and multispectral satellite imagery are employed to show the retrieval performance.

Finally, conclusions of this research are set out in Chapter 7. A number of areas to be pursued as further work are suggested at the end of this thesis.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Content-Based Image Retrieval (CBIR) is the process of searching for and retrieving images based on information extracted from the contents of the images. A number of CBIR systems have been developed by research groups and commercial companies around the world. Section 2.2 reviews the existing image retrieval systems in both commercial and research domains. The image features used in CBIR applications are discussed in Section 2.3. Section 2.4 focuses on image segmentation techniques used in CBIR applications. This is followed by a discussion on research issues related to image segmentation based on pixel-clustering. Several examples of such systems are introduced. In addition, a comparison between hybrid segmentation and pixel-clustering segmentation is presented at the end of this chapter.

2.2 Image Retrieval Systems

For more than a decade of intensive research, CBIR technology has moved out from the laboratory and into the marketplace. However, the commercial packages are still limited owing to a high technology barrier. Among the well-known image retrieval systems available as commercial packages are Query By Image Content (QBIC) from IBM, Visual RetrievalWare from Convera, VIR Image Engine from Virage Inc. and eVe from eVision. Besides these commercial packages, there are numerous prototype systems developed by researchers at universities and research institutes including Photobook, VisualSeek, MARS, Netra, Blobworld and Istorama. In the following section, some of the well-known CBIR systems that have been developed are reviewed.

2.2.1 Commercial Systems

(a) QBIC

Query by Image Content (QBIC) (Flickner *et al.*, 1995) from IBM is the first commercial CBIR system. The system offers retrieval by any combination of colour, texture, or shape, as well as by text. The colour features computed are the average colour of an object or the whole image in RGB, YIQ, Lab, and MTM coordinate, and a k-element colour histogram. The texture features used in QBIC are modified versions of the Tamura texture representation (Tamura *et al.*, 1978), *i.e.*, the combination of coarseness, contrast, and directionality features (Equitz and Niblack, 1994). The shape features in QBIC consist of shape area, circularity, eccentricity, major axis orientation, and a set of algebraic moment invariants (Faloutsos *et al.*, 1994; Scassellati *et al.*, 1994). All shapes are assumed to be non-occluded planar shapes allowing each shape to be represented as a binary image.

Image queries in QBIC can be formulated by selecting from a palette, specifying an example query image, or sketching a desired shape on the screen. QBIC allows combined type searches where text-based keywords and visual features are used in a single query. Currently, QBIC is available commercially as part of IBM products such as the DB2 Digital Library. The QBIC framework and techniques have substantial impact on later image retrieval systems.

(b) Visual RetrievalWare

Visual RetrievalWare is developed by Convera, which was formed by Excalibur Technologies Corporation and Intel's Interactive Media Services Division in 2000 (Convera, 2005). The Visual RetrievalWare SDK (software development kit) extends Convera's core expertise in advanced knowledge retrieval solutions to include visual search and retrieval through image-content matching ability. Visual RetrievalWare

offers a variety of image indexing and matching techniques based on the combination of colour, shape, texture features, brightness, colour layout, and aspect ratio of the image, and allows the user to adjust the weights associated with each feature. The feature indexing technique is based on the Adaptive Pattern Recognition Processing (APRP) technology, which is developed by the founder of Excalibur (Dowe, 1993). Based on neural network methods, it acts as a self-organizing system that automatically indexes binary patterns in digital information, creating a pattern-based memory that is optimised for the content of the data. Thus, it eliminates the costly labour of manually defining keywords and sorting and labeling information in database fields.

(c) VIR Image Engine

VIR (Visual Information Retrieval) Image Engine from Virage, Inc (Gupta, 1997) is another well-known commercial system. VIR Image Engine also supports visual queries based on primitive features such as colour, composition (colour layout), texture, and structure (shape characterization). The system goes one step further than QBIC as it supports arbitrary combinations of the above four atomic queries. The system allows a user to adjust the weights associated with each features according to his/her own emphasis.

Besides these universal primitives, VIR Image Engine uses an open framework for developers to accommodate domain specific primitives to solve specific image management problems, *e.g.* in satellite imaging, trademark, and face recognition retrieval applications. This makes it easy for developers to extend the system by building in new types of query interface, or additional customised modules to process specialised collections of images. Alternatively, the system is available as an add-on to existing database management systems such as Oracle or Informix.

