

**DEVELOPMENT OF SHAPE DESCRIPTORS BASED ON LEGENDRE  
POLYNOMIALS  
AND  
CONTENT-BASED RETRIEVAL OF SCOLIOSIS IMAGES**

Submitted to the  
Cochin University of Science and Technology  
*in partial fulfillment of the requirements for the award of the degree of*  
***Doctor of Philosophy***  
*In the Faculty of Technology*

by  
**Dinesh Kumar V. P.**

Under the guidance of  
**Dr. Tessamma Thomas**



DEPARTMENT OF ELECTRONICS  
COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY  
COCHIN, KERALA, INDIA 682022

APRIL 2008

Development of Shape Descriptors Based on Legendre Polynomials and Content-Based Retrieval of Scoliosis Images

**Ph.D. Thesis in the field of Image Processing**

**Author**

Dinesh Kumar V. P.  
Research Scholar  
Department of Electronics  
Cochin University of Science and Technology  
Cochin –682 022  
Kerala, India.  
e-mail:dkvp10@yahoo.com

**Research Advisor**

Dr. Tessamma Thomas  
Reader  
Department of Electronics  
Cochin University of Science and Technology  
Cochin –682 022  
Kerala, India.  
e-mail:tess@cusat.ac.in

April 2008

*Dedicated to...*

*Fond memories of my grandma*

DEPARTMENT OF ELECTRONICS  
COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY  
COCHIN-22

*C E R T I F I C A T E*

*This is to certify that this Thesis entitled Development of Shape Descriptors Based on Legendre Polynomials and Content-Based Retrieval of Scoliosis Images is a bonafide record of the research work carried out by Mr. Dinesh Kumar V.P. under my supervision in the Department of Electronics, Cochin University of Science and Technology. The results presented in this thesis or parts of it have not been presented for the award of any other degree.*

*Tessamma Thomas*

Dr. Tessamma Thomas  
(Supervising guide)  
Reader

Department of Electronics  
Cochin University of Science and Technology

Cochin-22  
16-04-2008

## DECLARATION

*I hereby declare that this Thesis entitled **Development of Shape Descriptors Based on Legendre Polynomials and Content-Based Retrieval of Scoliosis Images** is based on the original research work carried out by me under the supervision of **Dr. Tessamma Thomas** in the Department of Electronics, Cochin University of Science and Technology. The results presented in this thesis or parts of it have not been presented for the award of any other degree.*

Cochin-22  
16-04-2008



Dinesh Kumar V.P.

# Acknowledgement

I would like to express my heartfelt gratitude to my supervising guide Dr. Tessamma Thomas, Reader, Department of Electronics, Cochin University of Science and Technology for her valuable guidance as well as for her kind advice, constant encouragement and affectionate support.

I am greatly indebted to Dr. K. V. Menon, Lakeshore Hospital and Research Center, Maradu, Kerala, India, for his valuable suggestions in the work of indexing of scoliosis images and providing me important research materials. I am thankful to the anonymous reviewers of my publications for providing valuable suggestions and motivating comments.

Let me express my sincere gratitude to Prof. K. Vasudevan, Head of the Department of Electronics, Cochin University of Science and Technology, for extending the facilities in the department for my research work. Also, I would like to acknowledge all faculty members of the Department of Electronics, especially Prof. P. Mohanan, for their kind support and help.

I sincerely thank CSIR, India, for providing financial assistance during the period of my research by granting Senior Research Fellowship.

My sincere thanks to Dr. Miroslaw Bober, Visual Information Laboratory, Mitsubishi Electric, Guildford, U.K. for providing the MPEG-7 shape database for my research work. I am also thankful to Dr. L. R. Long, National Library of Medicine (NLM), for providing the NHANES database. I sincerely thank Amrita Institute of Medical Sciences & Research Center for providing the scoliosis image database and Dharma Systems, Inc., Nashua, for providing Dharma SDK Server for implementing the RDBMS support. I sincerely acknowledge the help of Mr. M. Vidyasagar and Mr. D. Bose, Dharma Systems Pvt Ltd, Bangalore.

I am grateful to Dr. P. S. John, Head, Department of Orthopedics, Medical College, Kottayam, for introducing me to the techniques of spinal surgery. I also thank Dr. Antony Joseph Thoppil, and Dr. James Chacko, Medical College, Kottayam, Dr. Binod Kumar Bijukche, Dr. Jim F. Vellara, Amrita Institute of Medical Sciences & Research Center, for their collaboration.

I would like to thank Dr. C. S. Sridhar for his help and motivations provided for my research. I thankfully remember my associations with Dr. V. P. Devassia and Dr. M. G. Mini, IHRD. With great sense of gratitude I thank the selfless assistance of Dr. Joe Jacob, Department of Physics, Newman College, Thodupuzha and Dr. P. B. Vinod Kumar, Rajagiri School of Engineering & Technology. I express my gratitude to Dr. V. Deepu Rajan, Nanyang Technological University, Singapore, for his valuable suggestions during the course of my research. I also thank Dr. A. Unnikrishnan, NPOL, for his encouraging words.

I thank all the research scholars of the department, especially Ms. Deepa Sankar, Mr. Gopikrishnan, Mr. Vinu Thomas, and Mr. Anil Lonappan, for their support. I would like to thank Ms. R. Ananda Reshmi, Mr. J. Linesh and Mr. Roopesh Katan who have collaborated with me. I take this chance to acknowledge Mr. Jimson John, Oracle Corporation and Mr. Benoy Jose, V.S.S.C, for their friendly and supportive attitude. I am also thankful to Dr. Sabu Hameed for helping me in collecting some important research papers. I extend my gratitude to all my friends and relatives for their whole hearted support.

I thank all administrative staff and librarian of the Department of Electronics and Cochin University of Science and Technology for their cooperation and support.

It is beyond words to express my gratitude to my mother and family for their help and encouragement. Without their help and sacrifice, I am sure I could not have accomplished this task.

**Dinesh Kumar V. P.**

---

# Contents

<b>1</b>	<b>Introduction .....</b>	<b>1</b>
1.1	Background.....	1
1.2	Challenges in CBIR .....	3
1.2.1	Patterns in Applications .....	4
1.2.2	Sensory Gap .....	4
1.2.3	Semantic Gap .....	4
1.3	Elements of CBIR.....	5
1.3.1	Feature Extraction.....	6
1.3.2	Dimension Reduction.....	7
1.3.3	Similarity Measure.....	7
1.3.4	Relevance Feedback.....	7
1.3.5	Storage and Indexing .....	7
1.4	Objectives .....	8
1.5	Contributions of the Thesis.....	9
1.5.1	Development of a New Shape Descriptor.....	9
1.5.2	Development of a Relevance Feedback Technique for CBIR .....	10
1.5.3	Development of a Scheme for CBIR of Scoliosis Images.....	10
1.5.4	Development of a Fast Algorithm for Discrete Wavelet Transform .....	10
1.5.5	Development of an Algorithm for Fast Computation of Wavelet Packet Transform on Massively Parallel Processors (MPP).....	11
1.6	Thesis Outline .....	11
	References .....	13
<b>2</b>	<b>Content-Based Image Retrieval – An Overview.....</b>	<b>15</b>
2.1	Introduction.....	15
2.2	MPEG-7 Overview .....	15
2.2.1	Scope.....	16
2.2.2	Multimedia Description Schemes.....	18
2.2.3	MPEG-7 Visual .....	18
2.3	Review of Shape Descriptors.....	20
2.3.1	Contour-Based Shape Descriptors.....	20
2.3.2	Region-Based Shape Descriptors .....	28
2.4	Review of Relevance Feedback Techniques.....	31
2.4.1	Classical Schemes.....	32
2.4.2	As an Optimization Problem .....	33
2.4.3	As a Learning Problem .....	34
2.4.4	As a Classification Problem.....	35
2.4.5	General Techniques .....	36
2.5	Review of Major CBIR Systems for Medical Images .....	37



2.5.1	General Systems .....	38
2.5.2	Specialized Systems.....	40
2.6	Similarity Measures.....	43
	References.....	44
<b>3</b>	<b>Scoliosis – An Overview .....</b>	<b>59</b>
3.1	Introduction.....	59
3.2	Human Spine Anatomy.....	59
3.3	X-ray Imaging of Spine .....	61
3.4	Scoliosis.....	62
3.4.1	Diagnosis .....	63
3.5	Types of Scoliosis.....	63
3.6	Basic Definitions .....	64
3.7	Radiographic Assessment of Scoliosis .....	66
3.7.1	Measurement of Curve .....	66
3.7.2	Measurement of Vertebral Rotation .....	68
3.8	Classification Schemes .....	68
3.8.1	Ponsetti and Friedman Scheme.....	69
3.9	Related Pathologies .....	70
3.9.1	Kyphosis.....	70
3.9.2	Lordosis .....	71
3.10	Literature Review .....	71
3.10.1	Spine Localization .....	71
3.10.2	Scoliosis Measurement.....	73
	References.....	74
<b>4</b>	<b>Basic Theory of Fundamental Tools.....</b>	<b>77</b>
4.1	Introduction .....	77
4.2	Classifiers .....	77
4.2.1	Probabilistic Neural Network (PNN).....	77
4.2.2	Learning Vector Quantization (LVQ) Network .....	78
4.2.3	Support Vector Machine (SVM) .....	79
4.2.4	Least Squares Support Vector Machine (Ls-SVM) .....	82
4.2.5	<i>k</i> -Means Clustering .....	84
4.3	Error-Correcting Output Coding (ECOC) .....	85
4.4	Performance Measurements.....	86
4.4.1	Recall and Precision Pair (RPP) .....	86
4.4.2	Average Normalized Modified Retrieval Rank (ANMRR).....	87
4.5	Statistical Tools .....	88
4.5.1	The McNemar Test.....	88
4.5.2	Kappa Statistics .....	89
4.5.3	Correlation Coefficient.....	90
	References.....	91
<b>5</b>	<b>Improved Legendre Moment Descriptor.....</b>	<b>93</b>
5.1	Introduction.....	93

5.2	MPEG-7 Shape Descriptors.....	93
5.2.1	Zernike Moment Descriptor (ZMD).....	94
5.2.2	Angular Radial Transformation Descriptor (ARTD).....	95
5.2.3	Curvature Scale Space Descriptor (CSSD).....	96
5.3	Motivations for Proposing ILMD.....	98
5.4	Legendre Moments.....	99
5.5	Development of ILMD.....	102
5.5.1	Translation.....	102
5.5.2	Rotation.....	102
5.5.3	Scaling.....	106
5.5.4	Comparison of LMD, ILMD1 and ILMD2.....	109
5.5.5	Selection of Optimum Parameters.....	113
5.5.6	Modifications to ILMD.....	114
5.6	Experimental Setup and Results.....	118
5.6.1	Comparison of Contour Representation Accuracy.....	118
5.6.2	Comparison of Region Representation Accuracy.....	124
5.7	Enhancement to ILMD for Affine Transformations.....	133
5.7.1	Retrieval Experiment and Results.....	137
5.8	Clustering Technique for CBIR.....	140
5.8.1	Experimental Setup and Results.....	140
5.9	Conclusions.....	142
	References.....	143
<b>6</b>	<b>Development of a Relevance Feedback Technique for CBIR.....</b>	<b>145</b>
6.1	Introduction.....	145
6.2	Selection of Optimum Shape Descriptor and Classifier.....	147
6.3	Experimental Setup and Results.....	147
6.3.1	Comparison of Classification Accuracy.....	147
6.3.2	Statistical Analysis.....	150
6.4	Retrieval Using Relevance Feedback.....	156
6.4.1	Method.....	156
6.4.2	Experimental Setup and Results.....	157
6.5	Conclusions.....	159
	References.....	159
<b>7</b>	<b>Content-Based Retrieval of Scoliosis Images.....</b>	<b>161</b>
7.1	Introduction.....	161
7.2	Motivations for Proposing CBIR of Scoliosis Images.....	161
7.3	Automatic Estimation of Position and Orientation of Spine.....	163
7.3.1	Methods.....	164
7.3.2	Experimental Setup and Results.....	169
7.3.3	Discussion.....	171
7.4	Measurement and Retrieval of Scoliosis Images.....	171
7.4.1	Acquisition of Land Mark Points.....	172
7.4.2	Rule Based Algorithm for Strategic Vertebra Selection.....	173
7.4.3	Measurement.....	178
7.4.4	A New Classification Scheme.....	180

7.4.5	Feature Descriptor and Distance Measure .....	181
7.4.6	Query Design .....	181
7.4.7	Experimental Setup .....	189
7.4.8	Results.....	190
7.4.9	Discussion.....	194
7.5	SpineSearch – The CBIR Software System .....	197
7.5.1	Design .....	197
7.5.2	Graphical User Interface .....	198
7.6	Conclusions.....	203
	References .....	204
<b>8</b>	<b>Conclusions and Future Work.....</b>	<b>207</b>
8.1	Conclusions .....	207
8.2	Future Work .....	210
	<b>Appendices .....</b>	<b>211</b>
<b>A</b>	<b>Development of a Modified FFT-Based Algorithm for DWT .....</b>	<b>213</b>
A.1	Introduction .....	213
A.2	Computational Structure for Fast Wavelet Transform (FWT).....	213
A.3	Proposed Modified FFT-Based Algorithm.....	215
A.4	Computational Complexity .....	217
A.5	Results and Discussions .....	219
A.6	Conclusion.....	220
	References .....	221
<b>B</b>	<b>Development of a Computational Structure for Fast Computation of Wavelet Packet Transform on MPP's .....</b>	<b>223</b>
B.1	Introduction .....	223
B.2	Wavelet Packet Transform Algorithms .....	224
B.2.1	The Filter Bank Tree WP Algorithm.....	224
B.2.2	Parallel Multiple Subsequence (PMS) Structure Based Algorithm .....	225
B.3	Algorithm Analysis.....	226
B.3.1	Computational Model and Assumptions .....	227
B.3.2	Data Distribution Strategy.....	227
B.3.3	Analysis of the Filter Bank Tree Algorithm.....	228
B.3.4	Analysis of the PMS Structure Based Algorithm.....	229
B.4	Analytical Results and Discussion .....	230
B.5	Conclusion .....	232
	References.....	232
	<b>List of Publications.....</b>	<b>235</b>
	<b>Resume .....</b>	<b>237</b>

# List of Figures

1.1	A conceptual framework for content-based image retrieval.....	5
2.1	Scope of MPEG-7.....	16
2.2	MPEG-7 main elements.....	17
2.3	Overview of the MPEG-7 Multimedia DSs.....	18
3.1	Nomenclature of vertebrae .....	60
3.2	Different axis of human spine.....	61
3.3	A case of scoliosis .....	62
3.4	Standard terminology description.....	64
3.5	Cobb method for measurement of scoliosis.....	67
3.6	Nash and Moe method of determining vertebral rotation.....	68
4.1	An example Probabilistic Neural Network.....	78
4.2	An example Learning Vector Quantization Network .....	79
4.3	SVM classification with a hyperplane that maximizes the separating margin between the two classes (indicated by data points marked by “x”s and “O”s). Support vectors are elements of the training set that lie on the boundary hyperplanes of the two classes .....	81
5.1	Block diagram of computing ZMD .....	95
5.2	Example of contour evolution and corresponding CSS [1] .....	97
5.3	Image Reconstruction of letter ‘E’ using Legendre moments and Zernike moments [10].....	99
5.4	Plot of Legendre polynomial function values.....	100
5.5	Various translated shapes and their proposed region of Legendre moment Computation.....	103
5.6	An example demonstrating four possible cases of major axis alignment .....	104
5.7	Flowchart of development of ILMD1 and ILMD2.....	110
5.8	Average precision-recall of 651 retrievals using LMD, ILMD1 and ILMD2 on MPEG-7 region shape whole database CE-2.....	111
5.9	Some examples of MPEG-7 region shape whole database CE-2.....	111
5.10	Retrieval accuracy of ILMD1 using Euclidean and City block distance measures on MPEG-7 CE-2 database.....	114
5.11	Plot of modified Legendre polynomial function values .....	115
5.12	Reconstructed images at various orders using Legendre moments and modified Legendre moments for (a) a small image (b) a larger image [22] .....	116
5.13	Average precision-recall of 420 retrievals using ILMD1, ZMD, ARTD and CSSD on Set A1 of MPEG-7 contour shape database CE-1 .....	120
5.14	Retrieval results of query image ‘butterfly-1’ on CE-1 Set A1 of CE-1 using	

	(a) ILMD1 (b) ZMD (c) ARTD (d) CSSD .....	120
5.15	Average precision-recall of 420 retrievals using ILMD1, ZMD, ARTD and CSSD on Set A2 of MPEG-7 contour shape database CE-1 .....	121
5.16	Retrieval results of query image ‘fly-1’ on CE-1 Set A2 of CE-1 using (a) ILMD1 (b) ZMD (c) ARTD (d) CSSD .....	121
5.17	Average precision-recall of 1400 retrievals using ILMD1, ZMD, ARTD and CSSD on Set B of MPEG-7 contour shape database CE-1 .....	122
5.18	Retrieval results of query image ‘fork-1’ on CE-1 Set B using (a) ILMD1 (b) ZMD (c) ARTD (d) CSSD .....	122
5.19	Average precision-recall of 200 retrievals using ILMD1, ZMD, ARTD and CSSD on Set C of MPEG-7 contour shape database CE-1 .....	123
5.20	Retrieval results of query image ‘bream-0’ on CE-1 Set C using (a) ILMD1 (b) ZMD (c) ARTD (d) CSSD .....	123
5.21	Average precision-recall of 100 retrievals using ILMD4, ZMD, and ARTD on Set A1 of MPEG-7 region shape database CE-2 .....	126
5.22	Retrieval results of query image ‘513’ on CE-2 Set A1 using (a) ILMD4 (b) ZMD (c) ARTD .....	127
5.23	Average precision-recall of 140 retrievals using ILMD4, ZMD, and ARTD on Set A2 of MPEG-7 region shape database CE-2 .....	127
5.24	Retrieval results of query image ‘277’ on CE-2 Set A2 using (a) ILMD4 (b) ZMD (c) ARTD .....	127
5.25	Average precision-recall of 330 retrievals using ILMD4, ZMD, and ARTD on Set A3 of MPEG-7 region shape database CE-2 .....	128
5.26	Retrieval results of query image ‘1484’ on CE-2 Set A3 using (a) ILMD4 (b) ZMD (c) ARTD .....	128
5.27	Average precision-recall of 330 retrievals using ILMD4, ZMD, and ARTD on Set A4 of MPEG-7 region shape database CE-2 .....	129
5.28	Retrieval results of query image ‘1470’ on CE-2 Set A4 using (a) ILMD4 (b) ZMD (c) ARTD .....	129
5.29	Average precision-recall of 682 retrievals using ILMD4, ZMD, and ARTD on Set B of MPEG-7 region shape database CE-2 .....	130
5.30	Retrieval results of query image ‘394’ on CE-2 Set B using (a) ILMD4 (b) ZMD (c) ARTD .....	130
5.31	Average precision-recall of 651 retrievals using ILMD4, ZMD, and ARTD on MPEG-7 whole region shape database CE-2 .....	131
5.32	Retrieval results of query image ‘2024’ on CE-2 whole database using (a) ILMD4 (b) ZMD (c) ARTD .....	131
5.33	Some examples of occlusion database .....	131
5.34	Average precision-recall of 50 retrievals using ILMD4, ZMD, and ARTD on occlusion database .....	132
5.35	Retrieval results of two query images on occlusion database using (a) ILMD4 (b) ZMD (c) ARTD .....	132
5.36	Average precision-recall of 1400 retrievals using ILMD4, ZMD, and ARTD on Set B of MPEG-7 shape database CE-1 as region shape .....	133
5.37	Retrieval results of query image ‘cattle-1’ on Set B of MPEG-7 shape database CE-1 as region shape using (a) ILMD4 (b) ZMD (c) ARTD .....	133
5.38	Two examples of image normalization through compacting .....	137

5.39	Average precision-recall of 200 retrievals using ILMD1 and EILMD1 on Set C of MPEG-7 contour shape database CE-1 .....	137
5.40	Retrieval results of query image 'bream-124' on CE-1 Set C using (a) ILMD1 (b) EILMD1.....	138
5.41	Average precision-recall of 330 retrievals using ILMD4 and EILMD4 on Set A4 of MPEG-7 region shape database CE-2.....	138
5.42	Retrieval results of query image '1970' on CE-2 Set A4 using (a) ILMD4 (b) EILMD4.....	139
5.43	Average precision-recall of 651 retrievals using ILMD4 and EILMD4 on MPEG-7 whole region shape database CE-2 .....	139
5.44	Retrieval results of query image '2842' on CE-2 whole database using (a) ILMD4 (b) EILMD4 .....	139
5.45	Histogram of accuracy of ILMD, ZMD and ARTD. (a – c) contour shape database CE-1 Set B and (d – f) Region shape database CE-2 A1.....	141
6.1	Query results of a shape for ILMD1 with SVM classifier with relevance feedback using CE1-B database.(a) retrieval result using city block distance (b) Feed back samples (i) relevant (ii) irrelevant (c) Images from query class correctly classified.....	158
6.2	Query results of a shape for ILMD4 with SVM classifier with relevance feedback using CE2 whole database.(a) retrieval result using city block distance (b) Feed back samples (i) relevant (ii) irrelevant (c) Images from query class correctly classified (d) Images from query class misclassified.....	158
7.1	(a) Original gray-scale image of a cervical spinal column. (b) Surface characteristic plot of the vertebra. (c) Six standard morphometric points marked on the histogram equalized image of the vertebra body .....	164
7.2.	(a) The Geometry of the Radon Transform. (b) The Radon Transform of a test image using 100 projections .....	167
7.3	The flowchart of the spine localization process.....	168
7.4	(a) Cervical spine image. (b) Intermediate result of segmentation using mathematical morphology. (c) A marker drawn using computed spine location and orientation .....	170
7.5	(a) Cervical spine image containing a metallic object. (b) Intermediate result of segmentation using mathematical morphology. (c) Image after removal of the metallic object using H-Maxima Transform. (d) A marker drawn using computed spine location and orientation .....	170
7.6	An example scoliosis image with landmark points marked.....	172
7.7	(a) Left (b) Right (c) Superior (d) Inferior end plate tilt angle plot of each vertebra of the image shown in Figure 7.6 .....	174
7.8	(a) Average left - right (L – R) end plate tilt angle plot (b) superior - inferior (S – I) end plate tilt angle plot. (c) Plot of the distance to the CSL of each vertebra (d) Plot of the Disc slope of each vertebra of the image shown in Figure 7.6.....	175
7.9	CBIR test results. First image in each set is the query image. (a) Retrieval results based on curve type (b) Retrieval results based on number of curves.....	195
7.10	Block diagram of SpineSearch CBIR system.....	197

7.11	Main window of the GUI of the SpineSearch.....	199
7.12	Window for retrieval using a specified PID .....	199
7.13	Interface to adjust image properties.....	200
7.14	Interface to mark landmark points .....	201
7.15	Result window of Cobb angle measurement .....	201
7.16	User interface for plotting scoliosis curve parameters.....	202
7.17	CBIR result window .....	203
A.1	(a) Elementary DWT cell for FWT and (b) FFT-based implementation of DWT cell .....	214
A.2	Proposed FFT-based algorithm (FS stands for Fourier-domain subsampling).....	216
B.1	Two level WP decomposition using filter bank tree algorithm .....	224
B.2	Second level WP decomposition using PMS structure.....	226
B.3	Comparison of scalability of filter bank tree and PMS algorithm at decomposition depth 6 and filter kernel size 16 for (a) different machine size and (b) different problem size .....	231
B.4	Comparison of performance (speedup) of filter bank tree and PMS algorithm at different decomposition depth and fixed problem size of 128 MB. (a) filter kernel size 16 on different machine size (b) number of PEs 1024 and different filter kernel size .....	231

# List of Tables

4.1	Error-correcting output codes for five-class problem with minimum Hamming distance of five.....	85
4.2	2x2 error matrix to calculate the statistical significance of differences in classification accuracy based on McNemar test for related samples.....	89
4.3	Kappa statistics computation matrix.....	90
5.1	List of Legendre moments up to order 6.....	102
5.2	First 10 Legendre moments of images shown in Figure 5.5.....	103
5.3	First 10 Legendre moments of images shown in Figure 5.6.....	105
5.4	ILMD1 and ILMD2 descriptors up to order 6 for shapes shown in Figure 5.9.....	112
5.5	Retrieval performance of ILMD1 with different orders on MPEG-7 CE-2.....	113
5.6	Retrieval performance of ILMD4 with different orders on MPEG-7 CE-2.....	118
5.7	ANMRR results of various descriptors on MPEG-7 contour database CE-1.....	124
5.8	ANMRR results of various descriptors on MPEG-7 region database.....	134
6.1	Classification accuracy of SVM and Ls-SVM for various kernel functions using CE1-B database.....	148
6.2	Classification accuracy of SVM and Ls-SVM for various kernel functions using CE2 whole database.....	149
6.3	Classification accuracy of various classifiers using CE1-B database.....	149
6.4	Classification accuracy of various classifiers using CE2 whole database.....	150
6.5	Z- scores of one descriptor against another for various classifiers using CE1-B.....	151
6.6	Z- scores of one classifier against another for various descriptors using CE1-B.....	152
6.7	Z-scores of descriptor-classifier combination using CE1-B.....	153
6.8	Z -scores of one descriptor against another for various classifiers using CE2 whole database.....	154
6.9	Z -scores of one classifier against another for various descriptors using CE2 whole database.....	155
6.10	Z-scores of descriptor-classifier combination using CE2 whole database.....	156
7.1	Computed spine orientation accuracy.....	169
7.2	Measurement error in apex level, end vertebra level and Cobb angle.....	191
7.3	Correlation coefficient results.....	192
7.4	Kappa statistics results.....	193
A.1	FFT-Based DWT algorithms: multiplication complexity per point.....	220
A.2	FFT-Based DWT algorithms: addition complexity per point.....	220



---

## Abbreviations

AB-SVM.	—	Asymmetric Bagging-based SVM
ABRS-SVM	—	Asymmetric Bagging and Random Subspace SVM
ACS	—	Active Contour Segmentation
AIS	—	Adolescent Idiopathic Scoliosis
ANMRR	—	Average Normalized Modified Retrieval Rank
AP	—	AnteroPosterior
ARTD	—	Angular Radial Transformation Descriptor
ART	—	Angular Radial Transformation
AR	—	AutoRegressive
ARBFN	—	Adaptive Radial Basis Function Network
ASM	—	Active Shape Modeling
ASSERT	—	Automatic Search and Selection Engine with Retrieval Tools
AV	—	Apical Vertebra
AVR	—	Average Rank
BAS	—	Beam Angle Statistics
BDA	—	Biased Discriminant Analysis
BiM	—	Binary format for MPEG-7
BSVM	—	Biased SVM
CAP	—	College of American Pathologists
CBIR	—	Content-Based Image Retrieval
CCD	—	Charge Coupled Device
CCS	—	Convex Contour Saliency
COBRA	—	Content-Based Retrieval Architecture
CS	—	Contour Saliencies
CSL	—	Central Sacral Line
CSM	—	Constrained Similarity Measure
CSS	—	Curvature Scale-Space
CT	—	Computed Tomography
CSVL	—	Central Sacral Vertical Line
D	—	Descriptor
DAG	—	Directed Acyclic Graph
DBMS	—	Database-Management Systems
DDL	—	Description Definition Language
DEXA	—	Dual X-ray Absorptiometry
DICOM	—	Digital Imaging and Communications in Medicine
DKBDA	—	Direct Kernel Biased Discriminant Analysis

