



FEATURE EXTRACTION AND SIMILARITY MEASURES FOR CONTENT BASED IMAGE RETRIEVAL

A Dissertation

Submitted to

**The Central Department of Computer Science and Information
Technology, Institute of Science and Technology
Tribhuvan University**

**In Partial Fulfillment of the Requirements for the degree of
Master of Science in Computer Science and
Information Technology**

By:

Laxman Thapa

October 2008



Tribhuvan University

Institute of Science and Technology

Central Department of Computer Science and Information Technology

Date: _____

Recommendation

I hereby recommend that the dissertation prepared under my supervision by **Mr. Laxman Thapa** entitled “**Feature Extraction And Similarity Measures for Content Based Image Retrieval**” be accepted as fulfilling in part requirements for the degree of masters of science. In my best knowledge this is an original work in computer science.

Prof. Dr. Shashidhar Ram Joshi

Head, Department of Electronics and Computer Engineering,
Institute Of Engineering, Pulchowk, Nepal
(Supervisor)



Tribhuvan University

Institute of Science and Technology

Central Department of Computer Science and Information Technology

We certify that we have read this dissertation work and in our opinion it is satisfactory in the scope and quality as a dissertation in the partial fulfillment for the requirement of Master of Science in Computer Science and Information Technology.

Evaluation Committee

Dr. Tanka Nath Dhamala

Head, Central Department of Computer
Science and Information Technology
Tribhuvan University

(External Examiner)

Prof. Dr. Shashidhar Ram Joshi

Head, Department of Electronics and Computer
Engineering, Institute Of Engineering,
Pulchowk, Nepal
(Supervisor)

(Internal Examiner)

Date: _____

ABSTRACT

Due to the enormous increase in image database sizes, as well as its vast deployment in various applications, the need for CBIR development arose. This thesis basically analyses the Content-based image retrieval (CBIR) techniques based on texture, shape and color. And implements CBIR based on texture and color as a case study.

In designing the CBIR, feature extraction and Similarity Measures require a great deal for retrieval of Images. The solution initially proposed was to extract the primitive features of a query image and compare them to those of database images. The image features under consideration were colour, texture and shape. Thus, using matching and comparison algorithms, the colour, texture and shape features of one image are compared and matched to the corresponding features of another image. This comparison is performed using colour, texture and shape distance matrices. In the end, these matrices are performed one after another, so as to retrieve database images that match to the query. The similarity between features is to be calculated using algorithms used by well known CBIR systems. For each specific feature there is a specific algorithm for extraction and another for matching. In this thesis, we have used histogram representation for the color extraction, and wavelet feature representation for the texture representation.

ACKNOWLEDGEMENT

I am extremely pleased to write this acknowledgement. In the first place my gratitude goes to Prof. Dr. Shashidhar Ram Joshi, the supervisor of my thesis, for his valuable guidance and expertise as well as for his kind advice, encouragement, and constant support. I am grateful to the Head of the Central Department of Computer Science and Information Technology, Dr. Tanka Nath Dhamala, for his help and support.

I am also grateful and thankful to all the respected teachers, Dr. Onkar P. Sharma (Marist College, USA), Dr. Subarna Shakya, Mr. Arun Timalisina, Mr. Sudarshan Karanjit, Mr. Min Bahadur Khati, Mr. Bishnu Gautam, Mr. Hemanta G.C., Mr. Dinesh Bajracharya and others for granting me broad knowledge and inspirations during my study period in Central Department of Computer Science and Information Technology (CDCSIT).

I would like to thank all the colleagues at the CDCSIT for the pleasant work atmosphere. I am indebted to my friends Mr. Bhogendra Mishra and Mr. Jagdish Bhatta for their contribution in helping me arranging the materials for documentation. My heartiest thanks go to my brother Mr. Pritam Raj Bista for contributing by writing mathematical functions and equations. I would also like to give gratitude to my friends Mr. Suraj Karki, Mr. Suresh Khatiwada and Mr. Sushil Nepal to encourage me to finish this dissertation in time. Finally, I must thank to all my family members for cooperating with me while I was doing this work

- Laxman Thapa

To my loving parents

Table of Contents

Details	Page No
Abstract	i
Acknowledgements	ii
Tables of Contents	iv
List of Figures	viii
List of Tables	ix
Abbreviations	x
CHAPTER I	1-6
1. INTRODUCTION	1
1.1 Content Based Image Retrieval	2
1.1.1 Feature Extraction	3
1.1.2 Similarities Measures	3
1.2 Motivation and Aim	4
1.3 Application of CBIR	4
1.4 CBIR Systems	5
1.5 Thesis Organization	6
CHAPTER II	7-11
2. BACKGROUND AND PROBLEM FORMULATION	7
2.1 Literature Review	7
2.1.1 CBIR System Architecture	8
2.1.2 Feature Extraction	8
2.1.3 Similarity Measures	10
2.1.4 Multilevel Image Indexing	11
CHAPTER III	12-21