(d) eVe (eVision Visual engine)

eVe (eVision Visual engine) from eVision is an advanced visual search engine that includes analysis, storage, indexing, and retrieval of images and video (eVision, 2001). eVision's premier product is a software development kit (SDK) named eVe SDK. The eVe SDK is a suite of Java component development tools that enable users to build, test, and publish their own eVe-powered, visual search applications. The toolkit supports format conversions, image analysis, automatic segmentation, and clustering indexes for very rapid search and retrieval.

The unique principle of eVision's technology, differentiating it from other commercial visual search engines, is its ability to automatically segment an image into relevant object regions and generate signatures that capture the colour, texture, shape, and object patterns. The ability to segment enables whole and partial image searches with an unparalleled search accuracy. The online demonstration of eVe can be found at http://www.evisionglobal.com/tech/demo.html.

2.2.2 Prototype Systems

Many prototype systems have been developed, mainly by academic institutions, in order to demonstrate the feasibility of content-based technology in searching and retrieval of images. Some of the best-known systems are described in the following sections.

(a) Photobook

The Photobook system (Pentland *et al.*, 1996) from Massachusetts Institute of Technology (MIT) is one of the earliest CBIR systems from the academic domain. Like other commercial systems, it aims to characterise images for retrieval by colour, texture,

and shape, as well as by keyword. Recent version of Photobook includes FourEyes, a feedback agent which selects and combines models based on examples from the user. This allows features relevant to a particular type of search to be computed at search time, giving a greater flexibility at the expense of speed. Further information on Photobook, together with an online demonstration, can be found at http://www-white.media.mit.edu/vismod/demos/photobook/.

Although Photobook itself never emerged a commercial product, its face recognition technology has been incorporated into the FaceID package from Viisage Technology (http://www.viisage.com/), now in use by several US police departments.

(b) VisualSEEK

VisualSEEK (Smith and Chang, 1997a) was developed at Columbia University. It offers searching by image region colour, shape, and spatial location, as well as by keyword. The major feature of VisualSEEK lies in the user interface design of a user query construction tool, called "Colour Region Queries" (Smith and Chang, 1997a). This tool provides a query grid and three regions, which allows users to assign their positions, sizes, colours, etc. The authors argued that such a design provides a method of "Joint Content-Based/Spatial Image Query", which will retrieve images according to the spatial relationship of the regions.

(c) MARS

MARS (Huang *et al.*, 1997), developed at the University of Illinois, is also based on the relevance feedback idea. The system characterizes each object within an image by a variety of features which include colour, texture, shape, and wavelet coefficients. Colour is represented using a 2D histogram over the HS coordinates of

the HSV space. Texture is represented by two histograms, one measuring the coarseness and the other one the directionality of the image, and one scalar defining the contrast. To extract the colour/texture layout, the image is divided into 5 × 5 sub-images. The shape of the boundary of the extracted object is represented by Fourier descriptors and chamfer descriptor. The system learns the best combination of these features as well as the similarity measures for the particular query by letting the user grades the retrieved images by their relevance. A demonstration of the MARS system can be viewed at http://jadzia.ifp.uiuc.edu:8001/.

(d) Netra

Netra (Ma and Manjunath, 1997) is a region-based image retrieval system developed at the University of California Santa Barbara. The Netra system partitions an image into 6 to 12 non-overlapping regions based on the edge flow algorithm which uses 'edges' in colour and texture features to detect homogeneous regions. The system then uses the extracted colour, texture, shape, and spatial location information to provide region-based searching based on local image properties. A web demonstration of Netra is available at http://vivaldi.ece.ucsb.edu/Netra.

(e) Blobworld

Blobworld (Carson *et al.*, 2002) is an image retrieval system developed at the University of California at Berkeley. The Expectation Maximization (EM) algorithm is employed to segment the image into regions (blobs) of uniform colour and texture. The colour is described by a histogram of 218 bins of colour coordinates in the Lab-space. Texture is represented by mean contrast and anisotropy over the region. Shape is represented by approximate area, eccentricity, and orientation. Instead of allowing users to compose queries using the whole image, Blobworld supports region-based

queries. Blobworld allows the user to view the internal representation of the submitted image and the query results, thereby allowing the user to know why some non-similar images are returned. The user can then modify his/her query accordingly. Figure 2.1 shows the user interface for the Blobworld system (http://elib.cs.berkeley.edu/photos/blobworld/).