## Abbreviations

---

MPEG-1	—	ISO Standard
MPEG-2	—	ISO Standard
MPEG-4	—	ISO Standard
MPP	—	Massively Parallel Processors
MRI	—	Magnetic Resonance Imaging
MRR	—	Modified Retrieval Rank
MTSVM	—	MultiTraining SVM
NHANES	—	National Health and Nutrition Examination Survey
NLM	—	National Library of Medicine
NMRR	—	Normalized Modified Retrieval Rank
OCS	—	Optimal Correspondent Subsequence
ODBC	—	Open DataBase Connectivity
OOPS	—	Object Oriented Programming System
OSH	—	Optimal Separating Hyperplane
PA	—	PosteroAnterior
PACS	—	Picture Archival and Communication System
PBR	—	Pathology Bearing Regions
PCA	—	Principal Component Analysis
PE	—	Processing Element
PET	—	Positron-Emission Tomography
PMS	—	Parallel Multiple Subsequence
PNN	—	Probabilistic Neural Network
PRS	—	Polar Raster Sampling
PSM	—	Partial Shape Matching
PUMC	—	Peking Union Medical College
RBF	—	Radial Basis Function
RDBMS	—	Relational DataBase Management System
RF	—	Relevance Feedback
ROI	—	Region of Interest
RPP	—	Recall and Precision Pair
RS-SVM	—	Random Subspace SVM
S – I	—	Superior – Inferior
SAD	—	Sum of Absolute Difference
SD	—	Shape Descriptor
SE	—	Structuring Element
SFD	—	Short-time Fourier Descriptor
SIMD	—	Single Instruction Multiple Data
SLO	—	Scanning Laser Ophthalmoscopy
SM	—	Shape Model
SNOMED	—	Systematized Nomenclature of Human and Veterinary Medicine
SOM	—	Self-Organizing Maps
SQL	—	Structural Query Language
SS	—	Segment Saliency

DM	—	Deformable Model
DPWT	—	Discrete Periodized Wavelet Transformation
DP	—	Dynamic Programming
DS	—	Description Schemes
DTW	—	Dynamic Time Warping
DWT	—	Discrete Wavelet Transform
ECOC	—	Error-Correcting Output Coding
EDT	—	Euclidean Distance Transform
EMR	—	Electronic Medical Records
EV	—	End Vertebrae
FD	—	Fourier Descriptor
FFT	—	Fast Fourier Transform
FRBFN	—	Fuzzy Radial Basis Function Network
FWT	—	Fast Wavelet Transform
GFD	—	Generic Fourier Descriptor
GHT	—	Generalized Hough Transform
GIFT	—	GNU Image Finding Tool
GLM	—	Gray Level Model
GoF	—	Group of Frames
GoP	—	Group of Pictures
GUI	—	Graphical User Interface
HMM	—	Hidden Markov Models
HRCT	—	High-Resolution Computed Tomography
HSV	—	Hue-Saturation-Value
IFFT	—	Inverse Fast Fourier Transform
IFT	—	Image Foresting Transform
IGDS	—	Image Guided Decision Support
ILMD	—	Improved Legendre Moment Descriptor
IRMA	—	Image Retrieval in Medical Applications
IRRL	—	Image Relevance Reinforcement Learning
ISO	—	International Standard Organization
KmED	—	Knowledge based MEDical Database
L – R	—	Left – Right
LEV	—	Lower End Vertebra
LFS	—	Lobular Feature Sets
LMD	—	Legendre Moment Descriptor
Ls-SVM	—	Least-Squares SVM
LVQ	—	Learning Vector Quantization
MAT	—	Medial Axis Transform
MDS	—	Multimedia Description Schemes
MIMD	—	Multiple Instruction, Multiple Data
MPEG	—	Moving Picture Experts Group
MPEG-7	—	ISO Standard

SVM	—	Support Vector Machine
TIFF	—	Tagged Image File Format
TSD	—	Tensor Scale Descriptor
UEV	—	Upper End Vertebra
UNL	—	Universidade Nova de Lisboa
US	—	UltraSound
VOI	—	Volume Of Interest
VOI-FIRS	—	Volume Of Interest based Functional Image Retrieval System
WebMIRS	—	Web-based Medical Information Retrieval System
WP	—	Wavelet Packet
WT	—	Wavelet Transform
WWW	—	World Wide Web
XML	—	Extensible Markup Language
ZCV	—	Zero Crossing Vertebra
ZMD	—	Zernike Moment Descriptor
1-D	—	one-Dimensional
2-D	—	two-Dimensional
3-D	—	three-Dimensional

---

# Chapter 1

## Introduction

---

### 1.1 Background

The advancement of technology has led to the explosive growth of digital material available to common man in all aspects of life including medicine. Nowadays, digital multimedia content are created with ease from a variety of sources, such as scanners, cameras, digital audio and video recorders and broadcast. The advent of internet revolutionized the distribution of information. The Web documents became more communicative over the years enriched with images, video, audio and music. Also, there is an increasing trend towards the digitization of medical imagery and the formation of adequate archives. Large-scale image databases collect various images like X-ray, computed tomography (CT), magnetic resonance imaging (MRI), ultrasound (US), nuclear medical imaging, endoscopy, microscopy, and scanning laser ophthalmoscopy (SLO). It is well acknowledged that medical image databases are a key component in diagnosis and preventive medicine [1, 2]. The resulting picture archiving and communication systems (PACS) are available across wards within a hospital setting and allow global access to shared resources. The wealth of information available freely on the web and on medical image databases posed a major problem for the end users: how to find the information needed ?

The challenge of efficient search for multimedia content has been addressed by the development of more and more sophisticated search engines, with the help of advances in the areas of annotation, indexing and natural language processing. A search engine is mainly a

database and the tools to generate that database and search it. However, in most of the early systems search is text-centric, and thus dependent on the keywords used to index the multimedia content. Some examples of text-based web search engines are Google, Yahoo, Altavista and Lycos. As the digital material on the web became inherently rich in images, audio and video sequences, tools to browse and search such audio-visual media were needed. A straightforward way of using the existing tools for information retrieval to index and search these collections of audiovisual material is to first annotate the audio-visual media by keywords and then use the text-based Database-Management Systems (DBMS) to retrieve it. Of various audio-visual data, images are the most important way of expression. Several approaches were proposed to use keyword annotations to index and retrieve images [3, 4, 5, 6, 7]. Comprehensive surveys in information retrieval can be found in [8, 9, 10].

The advantage of textual indexing of image is that it can provide user with key word searching, catalogue browsing and even with query interface such as Structural Query Language (SQL). However, it apparently has some limitations. One is that it is time consuming. When the database is large, it is almost impossible to manually annotate all the images. The other is visual features of image are difficult to be described using words. The keywords annotation for an image is subjective to the user who does it and will be application context sensitive. Due to the rich content of images, different persons may perceive them differently and annotate them with different keywords. A possible solution is to automatically index the images in a domain independent fashion based on their content. Retrieving images based on their content is called Content-Based Image Retrieval (CBIR). CBIR is introduced as a complement to the traditional textual indexing approach. Due to its extensive potential applications, CBIR has attracted a great amount of attention in the last decade and a review of early systems is presented in [11, 12]. Many research and commercial CBIR systems have been developed, including QBIC [13], MARS [14], Virage [15], Photobook [16], VisualSEEk [17], PicToSeek [18], and PicHunter [19].

Before the emergence of CBIR, medical images were also annotated with text, allowing the images to be accessed by text-based searching [20]. Thus, manual textual index entries are mandatory to retrieve medical images from digital archives, which are inefficient due to the following reasons [21].

- Manual annotations require too much time and are expensive to implement. As the number of images in a database grows, the difficulty in finding desired images increases. Muller *et. al.* reported that the University Hospital of Geneva produced approximately 12,000 medical images per day [22]. It is not feasible to manually annotate all attributes of the image content for large number of images.
- Manual annotations fail to deal with the discrepancy of subjective perception. Typically, a medical image usually contains several objects, which convey specific information. Nevertheless, different interpretations for a pathological area can be made by different radiologists. To capture all knowledge, concepts, thoughts, and feelings for the content of any image is almost impossible.
- The contents of medical images are difficult to be concretely described in words. For example, irregular organic shapes cannot easily be expressed in textual form, but people may expect to search for images with similar contents based on the examples they provide.

These problems limit the feasibility of text-based search for medical image retrieval. In an attempt to overcome these difficulties, content-based retrieval has been proposed to automatically access images with minimal human intervention [20, 23]. CBIR has been proposed by the medical community for inclusion into picture archiving and communication systems (PACS) [24].

CBIR has attracted tremendous research interest of industrial and academic community. Realizing the importance of content-based retrieval of multimedia data, the Moving Picture Experts Group (MPEG) of the International Standard Organization (ISO) has finally approved a standard MPEG-7 in July 2001 [25]. MPEG-7 was termed “Multimedia Content Description Standard,” quite different to earlier MPEG standards. MPEG-7 allows the interoperability between the devices and applications attempting to solve parts of content-based retrieval problem. Currently the standard does not address the content-based medical image retrieval.

## 1.2 Challenges in CBIR

In this section, major challenges in CBIR such as patterns in applications, the sensory gap, and the semantic gap are discussed.

### 1.2.1 Patterns in Applications

There are three broad categories of aims while using the CBIR system [11].

First category is *search by association*, in which the user arrives at a result by iterative refinement of the search based on the similarity of the specification by sketch [26] or by example images. Systems in this category typically are highly interactive. The result of the search can be manipulated interactively by relevance feedback [27, 28].

Another class of users aims the search for a specific image. The search may be for a precise copy of the image in mind. *Target search* may be for another image of the same object of which the user has an image. In some cases, target is interactively specified as similar to a group of given examples. These systems are suited to search for stamps, art, industrial components, and catalogues in general [13, 19].

The third class of applications, *category search*, aims at retrieving an arbitrary image representative of a specific class. It may be the case that the user has an example and the search is for other elements of the same class. Categories may be derived from labels or may emerge from the database [29, 30].

These classifications do not describe the whole story. Studies reveal that the range of queries is wider than just retrieving images based on the presence or absence of objects of simple visual characteristics [31, 32].

### 1.2.2 Sensory Gap

The *sensory gap* is the gap between the object in the real world and the information in a description derived from a recording of that scene. The *sensory gap* makes the description of objects an ill-posed problem. In visual search, explicit representation of the knowledge of the domain is important to alleviate the *sensory gap*.

### 1.2.3 Semantic Gap

The *semantic gap* is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation.



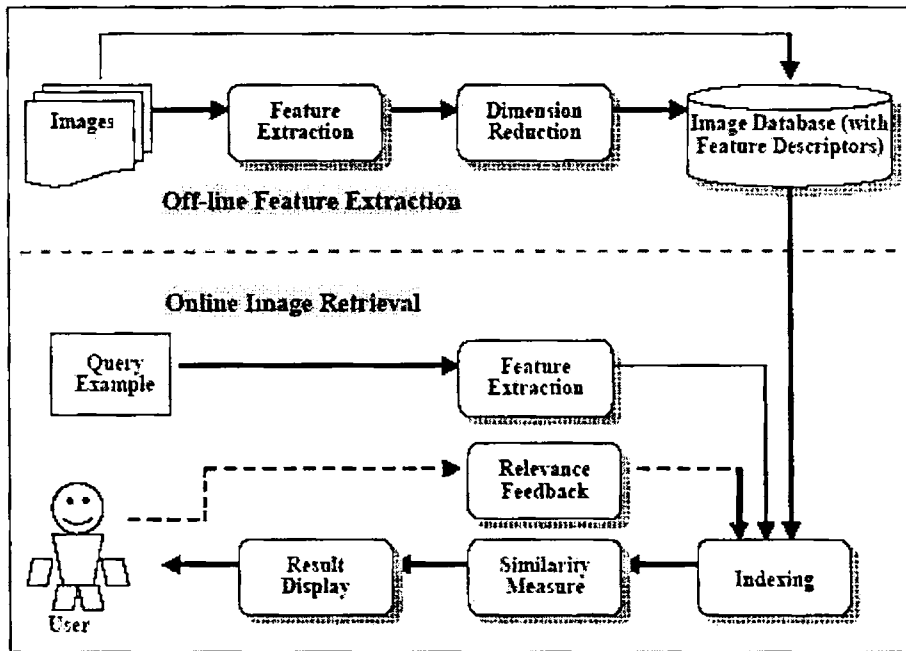


Figure 1.1 A conceptual framework for content-based image retrieval

The most immediate means of semantic characterization entail annotation by keywords or captions. The pivotal point in content-based retrieval is that the user seeks semantic similarity, but the database can only provide similarity by data processing.

### 1.3 Elements of CBIR

A typical CBIR system is divided into *off-line feature extraction* and *online image retrieval*. A conceptual framework for content-based image retrieval is illustrated in Figure 1.1. In offline feature extraction, the contents of the images in the database are extracted and described with a multi-dimensional feature vector called descriptor. The feature vectors of the image constitute a feature dataset stored in the database. In online image retrieval, the user can submit a query example to the retrieval system in search of desired images. The system represents this example with a feature vector. The distances (i.e., similarities) between the feature vectors of the query example and those of the media in the feature dataset are then computed and ranked. Retrieval is conducted by applying an indexing scheme to provide an efficient way of searching the image database. Finally, the system ranks the search results and then returns the results that are most similar to the query examples. If the user is not satisfied

with the search results, the user can provide relevance feedback to the retrieval system, which contains a mechanism to learn the user's information needs. The following sections will describe each component of the system.

### **1.3.1 Feature Extraction**

Representation of images needs to identify the most useful features for representing the contents of images and effective approaches for coding the attributes of the images. The methods for which can be divided into local color, local geometry, or the local texture. The image processing operations transpose the image data into another spatial data array. The purpose of image processing in image retrieval is to enhance aspects in the image data relevant to the query and to reduce the remaining aspects. The image is partitioned or segmented aiming at obtaining more selective features.

#### **1.3.1.1 Color**

Color is an important visual attribute for both human vision and computer processing. The color has superior discriminating potentiality of a three-dimensional domain compared to the single dimensional domain of gray-level.

#### **1.3.1.2 Shape**

The local shape refers to all properties that capture conspicuous geometric details in the image. Object shape feature provides powerful clue to object identity and functionality, and can even be used for object recognition. Shape often carries semantic information and humans can recognize characteristic objects solely from their shapes.

#### **1.3.1.3 Texture**

Perceptual attributes of texture such as directionality, regularity, and coarseness are important. Also, a quantitative characterization of homogenous texture regions is used for image retrieval.

### 1.3.2 Dimension Reduction

In an attempt to capture useful contents of an image and to facilitate effective querying of an image database, a CBIR system may extract a large number of features from the content of an image. Feature set of high dimensionality causes the “curse of dimension” problem in which the complexity and computational cost of the query increase exponentially with the number of dimensions [33]. The most widely-used technique for reduction of the dimensionality of a large feature set in image retrieval is principal component analysis (PCA). The goal of principal component analysis is to specify as much variance as possible with the smallest number of variables [34].

### 1.3.3 Similarity Measure

Similarity is an interpretation of the image based on the difference with another image. Selection of similarity metrics has a direct impact on the performance of content-based image retrieval. The kind of feature vectors selected determines the kind of measurement that will be used to compare their similarity.

### 1.3.4 Relevance Feedback

The main idea of relevance feedback is for the retrieval system to understand the user’s information needs. For a given query, the retrieval system returns initial results based on pre-defined similarity metrics. Then, the user is required to identify the positive examples by labeling those that are relevant to the query. The system subsequently analyzes the user’s feedback using a learning algorithm and returns refined results.

### 1.3.5 Storage and Indexing

For large image sets, computational performance cannot be ignored as an issue. In addition to the number of images, the dimension of the image vector can also be considerable.

Three classes of indexing methods are in use on large image databases: space partitioning, data partitioning, and distance-based techniques. In space-partitioning index techniques, the feature space is organized like a tree [35]. Data partitioning index techniques associate, with

each point in feature space, a region that represents the neighborhood of that feature vector. An R-tree is such a data partitioning structure to index hyper rectangular regions in  $M$ -dimensional space. The SS-tree [36] and its further improvement, the SR-tree [37], use the intersection of the minimum bounding hypersphere and minimum bounding hyperrectangle as the bounding region of a data element. Distance-based index structures are example-based space-partitioning techniques, very suited for query by example when feature space is metric. An example is M-tree [38], a more robust and scalable indexing strategy that uses the triangle-inequality of metric spaces, but retains the data partitioning properties of spatial access methods such as the R-tree and the SS-tree.

## 1.4 Objectives

From the above described background information, it is clear that the image database can be indexed either using textual information or content information. The research work presented in this thesis focuses on the content-based indexing and retrieval of images. In content-based indexing, from among the many active research topics as detailed in Section 1.3., feature extraction and relevance feedback are the topics of interest for this work. The images can be broadly classified into general images and medical images, based on their application. This study addresses CBIR of both types.

In general images, image content can be described using primitive features such as color, texture, shape or combination of them. Specifically, this research focuses on image indexing and retrieval using shape features. One main objective of this research is to develop promising shape descriptor(s) for image retrieval. The shape descriptor investigated and developed in this study can either be directly applied to particular applications such as trademark retrieval, object recognition etc. or be incorporated into any CBIR system to facilitate retrieval using combined image features. User interaction for improving the CBIR accuracy is important, and hence developing a novel scheme for relevance feedback is another major objective. Benchmarking the performance of the developed descriptor and relevance feedback scheme with state-of-the-art shape descriptors also forms a part of the study.

CBIR of medical images is strategically different from that of general images. Use of global image features such as color, texture, shape etc will assist only in general classification of images, for example, of a particular anatomical region. But in content-based medical image

retrieval more interest lies in obtaining images similar to the query image based on the pathology of interest. So, image processing and feature selection should be fine tuned for a specific pathology. In this research, the focus is on the deformity of spine called scoliosis, which is the lateral curvature of the spine. A major objective of the research is to identify and automatically extract features of a scoliosis image, and design a query scheme for CBIR of scoliosis spine images.

To summarize, the main objectives of the research are:

1. To develop a novel and more effective shape representation method.
2. To develop a novel scheme for relevance feedback for CBIR.
3. To develop a novel scheme for CBIR of spine radiographs with scoliosis.

## 1.5 Contributions of the Thesis

The contributions of the thesis are summarized as following.

### 1.5.1 Development of a New Shape Descriptor

A new shape descriptor called Improved Legendre Moment Descriptor (ILMD) based on orthogonal Legendre polynomial is proposed. Techniques for improvement of invariance to geometric transformations like translation, rotation, scaling and affine transformation is developed. Using Legendre polynomial a contour based shape descriptor, and with a combination of modified Legendre polynomial, a region based descriptor is developed. The MPEG-7 shape descriptors Angular Radial Transformation Descriptor (ARTD), Curvature Scale Space Descriptor (CSSD) and Zernike Moment Descriptor (ZMD) are implemented and the retrieval accuracy of the proposed descriptor is compared using MPEG-7 shape databases. The proposed shape descriptor is invariant to shape transformations and also robust to shape distortions and occlusions. Also, a scheme for CBIR is proposed and evaluated using  $k$ -means clustering. ILMD is better than both the contour shape descriptor and region shape descriptor proposed by MPEG-7.

Published or accepted papers related to this part of work are: List of Publications 1.1, 1.4, 2.2, 2.6

### **1.5.2 Development of a Relevance Feedback Technique for CBIR**

This research evaluates and compares the shape descriptors: ILMD, ZMD and ARTD, and classifiers: support vector machine (SVM), least-square SVM (Ls-SVM), probabilistic neural network (PNN) and linear vector quantization (LVQ) network to choose an ideal descriptor – classifier combination for relevance feedback. Statistical analysis using McNemar test is carried out to identify optimum shape descriptor – classifier combination for an efficient relevance feedback scheme. The SVM with RBF kernel using ILMD as pattern vector is found to provide an efficient CBIR system with relevance feedback.

Published or accepted papers related to this part of work are: List of Publications 1.3

### **1.5.3 Development of a Scheme for CBIR of Scoliosis Images**

This research proposes a novel scheme for CBIR of scoliosis images. An algorithm is proposed and evaluated for automatic determination of location and orientation of the spine in a digital radiograph using mathematical morphology. A rule based algorithm is developed to automatically measure the scoliosis from a given set of landmark points. The intra and inter observer reliability assessment of the proposed system is carried out using Kappa statistics and correlation coefficient. The proposed scheme of scoliosis measurement is found to be more reliable and repeatable. A feature descriptor for scoliosis image and query scheme is designed for CBIR of scoliosis images. A software system for CBIR of scoliosis images called “SpineSearch” is developed.

Published or accepted papers related to this part of work are: List of Publications 1.2, 1.5, 2.3, 2.4, 2.5.

### **1.5.4 Development of a Fast Algorithm for Discrete Wavelet Transform**

In the initial part of the research, wavelet transform based techniques were studied. In the process, a novel Fast Fourier Transform (FFT) based algorithm was developed for fast discrete wavelet transform computation. By virtue of the Fourier-space operations, significant saving in computational complexity is achieved. The computational complexity analysis of the algorithm was carried out and compared with a state-of-the-art FFT-based Fast Wavelet Transform (FWT) algorithm. The proposed algorithm was found to be computationally attractive.

Published or accepted papers related to this part of work are: List of Publications 2.7.

### **1.5.5 Development of an Algorithm for Fast Computation of Wavelet Packet Transform on Massively Parallel Processors (MPP)**

Wavelet packet (WP) transform based techniques were also studied, during initial part of the research. In the process, an extension to the Parallel Multiple Subsequence (PMS) structure for WP transform and an algorithm for its implementation on a massively parallel processors machine is proposed. Analytical comparison of the scalability and speedup of the PMS based algorithm and the popular filter bank tree algorithm was carried out in the Intel Paragon machine and the proposed algorithm was found to be better.

Published or accepted papers related to this part of work are: List of Publications 2.1.

## **1.6 Thesis Outline**

**Chapter 1** describes the background, challenges, basic elements of a CBIR system, and objectives of this research. Contributions of this research work are also summarized.

**Chapter 2** presents an overview of the MPEG-7 standard and reviews the literature on CBIR. The first part focuses on visual information description. In the second part, an account of the previous research work that has been carried out in the field of shape descriptors for CBIR by peer researchers is presented. An overview of relevance feedback techniques is presented and major medical CBIR systems are briefly reviewed.

**Chapter 3** is exclusively devoted for scoliosis. Anatomy of the spine, measurement of scoliosis, curve classification schemes etc. are introduced. A brief review of the literature pertinent to spine localization and scoliosis measurement is presented.

**Chapter 4** provides a summary of the fundamental tools used in the thesis. A review of the classifiers used – support vector machine (SVM), least square - support vector machine (Ls-SVM), probabilistic neural network (PNN), learning vector quantization network (LVQ), and  $k$ -means clustering – is presented. A description of the similarity and performance measurement techniques is provided. Also, statistical analysis tools used, McNemar, correlation coefficient, and Kappa statistics, in the thesis is described.

In **Chapter 5**, a novel general shape descriptor called Improved Legendre Moment Descriptor (ILMD) has been proposed. A brief introduction to the theory and implementation of MPEG-7 shape descriptors Zernike Moment Descriptor (ZMD), Angular Radial Transformation Descriptor (ARTD) and Curvature Scale Space Descriptor (CSSD) are provided. Motivations for proposing ILMD are mentioned. Techniques for improving the invariance properties of the Legendre moments and the derivations of the descriptors are outlined. Then, a comparative study of retrieval performance of the ILMD, ARTD, ZMD and CSSD using MPEG-7 shape database is presented. CBIR using  $k$ -means clustering algorithm is proposed and its experimental validation is detailed.

In **Chapter 6**, a novel scheme for relevance feedback for CBIR using SVM and ILMD has been proposed. A comparative study of the classification accuracy of SVM, LS-SVM, PNN and LVQ is reported for ILMD, ZMD and ARTD descriptors. Then statistical analysis results are detailed for selecting optimum classifier - descriptor combination. A relevance feedback technique is then reported.

In **Chapter 7**, a novel scheme for CBIR of scoliosis images is presented. The chapter begins with the description of motivations for the proposal. An algorithm is developed for automatic determination of location and orientation of the spine. A rule based algorithm for measurement of scoliosis is developed, with minimal human intervention. Inter and intra observer reliability experiments and statistical analysis of the results are presented. Then, feature descriptor for indexing of scoliosis image is designed. The implementation details such as database schema and query formulation, and experimental results are outlined. Details of the developed software system 'SpineSearch' is also given. The chapter ends with discussion and concluding remarks on the developed system.

A brief summary of the research work conducted and the important conclusions thereon are highlighted in **Chapter 8**. Suggestions for future research are also provided.

This thesis includes two appendices which describe some allied works carried out by the author during the tenure of research.

**Appendix A** proposes a novel fast algorithm for discrete wavelet transform (DWT) computation using FFT. Computational complexity analysis of the proposed algorithm and another FFT based algorithm is presented.



**Appendix B** proposes an extension to computational structure PMS, for fast WP transform and an algorithm for its implementation on MPPs. Analytical comparison of the scalability and speedup of the PMS based algorithm and popular filter bank tree algorithm in the Intel Paragon machine is reported.

## References:

- [1] Wong STC, and Huang HK, "Networked multimedia for medical imaging," *IEEE Multimedia*, vol. 4, no. 2, pp. 24–35, 1997.
- [2] Wong STC, and Tjandra D, "A digital library for biomedical imaging on the Internet," *IEEE Communications Magazine*, vol. 37, no. 1, pp. 84–91, 1999.
- [3] N. S. Chang and K. S. Fu, "Query-by-Pictorial," *IEEE Trans. on Software Engineering*, vol. 6, no. 6, 1980.
- [4] S. K. Chang and T. Kunii, "Pictorial Database Systems," *IEEE Computer*, pp. 13–21, November 1980.
- [5] D.C. Dimitroff, S.K. Chang, C.W. Yan and T. Arndt, "An Intelligent Image Database System," *IEEE Trans. on Software Engineering*, vol. 14, no. 5, pp. 681–688, 1988.
- [6] Lancaster F., *Indexing and abstracting in theory and practice*, 2nd ed., Library Association, London. 1998.
- [7] Rasmussen E., "Indexing images," *Annual Review of Information Science and Technology*, 32, pp. 169-196, 1997.
- [8] S.-K. Chang and A. Hsu, "Image Information Systems: Where Do We Go From Here?," *IEEE Trans. on Knowledge Data Engineering*, vol. 4, no. 5, pp. 431–442, 1992.
- [9] H. Tamura and N. Yokoya, "Image Database Systems: A Survey," *Pattern Recognition*, vol. 17, no. 1, pp. 29–49, 1984.
- [10] E. Riloff and L. Hollaar, "Text Databases and Information Retrieval," *ACM Computing Surveys*, vol. 28, no. 1, pp. 133-135, 1996.
- [11] Arnold W.M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain, "Content-Based Image Retrieval at the End of the Early Years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349-1380, 2000.
- [12] R. Veltkamp and M. Tanase, *Content-Based Image Retrieval Systems: A Survey*, Technical Report 34, Utrecht University, Information and Computing Sciences, Utrecht, The Netherlands, 2002.
- [13] M. Flickher, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by image and video content: the QBIC system," *IEEE Computer*, vol. 28, no. 9, pp. 23–32, 1995.
- [14] Y. Rui, T. S. Huang, and S. Mehrotra, "Content-based image retrieval with relevance feedback in MARS," in *Proc. IEEE Int. Conf. Image Processing*, Washington, DC, pp. 815–818, 1997.
- [15] G. Amarnath and J. Ramesh, "Visual information retrieval," *Commun. ACM*, vol. 40, no. 5, pp. 70–79, 1997.
- [16] A. Pentland, R. Picard, and S. Sclaroff, "Photobook: content-based manipulation of image databases," *Int. J. Comput. Vis.*, vol. 18, no. 3, pp. 233–254, June 1996.
- [17] J. R. Smith and S. F. Chang, "VisualSEEK: A fully automated contentbased image query system," in *Proc. ACM Multimedia*, pp. 87–98, Nov. 1996.
- [18] T. Gevers and A. W. M. Smeulders, "PicToSeek: combining color and shape invariant features for image retrieval," *IEEE Trans. Image Process.*, vol. 9, no. 1, pp. 102–119, 2000.