3. COLOR FEATURE EXTRACTION AND ANALYSIS	12
3.1 Definition	12
3.1.1 Color Space	12
a) RGB Color Space	13
b) CMY Color Space	13
c) CIE Color Space	13
d) HSV Color Space	14
3.2 Method of Representation	15
3.2.1 Color Histogram	15
3.2.2 Color Coherence Vector	19
3.2.3 Color Correlogram	20
3.2.4 Color Moment	21
CHAPTER IV	23-32
4. TEXTURE FEATURE EXTRACTION AND ANALYSIS	23
4.1 Definition	23
4.2 Method of Representation	24
4.2.1 Tamura Feature	24
4.2.2 Co-occurrence Matrix	28
4.2.3 Wavelet Transform	30
4.2.4 Gabor Filter	32
CHAPTER V	34-36
5. SHAPE FEATURE EXTRACTION AND ANALYSIS	34
5.1 Definition	34
5.2 Method of Representation	34
5.2.1 Fourier Descriptor	35
5.2.2 Moment Invariant	36
CHAPTER VI	37-38

6. SIMILARITIES MEASURES	37
6.1 Minkowski-Form Distance	37
6.2 Quadratic Form Distance	37
6.3 Mahalanobis Distance	38
6.4 Kullback-Leibler(KL) Divergence and Jeffray-Divergence(JD)	38
CHAPTER VII	40-50
7. IMPLEMENTATION	40
7.1 Implementation Tools	40
7.1.1 Matlab	40
7.1.2 Data Structures	41
7.2 Color Feature	41
7.2.1 Color Histogram	41
7.2.2 Quadratic Distance Matrix	42
7.2.3 Similarity Matrix	42
7.3 Texture Feature	45
7.3.1 Pyramid Structured Wavelet Transform	45
7.3.2 Energy Level	46
7.3.3 Euclidean Distance	47
7.4 Graphical User Interface(GUI)	47
7.5 Image Database	50
CHAPTER VIII	52-55
8. TESTING ANALYSIS	52
8.1 Color Extraction and Matching	53
8.2 Texture Extraction and Matching	54
8.3 Performace Evaluation	55
CHAPTER IX	56-57

9. CONCLUSION AND FURTHER RECOMMENDATION	56
9.1 Summary	56
9.2 Further Recommendation	57
References	58-62
Annex- Matlab code	63

List of Figures

Details	Page No.
Figure 1.1: Block diagram of Content Based Image Retrieval	3
Figure 3.1: RGB Color Space	13
Figure 3.2: HSV Color Space	14
Figure 3.3: Sample Image and its Corresponding Histogram	16
Figure 3.4: Two Images with similar Histogram	19
Figure 4.1: Example of Texture	23
Figure 4.2: Co-occurrence Matrix of Image	29
Figure 4.3: Classical Co-occurrence Matrix	29
Figure 4.4: Haar Wavelet Example	31
Figure 4.5: Daubechies Wavelet Example	32
Figure 5.1: Boundary-based and Region Based Shape	34
Figure 7.1: Color Histogram of Two Images	43
Figure 7.2: Minkowski Distance Approach	44
Figure 7.3: Quadratic Distance Approach	44
Figure 7.4: Similarity matrix A , with a diagonal of ones	45
Figure 7.5: Pyramid- structured Wavelet Transform	46
Figure 7.6: GUI Design	48
Figure 7.7: Menu Editor Specification for menu	49
Figure 7.8: Application Window at runtime	50
Figure 7.9: Image database	51
Figure 8.1: Query Image for searching	52
Figure 8.2: Color Results for the searching for 371.bmp	53
Figure 8.3: Texture Results for the searching for 371.bmp	54

List of Tables

Details	Page No.
Table 3.1: Colour Map and Number of pixels for the Image of Figure 3.3	17
Table 7.1: Colour Maps of two images.	43
Table 8.1: Color distance between query and results.	53
Table 8.2: Euclidean distance between query and results.	54

List of Abbreviations

CBIR	Content Based Image Retrieval
TBIR	Text Based Image Retrieval
QBIC	Query By Image Content
IBM	International Business Machine
MARS	Multimedia Analysis and Retrieval System
RGB	Red, Green, Blue
HSV	Hue, Saturation, and Value
SPCA	Shift-invariant Principal Component Analysis
FEM	Finite Element Method
HSB	Hue, Saturation, and Brightness
JPEG	Joint Photographic Experts Group
BMP	Bitmap Picture
GIF	Graphics Interchange Format
CMY	Cyan, Magenta, and Yellow
LCH	Local Colour Histograms
GCH	Global Colour Histograms
CCV	Color Coherence Vectors
SAR	Simultaneous Auto-Regressive
QF	Quadratic Form
KL	Kullback-Leibler
JD	Jeffrey Divergence