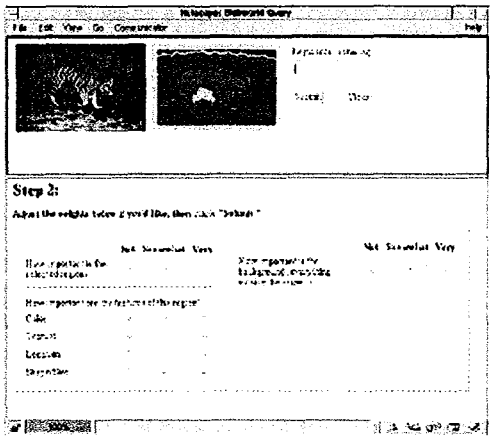


Figure 2.1: A screenshot of Blobworld system interface.

(f) Istorama

The Istorama (Kompatsiaris *et al.*, 2002) system, developed by the Informatics and Telematics Institute and Intrasoft North Greece, performs queries based on image regions obtained using an unsupervised k-means segmentation algorithm. The k-means algorithm is modified to take into account the coherence of the image regions. Based on the extracted regions, characteristic features are extracted using colour, texture, and shape/region boundary information. Like the Blobworld system, Istorama allows the user to view the segmentation mask of the query image, and to select a region and search similar image regions based on colour, region size, and its spatial location in the image. The user can change the weights associated with each feature to emphasise a specific feature he/she might be interested in. Figure 2.2 shows the

interface of Istorama, which can be reached in a publicly accessible web site at http://uranus.ee.auth.gr/Istorama/.

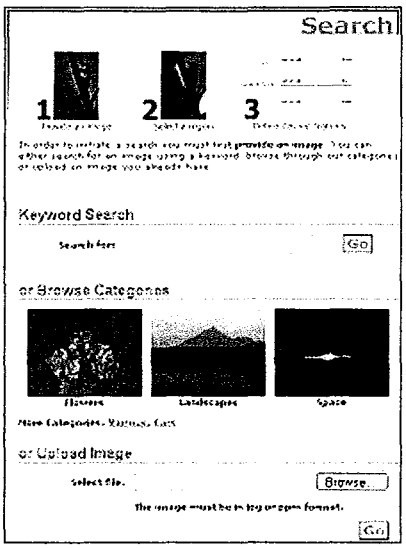


Figure 2.2: A screenshot of Istorama system interface.

2.3 Image Features

The image features used in CBIR applications can be classified into low-level features and high-level features (Johansson, 2000). The former include colour, texture, and edge features while the latter is application-dependent and may include, for example, shape, spatial location, and spatial relationship between image elements. High level features are often calculated from low-level features. Although they are not as general as low-level features, high level features are useful in solving specific image retrieval problems. In this research, the images features considered include two low-level features (colour and texture) and one high-level feature (shape). The following sections describe each of these features in detail.

2.3.1 Low-level features

(a) Colour

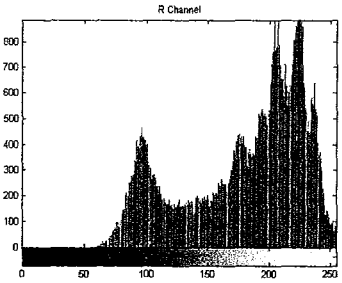
Colour is the most commonly used visual feature for indexing and retrieval of images (Swain and Ballard, 1991; Rui *et al.*, 1999; Schettini *et al.*, 2001). This is because the colour feature is robust to background complication and independent of image scaling, translation, and rotation. The key issues in colour feature extraction include the colour space, colour quantization, and colour matching.

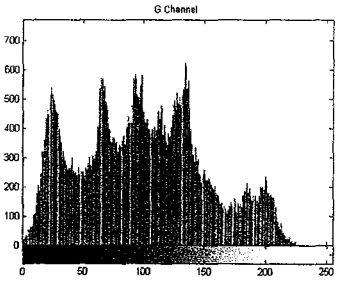
Typically, digital colour images are represented as red, green, and blue channels, or RGB images. In CBIR applications, the RGB colour space is rarely used for indexing and querying as it does not correspond well to human colour perception. Other spaces such as HSV (Hue, Saturation Value) or the CIELAB and Luv spaces are better with respect to human colour perception, and are more frequently used (Jain, 1989).