- [19] I. J. Cox, M. L. Miller, T. P. Minka, T. V. Papatomas, and P. N. Yianilos, "The Bayesian image retrieval system, PicHunter: theory, implementation, and psychophysical experiments," *IEEE Trans. Image Process.*, vol. 9, no. 1, pp. 20–37, 2000.
- [20] *Multimedia information retrieval and management: Technological fundamentals and applications*, Feng D., Siu W.C., and Zhang H.J., Eds., Springer, Berlin, 2003.
- [21] Wei C.-H., and Li C.-T., "Design of content-based multimedia retrieval," in *Encyclopedia of multimedia technology and networking*, M. Pagani, Ed., Hershey, PA: Idea Group Reference, 2005.
- [22] Muller H., Michous N., Bandon D., and Geissbuhler A., "A review of content-based image retrieval systems in medical applications — Clinical benefits and future directions," *Int'l J. of Medical Informatics*, vol. 73, no. 1, pp. 1-23, 2004.
- [23] Eakins J. P., "Towards intelligent image retrieval," *Pattern Recognition*, vol. 35, no. 1, pp. 3-14, 2002.
- [24] Lehmann T.M., Wein B.B., and Greenspan H., "Integration of Content-based Image Retrieval to Picture Archiving and Communication Systems," in *Proc. of Medical Informatics Europe (MIE 2003)*, Amsterdam, IOS Press, 2003.
- [25] J. M. Martinez, *MPEG-7 Overview*, Technical Report (version 10), ISO/IEC JTC1/SC29/WG11, <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>, October 2004.
- [26] J.M. Corridoni, A. del Bimbo, and P. Pala, "Image Retrieval by Color Semantics," *Multimedia Systems*, vol. 7, pp. 175-183, 1999.
- [27] G. Frederix, G. Caenen, and E.J. Pauwels, "PARISS: Panoramic, Adaptive and Reconfigurable Interface for Similarity Search," in *Proc. Int'l Conf. Image Processing*, 2000.
- [28] A. Hiroike, Y. Musha, A. Sugimoto, and Y. Mori, "Visualization of Information Spaces to Retrieve and Browse Image Data," in *Proc. Visual '99: Information and Information Systems*, pp. 155-162, 1999.
- [29] D.J. Swets and J. Weng, "Hierarchical Discriminant Analysis for Image Retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 386-401, May 1999.
- [30] M. Weber, M. Welling, and P. Perona, "Towards Automatic Discovery of Object Categories," in *Proc. Computer Vision and Pattern Recognition*, pp. 101-108, 2000.
- [31] P.G.B. Enser, "Pictorial Information Retrieval," *J. Documentation*, vol. 51, no. 2, pp. 126-170, 1995.
- [32] L. Armitage and P. Enser, "Analysis of User Need in Image Archives," *J. Information Science*, vol. 23, no. 4, pp. 287-299, 1997.
- [33] Egecioglu O., Ferhatosmanoglu H., and Ogras U., "Dimensionality reduction and similarity computation by inner-product approximations," *IEEE Trans. on Knowledge and Data Engineering*, vol. 16, no. 6, pp. 714-726, 2004.
- [34] Partridge M., and Calvo R. A., "Fast dimensionality reduction and simple PCA," *Intelligent Data Analysis*, vol. 2, no. 1-4, pp. 203-214, 1998.
- [35] L. Brown and L. Gruenwald, "Tree-Based Indexes for Image Data," *J. Visual Comm. and Image Representation*, vol. 9, no. 4, pp. 300-313, 1998.
- [36] D. White and R. Jain, "Similarity Indexing with the SS-Tree," in *Proc. 12th Int'l Conf. Data Eng.*, 1996.
- [37] N. Katayama and S. Satoh, "The SR-Tree: An Index Structure for High-Dimensional Nearest Neighbor Queries," in *Proc. Int'l Conf. Management of Data (SIGMOD)*, pp. 369-380, 1997.
- [38] P. Ciaccia, M. Patella, and P. Zezula, "M-Tree: An Efficient Access Method for Similarity Search in Metric Spaces," in *Proc. Very Large Data Bases Conf.*, 1997.

---

## Chapter 2

# Content-Based Image Retrieval – An Overview

---

## 2.1 Introduction

To facilitate effective search or filtering, multimedia content needs to be described efficiently. In the case of images, large interest in CBIR resulted in sheer increase in the published research work and the number of systems developed. Many features and descriptors were proposed by the industrial and academic research. A standardization of features becomes important to allow devices of different vendors to interoperate with each other. Acknowledging the industry demand and the availability of adequate technology, the MPEG of the ISO developed a standard called MPEG-7: the “Multimedia Content Description Interface”, to allow interoperability between devices and applications attempting to solve parts of the CBIR problem.

Section 2.2 presents an overview of the MPEG-7 focusing on the visual aspects. A review of important literature published for shape descriptors in the context of CBIR is presented in Section 2.3. Relevance feedback techniques employed for retrieval accuracy improvement of CBIR systems are reviewed in Section 2.4. In Section 2.5, a review of literature of the important CBIR systems for medical images is presented. The last section briefly reviews some of the commonly used similarity measures in CBIR systems.

## 2.2 MPEG-7 Overview

MPEG-7 [1, 2] is a standard for describing the multimedia content data. MPEG-7 tries to develop forms of audiovisual information representations that go beyond the simple waveform

or sample-based, compression-based (such as MPEG-1 and MPEG-2) or even object-based (such as MPEG-4) representations. These forms support some degree of interpretation of the information meaning, which can be passed onto, or accessed by, a device or a computer code. MPEG-7 is not aimed at a specific application in particular. It provides a rich set of standardized tools to describe multimedia content, allowing the development of a wide range of applications. Human users as well as automatic systems that process audiovisual information are within the scope of MPEG-7.

MPEG-7 offers a comprehensive set of audiovisual description tools (the metadata elements and their structure and relationships that are defined by the standard as descriptors (D) and Description Schemes (DS)). It specifies a Description Definition Language (DDL) to efficiently index and search materials with associated MPEG-7 data. These searches will permit inquiries based on scenes, motion and visual content as well as text-based queries. Indexing and searching is possible for AV material that has MPEG-7 data associated with it. This material may include: still pictures, graphics, 3D models, audio, speech, video, and information about how these elements are combined.

### 2.2.1 Scope

Figure 2.1 shows a highly abstract block diagram of a possible MPEG-7 processing chain. This chain includes feature extraction (analysis), the description itself, and the search engine (application). To fully exploit the possibilities of MPEG-7 descriptions, automatic extraction of features will be extremely useful. It is also clear that automatic extraction is not always possible, however. The higher the level of abstraction, the more difficult automatic extraction is, and interactive extraction tools will be of good use. However useful they are, neither automatic nor semi-automatic feature extraction algorithms are inside the scope of the standard. The main

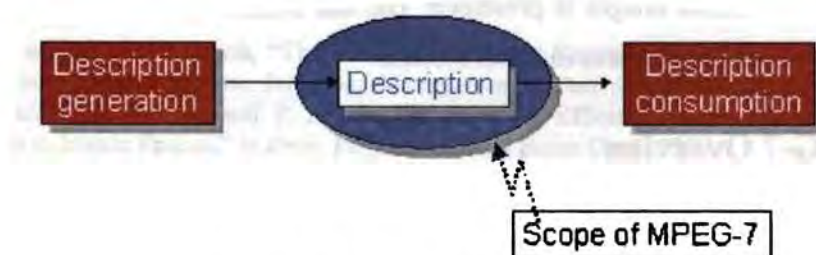


Figure 2.1 Scope of MPEG-7

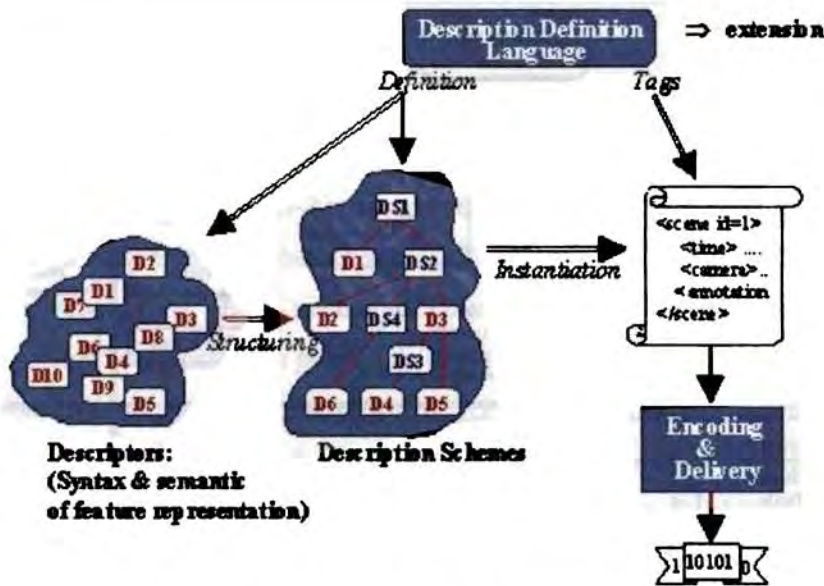


Figure 2.2 MPEG-7 main elements

reason is that their standardization is not required to allow interoperability, while leaving space for industry competition. Another reason not to standardize analysis is to allow making good use of the expected improvements in these technical areas. To guarantee interoperability, MPEG-7 specifies part of the extraction process of some of the low-level features. Also the search engines, filter agents, or any other program that can make use of the description, are not specified within the scope of MPEG-7.

Figure 2.2 shows the relationship among the different MPEG-7 elements introduced above. The DDL allows the definition of the MPEG-7 description tools, both Descriptors and Description Schemes, providing the means for structuring the Ds into DSs. The description tools are instantiated as descriptions. The MPEG-7 descriptions can be either in a textual form, based on the extension of Extensible Markup Language (XML) [3], suitable for editing, searching, filtering, and browsing. This format is not suitable for streaming and is sensitive to transmission errors. To overcome these shortcomings, MPEG-7 defines a second format: Binary format for MPEG-7 (BiM). This format is suitable to streaming and compression. Moreover, MPEG-7 defines coding and decoding tools for this format. For a given piece of AV content, both of its representations in XML and BiM format are equivalent and can be encoded and decoded losslessly [1].

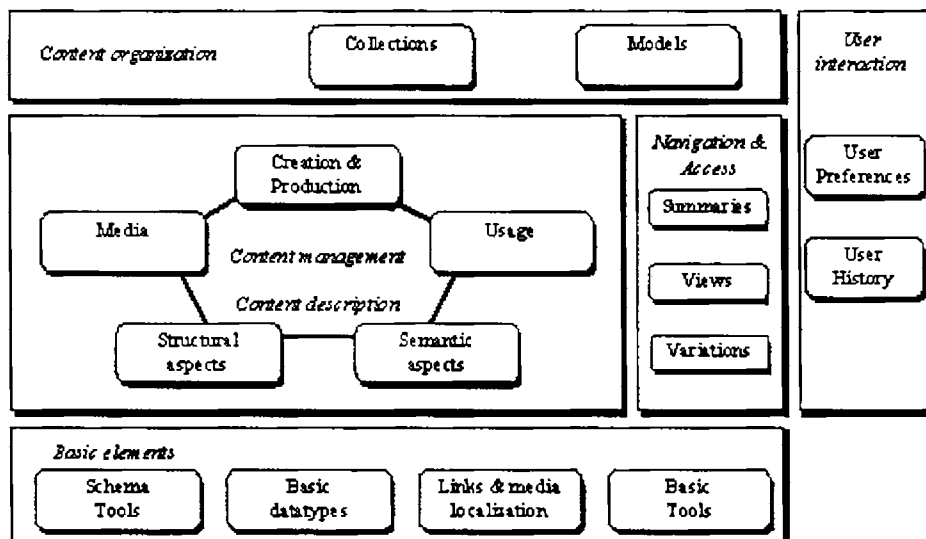


Figure 2.3 Overview of the MPEG-7 Multimedia DSs

## 2.2.2 Multimedia Description Schemes

MPEG-7 Multimedia Description Schemes (MDS) are metadata structures for describing and annotating AV content. MDS comprises the set of Description Tools (Descriptors and Description Schemes) dealing with generic as well as multimedia entities. Figure 2.3 provides an overview of the organization of MPEG-7 Multimedia DSs into the following areas: Basic Elements, Content Description, Content Management, Content Description, Content Organization, Navigation and Access, and User Interaction. The MPEG-7 DSs represent a set of description tools. For a particular application, these DSs can be used to describe multimedia content. A detailed overview of each of these functional areas is presented in the technical report by Martínez [2].

## 2.2.3 MPEG-7 Visual

MPEG-7 Visual Description Tools included in the standard consist of basic structures and Descriptors that cover the following basic visual features: Color, Texture, Shape, Motion, Localization, and Face recognition. Each category consists of elementary and sophisticated Descriptors.

### 2.2.3.1 Shape Descriptors

There are three shape Descriptors: Region Shape, Contour Shape, and Shape 3D. Two dimensional shape representation techniques can be classified into two categories: region-based and contour-based. Therefore, MPEG-7 has standardized two shape descriptors, namely Region Shape and Contour Shape, to describe 2D shapes. Both region-based and contour-based shape descriptors are intended for shape matching. They do not provide enough information to reconstruct the shape nor to define its position in the image.

The region-based descriptor is based on a 2D complex transform defined with polar coordinates on the unit disk, called Angular Radial Transformation (ART) [4]. The ART has separable basis functions along the angular and radial dimensions. Twelve angular and three radial basis functions are used. The descriptor represents the coefficients obtained by projection of the binary region onto the 35 ART basis functions. It is characterized by its small size, fast extraction time and matching. Thus this descriptor is suitable for tracking shapes in video data processing applications.

The Shape Contour descriptor captures characteristics of a shape based on its contour. It relies on the curvature scale-space (CSS) [5] representation, which tries to capture perceptually meaningful features of the shape. This descriptor represents the high curvature points by their position and value of the curvature.

It is common to represent 3D information as polygonal meshes. The MPEG-7 shape 3D descriptor provides an intrinsic shape description of 3D mesh models. It exploits some local attributes of 3D surfaces. The descriptor represents the 3D mesh shape spectrum, which is the histogram of the shape indexes [6] calculated over the entire mesh.

### 2.2.3.2 Other Descriptors

There are five Visual related Basic structures: the grid layout, time series, multiple view [7], the spatial 2D coordinates, and temporal interpolation. The seven color Descriptors are Color space, Color Quantization, Dominant Colors, Scalable Color, Color Layout, Color-Structure, and Group of Frames/Group of Pictures (GoF/GoP) Color. There are three texture Descriptors: homogeneous texture, edge Histogram, and texture browsing. Camera Motion, Motion Trajectory, Parametric Motion, and Motion Activity are the four motion Descriptors:

The two descriptors for localization are Region locator and Spatiotemporal locator. The MPEG-7 face descriptor is based on principal component analysis (PCA) [8, 9].

## **2.3 Review of Shape Descriptors**

Shape is an important visual feature and it is one of the basic features used to describe image content. Shape representation generally looks for effective and perceptually important shape features based on shape boundary information or boundary plus interior content. These various shape features are often evaluated by how accurately they allow one to retrieve similar shapes from a designated database. However, it is not sufficient to evaluate a representation technique only by the effectiveness of the features employed. In fact, MPEG-7 has set several principles to measure a shape descriptor, that is, good retrieval accuracy, compact features, general application, low computation complexity, robust retrieval performance and hierarchical coarse to fine representation [1]. In CBIR, good retrieval accuracy requires a shape descriptor be able to find perceptually similar shapes – rotated, translated, scaled and affinely transformed versions – from a database. The descriptor should also be able to find noise affected shapes, variously distorted shapes and defective shapes, which are tolerated by human beings when comparing shapes.

Many shape representation and description techniques have been developed in the past. A number of new techniques have been proposed in recent years. A comprehensive review of shape representation and description techniques is presented in [10-13]. In this section, we briefly review important shape description techniques in the context of CBIR. Shape representation and description techniques can be generally classified into two classes of methods: contour-based methods and region-based methods. The classification is based on whether shape features are extracted from the contour only or are extracted from the whole shape region.

### **2.3.1 Contour-Based Shape Descriptors**

Contour shape description techniques only exploit shape boundary information. The different methods are divided into structural approaches and global approaches. This sub-class



is based on whether the shape is represented by segments / sections (primitives) or represented as a whole.

### **2.3.1.1 Structural Techniques**

Structural approaches break the shape boundary into segments, called primitives using a particular criterion. The final representation is usually a string or a graph (or tree), the similarity measure is done by string matching or graph matching.

Many techniques exist utilizing Structural-based shape representation, such as chain code representation [14-16], polygon decomposition [17- 21], scale-space techniques [22, 23], smooth curve decomposition [24], shape invariants [25-28], and discrete segments [29]. They are not detailed here because of their lack of popularity.

### **2.3.1.2 Global Techniques**

In global techniques, shape is represented as a whole.

#### **2.3.1.2.1 Shape Signature**

A Shape signature represents a shape by a one dimensional function derived from shape boundary points. Many shape signatures exist, they include centroidal profile, complex coordinates, centroid distance, tangent angle, cumulative angle, curvature, area and chord-length [30-32].

#### **2.3.1.2.2 Directed Acyclic Graph**

A Directed Acyclic Graph (DAG) ordered tree is assigned to an object skeleton. The shape similarity measure is then based on the similarity of corresponding trees that is based on the matching algorithm for DAG ordered trees presented in [33].

#### **2.3.1.2.3 Elastic Matching**

Bimbo and Pala have proposed the use of elastic matching for shape based image retrieval [34]. According to this approach, a deformed template is generated as the sum of the original template and a warping deformation.

#### 2.3.1.2.4 Stochastic Method

Autoregressive (AR) modeling has been used for calculating shape descriptors [35 - 41]. Methods in this class are based on the stochastic modeling of a 1-D function obtained from the shape. The disadvantage of the AR method is that in the case of complex boundaries, a small number of AR parameters are not sufficient for an adequate description. Hidden Markov Models (HMMs) are also being explored as one of the possible shape modeling and classification frameworks [42 – 45]. HMM provides a probabilistic framework for training and classification. Arica and Vural [43] applied a circular HMM topology with 8 states to model the shape. Cai and Liu [44] applied a Fourier descriptor based HMM topology to classify the shapes. Bicego and Murino [45] proposed a curvature descriptor based on HMM. Curvatures are treated as mixtures of Gaussian and consequently an ergodic HMM is developed. Their work provides comprehensive results for classification with deformation, noise and occlusion. A weighted likelihood discriminant for minimum error shape classification is proposed in [46]. This uses a HMM for shape curvature as its 2-D shape descriptor.

#### 2.3.1.2.5 Curvature Scale Space (CSS)

Mokhtarian and Mackworth [47] proposed a scale space signature of the contour called Curvature Scale Space (CSS) [48–50]. The CSS descriptor is used in the MPEG-7 standard [51] and represents a multiscale organization of the curvature zero-crossing points of a planar curve. In this sense, the dimension of its feature vectors varies for different contours, thus a special matching algorithm is necessary to compare two CSS descriptors. A more detailed description of the technique is provided in Chapter 5.

#### 2.3.1.2.6 Fourier Descriptors

One of the most widely used shape description methods is Fourier Descriptor (FD) [30, 36, 52 – 65]. Conventional FD methods only deal with closed curve, however, Lin et al. and Mitchell et al. used FD to describe partial shapes [55, 58]. Arbter et al. introduced the affine-invariant FD to take into consideration of affine shape description [52]. Rauber proposed a UNL FD (named after Universidade Nova de Lisboa, Portugal) which is able to describe disjointed or articulated contour shape [60]. The UNL FD is acquired by applying 2-D Fourier transform on

the UNL transformed shape image. Even though a feature selection process is followed, the dimension of the feature vector acquired this way is very high. Richard and Hemami introduced a complex distance measurement, called the true distance measurement, for measuring the similarity between two set of FDs [61]. Since the true distance measurement requires two Fourier transforms for each matching, it involves 15 times more computation than a normal distance measurement. Rui et al. [62] proposed a distance measurement to classify similarity transformed characters using Fourier transformed coefficients. This distance measurement is the weighting sum of the variance of magnitude ratios and the variance of phase difference between two sets of Fourier coefficients. The Fourier coefficients are derived from Fourier reconstructed shape boundary rather than from original boundary. This is not different from FD derived from a smoothed boundary. Eichmann et al. proposed the use of a short-time Fourier descriptor (SFD) for shape description [66], however, Zhang and Lu have found that SFD is outperformed by conventional FD methods in shape retrieval [67]. This is because SFD cannot capture global shape features although it can capture local shape features more accurately. Zhang and Lu [32, 68] have found that for general shapes, the centroid distance function is the most desirable shape signature to derive FD. Kunttu and Lepisto proposed angular radius Fourier descriptor, which combines the boundary function with the directional angle of the boundary line [69].

### 2.3.1.2.7 Wavelet Descriptors

Wavelet descriptors are used in shape recognition and retrieval due to their multi-resolution nature and ability to maintain local shape features [70-79]. In [70, 71] a one-dimensional (1-D) discrete periodized wavelet transformation (DPWT) is applied to the contour for boundary representation. Li and Kuo [80] proposed a simplified method for defining a unique start point of the contour. A generalized uniqueness wavelet descriptor based on the generalized uniqueness property inhering in the one-dimensional (1-D) DPWT is proposed in [81]. Alferéz and Wang [82] have proposed geometric and illumination invariants that depend on the wavelet detail coefficients for object recognition. Tieng and Boles [83 - 87] have derived more than one affine invariant representation function by applying the dyadic WT to the contour of the shape. Also, Khalil and Bayoumi have derived a wavelet-based affine invariant function by using a dyadic wavelet transform [88-90]. In [91], Bala and Cetin proposed a computationally efficient method for computing Khalil's detail-based invariant functions by utilizing the decimated

wavelet transform. In [92], a new wavelet representation formula for rotation invariant feature extraction is proposed. Using the radial symmetry property feature vectors are generated, that are shown to be rotation invariant. Ibrahim *et. al.* proposed a new wavelet-based affine invariant functions for shape representation using approximation coefficients [93].

#### **2.3.1.2.8 Dynamic Programming**

Many dynamic programming (DP) based shape matching is reported in literature [94-99]. DP approaches are able to match the shapes part by part rather than point by point, and are robust to deformation and occlusion.

#### **2.3.1.2.9 Visual Parts**

A shape descriptor for non-rigid shapes with single closed contour is reported by Latecki *et. al.* [100, 101]. This is based on the correspondence of the visual parts of object contour and was a candidate descriptor in the MPEG-7 experiments.

#### **2.3.1.2.10 Multiscale Fractal Dimension**

The multiscale fractal dimension [102, 103] of a shape is computed based on the Euclidean distance transform (EDT) of its pixels. The EDT of these pixels is also related to their geometric Voronoi diagram [104], where each pixel defines an influence zone (discrete Voronoi region) composed by its closest image pixels. A new method for computing multiscale fractal dimension using the image foresting transform (IFT) is proposed in [105]. The IFT provides the simultaneous computation of the EDT and the discrete Voronoi regions in time usually proportional to the number of pixels [106], being more efficient than the method proposed in [102].

#### **2.3.1.2.11 Beam Angle Statistics (BAS)**

The BAS descriptor [107, 108] is based on the *beams* originated from a contour pixel. A beam is defined as the set of lines connecting a contour pixel to the rest of the pixels along the contour. At each contour pixel, the angle between a pair of lines is calculated, and the shape descriptor is defined by using the third-order statistics of all the beam angles in a set of

neighborhoods. The similarity between two BAS moment functions is measured by an optimal correspondent subsequence (OCS) algorithm [108].

#### 2.3.1.2.12 Tensor Scale Descriptor (TSD)

TSD [109] is a shape descriptor based on the tensor scale concept [110-112] — a morphometric parameter yielding a unified representation of local structure thickness, orientation, and anisotropy. That is, at any image point, its tensor scale is represented by the largest ellipse centered at that point and within the same homogeneous region. TSD is obtained by extracting the tensor scale parameters for the original image and then computing the ellipse orientation histogram. TSDs are compared by using a correlation-based distance function. A new descriptor based on tensor scale is proposed in [113] exploiting tensor scale orientation, which includes spatial information.

#### 2.3.1.2.13 Contour Saliences (CS)

The CS [114] computation uses the Image Foresting Transform [115] to compute the salience values of contour pixels and to locate salience points along the contour by exploiting the relation between a contour and its internal and external skeletons [116]. The contour salience descriptor consists of the salience values of salient pixels and their location along the contour, and on a heuristic matching algorithm as distance function. The convex contour salience (CCS) is the first descriptor of this category [105, 117]. The CCS of a contour is defined as the influence areas of its higher curvature convex points.

#### 2.3.1.2.14 Segment Saliences (SS)

The segment salience descriptor [114] is a variation of the contour salience descriptor which incorporates two improvements: the *salience values* of contour segments, in the place of salience values of isolated points, and another matching algorithm that replaces the heuristic matching by an optimum approach. The salience values along the contour are computed and the contour is divided into a predefined number of segments of the same size. The internal and external influence areas of each segment are computed by summing up the influence areas of their corresponding pixels.

#### 2.3.1.2.15 Shape Context

A shape descriptor, called the shape context [118], of a point  $p$  belonging to the contour of an object is a bivariate histogram in a log-polar coordinate system that gives the distribution of contour points in the surroundings of  $p$ . To compute the dissimilarity of two shapes from the point-wise dissimilarity matrix the Hungarian algorithm [119] is used.

#### 2.3.1.2.16 Distance Sets

Distance Set [120] is a rich local descriptor of an image point, which is determined by the spatial arrangement of image features around that point. A two-dimensional visual object is described by the set of (labeled) distance sets associated with the feature points of that object. Based on a dissimilarity measure between (labeled) distance sets and a dissimilarity measure between sets of (labeled) distance sets, the problem of shape matching is addressed.

#### 2.3.1.2.17 Curve Edit Distance

The correspondence between curves is based on a notion of an alignment curve which treats both curves symmetrically [121]. A similarity metric is defined based on the alignment curve using two intrinsic properties of the curve, namely, length and curvature. The optimal correspondence is found by an efficient dynamic-programming method both for aligning pairs of curve segments and pairs of closed curves, and is effective in the presence of a variety of transformations of the curve.

#### 2.3.1.2.18 Generative Models

Zhuowen Tu and Alan L. Yuille proposed an algorithm for shape matching and recognition based on a generative model on how one shape can be generated by the other [122]. This generative model allows for a class of transformations, such as affine and non-rigid transformations, and induces a similarity measure between shapes.

#### 2.3.1.2.19 Helmholtz's Equation

M. Zuliani *et. al.* proposed a new physically motivated curve descriptor based on the

solution of Helmholtz's equation [123]. The descriptor satisfies the six principles set by MPEG-7. Moreover this descriptor generalizes straightforwardly to three dimensional surfaces.

#### **2.3.1.2.20 WARP**

Ilaria Bartolini *et. al.* proposed a novel Fourier-based approach, called WARP, for matching and retrieving similar shapes [124]. The unique characteristics of WARP are the exploitation of the phase of Fourier coefficients and the use of the Dynamic Time Warping (DTW) distance to compare shape descriptors. While phase information provides a more accurate description of object boundaries than using only the amplitude of Fourier coefficients, the DTW distance permits to accurately match images even in the presence of (limited) phase shiftings.

#### **2.3.1.2.21 Morphological Curvature Scale Space**

Andrei C. Jalba *et. al.* proposed a multiscale, morphological method for the purpose of shape-based object recognition [125]. A connected operator similar to the morphological hat-transform is defined, and two scale-space representations are built, using the curvature function as the underlying one-dimensional signal. Each peak and valley of the curvature is extracted and described by its maximum and average heights and by its extent and represents an entry in the top or bottom hat-transform scale spaces.

#### **2.3.1.2.22 Poisson Based Shape Descriptor**

Lena Gorelick *et. al.* proposed a shape descriptor based on the Poisson equation [126]. The approach assigns, for every internal point of the silhouette, a value reflecting the mean time required for a random walk beginning at the point to hit the boundaries. This function can be computed by solving Poisson's equation, with the silhouette contours providing boundary conditions.

#### **2.3.1.2.23 Inner Distance**

The inner-distance is defined as the length of the shortest path between landmark points within the shape silhouette. It is articulation insensitive and more effective at capturing part

structures [127]. Three approaches are proposed for building shape descriptor using the inner-distance. The first method combines the inner-distance and multidimensional scaling to build articulation invariant signatures for articulated shapes. The second method uses the inner-distance to build a new shape descriptor based on shape contexts. The third one extends the second one by considering the texture information along shortest paths.

### **2.3.2 Region-Based Shape Descriptors**

In region based techniques, all the pixels within a shape region are taken into account to obtain the shape representation, rather than only use boundary information as in contour base methods. Similar to contour based methods, region based shape methods can also be divided into structural methods and global, depending on whether they separate shapes into sub parts or not.

#### **2.3.2.1 Structural Techniques**

The convex hull of a region is used as a region shape descriptor. The extracting of the convex hull can use both boundary tracing method [25] and morphological methods [31, 128]. The polygonal approximation is particularly attractive to smooth a boundary prior to partitioning.

Another important region-based shape descriptor is medial axis transform (MAT) proposed by Blum [129]. The medial axis is the locus of centers of maximal disks that fit within the shape. The skeleton can then be decomposed into segments and represented as a graph according to certain criteria. The medial axis obtained by Morse is computed from scale space [130].

#### **2.3.2.2 Global Techniques**

##### **2.3.2.2.1 Geometric Moment Invariants**

Hu published the first significant paper on the use of geometric moment invariants for 2D pattern recognition applications [131]. Using nonlinear combinations of the lower order moments, a set of moment invariants which has the desirable properties of being invariant under



translation, scaling and rotation, are derived. Since the values of the acquired moment invariants are usually very small, a normalization process, such as zscore normalization [132], is needed in the implementation. Geometric moment invariants have attracted wide attention [25, 128, 133-136] and have been used in many applications [10, 137-139]. The main problem with geometric moments is that only a few invariants derived from lower order moments is not sufficient to accurately describe shape and higher order invariants are difficult to derive.

#### 2.3.2.2.2 Algebraic Moment Invariants

Algebraic moment invariants have been introduced by Taubin and Cooper [140, 141]. The algebraic moment invariants are computed from the first  $m$  central moments and are given as the eigenvalues of predefined matrices whose elements are scaled factors of the central moments. Different from Hu's geometric moment invariants, the algebraic moment invariants can be constructed up to any arbitrary order and are invariant to affine transformations. Algebraic moment invariants tend to perform well as region-based descriptor and have poor performance as contour-based descriptor [142].

#### 2.3.2.2.3 Differential Geometry Based Techniques

Differential geometry based techniques are reported to recognize and locate partially occluded two-dimensional (2-D) objects [143]. For recognition of partially occluded three-dimensional (3-D) objects, the principal curvatures, mean curvature and Gaussian curvature are used as the local descriptions of the surfaces [144]. Dutta Majumder *et. al.* addressed the problem of constructing a similarity measure between shapes of 3-D objects by defining a shape distance between 3-D objects on the basis of certain characteristic planes of the objects [145]. A new shape based technique for classification and registration is proposed in [146,147].

#### 2.3.2.2.4 Generic Fourier Descriptor

Generic Fourier Descriptor (GFD) is proposed by Zhang and Lu in [148]. The GFD is acquired by applying a 2-D Fourier transform on a polar-raster sampled shape image. With an enhanced process, GFD can achieve retrieval performance on perspectively transformed shapes

as high as it achieves on similarity transformed shapes [149]. Recently a composite descriptor is derived from GFD of the shape region and the shape contour [150].

#### 2.3.2.2.5 Grid Based Method

The grid shape descriptor is proposed by Lu and Sajjanhar [151] and has been used in [152,153]. Basically, a grid of cells is overlaid on a shape and the grid is then scanned from left to right and top to bottom. The result is a bitmap. The cells covered by the shape are assigned 1 and those not covered by the shape are assigned 0. The shape can then be represented as a binary feature vector. The binary Hamming distance is used to measure the similarity between two shapes. Chakrabarti *et. al.* [152] improve grid descriptor by using an adaptive resolution representation.

The advantages of the grid descriptor are its simplicity in representation, conformance to intuition, and also agreement with shape coding method in MPEG-4. The main problem with this method is the major-axis based rotation normalization.

#### 2.3.2.2.6 Shape Matrix

Goshtasby proposed the use of a shape matrix which is derived from a circular raster sampling technique [154]. A polar raster of concentric circles and radial lines is overlaid in the center of the mass. The binary value of the shape is sampled at the intersections of the circles and radial lines. The shape matrix is formed so that the circles correspond to the matrix columns and the radial lines correspond to the matrix rows. Prior to the sampling, the shape is scale normalized using the maximum radius of the shape. The result matrix representation is invariant to translation, rotation, and scaling. Taza and Suen represent shape using a weighed shape matrix which gives more weight to peripheral samples [155].

#### 2.3.2.2.7 Shape Vector

Perui *et. al.* proposed a shape description based on the relative areas of the shape contained in concentric rings located in the shape center of the mass [156]. The area ratio descriptor ignores the shape distribution within the measured ring. As the result, it loses local information of a shape.

### **2.3.2.2.8 Zernike Moment Descriptor**

Teague [135] has proposed the use of orthogonal moments to recover the image from moments based on the theory of orthogonal polynomials, and has introduced Zernike moments, which allow independent moment invariants to be constructed to an arbitrarily high order. The complex Zernike moments are derived from Zernike polynomials. Zernike moment descriptor has been widely used as a shape descriptor [157-159]. It was major candidate as region-based descriptor in early MPEG-7 experiments [160]. A more detailed description of Zernike moment descriptor is presented in Chapter 5.

### **2.3.2.2.9 Angular Radial Transformation (ART)**

The 2-D Angular Radial Transformation (ART) [4] is the MPEG-7 proposed region-based shape descriptor [51]. It belongs to the broad Zernike Moment family and provides a compact and efficient way to express pixel distribution within a 2-D object region. The set of orthogonal moment basis is defined on a unit disc in polar coordinates. As the ART and Zernike moments belong to the same family, they have similar transformation invariance properties. A more detailed description of ART descriptor is presented in Chapter 5.

### **2.3.2.2.10 Polar Raster Sampling (PRS) Signature**

PRS [161] uses contour based method for region based shape representation and retrieval. A polar raster grid is overlaid over the shape image, and the number of shape pixels on each of the concentric circles and on each of the diameters of the polar sampling grid is computed. The number of pixels is a function of the radius and the angle. The PRS is translation, rotation and scaling invariant and computationally less expensive.

## **2.4 Review of Relevance Feedback Techniques**

CBIR systems operate on features extracted automatically from the images, such as color, texture and shape. The retrieval accuracy is often affected due to the semantic gap in feature extraction. Furthermore, the perception of similarity is user and context subjective. Combining different features – color, texture or shape – to form a coherent query representing the image

sought is a difficult task. In CBIR systems, image is represented using a high dimensional feature vector and uses feature dependent similarity measures, which are not simple to tune manually.

These problems are addressed by relevance feedback (RF) techniques to allow the system to learn from the users' interaction with the retrieval results. Relevance feedback was introduced in the late 1960's for relevance document retrieval [162] and was found to be effective in text-based information retrieval [163, 164]. In these systems, RF allows the user to interact with the retrieval results of a query by selecting terms from the documents he considers relevant or irrelevant to modify the original query. Therefore, the key issue in relevance feedback is how to learn from the user's feedback and use the positive and negative examples to refine the query or adjust the similarity measure. In general, relevance feedback process in CBIR is as following. For a given query, the CBIR system first retrieves a list of ranked images according to a predefined similarity metrics, often defined by the distance between query vector and feature vectors of images in a database. Then, the user selects a set of positive and/or negative examples from the retrieved images, and the system will refine the query and retrieve a new list of images. Hence, the key issue in relevance feedback approaches is how to incorporate positive and negative examples in query and/or the similarity refinement.

Initially developed in document retrieval, relevance feedback was transformed and introduced into content-based multimedia retrieval, mainly CBIR, during early and mid 1990s [165-167]. RF has been an active research area in CBIR and many new techniques are proposed in the literature. A comprehensive review of RF techniques is presented in [168]. An overview of important techniques of RF in CBIR is presented in this section.

### **2.4.1 Classical Schemes**

The early relevant feedback schemes for CBIR have been mainly adopted from text document retrieval researches and can be classified into two approaches: query point movement (query refinement) and re-weighting (similarity measure refinement). Both have been built based upon the vector model in information retrieval theory [162, 169, 170].

The query point movement method essentially tries to improve the estimate of the "ideal query point" by moving it towards good example points and away from bad example points. The frequently used technique to iteratively improve this estimation is the Rocchio's formula

[162]. This is the technique implemented in MARS system [171]. Two factors, “component importance” and “inverse collection importance”, were proposed for images in accordance to the factors “term frequency” and “inverse document frequency” in text retrieval. The vector space model was used for relevance feedback. They also used Gaussian normalization to put equal emphasis on each feature component, and then used the inverse of the standard deviation of each component for the images in the relevant feedback set as weights. They concluded that the approach adopted from text retrieval performed better than Gaussian normalization but the latter was more robust to unknown feature components. Another way of using the feedback to refine the similarity measure was proposed in [172], where a set of similarity measures is pre-defined, and the system selected the similarity measure which minimized the sum of the differences between the ranks of the retrieved images and the ranks of the relevant images selected by the user. A simple algorithm based on the re-weighting method is described in the ImageRover system [173]. This algorithm automatically selects appropriate Minkowski distance metrics that minimize the mean distance between the relevant images specified by the user.

Another implementation of point movement strategy is using the Bayesian method, such as the work by Vasconcelos and Lippman [174] wherein Bayesian learning is used to incorporate user’s feedback to update the probability distribution of all the images in the database.

#### 2.4.2 As an Optimization Problem

Some researchers considered RF as a distance optimization problem whose solutions are the parameters that make it possible to find the ideal query, weight the features, and transform the feature space into a new one that corresponds better to the user. The MindReader retrieval system designed by Ishikawa *et. al.* [175] formulates a minimization problem on the parameter estimating process. Unlike traditional retrieval systems whose distance function can be represented by ellipses aligned with the coordinate axis, the MindReader system proposed a distance function that is not necessarily aligned with the coordinate axis. Therefore, it allows for correlations between attributes in addition to different weights on each component. A further improvement over the MindReader approach is given in [176], where, optimal query estimation and weighting functions are derived by a unified framework, based on the minimization of the total distances of the positive examples from the updated query. The weighted average and a whitening transform in the feature space were found to be the optimal solutions. Discriminant

Analysis-based methods either find a low-dimensional subspace of the feature space, such that the positive and negative samples are well separated after projection to this subspace [177] or define a  $(1+x)$ -class problem (biased discriminant analysis (BDA)) and find a subspace within which to discriminate the one positive class and the unknown number of negative samples classes [178]. More recently, the direct kernel biased discriminant analysis (DKBDA) was developed and reported to outperform the BDA in both linear space and the nonlinear kernel space [179].

### 2.4.3 As a Learning Problem

Other researchers consider RF as a learning problem in which samples fed back by the user are used to train a model, which is then used for retrieval. Probabilistic Model-based methods use entropy to minimize the expected number of iterations [180, 181]. A Bayesian framework is used in [180] to predict what target image users want, given the action they undertook. This is done via a probability distribution over possible image targets, rather than by refining a query. Furthermore, this model tries to minimize the number of feedback iterations by maximizing the information obtained from a user at each feedback iteration using an entropy-minimizing algorithm. Chiou-Ting Hsu *et. al.* proposed a generalized Bayesian framework for RF in CBIR [182]. The proposed feedback technique is based on the Bayesian learning method and incorporates a time-varying user model into the formulation. In [183], the authors used decision tree learning. They proposed an RF model which, for each retrieval iteration, learns a decision tree to uncover a common thread uniting all images marked as relevant. This tree is then used as a model for inferring which of the unseen images the user would most likely desire. Neural networks have been adopted in interactive image retrieval in view of their learning capability and generalization power [184-189]. Radial basis function (RBF) networks have been used to determine the nonlinear relationship between features so that a more accurate similarity comparison between images can be supported [184]. In [185], self-organizing maps (SOMs) are used to measure similarity between images. A separate SOM is trained for each feature vector type, then the system adapts to the user's preferences by returning him more images from those SOMs where his responses have been most densely mapped. An adaptive radial basis function network (ARBFN) has been proposed for interactive image retrieval [186]. A radial-basis function (RBF) network for implementing an adaptive metric which progressively models the

notion of image similarity through continual relevance feedback from users is proposed in [187]. A fuzzy radial basis function network (FRBFN) has been proposed to learn the users' fuzzy perceptions of visual contents using fuzzy relevance feedback [188,189]. It provides a natural way to model the user interpretation of image similarity. Peng-Yeng Yin *et. al.* proposed an image relevance reinforcement learning (IRRL) model for integrating existing RF techniques in a CBIR system [190]. Various integration schemes are presented and a long-term shared memory is used to exploit the retrieval experience from multiple users.

#### 2.4.4 As a Classification Problem

More recently, some researchers have considered RF as a classification problem in which sample images provided by the user are employed to train a classifier, which is then used to classify the database into images that are relevant to the query and those that are not. In [191], the authors developed a Bayesian model which supports image classes that assign a high membership probability to positive example images and penalizes classes that assign a high membership probability to negative example images. In [192], Meilhac *et al.* consider that the image collection is made up of relevant images, among which the user chooses the positive examples, and non-relevant images, among which the user chooses the negative examples. They use a Bayesian model in which they try to estimate the distribution of relevant images and simultaneously minimize the probability of retrieving non-relevant images. Bayesian classifier has also been developed to perform retrieval based on the feedback samples. Positive examples are used to estimate a Gaussian distribution that represents the desired images for a given query, while the negative examples are used to modify the ranking of the retrieved candidates. In [193], Su *et al.* present a Bayesian classifier in which the positive example is used to estimate the Gaussian distribution that represents the class of sought images, while the negative example is used to modify the ranking of the retrieved candidates. They use principal component analysis (PCA) to perform a dimension reduction and work in proper subspaces. The features can be selected by the boosting technique in which a strong classifier is obtained as a weighted sum of weak classifiers along the different feature dimensions [194]. Support Vector Machines (SVM) have also been widely used in RF. SVM-based methods [195] either estimate the density of positive instances [196] or regard RF as a classification problem with the positive and negative samples as training sets [197-203]. In [199], the authors combine the random subspace method

with SVM in order to improve the performance of the classifier. SVM active learning [200], which plays an important role in CBIR RF research, selects the samples near the SVM boundary and queries the user for labels. The points near the SVM boundary are regarded as the most informative images while the most-positive images are the farthest ones from the boundary on the positive side. Recently, SVM active learning is also combined with a multimodal concept-dependent process for CBIR, Constrained Similarity Measure (CSM) [204]. CSM-based SVM [201] learns a boundary that separates all the images in the database into two clusters and the images inside the boundary are ranked by their Euclidean distances to the query image. Derived from one-class SVM [196], in [205] a Biased SVM (BSVM) is proposed, which can better model the relevance feedback problem and reduce the performance degradation caused by the imbalanced data set problem, i.e., the number of the positive feedback samples is much less than the number of the negative feedback samples. Tao *et. al* proposed an asymmetric bagging-based SVM (AB-SVM) to address the imbalanced data set problem. To address the over-fitting because the number of feature dimensions is much higher than the size of the training set, the random subspace method and SVM are combined for relevance feedback, which is named random subspace SVM (RS-SVM). Finally, by integrating AB-SVM and RS-SVM, an asymmetric bagging and random subspace SVM (ABRS-SVM) is proposed [206]. The small sample training problem is addressed by Kui Wu *et. al* by pseudo-labeling using a fuzzy rule by proposing fuzzy SVM [207]. Jing Li *et. al.* proposed a new machine learning technique, multi-training SVM (MTSVM), which combines the merits of the co-training technique and a random sampling method in the feature space to address the over-fitting and small sample training problem [208].

#### 2.4.5 General Techniques

Ye Lu *et. al.* developed a relevance feedback framework to take advantage of the semantic contents of images in addition to low-level features by forming a semantic network on top of the keyword association on the images [209]. Azadeh Kushki *et. al.* proposed method for interactive image retrieval using query feedback [210]. Query feedback learns the user query as well as the correspondence between high-level user concepts and their low-level machine representation by performing retrievals according to multiple queries supplied by the user during the course of a



retrieval session. Junwei Han *et. al.* reported a framework for effective image retrieval by memory learning [211]. It forms a knowledge memory model to store the semantic information by simply accumulating user-provided interactions. A learning strategy is then applied to predict the semantic relationships among images according to the memorized knowledge. Wei Jiang *et. al.* investigated online feature selection in the relevance feedback learning process to improve the retrieval performance of the region-based image retrieval system [212]. Kherfi *et. al.* presented a new RF framework based on a feature selection algorithm that nicely combines the advantages of a probabilistic formulation with those of using both the positive examples and negative examples [213]. Through interaction with the user, the algorithm learns the importance he assigns to image features, and then applies the results obtained to define similarity measures that correspond better to his judgment. In CBIR, it is crucial to effectively discover users' concept patterns through an acquired understanding of the subjective role played by humans in the retrieval process. Chen *et. al.* proposed multiple instance learning for RF to discover users concept patterns [214]. Hoi *et. al.* proposed a unified framework for log-based relevance feedback that integrates the log of feedback data into the traditional relevance feedback schemes to learn effectively the correlation between low-level image features and high-level concepts [215]. In RF all positive feedbacks share a homogeneous concept while negative feedbacks do not. Dacheng Tao *et. al.* proposed a scheme for negative sample analysis in RF, in which positive samples are described by a single hypersurface and negative samples are split into a number of subsets, each one of which has a simple distribution [216]. An orthogonal complement component analysis technique is also proposed in [217] to address this issue.

## 2.5 Review of Major CBIR Systems for Medical Images

Although CBIR of general images has matured and a standard (MPEG-7) has been developed for coordination of research, content-based retrieval for medical images is still in its infancy. CBIR has been proposed for various medical applications. However, due to the nature of medical images, content-based retrieval for medical images is still faced with challenges.

Tagare *et. al.* pointed out some of the unique challenges confronting retrieval engines with medical image collections [218]. Low resolution and strong noise are two common characteristics in most medical images [219]. With these characteristics, medical images cannot be precisely segmented and extracted for the visual content of their features. In addition,

medical images obtained from different scanning devices may display different features, though some approaches to image correction and normalization have been proposed [220]. Medical images are digitally represented in a multitude of formats based on their modality and the scanning device used [221]. Another characteristic of medical images is that many images are represented in gray level rather than color. Even with the change of intensity, monochrome may fail to clearly display the actual circumstance of lesion area. Furthermore, medical image interpretation is a complex and poorly understood process.

Common systems for CBIR have only a rudimentary understanding of image content. Such systems make no distinction between important and unimportant features or between multiple objects in the image. The features used for automated indexing characterize the entire image rather than unique regions or objects. In contrast, queries of medical or diagnostic relevance include searching for organs, their relative locations, and other distinct features such as morphological appearances. Therefore, common CBIR-systems cannot guarantee a meaningful query completion when used within the medical context [222]. Therefore, the results are rather poor when common CBIR-systems are used to retrieve medical images [223, 224]. For CBIR of medical images, specialized solutions must be developed depending on the image modality, anatomical region and pathology of interest. Often, human-in-the-loop (a physician-in-the-loop, more specifically) approach in which the human delineates the pathology bearing regions (PBR) and a set of anatomical landmarks of the image, at the time the image is entered into the database [225]. The content-based retrieval systems for medical images can be broadly classified into two: systems for general medical categorization of images and specialized systems for images of specific organ and modality. Although content-based image retrieval has frequently been proposed for use in medical image management, only a few content-based retrieval systems have been developed specifically for medical images. A detailed review of medical CBIR systems is given in [226, 227]. This section gives an overview of the currently available literature on content-based image retrieval in the medical domain.

### **2.5.1 General Systems**

This section provides an overview of non-specialized, in terms of image type, CBIR systems in the medical field.

### 2.5.1.1 KmED

Knowledge based medical database (KmED) system utilizes semantic modeling focusing on object shapes and spatial relationship between them [228]. A knowledge-based approach to retrieve medical images by feature and content with spatial and temporal constructs was developed by Chu *et. al.* [229]. Selected objects of interest in a medical image (e.g., x-ray, MR image) are segmented, and contours are generated from these objects. Features (e.g., shape, size, texture) and content (e.g., spatial relationships among objects) are extracted and stored in a feature and content database. A knowledge-based semantic image model is proposed that consists of four layers (raw data layer, feature and content layer, schema layer, and knowledge layer) to represent the various aspects of the characteristics of an image. The model provides a mechanism for accessing and processing spatial, evolutionary, and temporal queries.

### 2.5.1.2 COBRA

COBRA [230] (COnent-Based Retrieval Architecture) is an open architecture for PACS based on the widely used health care and technology standards. In addition to regular PACS components, COBRA includes additional components to handle representation, storage, and content-based similarity retrieval. Within COBRA, an anatomy classification algorithm is introduced to automatically classify PACS studies based on their anatomy. Such a classification allows the use of different segmentation and image-processing algorithms for different anatomies. COBRA uses primitive retrieval criteria such as color, texture, shape, and more complex criteria including object-based spatial relations and regions of interest.

### 2.5.1.3 medGIFT

The medGIFT [231,232] is based on the open source image retrieval system (GIFT - GNU Image Finding Tool) for the retrieval of medical images in the medical case database system CasImage that is used in daily, clinical routine in the university hospitals of Geneva. The CasImage system, works on a variety of images from CT, MRI, and radiographs, to color photos and is integrated to PACS [227, 233]. The medGIFT retrieval system extracts global and regional color and texture features, including 166 colors in the HSV color space, and Gabor

filter responses in four directions each at three different scales. Combinations of textual labels and visual features are used for medical image retrieval.

#### **2.5.1.4 IRMA**

The IRMA (Image Retrieval in Medical Applications) system supports image database of varying image modalities [224, 234-237]. This system is implemented as a platform for content-based image retrieval in medical applications. To enable complex content understanding the IRMA concept is based on a conceptual and algorithmic separation of seven processing steps

- categorization with respect to image modality, anatomic region, function system, and body orientation using global image features
- registration in geometry and contrast for each likely category
- feature extraction using local features
- feature selection and combination with respect to category and query content
- indexing resulting in a hierarchical multi-scale blob representation
- identification of blobs by linking a-priori knowledge to image content
- retrieval processed on the abstract blob-level

### **2.5.2 Specialized Systems**

#### **2.5.2.1 ASSERT**

The ASSERT (Automatic Search and Selection Engine with Retrieval Tools) system uses a physician-in-the-loop approach for retrieving images of High-Resolution Computed Tomography (HRCT) of the lung [238-241]. This approach requires users to delineate the pathology-bearing regions and identify certain anatomical landmarks for each image. This system extracts 255 features of texture, shape, edges, and gray-scale properties in the pathology-bearing regions. In this system, lobular feature sets (LFS) on HRCT images are translated into an index for archiving and retrieval. A multi-dimensional hash table for the LFS classes is constructed for the system. A decision tree algorithm is used to construct a minimum entropy partition of the feature space where the LFS classes reside. After translating a decision tree to a hash table, the system prunes the set of retrieved LFS classes and candidate images.

### **2.5.2.2 WebMIRS**

The WebMIRS (Web-based Medical Information Retrieval System) will allow access to databases containing text and images and will allow database query by standard Structured Query Language (SQL), by image content, or by a combination of the two [242-249]. The Lister Hill National Center for Biomedical Communications, an R&D division of the U.S. National Library of Medicine, maintains a collection of 17,000 digitized spine X-rays from the Second National Health and Nutritional Examination Survey (NHANES II). The WebMIRS system is evaluated using this database, and the pathology of interest was anterior osteophyte.

This system contains the Active Contour Segmentation (ACS) tool, which allows the users to create a template by marking points around the vertebra. If the segmentation of a template is accepted, the ACS tool will estimate the location of the next vertebra, place the template on the image, and then segment it. In data representation, a polygonal approximation process is applied for eliminating insignificant shape features and reducing the number of data points. The data obtained in the polygonal approximation process represent the shape of vertebra. Then, the approximated curve of the vertebra is converted to tangent space for similarity measurement. Dynamic Programming (DP) in Partial Shape Matching (PSM) techniques are used for shape based retrieval [250].

A linear weight-updating approach for RF has been proposed for improving the retrieval accuracy of vertebra shapes [251].

### **2.5.2.3 I-Browse**

I-Browse project [252] was aimed at supporting intelligent retrieval and browsing of histological images, obtained along the gastrointestinal tract. With the help of knowledge bases and reasoning engines, high-level semantic attributes of images are obtained and textual annotation of images are automatically generated in this system.

### **2.5.2.4 IGDS**

Image Guide IGDS [253] (Image Guided Decision Support) is a CBIR system in the domain of cytopathology images. In this system, a classification-based approach is performed to detect different types of blood cells based on the properties of cell nucleus.

### 2.5.2.5 VOI-FIRS

VOI-FIRS (Volume Of Interest based Functional Image Retrieval System) is based on multidimensional feature extraction and retrieval [254]. It is volume of interest (VOI) based content-based retrieval of four-dimensional (three spatial and one temporal) dynamic positron-emission Tomography (PET) images. By segmenting the images into VOIs consisting of functionally similar voxels (e.g., a tumor structure), multidimensional visual and functional features were extracted and used as region-based query features.

### 2.5.2.6 Other Work

Korn *et. al.* described a system for fast and effective retrieval of tumor shapes in mammogram X-rays [255]. Other works for retrieval of digital mammogram images are reported in [256-258]. Ultrasound images of the breast are used in [259]. Recently, Issam El-Naqa *et. al.* proposed a similarity learning approach to CBIR of digital mammograms based on the cluster of micro calcifications [260]. Neural networks and support vector machines are used to predict the user's notion of similarity. A hierarchal learning approach is used, which consists of a cascade of a binary classifier and a regression module to optimize retrieval effectiveness and efficiency.

Swett and Miller developed a rule-based expert system was developed to display chest radiographs from a library of images as illustrative examples for helping radiologists' diagnosis [261]. Kelly and Cannon proposed a retrieval method based on texture, and shape analysis was applied for search and retrieval of a database containing pulmonary computed tomography (CT) images [262]. Guimond and Subsol described an algorithm for retrieval of 3-D magnetic resonance images based on anatomical structure matching [263]. Liu and Dellaert developed a similarity metric based on Bayes decision theory for retrieval of neuroradiological CT images [264]. In [265] and [266], a technique was developed that reduces high-dimensional data to a two-dimensional feature space in which images that are close to each other are selected for purposes of visualizing relationships in the data. Kawata *et. al.* developed a retrieval method using correlation coefficients in a database of pulmonary nodules represented by the joint histogram of the pattern CT density and 3-D curvature shape index [267]. Zhao *et. al.* used Gabor texture descriptor for retrieval of liver CT images [268]. A CBIR system for osteo-

articular MRI application based on a semi-supervised learning method using mixture models is proposed by Najjar *et. al.* [269].