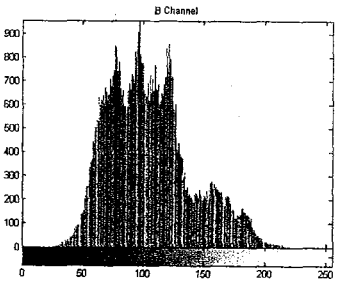
Colour quantization is an important process of reducing the colour resolution of an image (Smith and Chang, 1996). This can be done by performing sampling on each colour channel. For example, grayscale images are quantised at 256 levels, and require 1 byte (8 bits) for the representation of each pixel. Binary images are produced when they are quantised to 0 and 1. Hence, binary images only require 1 bit per pixel.

A common method to represent colour information of an image is to transform the image into a histogram. A colour histogram is computed by discretising the colours within the image and counting the number of pixels of each colour, as illustrated in Figure 2.3.

(a)







(b) (c) (d)

Figure 2.3: The RGB colour histograms. (a) Lena image and its (b) red channel (c) green channel, and (d) blue channel histogram.

To find perceptually similar images based on the colour histogram requires the use of similarity measures. Swain and Ballard (1991) proposed histogram intersection, an L1 metric, as the similarity measure for the colour histogram. To measure similarity between close but not identical colours, the L2-based metric has been proposed (Ioka, 1989; Niblack, 1994). Furthermore, considering that most colour histograms are very sparse and thus sensitive to noise, Stricker and Orengo (1996) proposed using the cumulated colour histogram. Their results demonstrated the advantages of the proposed approach over the conventional colour histogram approach. Besides colour histogram, other descriptors of colour feature used in CBIR include colour moments (Stricker and Orengo, 1996), colour sets (Smith and Chang, 1996), colour coherence vector (Pass and Zabih, 1996), and colour correlogram (Huang *et al.*, 1997). These descriptors have been shown to be effective when compared with the conventional colour histogram approach.

(b) Texture

Texture is a repeated pattern of elementary shapes occurring on an object's surface. Texture may appear to be regular and periodic, random, or partially periodic. Figure 2.4 illustrates some examples of textured surfaces.

(a) (b) (c)

Figure 2.4: Examples of textured surfaces. (a) Brick, (b) wood, and (c) gravel.

Texture feature is useful in distinguishing areas of images with similar colour such as sky and sea, or grass and leaf (Eakins and Graham, 1999; Korfhage, 1997; Ma and Manjunath, 1998). Complex scene images such as aerial images, biomedical images, and outdoor scene images are textured images. These images are generally hard to be categorised using keywords alone owing to limited vocabulary for texture description. Hence, an effective and efficient texture extraction method is very useful in categorising these images.

A variety of techniques have been developed to describe texture feature in an image. The techniques can generally be classified into three major categories, namely, statistical, structural, and spectral (Haralic, 1979; Chen, 1982). In statistical approaches, statistical analyses based on co-occurrence matrices (Haralick *et al.*, 1973), primitive length (Galloway, 1975), edge frequency (Davis and Mitiche, 1980), and spatial frequency (Krumm and Shafer, 1990) are employed to discriminate different textures. Many simple features such as energy, entropy, coarseness, contrast, homogeneity, correlation, cluster tendency, anisotropy, phase, roughness, directionality,

flames, stripes, repetitiveness, and granularity are derived. Statistical methods are useful if texture primitive sizes are comparable with the pixel sizes.

For structural approaches, "texture primitive", the basic element of texture, is used to form more complex texture patterns by grammar rules which specify the generation of texture pattern (Haralick *et al.*, 1979). On the other hand, in spectral approaches, the textured image is transformed into the frequency domain. Then, extraction of texture features can be done by analysing the power spectrum. In addition to the aforementioned methods, multiresloution-based approaches have been proposed to characterise texture features in an image (Laine and Fan, 1993).

2.3.2 High level-features

Shape

Shape is an important feature in object representation and recognition. Unlike texture, shape is a well-defined concept. It is widely believed that humans recognize natural objects by their shape (Liu, 2002). Shape representation can generally be classified into two broad categories, *i.e.*, boundary-based and region-based (Alaya *et al.*, 1999).

In boundary-based techniques, only the outer boundary or outline of an object is considered. Examples of this technique include the use of chain codes (Freeman and Davis, 1977), Fourier descriptors (Persoon and Fu, 1977), and simple geometric border representations, such as curvature, bending energy, boundary length, and signature (Mehtre *et al.*, 1997; Loncaric, 1998). The region-based technique considers the properties of the entire object region. Various descriptors are available for describing an object region, *e.g.* area, Euler number, eccentricity, elongation, and compactness