## 2.6 Similarity Measures

In CBIR systems, image features are in general organized into  $n$ -dimensional feature vectors. Thus the query image and the database images can be compared by evaluating the distance between their corresponding feature vectors. Both metric and non-metric measures have been used in CBIR systems. Statistical distances such as the Mahalanobis distance have also been used [270].

Often different image features are indexed separately, thus similarity scores can be computed independently for each feature. Then the overall similarity score is obtained as a linear combination of these scores [271, 272]. The weights of this linear combination may be specified by the user or automatically adjusted by the system based on the feedback of the user [273].

Specific distances have been defined for specific features: e.g. for histograms commonly used measures are the histogram difference, histogram intersection [274, 275] or the quadratic distance [273, 276, 277]. The latter tries to account for the perceptual difference between any pair of bins in the histogram. The Hausdorff distance has been used to compare histograms as well as shapes in [278].

Many similarity measures are based on the  $L_p$  distance between two points in the  $n$ -dimensional feature space. For two points  $x, y$  the  $L_p$  distance is defined as

$L_p(x, y) = \left( \sum_{i=0}^n |x_i - y_i|^p \right)^{1/p}$ , called Minkowski distance. For  $p = 2$  we get the Euclidean distance and for  $p = 1$  we get the Manhattan, city block, or taxicab distance. The  $L1$  and  $L2$  norms are analyzed in [272] and [279], and their performances are compared.

The retrieval performance of a system depends on the agreement between the similarity measure used and human judgments of similarity, since the end consumer of CBIR results is a human. Therefore, several measures in accordance with the human perception have been developed in [280-288]. A good review of similarity measures for shape matching is presented in [289].

## References:

- [1] *Introduction to MPEG-7: Multimedia Content Description Interface*, B. S. Manjunath, P. Salembier, and T. Sikora, Eds., John Wiley & Sons Ltd., 2002.
- [2] J. M. Martínez, *MPEG-7 Overview (version 10)*, Technical Report, ISO/IEC JTC1/SC29/WG11 N6828, Palma de Mallorca, October 2004.
- [3] D. C. Fallside, *XML Schema Part 0: Primer*, Technical report, W3C Recommendation, <http://www.w3.org/TR/xmlschema-0>, May 2001.
- [4] W.-Y. Kim, Y.-S. Kim, "A new Region-Based Shape Descriptor," ISO/IEC MPEG99/M5472, Maui, Hawaii, Dec. 1999.
- [5] S. Abbasi, F. Mokhtarian, and J. Kittler, "Curvature Scale Space Image in Shape Similarity Retrieval," *Springer Journal of Multimedia Systems*, vol. 7, no. 6, pp. 467–476, 1999.
- [6] J. J. Koenderink and A. J. van Doorn, "Surface shape and curvature scale," *Image and Vision Computing*, vol. 10, no. 8, pp. 557–565, 1992.
- [7] K. Muller, J. -R. Ohm, J. Cooper, and M. Bober, "Results of 2D/3D shape core experiments MS-4," *ISO/IEC MPEG99/M6190*, Beijing, China, July 2000.
- [8] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 1, pp. 103-108, 1990.
- [9] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, pp. 72-86, 1991.
- [10] B. M. Mehtre, M. S. Kankanhalli, and W. F. Lee, "Shape Measures for Content Based Image Retrieval: A Comparison," *Information Processing and Management*, vol. 33, no. 3, pp. 319–337, 1997.
- [11] Loncaric S, "A survey of shape analysis techniques," *Pattern Recognition*, vol. 31, no. 8, pp. 983-1001, 1998.
- [12] L.F. Costa and R.M. Cesar Jr. *Shape Analysis and Classification: Theory and Practice*. CRC Press, Boca Raton, FL, USA, 2001.
- [13] Dengsheng Zhang, and Guojun Lu, "Review of shape representation and description techniques," *Pattern Recognition*, vol. 37, pp. 1–19, 2004.
- [14] Freeman H., "On encoding arbitrary geometric configurations," *IRE Trans. Electronic Computers*, vol. 10, pp. 260-268, 1961.
- [15] H. Freeman, and A. Saghri, "Generalized chain codes for planar curves," in *Proc. of the Fourth Int'l Joint Conf. on Pattern Recognition*, Kyoto, Japan, pp. 701–703, 1978.
- [16] J. Iivarinen, and A. Visa, "Shape recognition of irregular objects," In *Intelligent Robots and Computer Vision XV: Algorithms, Techniques, Active Vision, and Materials Handling, Proc. SPIE 2904*, D.P. Casasent (Ed.), pp. 25–32., 1996
- [17] W.I. Groskey, and R. Mehrotra, "Index-based object recognition in pictorial data management," *Comput. Vision Graphics Image Process.*, vol. 52, pp. 416–436, 1990.
- [18] W.I. Groskey, P. Neo, and R. Mehrotra, "A pictorial index mechanism for model-based matching," *Data Knowledge Eng.*, vol. 8, pp. 309–327, 1992.
- [19] R. Mehrotra, and J.E. Gary, "Similar-shape retrieval in shape data management," *IEEE Comput.*, vol. 28, no. 9, pp. 57–62, 1995.
- [20] Ray B., and K. Ray, "A new split-and-merge technique for polygonal approximation of chain coded curves," *Pattern Recognition Letters*, vol. 16, pp. 161-169, 1995.
- [21] Jain R., R. Kasturi and B.G. Schunck, *Machine Vision*, McGraw Hill, pp: 188-210, 1995.
- [22] Witkin A.P., Scale-space filtering, in *Proc. 8<sup>th</sup> Int'l. Joint Conf. Artificial Intelligence*, pp. 1019-1022, 1983.
- [23] G. Dudek, and J.K. Tsotsos, "Shape representation and recognition from multiscale curvature," *Comput. Vision Image Understanding*, vol. 68, no. 2, pp. 170–189, 1997.



- [24] S. Berretti, A.D. Bimbo, and P. Pala, "Retrieval by shape similarity with perceptual distance and effective indexing," *IEEE Trans. Multimedia*, vol. 2, no. 4, pp. 225–239, 2000.
- [25] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis and Machine Vision*, Chapman & Hall, London, UK, NJ, pp. 193–242, 1993.
- [26] C.-L. Huang, and D.-H. Huang, "A content-based image retrieval system," *Image Vision Comput.*, vol. 16, pp. 149–163, 1998.
- [27] S.Z. Li, Shape matching based on invariants, in *Shape Analysis, Progress in Neural Networks*, O. Omidvar (Ed.), Vol. 6, Ablex, Norwood, NJ, pp. 203–228, 1999.
- [28] D.M. Squire, and T.M. Caelli, "Invariance signature: characterizing contours by their departures from invariance," *Comput. Vision Image Understanding*, vol. 77, pp. 284–316, 2000.
- [29] Nghan D. Salih, David Chek Ling Ngo, and Hakim Mellah, "2D Object Description with Discrete Segments," *Journal of Computer Science*, vol. 2, no. 7, pp. 572-576, 2006.
- [30] P.J. Van Otterloo, *A Contour-Oriented Approach to Shape Analysis*, Prentice-Hall International (UK) Ltd, Englewood Cliffs, NJ, pp. 90–108, 1991.
- [31] E.R. Davies, *Machine Vision: Theory, Algorithms, Practicalities*, Academic Press, New York, pp. 171–191, 1997.
- [32] D.S. Zhang, and G. Lu, "A comparative study of Fourier descriptors for shape representation and retrieval," in *Proc. of the Fifth Asian Conf. on Computer Vision (ACCV02)*, Melbourne, Australia, pp. 646–651, 2002.
- [33] I.-J. Lin and S. Y. Kung, "Coding and comparison of DAGs as a novel neural structure with application to on-line handwritten recognition." *IEEE Trans. Signal Processing*, vol. 45, No. 11, pp. 2701-2708, 1997.
- [34] A. Del Bimbo, and P. Pala, "Visual image retrieval by elastic matching of user sketches," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 121–132, 1997.
- [35] K.L. Kashyap, and R. Chellappa, "Stochastic models for closed boundary analysis: representation and reconstruction," *IEEE Trans. Inform. Theory*, vol. 27, pp. 627–637, 1981.
- [36] R. Chellappa, and R. Bagdazian, "Fourier coding of image boundaries," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 1, pp. 102–105, 1984.
- [37] S.R. Dubois, and F.H. Glanz, "An autoregressive model approach to two-dimensional shape classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, pp. 627–637, 1986.
- [38] M. Das, M.J. Paulik, and N.K. Loh, "A bivariate autoregressive modeling technique for analysis and classification of planar shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 1, pp. 97–103, 1990.
- [39] K. Eom, and J. Park, "Recognition of shape by statistical modeling of centroidal profile," in *Proc. of the Tenth Int'l Conf. on Pattern Recognition*, vol. 1, Atlantic City, NJ, pp. 860–864, 1990.
- [40] Y. He, and A. Kundu, "2—D shape classification using hidden Markov model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 11, pp. 1172–1184, 1991.
- [41] I. Sekita, T. Kurita, and N. Otsu, "Complex autoregressive model for shape recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, pp. 489–496, 1992.
- [42] A. Fred, J. Marques, and P. Jorge, "Hidden markov models vs. syntactic modeling in object recognition," in *Proc. of IEEE Int'l Conf. of Image Processing*, vol. 1, pp. 893– 896, 1997.
- [43] N. Arica and F. Yarman-Vural, "A shape descriptor based on circular hidden markov model," in *Proc. of IEEE Int'l Conf. on Pattern Recognition*, vol. 1, 2000, pp. 924–927.
- [44] J. Cai and Z.-Q. Liu, "Hidden markov models with spectral features for 2d shape recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 12, pp. 1454–1458, 2001.
- [45] M. Bicego and V. Murino, "Investigating hidden markov models' capabilities in 2d shape classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 281–286, 2004.
- [46] Ninad Thakoor, and Jean Gao, "Shape Classifier based on Generalized Probabilistic Descent Method with Hidden Markov Descriptor," in *Proc. Int'l Conf. on Computer Vision (ICCV2005)*, vol. 1, pp. 495- 502, 2005.

- [47] F. Mokhtarian, and A. Mackworth, "Scale-based description and recognition of planar curves and two-dimensional shapes," *IEEE Pattern Anal. Mach. Intell.*, vol. 8, no. 1, pp. 34–43, 1986.
- [48] F. Mokhtarian, S. Abbasi, and J. Kittler, "Efficient and robust retrieval by shape content through curvature scale space", in *image databases and multi media search, Proc. of the first international workshop IDB-MMS'96*, Amsterdam, the Netherlands, pages 35-42, 1996.
- [49] S. Abbasi, F. Mokhtarian, and J. Kittler, "Enhancing CSS-based Shape Retrieval for Objects with Shallow Concavities," *Image and Vision Computing*, vol. 18, no. 3, pp. 199–211, 2000.
- [50] F. Mokhtarian and S. Abbasi, "Shape Similarity Retrieval Under Affine Transforms," *Pattern Recognition*, vol. 35, no. 1, pp. 31–41, 2002.
- [51] M. Bober, "MPEG-7 Visual Shape Descriptors," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 11, no. 6, pp. 716–719, 2001.
- [52] K. Arbter, "Affine-invariant Fourier descriptors", in *From Pixels to Features*, J.C. Simon, Ed., Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1989, pp. 153–164.
- [53] H. Kauppinen, T. Seppanen, and M. Pietikainen, "An experimental comparison of autoregressive and Fourier-based descriptors in 2D shape classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 2, pp. 201–207, 1995.
- [54] A. Krzyzak, S.Y. Leung, and C.Y. Suen, "Reconstruction of two dimensional patterns from Fourier Descriptors," *Mach. Vision Appl.*, vol. 2, pp. 123–140, 1989.
- [55] C.C. Lin, and R. Chellappa, "Classification of partial 2D shapes using Fourier descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no. 5, pp. 686–690, 1987.
- [56] C.S. Lin, and C.L. Hwang, "New forms of shape invariants from elliptic Fourier descriptors," *Pattern Recognition*, vol. 20, no. 5, pp. 535–545, 1987.
- [57] F.J.S. Marine, "Automatic recognition of biological shapes with and without representation of shape," *Artif. Intell. Med.*, vol. 18, pp. 173–186, 2000.
- [58] O.R. Mitchell, and T.A. Grogan, "Global and partial shape discrimination for computer vision," *Opt. Eng.*, vol. 23, no. 5, pp. 484–491, 1984.
- [59] E. Persoon, and K.S. Fu, "Shape discrimination using Fourier descriptors," *IEEE Trans. System Man Cybernet.*, vol. SMC-7, no. 3, pp. 170–179, 1977.
- [60] T.W. Rauber, *Two-dimensional shape description*, Technical Report: GR UNINOVA-RT-10-94, University Nova de Lisboa, Portugal, 1994.
- [61] C.W. Richard, and H. Hemami, "Identification of three dimensional objects using Fourier descriptors of the boundary curve," *IEEE Trans. System Man Cybernet*, vol. SMC-4, no. 4, pp. 371–378, 1974.
- [62] Y. Rui, A.C. She, and T.S. Huang, "A modified Fourier descriptor for shape matching in MARS," in *Image Databases and Multimedia Search*, A.W.M. Smeulders, R. Jain, Eds., World Scientific Publishing Co., Singapore, 1997, pp. 165–177.
- [63] C.T. Zahn, and R.Z. Roskies, "Fourier descriptors for plane closed curves," *IEEE Trans. Comput.*, vol. C-21, no. 3, pp. 269–281, 1972.
- [64] T.P. Wallace, and P. Wintz, "An efficient three-dimensional aircraft recognition algorithm using normalised fourier descriptors," *Comput. Graphics Image Process.*, vol. 13, pp. 99–126, 1980.
- [65] Rafiei, D. and A. Mendelzon, "Efficient retrieval of similar shapes," *The VLDB J.*, vol. 11, pp. 17-27, 2002..
- [66] G. Eichmann, et al., "Shape representation by Gabor expansion," in *Hybrid Image and Signal Processing II, SPIE Vol. 1297*, Orlando, Florida, USA 1990, pp. 86–94.
- [67] D.S. Zhang, and G. Lu, "A comparison of shape retrieval using Fourier descriptors and short-time Fourier descriptors," in *Proc. of the Second IEEE Pacific-Rim Conference on Multimedia (PCM01)*, Beijing, China, pp. 855–860, 2001.
- [68] D. S. Zhang and G. Lu. "Shape Based Image Retrieval Using Generic Fourier Descriptors", *Int'l J. of Signal Processing and Image Communication*, vol. 17, no. 10, pp. 825-848, 2002.
- [69] Kunttu I., and Lepisto L, "Shape-based retrieval of industrial surface defects using angular radius Fourier descriptor," *IET Image Processing*, vol. 1, no. 2, pp. 231 – 236, 2007.

- [70] G. C. H. Chuang and C. C. J. Kuo, "Wavelet descriptor of planar curves: Theory and applications," *IEEE Trans. on Image Processing*, vol. 5, pp. 56–70, 1996.
- [71] R. S. Kashi, B. K. Pratima, R. S. Nowakowski, and T. V. Papathomas, "2-D shape representation and averaging using normalized wavelet descriptors," *Simulation*, vol. 66, pp. 164–178, 1996.
- [72] Y. Tang and J. Liu, "Wavelet-Based Rotationally Invariant Target Classification," in *Signal Processing, Sensor Fusion, and Target Recognition, Proc. SPIE*, vol. 3068, no. 6, pp. 102-112, 1997.
- [73] D. Shen and H. Ip, "Discriminative Wavelet Shape Descriptors for Recognition of 2-D Patterns," *Pattern Recognition*, vol. 32, pp. 151-165, 1999.
- [74] L. Feng and T.D. Bui, "Classification of Similar 2-D Objects by Wavelet- Sparse-Matrix (WSM) Method," *Int'l J. Pattern Recognition and Artificial Intelligence*, vol. 15, no. 2, pp. 329-345, 2001.
- [75] K. Tsang, "Recognition of 2-D Standalone and Occluded Objects Using Wavelet Transform," *Int'l J. Pattern Recognition and Artificial Intelligence*, vol. 15, no. 4, pp. 691-705, 2001.
- [76] M. Mandal, S. Panchanathan, and T. Aboulnasr, "Illumination Invariant Image Indexing Using Moments and Wavelets," *J. Electronic Imaging*, vol. 7, no. 2, pp. 282-293, 1998.
- [77] H.S. Yang, S.U. Lee, and K.M. Lee, "Recognition of 2D object contours using starting-point-independent wavelet coefficient matching," *J. Visual Commun. Image Represent.*, vol. 9, no. 2, pp. 171–181, 1998.
- [78] A. Quddus, F. A. Cheikh, and M. Gabbouj, "Wavelet-based multi-level object retrieval in contour images," in *Proc. Very Low Bit Rate Video Coding (VLBV99) Workshop*, Kyoto, Japan, pp. 43–46, 1999.
- [79] J.R. Ohm, F.B. Bunjamin, W. Liebsch, B. Makai, K. Muller, A. Somlic, and D. Zier, "A set of visual feature descriptors and their combination in a low-level description scheme," *Signal Process. Image Commun.*, vol. 16, pp. 157–179, 2000.
- [80] J. Li and C. C. J. Kuo, "Automatic target shape recognition via deformable wavelet templates," in *Proc. SPIE Int. Symp. Aerospace/Defense Sensing Controls*, Orlando, FL, pp. 8–12, 1996.
- [81] King-Chu Hung, "The Generalized Uniqueness Wavelet Descriptor for Planar Closed Curves," *IEEE Transactions on Image Processing*, vol. 9, no. 5, 2000.
- [82] R. Alferez and Y. Wang, "Geometric and Illumination Invariants for Object Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 6, pp. 505-536, 1999.
- [83] Q. Tieng and W. Boles, "An Application of Wavelet Based Affine Invariant Representation," *Pattern Recognition Letters*, vol. 16, no. 12, pp. 1287-1296, 1995.
- [84] Q. Tieng and W. Boles, "Wavelet Based Affine Invariant Representation: A Tool for Recognizing Planar Objects in 3-D Space," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 8, pp. 846-857, 1997.
- [85] Q. Tieng and W. Boles, "Complex Daubechies Wavelet Based Affine Invariant Representation for Object Recognition," in *Proc. IEEE Int'l Conf. Image Processing*, vol. 1, pp. 198-202, 1994.
- [86] Q. Tieng and W. Boles, "Object Recognition Using an Affine Invariant Wavelet Representation," in *Proc. Second Australian and New Zealand Conf. Intelligent Information Systems*, pp. 307-311, 1994.
- [87] Q. Tieng, and W. Boles, "Recognition of 2D object contours using the wavelet transform zero-crossing representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 8, pp. 910–916, 1997.
- [88] M. Khalil and M. Bayoumi, "Affine Invariant Object Recognition Using Dyadic Wavelet Transform," in *Proc. 2000 Canadian Conf. Electrical and Computer Eng.*, vol. 1, pp. 421-425, 2000.
- [89] M. Khalil and M. Bayoumi, "Affine Invariants for Object Recognition Using the Wavelet Transform," *Pattern Recognition Letters*, vol. 23, pp. 57-72, 2002.
- [90] M. Khalil and M. Bayoumi, "A Dyadic Wavelet Affine Invariant Function for 2-D Shape Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1152-1164, 2001.

- [91] E. Bala and A. Cetin, "Computationally Efficient Wavelet Affine Invariant Functions for Shape Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 1095-1099, 2004.
- [92] Ch.S. Sastry, Arun K. Pujari, B.L. Deekshatulu, and C. Bhagvati, "A wavelet based multiresolution algorithm for rotation invariant feature extraction," *Pattern Recognition Letters*, vol. 25, no. 16, pp. 1845-1855, 2004.
- [93] Ibrahim El Rube, Maher Ahmed, and Mohamed Kamel, "Wavelet Approximation-Based Affine Invariant Shape Representation Functions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 323-327, 2006.
- [94] J.W. Gorman, O. R. Mitchell, and F. P. Kuhl, "Partial shape recognition using dynamic programming," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, no. 2, pp. 257-266, 1988.
- [95] N. Ansari and E. J. Delp, "Partial shape recognition: A landmark-based approach," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 5, pp. 470-483, 1990.
- [96] Y. Gdalyahu and D. Weinshall, "Flexible syntactic matching of curves and its application to automatic hierarchical classification of silhouettes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 12, pp. 1312-1328, 1999.
- [97] Evangelos Milios and Euripides G. M. Petrakis "Shape Retrieval Based on Dynamic Programming", *IEEE Trans. on image processing*, vol 9, no 1, pp 141-147, 2001.
- [98] E. G. M. Petrakis, A. Diplaros, and E. Milios, "Matching and retrieval of distorted and occluded shapes using dynamic programming," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 11, pp. 1501-1516, 2002.
- [99] T. Adamek and N. E. O'Connor, "A multiscale representation method for non rigid shapes with a single closed contour," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 14, no. 5, pp. 742-753, 2004.
- [100] L. J. Latecki and R. Lakamper, "Contour-based shape similarity," in *Proc. of Int'l. Conf. on Visual Information Systems*, D. P. Huijsmans and A. W. M. Smeulders, Eds, vol. LNCS 1614, 1999, pp. 617-624.
- [101] L.J. Latecki, R. Lakamper, and U. Eckhardt, "Shape Descriptors for Non-Rigid Shapes with a Single Closed Contour," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 424-429, 2000.
- [102] L. da F. Costa, A.G. Campos, and E.T.M. Manoel, "An integrated approach to shape analysis: results and perspectives," in *Int'l Conference on Quality Control by Artificial Vision*, Le Creusot, France, pp. 23-34, 2001.
- [103] L. da F. Costa, and E.T.M. Manoel, "A shape analysis framework for neuromorphometry," *Network*, vol. 13, pp. 283-310, 2002.
- [104] F.P. Preparata, and M.I. Shamos, *Computational Geometry: An Introduction*, Springer, Berlin, 1985.
- [105] R. da S. Torres, A.X. Falcão, and L. da F. Costa, "A graph-based approach for multiscale shape analysis," *Pattern Recognition*, vol. 37, pp. 1163 - 1174, 2004.
- [106] A.X. Falcão, L. da F. Costa, and B.S. da Cunha, "Multiscale skeletons by image foresting transform and its applications to neuromorphometry," *Pattern Recognition*, vol. 35, no. 7, pp. 1571-1582, 2002.
- [107] N. Arica, and F.T.Y. Vural, "A perceptual shape descriptor," in *Int'l Conf. on Pattern Recognition*, pp. 375-378, 2002.
- [108] N. Arica and F. T. Y. Vural, "BAS: A Perceptual Shape Descriptor Based on the Beam Angle Statistics," *Pattern Recognition Letters*, vol. 24, no. 9-10, pp. 1627-1639, 2003.
- [109] P. Miranda, R. da S. Torres, and A. X. Falcão, "TSD: A Shape Descriptor Based on a Distribution of Tensor Scale Local Orientation," in *XVIII Brazilian Symposium on Computer Graphics and Image Processing*, Natal, RN, Brazil, pp. 139-146, 2005.
- [110] P.K. Saha, *Tensor Scale: A Local Morphometric Parameter with Applications to Computer Vision and Image Processing*, Technical Report 306, Medical Image Processing Group, Department of Radiology, University of Pennsylvania, September 2003.

- [111] P. K. Saha, J. C. Gee, Z. Xie, and J. K. Udupa, "Tensor scale based image registration," in *Proc. of SPIE: Medical Imaging*, vol. 5032, pp. 743-753, 2003.
- [112] P. K. Saha and J. K. Udupa, "Tensor scale-based fuzzy connectedness image segmentation," in *Proc. of SPIE: Medical Imaging*, vol. 5032, pp. 1580-1590, 2003.
- [113] Fernanda A. Andaló, Paulo A. V. Miranda, Ricardo da S. Torres and Alexandre X. Falcão, "A new shape descriptor based on tensor scale," in *Proc. 8th Int'l Symposium on Mathematical Morphology (ISMM'07)*, Rio de Janeiro, Brazil, 2007.
- [114] R. da S. Torres and Alexandre X. Falcão, "Contour Saliency Descriptors for Effective Image Retrieval and Analysis," *Image and Vision Computing*, vol. 25, no. 1, pp. 3-13, 2006.
- [115] A. X. Falcão, J. Stolfi, and R. A. Lotufo, "The Image Foresting Transform: Theory, Algorithms, and Applications," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 26, no. 1, pp. 19-29, 2004.
- [116] M. Leyton, "Symmetry-Curvature Duality," *Computer Vision, Graphics, and Image Processing*, vol. 38, no. 3, pp. 327-341, 1987.
- [117] R. da S. Torres, E. M. Picado, A. X. Falcão, and L. da F. Costa, "Effective Image Retrieval by Shape Saliencies," in *Proc. of the Brazilian Symposium on Computer Graphics and Image Processing*, Sao Carlos, SP, Brazil, pp. 49-55, 2003.
- [118] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape context," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 4, pp. 509-522, 2002.
- [119] C. Papadimitriou and K. Stieglitz, *Combinatorial Optimization*, Englewood Cliffs, N.J.: Prentice-Hall, 1982.
- [120] C. Grigorescu and N. Petkov, "Distance Sets for Shape Filters and Shape Recognition," *IEEE Trans. Image Processing*, vol. 12, no. 10, pp. 1274-1286, 2003.
- [121] T. B. Sebastian, P. N. Klein, and B. B. Kimia, "On aligning curves," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, no. 1, pp. 116-125, 2003.
- [122] Z. Tu and A.L. Yuille, "Shape Matching and Recognition-Using Generative Models and Informative Features," in *Proc. European Conf. Computer Vision*, vol. 3, pp. 195-209, 2004.
- [123] M. Zuliani, C.S. Kenney, S. Bhagavathy, and B.S. Manjunath, "Drums and Curve Descriptors," in *Proc. British Machine Vision Conf.*, 2004.
- [124] Ilaria Bartolini, Paolo Ciaccia, and Marco Patella, "WARP: Accurate Retrieval of Shapes Using Phase of Fourier Descriptors and Time Warping Distance," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 27, no. 1, pp. 142 - 147, 2005.
- [125] A.C. Jalba, M.H.F. Wilkinson, and J.B.T.M. Roerdink, "Shape Representation and Recognition through Morphological Curvature Scale Spaces," *IEEE Trans. Image Processing*, vol. 15, no. 2, pp. 331-341, 2006.
- [126] Lena Gorelick, Meirav Galun, Eitan Sharon, Ronen Basri, and Achi Brandt, "Shape Representation and Classification Using the Poisson Equation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 28, no. 12, pp. 1991-2005, 2006.
- [127] Haibin Ling, and David W. Jacobs, "Shape Classification Using the Inner-Distance," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 29, no. 2, pp. 286-299, 2007.
- [128] [17B] R.C. Gonzalez, and R.E. Woods, *Digital Image Processing*, Pearson Education, Inc., India, 2005.
- [129] H. Blum, "A transformation for extracting new descriptors of shape," in *Models for the Perception of Speech and Visual Forms*, W. Whaten-Dunn, Ed., MIT Press, Cambridge, MA, pp. 362-380, 1967.
- [130] B.S. Morse, *Computation of object cores from grey-level images*, Ph.D. Thesis, University of North Carolina at Chapel Hill, 1994.
- [131] M. K. Hu, "Visual Pattern Recognition by Moment Invariants," *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179-187, 1962.
- [132] R.L. Kennedy, Y. Lee, B.V. Roy, C.D. Reed, and R.P. Lippmann, *Solving Data Mining Problems Through Pattern Recognition*, Prentice-Hall, PTR, Upper Saddle River, NJ, pp. 11-16, 1998.

- [133] B. Jahne, *Digital Image Processing—Concepts, Algorithms and Scientific Applications*, Springer, Berlin, Heidelberg, pp. 509–512, 1997.
- [134] S.X. Liao, and M. Pawlak, “On image analysis by moments,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 3, pp. 254–266, 1996.
- [135] M.R. Teague, “Image analysis via the general theory of moments,” *J. Opt. Soc. Am.*, vol. 70, no. 8, pp. 920–930, 1980.
- [136] C.-H. Teh, and R.T. Chin, “On image analysis by the methods of moments,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 4, pp. 496–513, 1988.
- [137] A.S. Dudani, K.J. Breeding, and R.B. McGhee, “Aircraft identification by moment invariants,” *IEEE Trans. Comput.*, vol. C-26, no. 1, pp. 39–46, 1977.
- [138] S.O. Belkasim, M. Shridhar, and M. Ahmadi, “Pattern recognition with moment invariants: a comparative study and new results,” *Pattern Recognition*, vol. 24, no. 12, pp. 1117–1138, 1991.
- [139] R.J. Prokop, and A.P. Reeves, “A survey of moment-based techniques for unoccluded object representation recognition,” *Graph. Models Image Process.*, vol. 54, pp. 438–460, 1992.
- [140] G. Taubin, and D.B. Cooper, “Recognition and positioning of rigid objects using algebraic moment invariants,” in *Proc. SPIE Conference on Geometric Methods in Computer Vision*, vol. 1570, University of Florida, Florida, USA, pp. 175–186, 1991.
- [141] G. Taubin, and D.B. Cooper, “Object recognition based on moment or Algebraic,” in *Geometric Invariance in Computer Vision*, J. Mundy, A. Zisserman, Eds., MIT Press, Cambridge, MA, 1992, pp. 375–397.
- [142] B. Scassellati, S. Slexopoulos, and M. Flickner, “Retrieving images by 2D shape: a comparison of computation methods with human perceptual judgments,” in *SPIE Proceedings on Storage and Retrieval for Image and Video Databases II*, vol. 2185, San Jose, CA, USA, pp. 2–14, 1994.
- [143] K. S. Ray and D. Dutta Majumder, “Application of differential geometry to recognize and locate partially occluded objects,” *Pattern Recognition Letters*, vol. 9, no. 6, pp. 351–360, 1989.
- [144] K. S. Ray and D. Dutta Majumder, “Recognition and positioning of partially occluded 3-D objects,” *Pattern Recognition Letters*, vol. 12, no. 2, pp. 93–108, 1991.
- [145] D. Dutta Majumder, D. K. Banerjee and S. K. Parui, “A shape metric for 3-D objects,” *Indian J. Pure Appl. Math.*, vol. 25, no. 1 & 2, pp. 95–111, 1994.
- [146] D. Dutta Majumder, “A Study on a Mathematical Theory of Shapes In Relation to Pattern Recognition and Computer Vision,” *Indian J. of Theoretical Physics*, vol. 43, no. 4, pp. 19-30, 1995.
- [147] D Dutta Majumder and M Bhattacharya, “A New Shape Based technique For Classification and registration: Application to Multimodal Medical Images,” *Int’l J. of Image processing and Communication*, vol.4, no.3-4, pp. 5-70, 1998.
- [148] D.S. Zhang, and G. Lu, “Generic Fourier descriptor for shape-based image retrieval,” in *Proc. of IEEE Int’l Conf. on Multimedia and Expo (ICME2002)*, vol. 1, Lausanne, Switzerland, pp. 425–428, 2002.
- [149] D.S. Zhang, and G. Lu, “Enhanced generic Fourier descriptor for object-based image retrieval,” in *Proc. of the IEEE Int’l. Conf. on Acoustics, Speech, and Signal Processing (ICASSP2002)*, vol. 4, Orlando, FL, USA, pp. 3668–3671, 2002.
- [150] Atul Sajjanhar, Guojun Lu, Dengsheng Zhang, and Wanlei Zhou, “A Composite Descriptor for Shape Retrieval,” in *Proc. of 6th IEEE/ACIS Int’l Conf. on Computer and Information Science (ICIS 2007)*, pp. 795 – 800, 2007.
- [151] G.J. Lu, and A. Sajjanhar, “Region-based shape representation and similarity measure suitable for content-based image retrieval,” *Multimedia Syst.*, vol. 7, no. 2, pp. 165–174, 1999.
- [152] K. Chakrabarti, M.O. Binderberger, K. Porkaew, and S. Mehrotra, “Similar shape retrieval in MARS,” in *Proc. of IEEE Int’l Conf. on Multimedia and Expo*, vol. 2, New York, USA, pp. 709–712, 2000.
- [153] T.S. Huang, S. Mehrotra, and K. Ramachandran, “Multimedia analysis and retrieval system (MARS) project,” in *Proc. of 33rd Annual Clinic on Library Application of Data Processing—*

- Digital Image Access and Retrieval*, University of Illinois at Urbana-Champaign, Illinois, USA, pp. 101–117, 1996.
- [154] A. Goshtasby, "Description and discrimination of planar shapes using shape matrices," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 7, pp. 738–743, 1985.
- [155] A. Taza, and C. Suen, "Discrimination of planar shapes using shape matrices," *IEEE Trans. Syst. Man Cybern.*, vol. 19, pp. 1281–1289, 1989.
- [156] S. Parui, E. Sarma and D. Majumder, "How to Discriminate Shapes Using the Shape Vector," *Pattern Recognition Letter*, vol. 4, pp. 201-204, 1986.
- [157] Y.-S. Kim, and W.-Y. Kim, "Content-based trademark retrieval system using a visually salient feature," *Image Vision Comput.*, vol. 16, pp. 931–939, 1998.
- [158] A. Khotanzan and Y. H. Hong, "Invariant Image Recognition by Zernike Moments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 5, pp. 489-487, 1990.
- [159] W.-Y. Kim, and Y.-S. Kim, "A region-based shape descriptor using Zernike moments," *Signal Process. Image Commun.*, vol. 16, pp. 95–102, 2000.
- [160] *MPEG-7 Visual part of experimentation model version 5.0*, S. Jeannin, Ed., ISO/IEC JTC1/SC29/WG11/N3321, Nordwijkerhout, March, 2000.
- [161] Dengsheng Zhang and Melissa Chen Yi Lim, "An Efficient and Robust Technique for Region Based Shape Representation and Retrieval," in *Proc. of 6th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2007)*, 2007.
- [162] J. J. Rocchio Jr., "Relevance Feedback in Information Retrieval," in *The SMART Retrieval System: Experiments in Automatic Document Processing*, Prentice Hall, Englewood Cliffs, pp. 313–323, 1971.
- [163] G. Salton. *Automatic Text Processing*. Addison-Wesley Publishing Company, 1989.
- [164] D. Harman, "Relevance feedback revisited," in *Proc. of 15th Annu. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval*, pp. 1–10, 1992.
- [165] Kurita T., and Kato T., "Learning of personal visual impression for image database systems," in *Proc. of IEEE Int'l Conf. Document Analysis and Recog.*, pp. 451-459, 1993.
- [166] Picard R. W., Minka T.P. and Szummer M., "Modeling user subjectivity in image libraries", in *Proc. of Int'l Conf. on Image Processing*, Lausanne, pp. 777-780, 1996.
- [167] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: A power tool for interactive content-based image retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 9, pp. 644–655, 1998.
- [168] X. S. Zhou and T. S. Huang, "Relevance feedback in image retrieval: A comprehensive review," *Multimedia Syst.*, vol. 8, pp. 536–544, 2003.
- [169] G. Salton and M. McGill, *Introduction to Modern Information Retrieval*, McGraw-Hill Book Company, 1983.
- [170] Buckley C., and Salton G. "Optimization of Relevance Feedback Weights," in *Proc of SIGIR '95*, 1995.
- [171] Rui Y., Huang T. S., and Mehrotra S., "Content-Based Image Retrieval with Relevance Feedback in MARS," in *Proc. IEEE Int'l Conf. on Image proc.*, Santa Barbara, California, USA, pp. 815–818, 1997.
- [172] Y. Rui, T.S. Huang, S. Mehrotra, and M. Ortega, "Automatic Matching Tool Selection via Relevance Feedback in MARS," in *Proc. of the 2<sup>nd</sup> International Conference on Visual Information Systems*, San Diego, California, pp. 109–116, 1997.
- [173] S. Sclaroff, L. Taycher, and M. L. Cascia, "ImageRover: a content-based image browser for the World Wide Web," in *Proc. of IEEE Workshop on Content-based Access of Image and Video Libraries*, 1997.
- [174] Vasconcelos N., and Lippman A., "A Bayesian Framework for Content-Based Indexing and Retrieval", in *Proc. of DCC '98*, Snowbird, Utah, 1998.

- [175] Ishikawa Y., Subramanya R., and Faloutsos C., “Mindreader: Querying Databases Through Multiple Examples,” in *Proc. of the 24th International Conference Very Large Data Bases*, New York City, USA, pp. 218–227, 1998.
- [176] Y. Rui and T. S. Huang, “A Novel Relevance Feedback Technique in Image Retrieval,” In *ACM Multimedia*, vol. 2, pp. 67–70, 1999.
- [177] D. Swets and J. Weng, “Discriminant Analysis and Eigenspace Partition Tree for Face and Object Recognition from Views,” in *Proc. IEEE Int’l Conf. Automatic Face and Gesture Recognition*, pp. 192–197, 1996.
- [178] X. Zhou and T. Huang, “Small Sample Learning during Multimedia Retrieval Using Biasmap,” in *Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 11–17, 2001.
- [179] D. Tao, X. Tang, X. Li, and Y. Rui, “Direct Kernel Biased Discriminant Analysis: A New Content-Based Image Retrieval Relevance Feedback Algorithm,” *IEEE Trans. Multimedia*, vol. 8, no. 4, pp. 716–727, 2005.
- [180] I. Cox, L. Miller, P. Minka, V. Paphomas, and P. Yianilos, “The Bayesian Image Retrieval System, PicHunter: Theory, Implementation and Psychophysical Experiments,” *IEEE Trans. Image Processing*, vol. 9, no. 1, pp. 20–37, 2000.
- [181] N. Vasconcelos, and A. Lippman, “Bayesian Relevance Feedback for Content-Based Image Retrieval,” in *Proc. IEEE Workshop Content-Based Access of Image and Video Libraries*, 2000.
- [182] Chiou-Ting Hsu, and Chuech-Yu Li, “Relevance feedback using generalized Bayesian framework with region-based optimization learning,” *IEEE Trans. on Image Processing*, vol. 14, no. 10, pp. 1617 – 1631, 2005.
- [183] S. D. MacArthur, C. E. Brodley, and C.-R. Shyu, “Relevance feedback decision trees in content-based image retrieval,” in *Proc. of the IEEE Workshop on Content-based Access of Image and Video Libraries*, Hilton Head, SC, 2000.
- [184] H.K. Lee, and S.I. Yoo, “A neural network-based image retrieval using nonlinear combination of heterogeneous features,” *Int’l J. of Computational Intelligence and Applications*, vol. 1, no. 2, pp. 137–149, 2001.
- [185] J. Laaksonen, M. Koskela, and E. Oja, “PicSom-self-organizing image retrieval with MPEG-7 content descriptions,” *IEEE Trans. Neural Network*, vol. 13, no. 4, pp. 841–853, 2002.
- [186] P. Muneesawang, and L. Guan, “Automatic machine interactions for content-based image retrieval using a self-organizing tree map architecture,” *IEEE Trans. Neural Networks*, vol. 13, no. 4, pp. 821–834, 2002.
- [187] P. Muneesawang, and L. Guan, “An interactive approach for CBIR using a network of radial basis functions,” *IEEE Trans. Multimedia*, vol. 6, no. 5, pp. 703 – 716, 2004.
- [188] K.H. Yap, and K. Wu, “Fuzzy relevance feedback in content-based image retrieval systems using radial basis function network,” in *Proc. of IEEE Int’l Conf. Multimedia and Expo*, Amsterdam, The Netherlands, pp. 177–180, 2005.
- [189] K.H. Yap, and K. Wu, “A soft relevance framework in content-based image retrieval systems,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 15, no. 12, pp. 1557–1568, 2005.
- [190] Peng-Yeng Yin, Bir Bhanu, Kuang-Cheng Chang, and Anlei Dong, “Integrating Relevance Feedback Techniques for Image Retrieval Using Reinforcement Learning,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1536–1551, 2005.
- [191] N. Vasconcelos, and A. Lippman, “Learning from user feedback in image retrieval systems,” *Neur. Inf. Process. Syst.*, 1999.
- [192] C. Meilhac, and C. Nastar, “Relevance feedback and category search in image databases,” in *Proc. IEEE Int. Conf. Multimedia Computing and Systems*, Florence, Italy, pp. 512–517, 1999.
- [193] Z. Su, H.J. Zhang, S. Li, and S.P. Ma, “Relevance feedback in content-based image retrieval: Bayesian framework, feature subspaces, and progressive learning,” *IEEE Trans. Image Processing*, vol. 12, no. 8, pp. 924–937, 2003.
- [194] K. Tieu, and P. Viola, “Boosting Image Retrieval,” *Int’l J. Computer Vision*, vol. 56, no. 1–2, pp. 17–36, 2004.



- [195] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, 1995.
- [196] Y. Chen, X. Zhou, and T. Huang, "One-Class SVM for Learning in Image Retrieval," in *Proc. IEEE Int'l Conf. Image Processing*, pp. 815-818, 2001.
- [197] P. Hong, Q. Tian, and T. S. Huang, "Incorporate Support Vector Machines to Content-based Image Retrieval with Relevant Feedback," in *Proc. IEEE Int'l Conf. Image Processing*, Vancouver, Canada, pp. 750-753, 2000.
- [198] L. Zhang, F. Lin, and B. Zhang, "Support Vector Machine Learning for Image Retrieval," in *Proc. IEEE Int'l Conf. Image Processing*, Thessaloniki, Greece, pp. 721-724, 2001.
- [199] D. Tao, and X. Tang, "Random sampling based SVM for relevance feedback image retrieval," in *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, Washington, DC, 2004.
- [200] S. Tong, and E. Chang, "Support Vector Machine Active Learning for Image Retrieval," in *Proc. ACM Int'l Conf. Multimedia*, pp. 107-118, 2001.
- [201] G. Guo, A. Jain, W. Ma, and H. Zhang, "Learning Similarity Measure for Natural Image Retrieval with Relevance Feedback," *IEEE Trans. Neural Networks*, vol. 13, no. 4, pp. 811-820, 2002.
- [202] D. Tao, and X. Tang, "Orthogonal Complement Component Analysis for Positive Samples in SVM Based Relevance Feedback Image Retrieval," in *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 586-591, 2004.
- [203] L. Wang, Y. Gao, K.-L. Chan, P. Xie, and W.-Y. Yau, "Retrieval with Knowledge-Driven Kernel Design: An Approach to Improving SVM-Based CBIR with Relevance Feedback," in *Proc. IEEE Int'l Conf. Computer Vision*, vol. 2, pp. 1355-1362, 2005.
- [204] K. Goh, E. Chang, and W. Lai, "Concept-Dependent Multimodal Active Learning for Image Retrieval," in *Proc. ACM Int'l Conf. Multimedia*, pp. 564-571, 2004.
- [205] C. Hoi, C. Chan, K. Huang, M.R. Lyu, and I. King, "Biased Support Vector Machine for Relevance Feedback in Image Retrieval," in *Proc. Int'l Joint Conf. Neural Networks*, 2004.
- [206] Dacheng Tao, Xiaou Tang, Xuelong Li, Xindong Wu, "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 7, pp. 1088-1099, 2006.
- [207] Kui Wu, Kim-Hui Yap, "Fuzzy SVM for content-based image retrieval: a pseudo-label support vector machine framework," *IEEE Computational Intelligence Magazine*, vol. 1, no. 2, pp. 10 - 16, 2006.
- [208] Li J., Allinson N., Tao D., Li X, "Multitraining Support Vector Machine for Image Retrieval," *IEEE Trans. Image Processing*, vol. 15, no. 11, pp. 3597-3601, 2006.
- [209] Ye Lu, Hongjiang Zhang, Liu Wenyin, and Chunhui Hu, "Joint Semantics and Feature Based Image Retrieval Using Relevance Feedback," *IEEE Trans. Multimedia*, vol. 5, no. 3, pp. 339-347, 2003.
- [210] Azadeh Kushki, Panagiotis Androustos, Konstantinos N. Plataniotis, and Anastasios N. Venetsanopoulos, "Query Feedback for Interactive Image Retrieval," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 14, no. 5, pp. 644-655, 2004.
- [211] Junwei Han, King N. Ngan, Mingjing Li, and Hong-Jiang Zhang, "A Memory Learning Framework for Effective Image Retrieval," *IEEE Trans. Image Processing*, vol. 14, no. 4, pp. 511-524, 2005.
- [212] Wei Jiang, Guihua Er, Qionghai Dai, Jinwei Gu, "Similarity-based online feature selection in content-based image retrieval," *IEEE Trans. Image Processing*, vol. 15, no. 3, pp. 702-712, 2006.
- [213] Kherfi M.L., Ziou D, "Relevance feedback for CBIR: a new approach based on probabilistic feature weighting with positive and negative examples," *IEEE Trans. Image Processing*, vol. 15, no. 4, pp. 1017-1030, 2006.
- [214] S.-C. Chen, S.H. Rubin, M.-L. Shyu, C. Zhang, "A Dynamic User Concept Pattern Learning Framework for Content-Based Image Retrieval," *IEEE Trans. Systems, Man and Cybernetics, Part C: Applications and Reviews*, vol. 36, no. 6, pp. 772-783, 2006.

- [215] Hoi S.C.H., Lyu M.R., Jin R., “A unified log-based relevance feedback scheme for image retrieval,” *IEEE Trans. Knowledge and Data Eng.*, vol. 18, no. 4, pp. 509 – 524, 2006.
- [216] Dacheng Tao, Xuelong Li, Maybank, S.J., “Negative Samples Analysis in Relevance Feedback,” *IEEE Trans. Knowledge and Data Eng.*, vol. 19, no. 4, pp. 568 – 580, 2007.
- [217] Tao D., Tang X., Li X., “Which Components Are Important for Interactive Image Searching?,” *IEEE Trans. Circuit Syst. Video Technol.*, 2007.
- [218] H. D. Tagare, C. C. Jaffe, and J. Dungan, “Medical Image Databases: A Content-Based Retrieval Approach,” *J. Amer. Med. Inf. Assoc.*, vol. 4, pp. 184–198, 1997.
- [219] Glatard T., Montagnat J., and Magnin I. E., “Texture based medical image indexing and retrieval: Application to cardiac imaging,” in *Proc. of the ACM SIGMM International Workshop on Multimedia Information Retrieval*, pp. 135-142, 2004.
- [220] Buhler P., Just U., Will E., Kotzerke J., and van den Hoff J., “An accurate method for correction of head movement in PET,” *IEEE Trans. Medical Imaging*, vol. 23, no. 9, pp. 1176-1185, 2004.
- [221] Wong S., and Hoo K. S., *Medical imagery, in Image databases: Search and retrieval of digital imagery*, V. Castelli, and L. D. Bergman, Eds., New York: John Wiley & Sons Inc. , 2002, pp. 83-105.
- [222] M. Nappi, G. Polese, and G. Tortora, “FIRST: Fractal Indexing and Retrieval System for Image Databases,” *Imag. Vis. Comp.*, vol. 6, pp. 1019-1031, 1998.
- [223] A. Pentland, R. W. Picard, and S. Scarloff, “Photobook: Content-Based Manipulation of Image Databases,” in *Proc. SPIE*, vol. 2185, pp. 34-47, 1994.
- [224] T. M. Lehmann, B. Wein, J. Dahmen, J. Bredno, F. Vogelsang, M. Kohnen, “Content based Image Retrieval in Medical Applications: A Novel Multi-Step Approach,” in *Proc. SPIE*, vol. 3972, pp. 312-320, 2000.
- [225] Shyu C.R., Brodley C.E., Kak A.C., Kosaka A., Aisen A., Broderick L., “Local versus global features for content-based image retrieval,” in *Proc. IEEE Workshop on Content-Based Access of Image and Video Libraries*, pp. 30 – 34, 1998
- [226] L. H. Y. Tang, R. Hanka, and H. H. S. Ip, “A review of intelligent content-based indexing and browsing of medical images,” *Health Informatics Journal*, vol. 5, pp. 40-49, 1999.
- [227] Henning Müller, Nicolas Michoux, David Bandon and Antoine Geissbuhler, “Review of Content-Based Image Retrieval Systems in Medical Applications - Clinical Benefits and Future Directions,” *Int’l J. of Medical Informatics*, vol. 73, no. 1, pp. 1-23, 2004.
- [228] W.W. Chu, C. C. Hsu, A. F. C’ardenas, and R. K. Taira, “KMeD: a knowledge-based multimedia medical distributed database system,” *Information Systems*, vol. 20, no. 2, pp. 75 – 96, 1995.
- [229] W.W. Chu, C. C. Hsu, A. F. C’ardenas, and R. K. Taira, “Knowledge-based image retrieval with spatial and temporal constructs,” *IEEE Trans. Knowl. Data Eng.*, vol. 10, no. 6, pp. 872–888, 1998.
- [230] E. A. El-Kwae, H. Xu, and M. R. Kabuka, “Content-based retrieval in picture archiving and communication systems,” *J. Digital Imaging*, vol. 13, no. 2, pp. 70–81, 2000.
- [231] Squire D. M., Muller H., Muller W., Marchand-Maillet S., and Pun T., Design and evaluation of a content-based image retrieval system, in *Design and management of multimedia information systems: Opportunities and challenges*, S. M. Rahman, Ed., Hershey, PA: Idea Group Publishing, 2001, pp. 125-151.
- [232] H.Müller, A. Rosset, J. P. Vallée, and A. Geissbuhler, “Comparing feature sets for content-based medical information retrieval,” in *Proc. SPIE Med. Imaging*, San Diego, CA, vol. 5351, pp. 99–109, 2004.
- [233] Rosset A., Ratib O., Geissbuhler A., and Vallee J. P., “Integration of a multimedia teaching and reference database in a PACS environment,” *Radiographics*, vol. 22, no. 6, pp. 1567-1577, 2002.
- [234] Lehmann T. M., Wein B. B., Keyzers D , Jörg Bredno, Guld M. O., Schubert H., and Michael Kohnenb, “Image Retrieval in Medical Applications The IRMA-Approach,” in *Proc. VISIM Workshop: Information Retrieval and Exploration in Large Medical Image Collections, Fourth*

- Int'l Conf. on Medical Image Computing and Computer-Assisted Intervention*, Utrecht, The Netherlands, October 2001.
- [235] J. Bredno, M. Kohnen, J. Dahmen, F. Vogelsang, B. Wein, and T. M. Lehmann, "Synergetic Impact Obtained by a Distributed Developing Platform for Image Retrieval in Medical Applications (IRMA)," in *Proc. SPIE*, vol. 3972, pp. 321-331, 2000.
- [236] Lehmann T. M., Guld M. O., Keysers D., Deselaers T., Schubert H., Wein B. B., and Spitzer K., "Similarity of medical images computed from global feature vectors for content-based retrieval," in *Lecture Notes in Artificial Intelligence*, pp. 989-995, 2004.
- [237] Lehmann T. M., Guld M. O., Thies C., Plodowski B., Keysers D., Ott B., and Schubert H., "IRMA — Content-based image retrieval in medical applications," in *Proc. of the 14th World Congress on Medical Informatics*, pp. 842-848, 2004.
- [238] Shyu C., Brodley C., Kak A., Kosaka A., Aisen A., and Broderick L., "ASSERT: A physician-in-the-loop content-based image retrieval system for HRCT image databases," *Computer Vision and Image Understanding*, vol. 75, no. 1/2, pp. 111-132, 1999.
- [239] Brodley C. E., Kak A., Dy J., Shyu C. R., Aisen A., and Broderick L., "Content-based retrieval from medical image database: A synergy of human interaction, machine learning, and computer vision," in *Proc. of the Sixteenth National Conference on Artificial Intelligence*, pp. 760-767, 1999.
- [240] Shyu C.-R., Pavlopoulou C., Kak A. C., Brodly C. E., and Broderick L., "Using human perceptual categories for content-based retrieval from a medical image database," *Computer Vision and Image Understanding*, vol. 88, no. 3, pp. 119-151, 2002.
- [241] J. G. Dy, C. E. Brodley, A. Kak, L. S. Broderick, and A. M. Aisen, "Unsupervised feature selection applied to content-based retrieval of lung images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 3, pp. 373-378, 2003.
- [242] Long L. R., Pillemer S. R., Lawrence R. C., Goh G-H, Neve L, Thoma G. R., "World Wide Web platform-independent access to biomedical text/image databases," in *Proc. of SPIE Medical Imaging: PACS Design and Evaluation: Engineering and Clinical Issues*, vol. 3339, San Diego, CA, pp. 52-63, 1998.
- [243] Long L. R., Pillemer S. R., Lawrence R. C., Goh G-H, Neve L, Thoma G. R., "WebMIRS: Web-based Medical Information Retrieval System," in *Proc. of SPIE Storage and Retrieval for Image and Video Databases VI*, vol. 3312, San Jose, CA, pp. 392-403, 1998.
- [244] Antani S., Long L. R., Thoma G. R., and Lee D. J., "Evaluation of shape indexing methods for content-based retrieval of X-ray images," in *Proc. of IS&T/SPIE 15th Annual Symposium on Electronic Imaging, Storage, and Retrieval for Media Databases*, pp. 405-416, 2003.
- [245] Long L. R., Antani S., Lee D. J., Krainak D. M., Thoma G. R., "Biomedical Information from a National Collection of Spine X-rays: Film to Content-based Retrieval," in *Proc. of the IS&T/SPIE Medical Imaging 2003 – PACS and Integrated Medical. Information Systems: Design and Evaluation*, Huang H. K, Ratib O. M., Eds., San Diego, CA, vol. 5033, pp. 70-84, 2003.
- [246] S. Antani, L. R. Long, G. R. Thoma, "A biomedical information system for combined content-based retrieval of spine x-ray images and associated text information," in *Proc. of the 3rd Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP 2002)*, Ahamdabad, India, 2002.
- [247] Antani S., Lee D. J., Long L. R., and Thoma G. R., "Evaluation of shape similarity measurement methods for spine X-ray images," *Journal of Visual Communication and Image Representation*, vol. 15, no. 3, pp. 285-302, 2004.
- [248] Antani S., Xu X., Long L. R., and Thoma G. R., "Partial shape matching for CBIR of spine X-ray images," in *Proc. of IS&T/SPIE Electronic Imaging — Storage and Retrieval Methods and Applications for Multimedia 2004*, pp. 1-8, 2004.
- [249] Sameer Antani, L. Rodney Long, George R. Thoma, "Applying Vertebral Boundary Semantics to CBIR of Digitized Spine X-ray Images," in *Proc. IS&T/SPIE Electronic Imaging Science and*

- Technology, Conf. on Storage and Retrieval Methods and Applications for Multimedia 2005*, Lienhart RW, Babaguci N, Chang EY, Eds., San Jose, California, USA, SPIE Vol. 5682, pp. 98-107, 2005.
- [250] Xiaoqian Xu, D.J. Lee, S. Antani, and L.R. Long, "Partial Shape Matching of Spine X-ray Shapes Using Dynamic Programming", in *Proc. of the 17th IEEE Symposium on Computer-Based Medical Systems*, pp. 97-102, 2004.
- [251] Xu X., Lee D. J., Antani S., Long, L. R., "Relevance feedback for spine X-ray retrieval," in *Proc. of the 18th IEEE Symposium on Computer-Based Medical Systems*, pp. 197-202, 2005.
- [252] L. H. Tang, R. Hanka, H. H. S. Ip, and R. Lam. "Extraction of semantic features of histological images for content based retrieval of images" in *Proc. of SPIE -- Volume 3662 Medical Imaging 1999: PACS Design and Evaluation: Engineering and Clinical Issues*, G. James Blaine, Steven C. Horii, Eds., pp. 360-368, 1999.
- [253] D. Comaniciu, P. Meer, and D. Foran, "Image guided decision support system for pathology," *Mach. Vis. Appl.*, vol. 11, pp. 213-224, 1999.
- [254] Jinman Kim, Weidong Cai, Dagan Feng, and Hao Wu, "A New Way for Multidimensional Medical Data Management: Volume of Interest (VOI)-Based Retrieval of Medical Images With Visual and Functional Features," *IEEE Trans. Inf. Tech. in Biomed.*, vol. 10, no. 3, pp. 598-607, 2006.
- [255] P. Korn, N. Sidiropoulos, C. Faloutsos, E. Siegel, and Z. Protopapas, "Fast and effective retrieval of medical tumor shapes," *IEEE Trans. Knowl. Data Eng.*, vol. 10, no. 6, pp. 889-904, 1998.
- [256] A. S. Constantinidis, M. C. Fairhurst, and A. F. R. Rahman, "A new multi-expert decision combination algorithm and its application to the detection of circumscribed masses in digital mammograms," *Pattern Recognition*, vol. 34, pp. 1527-1537, 2001.
- [257] S. Baeg, and N. Kehtarnavaz, "Classification of breast mass abnormalities using denseness and architectural distortion," *Electronic Letters on Computer Vision and Image Analysis*, vol. 1, no. 1, pp. 1-20, 2002.
- [258] F. Schnorrenberg, C. S. Pattichis, C. N. Schizas, and K. Kyriacou, "Content-based retrieval of breast cancer biopsy slides," *Technology and Health Care*, vol. 8, pp. 291-297, 2000.
- [259] W.-J. Kuo, R.-F. Chang, C. C. Lee, W. K. Moon, and D.-R. Chen, "Retrieval technique for the diagnosis of solid breast tumors on sonogram," *Ultrasound in Medicine and Biology*, vol. 28, no. 7, pp. 903-909, 2002.
- [260] Issam El-Naqa, Yongyi Yang, Nikolas P. Galatsanos, Robert M. Nishikawa, and Miles N. Wernick, "A Similarity Learning Approach to Content-Based Image Retrieval: Application to Digital Mammography," *IEEE Trans. Medical Imaging*, vol. 23, no. 10, pp. 1233-1244, 2004.
- [261] H. A. Swett and P. L. Miller, "ICON: a computer-based approach to differential diagnosis in radiology," *Radiology*, vol. 163, pp. 555-558, 1987.
- [262] P. M. Kelly and T. M. Cannon, "CANDID: comparison algorithm for navigating digital image databases," in *Proc. Int'l. Working Conf. Scientific and Statistical Database Management*, pp. 252-258, 1994.
- [263] A. Guimond and G. Subsol, "Automatic MRI database exploration and applications," *Pattern Recogn. Artif. Intell.*, vol. 11, no. 8, pp. 1345-1365, 1997.
- [264] Y. Liu and F. Dellaert, "A classification based similarity metric for 3D image retrieval," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, pp. 800-805, 1998.
- [265] J. Sklansky, E. Tao, M. Bazargan, C. Ornes, R. Murchison, and S. Teklehaimanot, "Computer-aided, case-based diagnosis of mammographic regions of interest containing microcalcifications," *Academic Radiol.*, vol. 7, no. 6, pp. 395-406, 2000.
- [266] C. Ornes and J. Sklansky, "A visual neural classifier," *IEEE Trans. Systems, Man, and Cybernetics*, pt. B, vol. 28, pp. 620-625, 1998.
- [267] Y. Kawata, N. Niki, H. Ohmatsu, M. Kusumoto, R. Kakinuma, K. Mori, H. Nishiyama, K. Eguchi, M. Kaneko, and N. Moriyama, "Three-dimensional CT image retrieval in a database of

- pulmonary nodules,” in *Proc. of IEEE Int’l Conf. on Image Processing*, Rochester, NY, pp. 149–152, 2002.
- [268] Zhao C.G., Cheng H.Y., Huo Y.L., Zhuang T.G., “Liver CT-image retrieval based on Gabor texture,” in *Proc. of the 26th Annual Int’l Conf. of the Engineering in Medicine and Biology Society*, vol. 1, pp. 1491–1494, 2004.
- [269] Najjar M., Ambroise C., Cocquerez J.-P., Cotton A., Eltabach M., “A Content-Based Image Retrieval System for Osteo-Articular Applications,” in *Information and Communication Technologies (ICTTA ’06)*, vol. 2, pp. 1726–1731, 2006.
- [270] C. Schmid and R. Mohr, “Local Grayvalue Invariants for Image Retrieval,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 5, pp. 530–535, 1997.
- [271] J. R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Humphrey, R. Jam, and C. F. Shu, “The Virage Search Engine: An Open Framework for Image Management,” in *Proc. of the SPIE Storage and Retrieval for Still Image and Video Databases IV*, vol. 2670, pp. 77–87, 1996.
- [272] M. Stricker and M. Orengo, “Similarity of Color Images,” in *Storage and Retrieval for Image and Video Databases III (I&ST/SPIE)*, San Jose, CA, USA, vol. 2420, pp. 381–392, 1995.
- [273] F. Alaya Cheikh, B. Cramariuc, and M. Gabbouj, “Relevance Feedback for Shape Query Refinement,” in *Proc. of the IEEE Int’l Conf. on Image Processing (ICIP 2003)*, Barcelona, Spain, vol. 1, pp. 745–748, 2003.
- [274] P. Kerminen and M. Gabbouj, “The Visual Goodness Evaluation of Color based Retrieval Processes,” in *Proc. of the EUSIPCO 2000*, Tampere, Finland, pp. 2153–2156, 2000.
- [275] M. Swain and D. Ballard, “Colour indexing,” *Int’l J. Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.
- [276] J. R. Smith, *Integrated Spatial and Feature Image Systems: Retrieval, Analysis and Compression*. Ph.D. thesis, Graduate School of Arts and Sciences, Columbia University, Columbia, USA, 1997.
- [277] J. R. Smith and S.-F. Chang, “Local Colour and Texture Extraction and Spatial Query,” in *Proc. of the IEEE Int’l Conf. on Image Processing*, Lausanne, Switzerland, vol. 3, pp. 1011–1014, 1996.
- [278] E. Saber and A. M. Tekalp, “Region-based Shape Matching for Automatic Image Annotation and Query-by-Example,” *Journal of Visual Communication and Image Representation*, vol. 8, no. 1, pp. 3–20, 1997.
- [279] M. A. Stricker, “Bounds for the Discrimination Power of Colour Indexing Techniques,” in *Storage and Retrieval for Image and Video Databases II, (I&ST/SPIE)*, San Jose, CA, USA, vol. 2185, pp. 15–24, 1994.
- [280] S. Santini and R. Jain, “Similarity Measures,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 9, pp. 871–883, 1999.
- [281] I. Shmulevich, B. Cramariuc, and M. Gabbouj, “A Framework for Ordinal based Image Correspondence,” in *Proc. of the X European Signal Processing Conf., EUSIPCO-2000*, Tampere, Finland, pp. 1389–1392, 2000.
- [282] F. Alaya Cheikh, B. Cramariuc, M. Partio, P. Reijonen, and M. Gabbouj, “Shape Similarity Estimation using Ordinal Measures,” in *Proc. of the Int’l Workshop on Very Low Bitrate Video Coding (VLBV01)*, Athens, Greece, pp. 44–49, 2001.
- [283] F. Alaya Cheikh, B. Cramariuc, M. Partio, P. Reijonen, and M. Gabbouj, “Evaluation of Shape Correspondence Using Ordinal Measures,” in *Proc. of the IS&T/SPIE Electronic Imaging 2002 Symposium, Conf. on Storage and Retrieval for Media Databases 2002* M. M. Yeung, C.-S. Li, and R. W. Lienhart, Eds, vol. 4676, pp. 22–30, 2002.
- [284] F. Alaya Cheikh, B. Cramariuc, M. Partio, P. Reijonen, and M. Gabbouj, “Ordinal-Measure Based Shape Correspondence,” *Eurasip Journal of Applied Signal Processing, Special Issue on Image Analysis for Multimedia Interactive Services - Part I*, vol. 4, pp. 362–371, 2002.

- [285] F. Alaya Cheikh, A. Quddus, and M. Gabbouj, "Contour Based Object Recognition Using Wavelet-Transform.," in *Proc. of the X European Signal Processing Conf., EUSIPCO 2000*, Tampere, Finland, vol. 2, pp. 2141–2144, 2000.
- [286] F. Alaya Cheikh, A. Quddus, and M. Gabbouj, "Shape Recognition based on Wavelet-Transform Modulus Maxima," in *Proc. of the 7th IEEE Int'l Conf. on Electronics, Circuits and Systems (ICECS2K)*, Beirut, Lebanon, pp. 461–464, 2000.
- [287] A. Quddus, F. Alaya Cheikh, and M. Gabbouj, "Wavelet-Based Multi-level Object Retrieval in Contour Images," in *Proc. of the Very Low Bit rate Video Coding (VLBV'99) workshop*, Kyoto, Japan, pp. 43–46, 1999.
- [288] M. Trimeche, F. Alaya Cheikh, and M. Gabbouj, "Similarity Retrieval of Occluded Shapes Using Wavelet-Based Shape Feature," in *Proc. of the SPIE Conf. on Internet and Multimedia Management Systems*, Boston, USA, vol. 4210, pp. 281–289, 2000.
- [289] R. C. Veltkamp, "Shape Matching: Similarity Measures and Algorithms," in *Proc. of the Int'l Conf. on Shape Modeling and Applications*, Genova, Italy, pp. 188–197, 2001.

---

## **Chapter 3**

# **Scoliosis – An Overview**

---

### **3.1 Introduction**

All spines have curves, but occasionally the spine twists and develops curves in the wrong direction - sideways. It is a condition known as scoliosis. Scoliosis occurs relatively frequently in the general population, and its frequency depends upon the magnitude of the curve being described. Scoliosis of greater than 25 degrees has been reported in about 3-5/1000 persons in the United States [1]. If scoliosis is neglected, the curves may progress dramatically, creating significant physical deformity and even cardiopulmonary problems. Generally, scoliosis is treated by orthopedic surgeons with special training in spinal problems. Since a part of the work is to construct a system for measurement and analysis of scoliosis, a brief introduction to the pathology is presented here. Important literature related to spine localization and scoliosis measurement is reviewed in this chapter.

### **3.2 Human Spine Anatomy**

The human spine consists of number of small bones (vertebrae) and joints (intervertebral disks) together to form a flexible spinal column. A normal adult spine contains 33 vertebrae, and 23 intervertebral disks. A vertebra is composed of a ventrally placed body and a dorsal arch. The dimensions of the vertebral body and disk gradually increase from cervical to lumbar spine. The vertebral body is responsible for carrying weight and protecting the spinal cord and the

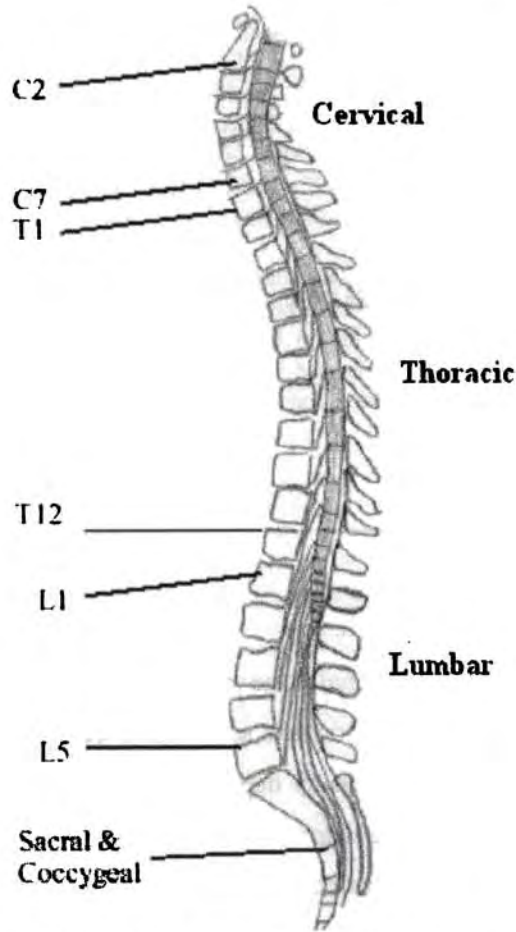


Figure 3.1 Nomenclature of vertebrae

nerve roots. They are attached firmly to each other by intervertebral disks.

Classically the vertebral column is divided into several regions. Seven vertebrae starting from the top of the spine belong to the Cervical spine (C1-C7); twelve vertebrae after cervical spines are called the Thoracic spine (T1-T12); five (or six) bottom ones belong to the Lumbar spine (L1-L5); five (fused) vertebrae form the Sacrum spine (S1-S5); and the last three belong to the Coccyx spine. A normal adult spine also has cervical, thoracic, lumbar and sacral curves along the vertebral column as we can see in Figure 3.1.

The lumbar spine's shape has a lordotic curve. The lordotic shape is like a backward "C". If you think of the spine as having an "S"-like shape, the lumbar region would be the bottom of the "S". The vertebrae in the lumbar spine area are the largest of the entire spine.



The vertebrae of thoracic spine connect to the ribs and form part of the back wall of the thorax (the ribcage area between the neck and the diaphragm). Very narrow and thin intervertebral discs present at this part of the spine allow only much less movement between these vertebrae than in the lumbar or cervical spine. The thoracic spine's curve is called kyphotic because of its shape, which is a regular "C" curve with the opening of the C in the front.

The cervical spine starts just below the skull and ends just above the thoracic spine. The cervical spine has a lordotic curve (a backward C-shape) - just like the lumbar spine. The cervical spine is much more mobile than both of the other spinal regions. Two vertebrae in the cervical spine, the atlas and the axis, differ from the other vertebrae because they are designed specifically for rotation.

### 3.3 X-ray Imaging of Spine

X-ray imaging is a painless method of using radioactive materials to capture images of bone. During X-ray imaging, patient will be asked to hold certain still positions in a standing or lying posture, while pictures of spine are taken.

The three axis of human spine is shown in Figure 3.2. They are coronal plane, sagittal plane and axial plane. In X-ray imaging, common views taken are

- Anteroposterior (AP) view: front-back view of spine (coronal plane)
- Posteroanterior (PA) view: back-front view of spine (coronal plane)

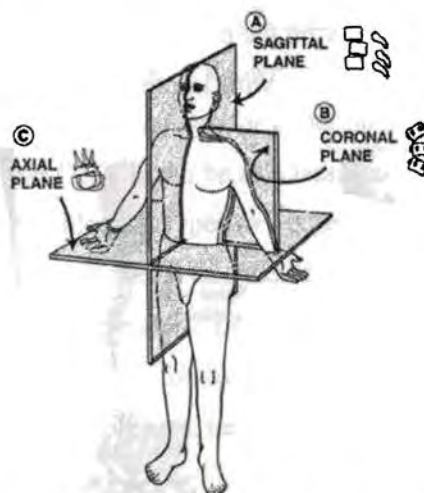


Figure 3.2 Different axis of human spine

- Lateral view: side-to-side view of spine (sagittal plane)

### 3.4 Scoliosis

Scoliosis is lateral curvature of the spine. The normal spine is straight in the coronal plane and has 'S' shaped curvature in the sagittal plane. But in the case of scoliosis, the spine is curved in the coronal plane as shown in Figure 3.3. The following are a few typical physical features/symptoms of scoliosis:

- Asymmetry in the shoulders, trunk, scapula, and waistline - one shoulder or hip will be higher than the other
- A prominent shoulder blade - one is higher than the other and sticks out further
- Symptoms are most noticeable when bending over
- Rib hump - which occurs when scoliosis causes the chest to twist and throws off the alignment of the shoulder blade; this causes a hump on the back at the ribs or near the waist when the patient bends
- One arm hangs longer than the other because of a tilt in the torso

Scoliosis is a complicated deformity that is characterized by both lateral curvature and vertebral rotation. As the disease progresses, the vertebrae and spinous processes in the area of the major curve rotate toward the concavity of the curve. On the concave side of the curve, the ribs are close together. On the convex side, they are widely separated. Vertebral canal is narrower on the concave side. Physiological changes include: 1) Decrease in lung vital capacity due to a compressed intrathoracic cavity on the convex side and 2) With left scoliosis, the heart is displaced downward; and in conjunction with intrapulmonary obstruction, this can result in

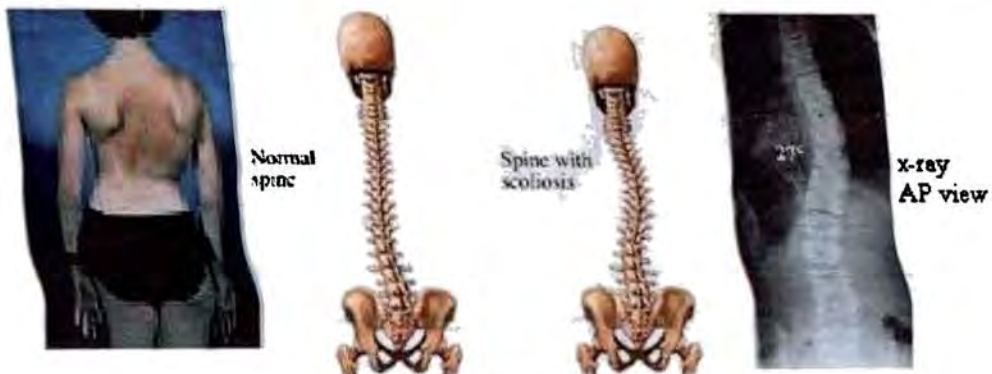


Figure 3.3 A case of scoliosis

right cardiac hypertrophy.

### 3.4.1 Diagnosis

X-ray images are the most important evidence of scoliosis. For diagnosing scoliosis the following images will be taken:

- Three-foot standing AP or PA
- Three-foot standing lateral
- Lateral bend (sideways view while bending forward at the waist)
- Traction films - traction is when your spine is pulled and held in a particular position; these films are only occasionally taken

## 3.5 Types of Scoliosis

The scoliosis curves are broadly classified into two: structural or nonstructural. Structural curves are those in which lateral bending of the spine is asymmetric or the involved vertebrae are fixed in a rotated position or both. These are the curves the patient cannot correct by lateral bending. Nonstructural curves, in contrast, are those in which intrinsic changes in the spine or its supporting structures are absent. In these curves, lateral bending is symmetric and the involved vertebrae are not fixed in the rotated position. Generally a nonstructural curve requires no treatment or any treatment is directed toward its cause, which is not located in the spine itself.

The structural scoliosis can be further classified into idiopathic, congenital, and paralytic [2]. When the scoliosis does not have a known cause, the condition is called idiopathic scoliosis. Idiopathic genetic scoliosis accounts for about 80 % of all cases of the disorder, and has a strong female predilection (7:1). It can be subclassified into infantile (< 3 years), juvenile (between 4 and 10 years) and adolescent types (> 10 years), depending upon the age of onset. Idiopathic scoliosis in adolescents is called adolescent idiopathic scoliosis (AIS). Scoliosis can result from congenital vertebral anomalies. Discovery of these anomalies should prompt a workup for other associated cardiac, genitourinary or vertebral anomalies. Paralytic scoliosis is generally caused by neuromuscular diseases, neurofibromatosis or spinal cord tumors. Scoliosis also occurs due to trauma.

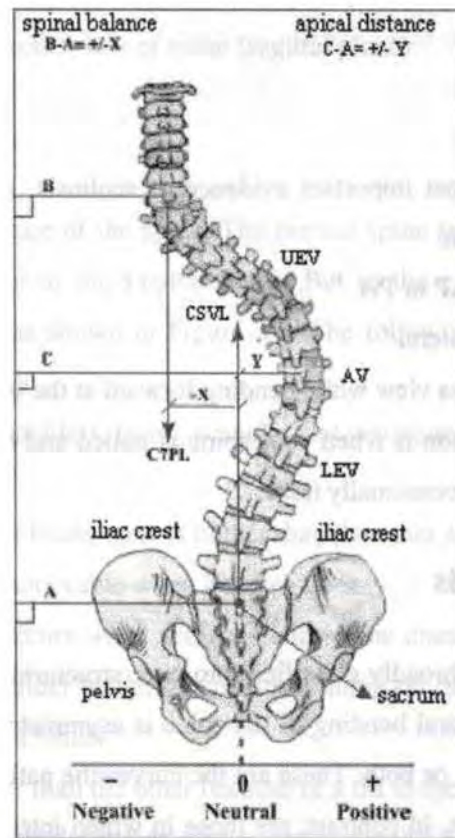


Figure 3.4 Standard terminology description

## 3.6 Basic Definitions

Basic definitions of the standard terminology used in scoliosis measurement are given below. The defined terms are marked in an AP view scoliosis image in Figure 3.4.

### 3.6.1 Central Sacral Line (CSL)

CSL is called the spinal axis. It is the line passing through the centroid of the sacrum and perpendicular to the line joining the top points of iliac crests. In normal cases the pelvis is aligned parallel to the horizontal axis and CSL is a vertical line passing through the centroid of sacrum. This CSL is called central sacral vertical line (CSVL). This is based on the assumption that the leg lengths are equal.

### **3.6.2 Pelvic Inclination**

The inclination of the line joining the top – mid points of the iliac crests to horizontal axis is called pelvic inclination.

### **3.6.3 Apical Vertebra (AV)**

It is the center vertebra of each of the curve, also called curve apex. The curve apex could be a vertebra or a disk between the vertebrae, if two vertebrae are located in the curve center. By definition, AV has the following properties: most horizontal vertebra with respect to the horizontal axis, maximally rotated vertebra, most wedged vertebra, and maximally displaced vertebra from CSL

### **3.6.4 Apical Distance**

Apical distance is the perpendicular distance from centroid of AV to the CSL.

### **3.6.5 End Vertebrae**

They are the vertebra in the upper and lower limit of the curve and with maximum endplate tilt angle. The vertebra at the upper limit of the curve is called upper end vertebra (UEV) and at the lower limit is lower end vertebra (LEV). They are also called superior end vertebra and inferior end vertebra. The disc above the UEV and below the LEV will have maximum parallel end plates.

### **3.6.6 Spinal Balance (Coronal Balance)**

Spinal balance is the horizontal distance between sagittal C7 plumbline (a vertical line passing through the centroid of C7) and CSL. This is an indication of the amount of shoulder elevation due to scoliosis.

### **3.6.7 Tilt Angle of a vertebra**

The angle made by the inferior end plate of a vertebra with the horizontal axis.

## 3.7 Radiographic Assessment of the Scoliosis

### 3.7.1 Measurement of Curve

The radiographic assessment of the scoliosis begins with erect anteroposterior (AP) and lateral views of the entire spine (occiput to sacrum). A lateral view of the lumbar spine is also taken to look for the presence of spondylolysis or spondylolisthesis (prevalence in the general population is about 5 %). The scoliotic curve is then measured from the AP view. The Cobb method [3, 4] is the most commonly used method (used by the Scoliosis Research Society) for scoliosis measurement [2]. The Cobb method has several advantages over other methods, including the fact that it is more consistent while measured by several different examiners.

#### 3.7.1.1 Cobb Method

Cobb angle is the most important parameter in scoliosis diagnosis and is reported along with the location of the curve. This method consists of three steps: (1) locating the superior end vertebra (upper limit of curve), (2) locating the inferior end vertebra (lower limit of curve), and (3) drawing intersecting perpendiculars from the superior surface of the superior end vertebra and from the inferior surface of the inferior end vertebra as shown in Figure 3.5. The angle of deviation of these perpendiculars from a straight line is the angle of the curve. The end vertebrae of the curve are the ones that tilt the most into the concavity of the curve being measured. If the endplates cannot be easily seen, these lines can be drawn along the top or bottom or the pedicles. Generally, as one moves away from the apex of the curve being measured, the next intervertebral space inferior to the inferior end vertebra or superior to the superior end vertebra is wider on the concave side of the curve being measured. Within the curve being measured the intervertebral spaces are usually wider on convex side and narrower on the concave side. When significantly wedged, the vertebrae themselves rather than the intervertebral spaces may be wider on the convex side of the curve and narrower on the concave side. Appelgren and Willner modified the Cobb method by dividing the Cobb angle into two parts, as the sum of the angles between each end vertebra and the horizontal plane [5]. Usually, the original Cobb technique is used for measurements.

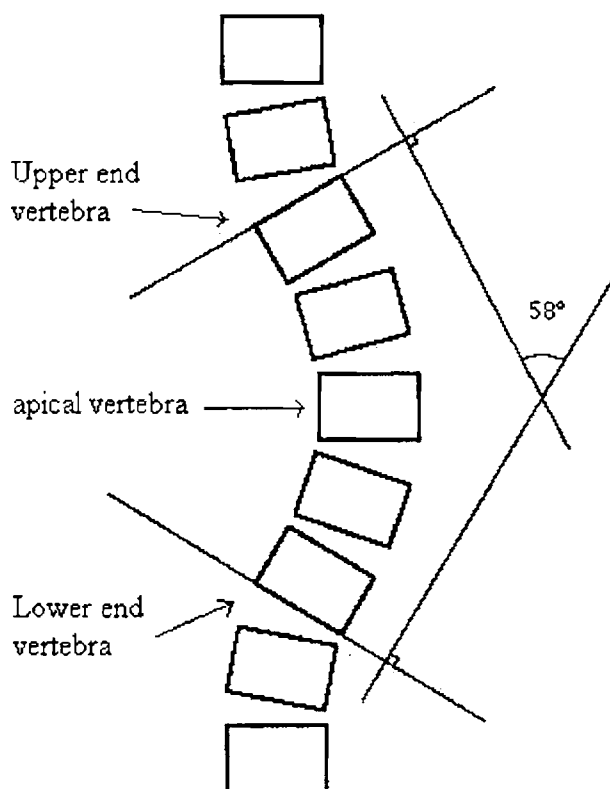


Figure 3.5 Cobb method for measurement of scoliosis

### 3.7.1.2 Other Techniques

Other similar measurement techniques are: 1) Anterior Vertebral body technique; where scoliosis is measured in terms of the intersecting angle of the lines paralleling the anterior aspect of each of the vertebral bodies, 2) Posterior Vertebral body Technique; where the intersecting angle of the lines paralleling the posterior aspect of each of the vertebral bodies, 3) Cobb-Posterior Vertebral Technique; where the intersecting angle of the lines drawn along the superior end plate of the superior vertebral body and along the posterior aspect of the inferior vertebral body is measured [6]. In Risser-Ferguson method, straight lines are drawn from the middle of the end vertebra to the middle of the vertebrae at the apex of the curve [7]. This method is not frequently used. A technique useful when evaluating minimal curvatures, which are often difficult to measure with the currently accepted methods, was introduced in 1978 by Adam Greenspan and is more accurate in measuring the deviation of each involved vertebra [8]. This technique, called the "scoliotic index," measures the deviation of each involved vertebra

from the vertical spinal line, as determined by a point at the center of the vertebra immediately above the upper end-vertebra of the curve, and the center of the vertebra immediately below the lower end-vertebra.

The accuracy and precision of the Cobb method compared to other techniques in measuring coronal plane abnormalities have been well documented [9, 10].

### 3.7.2 Measurement of Vertebral Rotation

The position of the pedicles on the x-ray image indicates the degree of vertebral rotation, which Nash and Moe divided into five grades as shown in Figure 3.6. If the pedicles are equidistant from the sides of the vertebral bodies, there is no vertebral rotation (grade 0). The grade then increase up to grade 4 rotation, which indicates that the pedicle is past the center of

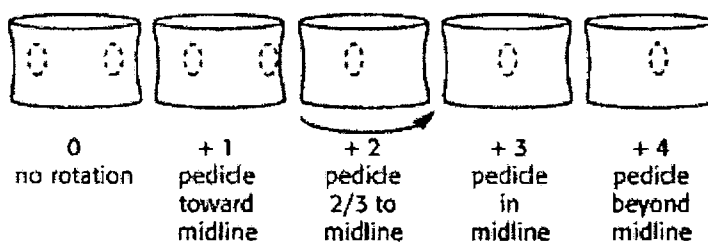


Figure 3.6 Nash and Moe method of determining vertebral rotation

the vertebral body [2].

## 3.8 Classification Schemes

In idiopathic scoliosis most of the characteristic features of the primary curve or curves are present at the onset of deformity and they rarely change. As the primary curve increases, one or two vertebrae may be added to it, but its apex, location and the direction of rotation of the included vertebrae remain unchanged. A curve classification scheme is useful in predicting the natural history based on the grouping and also in assignment of therapeutic strategies. Classification is usually taken as a factor in deciding the extent of spinal arthrodesis. Usually curves are identified in the medical community by their curve pattern type and are an integral part in scoliosis description. Many curve pattern classification schemes are in use today and new schemes are frequently proposed. Most important schemes are

- Ponsetty and Friedman scheme [11]



- King *et. al.* classification [12]
- Lenke *et. al.* classification [13]
- Peking Union Medical College (PUMC) method [14]
- SRS classification [15]

The Ponsetty and Friedman scheme is used in this thesis and hence its brief review is provided here.

### **3.8.1 Ponsetti and Friedman Scheme**

Ponsetti and Friedman found that spinal curves form five main patterns that behaved differently. Also, a sixth pattern was described by Moe.

#### **3.8.1.1 Single Major Lumbar Curve**

This was described as the most benign and least deforming of all curves. It can however cause marked distortion of the waistline. It usually contained five vertebrae, T11 to L3, with apex at L1 or L2.

#### **3.8.1.2 Single Major Thoracolumbar Curve**

It usually included six to eight vertebrae, and extended from T6 or T7 to L1 or L2. Its apex was at T11 or T12. Curves of this type produce more cosmetically objectionable deformities than thoracic or lumbar curves of the same magnitude, especially when the curves are long.

#### **3.8.1.3 Combined Thoracic and Lumbar Curve (Double Major Curve)**

In these patients, the two curves being present from the onset and essentially equal. The thoracic curve is usually to the right and includes five or six vertebrae from T5 or T6 to T10 or T11. Its apex is at T7 or T8. The lumbar curve is usually to the left and includes five or six vertebrae from T10 or T11 to L3 or L4. Its apex is at L2 or L2. Often a neutral or unrotated vertebra was common to the adjacent ends of the curves. The prognosis as to cosmetics is good. Deformity of the back and decrease in the vital capacity is less severe than in single thoracic curve.

### 3.8.1.4 Single Major Thoracic Curve

Its onset is earlier than any other type of curve. It usually includes six vertebrae from T5 or T6 to T11 or T12 and has its apex at T8 or T9. Because of the thoracic location of this curve, marked rotation of the involved vertebrae is present. The curve produces prominence of the ribs on its convex sides and depression of the ribs on its concave side and elevation of one shoulder, resulting in an unpleasant deformity.

### 3.8.1.5 Cervicothoracic Curve

Although this type of curves never seemed to become large, the deformity is unsightly because of the elevated shoulder. The deformed thorax could be poorly disguised by clothing. The apex is usually at T3 with the curve extending from C7 or T1 to T4 or T5.

### 3.8.1.6 Double Major Thoracic Curves

This pattern was described by Moe. It consists of a short upper thoracic curve often extending from T1 to T5 or T6 with considerable rotation of the vertebrae and other structural changes in combination with a lower thoracic curve extending from T6 to T12 or L1. The upper curve is usually to the left and lower curve is usually to the right. The appearance of patients with this curve is usually better than with a single thoracic curve, but asymmetry in neckline is evident.

## 3.9 Related Pathologies

This section briefly introduces other pathologies related to scoliosis: kyphosis and lordosis. They are also deformities due to the curvature of spine.

### 3.9.1 Kyphosis

As introduced in Section 3.2 the thoracic spine has a curvature in the sagittal plane. Normal thoracic kyphosis is in the range of 20° to 50° [16]. But, if the curve in a person's thoracic spine is more than 50 degrees, it is considered abnormal and the spinal deformity is called kyphosis.

Sometimes this deformity is described as "round back posture" or "hunchback". The kyphosis is measured using Cobb technique in lateral x-rays of the spine.

### 3.9.2 Lordosis

The cervical spine (neck) and lumbar spine (lower back) have lordotic curves (like a backward "C") in sagittal plane. The direction of this curve is opposite to that of kyphosis. Hardacker suggests that the normal range of cervical lordosis is  $40^\circ$  (+/-)  $9.7^\circ$  [17]. Lumbar lordosis ranges from  $31^\circ$  to  $79^\circ$  in normal situations depending from where you chose to measure it [18]. The lordosis is measured using Cobb technique in lateral x-rays of the spine. A value exceeding normal limits is a spinal deformity called lordosis.

## 3.10 Literature Review

This section reviews important literature related to spine localization and scoliosis measurement of digitized spine x-ray images.

### 3.10.1 Spine Localization

Spine localization refers to estimation of location and position of spine in digitized x-ray images. The methods proposed can be divided into two categories: semi-automated and fully-automated. Semi-automated techniques demands human intervention at some point during segmentation, while fully-automated techniques do not involve human intervention.

#### 3.10.1.1 Semi-automated Techniques

In the literature, some human assisted methods are reported for the segmentation of the vertebrae by using Active Contour Segmentation (ACS) technique [19]. A fundamental and comprehensive treatment of the whole field of active shape modeling (ASM), which has given technical direction to a number of research efforts was provided by cootes *et. al.* [20]. A semi-automated method has been proposed for the segmentation of lumbar spine dual x-ray absorptiometry (DEXA) images [21]. The user manually identifies two "anchor points" for placing a template. The template then deforms by ASM, maintaining invariance of the anchor

points, which are placed at the top and bottom of a column of vertebrae. Another ASM based technique uses manually selected boundary points of the vertebra to form the *shape model* and *gray-scale model* created by sampling the *gray-scale profile* (or gray-scale difference profile) along normal to each of these points [22]. In a test set of 40 cervical spine images, for 16 of the cases, the results showed a mean point-to-point error of less than 1/10 inch. Strong sensitivity of convergence to initial template positioning was observed in some cases [23].

### 3.10.1.2 Fully-automated Techniques

The fully automated methods proposed in the literature can be broadly classified into two groups. 1) based on landmark points like skull, shoulder etc to determine the characteristic curves assumed to lie in the spine region [24–26] and 2) a template matching based method for spine pose estimation [27, 28]. The first category includes a method proposed by Zamora *et. al.* which determines the approximate spine axis location, based on line integrals of image gray-scale [29]. He reported an orientation error in his algorithm of less than 15 degrees, for 34 of the cases in a test set of 40 cervical spine images. Also, Dynamic programming methods are proposed for spine axis localization in the region of interest computed from basic landmark points [30]. The algorithm was tested in a test set of 48 images and the landmark points and spine axis could be successfully computed for 46 cases. In the template matching method, a customized implementation of the Generalized Hough Transform (GHT) is used for the object localization [27]. A template, which represents the shape of the object of interest (spine), is matched to a target image for different values of position, orientation, and scale. The output of GHT is the pose of the object within the image. For the cervical images, the GHT template represents the interior endplate of C2 and C3 through C6, and for the lumbar images, the template represents L1 through L5.

Another solution to the problem of vertebrae segmentation in digitized x-ray images is a hierarchical approach that combines three different methodologies [25, 26]. The first module is a customized Generalized Hough Transform (GHT) algorithm that is used to find an estimate of vertebral pose within target images. The second module is a customized version of Active Shape Models (ASM) that is used to combine gray-level values and edge information in order to find vertebral boundaries. ASM is a technique that captures the variability of shape and local gray-level values from the training set of images and builds two models, one for shape (SM) and

for gray-level values (GLM). Segmentation with ASM is achieved by iteratively deforming the SM towards the boundaries of the objects of interest as guided by GLM. It requires GLM to be a good representative of the gray-level values of the images around the shapes of interest and the deformation of SM does not allow for local deformations of shape at key parts of the objects. The third module is a customized DM (Deformable Model) approach based on the minimization of external and internal energies, that allow the capture of fine details such as vertebral corners [26]. The local deformation shortcoming of ASM is addressed by building a number of DMs along the local shapes of interest, such as vertebral corners, and deforming them according to an energy minimization approach. Thus, the ASM module needs to be correctly initialized with the location and orientation of the spine for accurate segmentation.

### 3.10.2 Scoliosis Measurement

The scoliosis is evaluated by measuring the Cobb angle of the curve. Accuracy and reliability in Cobb angle measurement is an important issue. The poor reproducibility of Cobb angle measurement commonly raises uncertainty about the validity of identified changes or progression in curvature [9, 31]. A 5° increase in Cobb angle measurements between two follow-up visits can suggest a curvature progression, and this may lead to changes in the treatment plan [32]. There are manual and computer-assisted techniques in use for scoliosis measurement. A number of publications are available in the literature reporting the measurement error and inter and intra observer reliability.

Manual measurement of the Cobb angle between involved vertebrae on spinal radiographs is done using simple protractors or other devices [33]. Numerous studies have focused on the error inherent in manual measurements of spinal alignment using plain film radiographs [9, 10, 34-37]. Carman *et. al* [9] reported an average difference of 3.8° (95% of differences less than 8.0°) in repeated measurements by 5 readers on 8 radiographs. They inferred from analysis of variance components that the overall standard deviation was 2.97°. Average intraobserver standard deviation (SD) has been reported as 3.5 ° and interobserver SD range from 2.8 ° to 7.2° [10, 38]. Cumulative errors ranging from 5 ° to 10 ° have been reported for successive measurements [39]. Oda *et al* reported that 5 surgeons, measuring 50 radiographs, had an average error of 9° (calculated as twice the standard deviation) and that the main error source was in identifying end vertebrae [40]. High levels of variability have also been reported when

the end vertebrae of the curve are pre-selected. Morrissy *et al.*[10] reported an intra-subject variability of 2.8 ° and an inter-subject variability of 6.3 ° under conditions of pre-selection. These findings indicated that a change in a Cobb angle measurement of less than 10° cannot be interpreted with confidence as a real change [38]. Goldberg *et al* [41] showed interobserver variability of 2.5° and intraobserver reliability of 1.9° in a study by 4 evaluators of the primary curve identified in 30 radiographs. They also reported that the interclass correlation coefficient for the Cobb angle was 0.98. The interobserver standard deviation was 2.8° and the intraobserver standard deviation was 1.8° in a study by Ylikoski and Tallroth [42] in consecutive measurements of Cobb angle of 30 untreated patients having a mean Cobb angle of 24.4° by 2 readers using a specially designed angle-measuring instrument (“Plurimeter”).

Computer-assisted methods have enabled the evaluation of spinal curves with greater accuracy and lower measurement error compared to manual techniques [43]. Shea *et al.* reports that for manual measurements the intraobserver error was 3.3 ° and for computer-assisted measurement 2.6 ° [43].

Nachiappan *et al.* [44] proposed a system in which the vertebral column is subdivided into a number of segments and the observer marks the lateral and medial intersecting points. The maximum angle is taken as the Cobb angle. Recently, Stokes *et al.* proposed a new system in which the user marks standard landmark points and the computer logic automatically measures the Cobb angle and classifies the curve [38, 45]. In that study of patients with larger (preoperative) scoliosis, the average sample standard deviations of the Cobb angle were (intraobserver) 2.0° for upper and lower curves, and (interobserver) 2.5° and 2.6° for upper and lower curves, respectively.

## References:

- [1] <http://www.rad.washington.edu/mskbook/scoliosis.html>
- [2] Allen S. Edmonson, Scoliosis, in *Campbell's operative orthopaedics (Edition 8)*, A. H. Crenshaw, Ed, vol. 5, St Louis, Mosby Year Book, pp. 3611-3614, 1992.
- [3] Cobb J. R., “The treatment of Scoliosis,” *Connecticut Med. J.*, vol. 7, pp. 467, 1943.
- [4] Cobb J. R., “Outline for the study of scoliosis,” *AAOS Instr Course Lect.*, vol. 5, pp. 261-275, 1948.
- [5] Appelgren G. and Willner S., “End vertebra angle- a roentgenographic method to describe a scoliosis: a follow up study of idiopathic scoliosis treated with Boston brace, *Spine*, vol. 15, no. 2, pp. 71-74, 1990.
- [6] Thomas C. Schuler, Brian R. Subach, Charles L. Branch, Kevin T. Foley, J. Kenneth Burkus and the Lumbar Spine Study Group, “Segmental Lumbar Lordosis: Manual Versus Computer-

- Assisted Measurement Using Seven Different Techniques,” *J. Spinal Disorders and Techniques*, vol 17, no. 5, pp. 372-379, Oct 2004.
- [7] Risser J. C., and Ferguson A. B., “Scoliosis: its prognosis,” *J Bone Joint Surg.*, vol. 18, pp. 667-670, 1936.
- [8] Adam Greenspan, “Scoliotic Index: A Comparative evaluation of Methods of the Measurement of Scoliosis,” *Bulletin of the Hospital for Joint Diseases*, vol. 39, no. 2, pp 117:125, Oct 1978.
- [9] Carman D. L., Browne R. H. and Birch J. G., “Measurement of scoliosis and kyphosis radiographs: intraobserver and interobserver variation,” *J Bone Joint Surg Am.*, vol. 72, pp. 328–333, 1990.
- [10] Morrissy R. T., Goldsmith G. S., Hall E. C., Kehl D. and Cowie G. H., “Measurement of the Cobb angle on radiographs of patients who have scoliosis: evaluation of intrinsic error,” *J Bone Joint Surg Am.*, vol. 72, pp. 320–327, 1990.
- [11] Ponsetti I. V. and Friedman B., “Prognosis in Idiopathic Scoliosis,” *J Bone Joint Surg. Am.*, vol. 32, pp. 381-395, 1950.
- [12] King H. A., Moe J. H., Bradford D. S., and Winter R. B., “The selection of fusion levels in thoracic idiopathic scoliosis,” *J. Bone Joint Surg. Am.*, vol. 65, no. 9, pp. 1302–1313, 1983.
- [13] Lenke L. G., Betz R. R., Harms J., Bridwell K. H., Clements D. H., Lowe T. G. and Blanke K., “Adolescent idiopathic scoliosis: a new classification to determine extent of spinal arthrodesis,” *J. Bone Joint Surg. Am.*, vol. 83, no. 8, pp. 1169-81, 2001.
- [14] Qiu G., Zhang J., Wang Y., Xu H., Zhang J., Weng X., Lin J., Zhao Y., Shen J., Yang X., Luk K. D., Lu D. and Lu W. W., “A new operative classification of idiopathic scoliosis: A Peking Union Medical College Method,” *Spine*, vol. 30, no. 12, pp. 1419-1426, 2005.
- [15] Lowe T., Berven S. H., Schwab F. J. and Bridwell K. H., “The SRS classification for adult spinal deformity: building on the King/Moe and Lenke classification systems,” *Spine*, vol. 31, no. 19S, pp. 119-125, 2006.
- [16] Bernhardt M. and Bridwell K. H., “Segmental Analysis of the Sagittal Plane Alignment of the Normal Thoracic and Lumbar Spines and Thoracolumbar Junction,” *Spine*, vol. 14, pp. 717-721, 1989.
- [17] Hardacker J. W., Shuford R. F., Capicotto R. N. and Pryor P. W., “Radiographic Standing Cervical Segmental Alignment in Adult Volunteers Without Neck Symptoms,” *Spine*, vol. 22, pp. 1472-1480, 1997.
- [18] Wambolt A and Spencer D. L., “A segmental analysis of the distribution of lumbar lordosis in the normal spine,” *Orthopaedic Transactions*, vol. 11, pp. 92-93, 1987.
- [19] Xu X, Lee D. J., Antani S. and Long L. R., “Localizing contour points for indexing an x-ray image retrieval system,” in *Proc. of 16th IEEE Symposium on Computer-Based Medical Systems*, IEEE Computer Society, Los Alamitos, pp. 169-174, 2003.
- [20] Cootes T.F. and Taylor C. J., “Statistical Models of Appearance for Medical Image Analysis and Computer Vision,” in *Proc. of SPIE*, vol. 4322, pp. 236-248, 2001.
- [21] Smyth P. P., Taylor C. J. and Adams J. E., “Vertebral shape: automatic measurements with active shape models,” *Radiology*, vol. 211, pp. 571-578, 1999.
- [22] Long L. R. and Thoma G. R., “Use of shape models to search digitized spine x-rays,” in *Proc. of IEEE Computer Based Medical Systems 2000*, Houston, TX, pp 255-260, 2000.
- [23] Long L. R. and Thoma G. R., “Computer Assisted Retrieval of Biomedical Image Features from Spine X-rays: Progress and Prospects,” in *Proc. of 14<sup>th</sup> Symposium on Computer-Based Medical Systems*, Bethesda, MD, pp 46-50, 2001.
- [24] Long L. R. and Thoma G. R., “Feature indexing in a database of digitized x-rays,” in *Proc. of SPIE*, vol. 4315, pp. 393-403, 2001.
- [25] Thoma G. R., *Annual Report - Communications Engineering Branch*, LHNCBC, NLM, 2002.
- [26] Zamora G., Sari-Sarraf H. and Long L. R., “Hierarchical Segmentation of Vertebrae from X-ray Images,” in *Proc. of SPIE Medical Imaging: Image Processing*, San Diego, CA, February 15-20, 2003.

- [27] Tezmol A., Sari-Sarraf H., Mitra S. and Long R., "Customized Hough transform for robust segmentation of the cervical vertebrae from x-ray images," in *Proc. of 5<sup>th</sup> IEEE Southwest Symposium on Image Analysis and Interpretation*, Santa Fe, New Mexico, pp 224-228, 2002.
- [28] Yalin Zheng, Mark S. Nixon and Robert Allen, "Automatic lumbar vertebrae segmentation in fluoroscopic images via optimized Concurrent Hough Transform," in *Proc. of the 23<sup>rd</sup> Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Istanbul, Turkey, pp 2653-2656, 2001.
- [29] Zamora G., Sari-Sarraf H., Mitra S. and Long R., "Estimation of Orientation and Position of Cervical Vertebrae for Segmentation with Active Shape Models," in *Proc. of SPIE*, vol. 4322, pp. 378-387, 2000.
- [30] Long L. R. and Thoma G. R., "Landmarking and feature localization in spine x-rays," *J. Electronic Imaging*, vol. 10, pp. 939-956, 2001.
- [31] Jeffries B. F., Tarlton M., DeSmet A. A., Dwyer S. J. and Brower A. C., "Computerized measurement and analysis of scoliosis. A more accurate representation of the shape of the curve," *Radiology*, vol. 134, pp. 381-385, 1980.
- [32] Beauchamp M., Labelle H., Grimard G., Stanciu C., Poitras B. and Dansereau J., "Diurnal variation of Cobb angle measurement in adolescent idiopathic scoliosis," *Spine*, vol. 18, pp. 1581-1583, 1993.
- [33] Whittle M.W. and Evans M., "Instrument for measuring the Cobb Angle in Scoliosis," *Lancet*, vol. 414, Feb 1979.
- [34] Chen Y., "Vertebral centroid measurement of lumbar lordosis compared with the Cobb technique," *Spine*, vol. 24, pp. 1786-1790, 1999.
- [35] Chernukha K. V., Daffner R. H. and Reigel D. H., "Lumbar lordosis measurement; a new method versus Cobb technique," *Spine*, vol. 23, pp. 74-79, 1998.
- [36] Facanha-Filho F. A., Winter R. B., Lonstein J. E., Steven Koop, Thomas Novacheck, E. A. L'Heureux and Cheryl A. Noren, "Measurement accuracy in congenital scoliosis," *J Bone Joint Surg Am.*, vol. 83, pp. 42-45, 2001.
- [37] Harrison D. E., Harrison D. D., Cailliet R., Janik T. J. and Holland B., "Radiographic analysis of lumbar lordosis," *Spine*, vol. 26, no. 11, pp. E235-E242, 2001.
- [38] Stokes I. A. F. and Aronsson D. D., "Computer-assisted algorithms improve reliability of King classification and Cobb angle measurement of scoliosis," *Spine*, vol. 31, no. 6, pp. 665-670, 2006.
- [39] Berliner L., Somchai Kreang-Arekul and Kaufman L., "Scoliosis Evaluation by Direct Digital Radiography and Computerized Post-Processing," *J. Digital Imaging*, vol. 15 (Suppl. 1), pp. 270-274, 2002.
- [40] Oda M., Rauh S., Gregory P. B., Silverman F. N. and Bleck E. E., "The significance of roentgenographic measurement in scoliosis," *J Pediatr Orthop*, vol. 2, pp. 378-382, 1982.
- [41] Goldberg M. S., Poitras B, Mayo N. E. , Labelle H, Bourassa R and Cloutier R., "Observer variation in assessing spinal curvature and skeletal development in adolescent idiopathic scoliosis," *Spine*, vol. 13, pp. 1371-1377, 1988.
- [42] Ylikoski M. and Tallroth K., "Measurement variations in scoliotic angle, vertebral rotation, vertebral body height, and intervertebral disc space height," *J. Spinal Disorders and Techniques*, vol. 3, pp. 387-391, 1990.
- [43] Shea K. G., Stevens P. M., Nelson M., Smith J. T., Masters K. S. and Yandrow, "A comparison of manual versus computer-assisted radiographic measurement," *Spine*, vol. 23, pp. 551-555, 1998.
- [44] Nachiappan C., Dangerfield P. H., Giakas G., Cochrane T. and Dorgan J. C., "Computer-assisted Cobb measurement of scoliosis," *Eur Spine J.*, vol. 11, pp. 353-357, 2002.
- [45] Stokes I. A. F. and Aronsson D. D., "Identifying Sources of Variability in Scoliosis Classification Using a Rule-Based Automated Algorithm," *Spine*, vol. 27, no. 24, pp. 2801-2805, 2002.



---

## Chapter 4

# Basic Theory of Fundamental Tools

---

### 4.1 Introduction

This chapter provides a brief overview of the theory of fundamental tools used in this research. A brief review of the various classifiers used in the thesis is presented. A technique for multi-class support for binary classifiers is detailed next. Various performance measurement schemes employed in this research and statistical tools used for analysis are outlined.

### 4.2 Classifiers

Neural networks are frequently employed to classify patterns based on learning from examples. While different neural network paradigms employ different learning rules, all these paradigms determine pattern statistics from a set of training samples and then classify new patterns on the basis of these statistics.

#### 4.2.1 Probabilistic Neural Network (PNN)

The PNN was developed by Donald Specht following an approach developed in statistics, called Bayesian classifiers [1, 2, 3]. The PNN uses a supervised training set to develop distribution functions within a pattern layer. These functions, in the recall mode, are used to estimate the likelihood of an input feature vector being part of a learned category or class. The learned patterns can also be combined or weighted, with the a priori probability, also called the

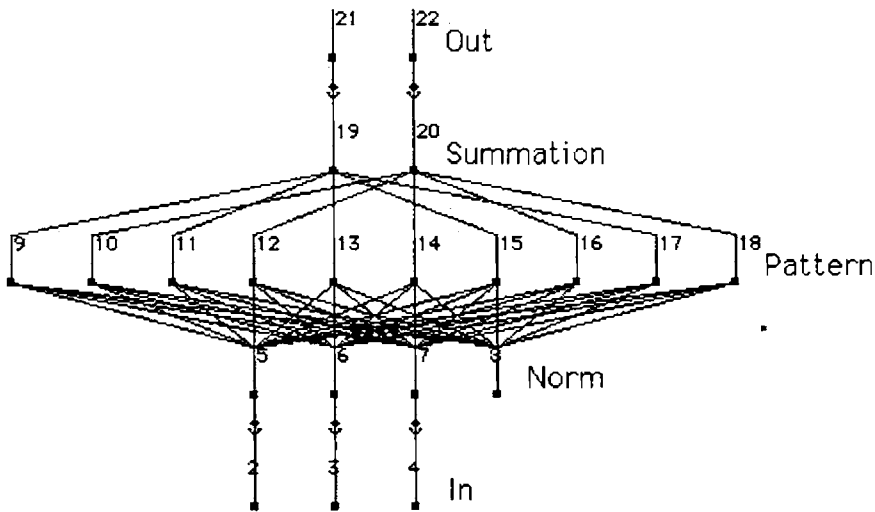


Figure 4.1 An example Probabilistic Neural Network

relative frequency, of each category to determine the most likely class for a given input vector. If the relative frequency of the categories is unknown, then all categories can be assumed to be equally likely and the determination of category is solely based on the closeness of the input feature vector to the distribution function of a class.

An example of a probabilistic neural network is shown in Figure 4.1. This network has three layers. The network contains an input layer which has as many elements as there are separable parameters needed to describe the objects to be classified. It has a pattern layer, which organizes the training set such that each input vector is represented by an individual processing element. And finally, the network contains an output layer, called the summation layer, which has as many processing elements as there are classes to be recognized. Each element in this layer combines via processing elements within the pattern layer which relate to the same class and prepares that category for output. Sometimes a fourth layer is added to normalize the input vector, if the inputs are not already normalized before they enter the network. The input vector must be normalized to provide proper object separation in the pattern layer.

#### 4.2.2 Learning Vector Quantization (LVQ) Network

Learning Vector Quantization (LVQ) network topology was originally suggested by Tuevo Kohonen in the mid 80's [4]. Topologically, this network contains an input layer, a single Kohonen layer and an output layer. An example network is shown in Figure 4.2. The output

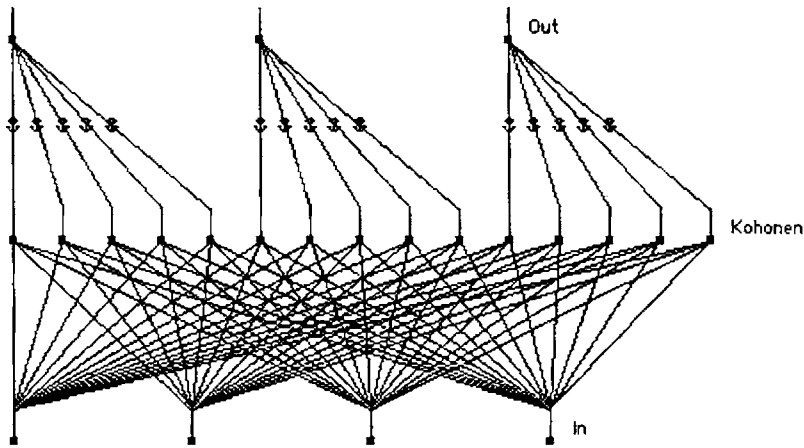


Figure 4.2 An example Learning Vector Quantization Network

layer has as many processing elements as there are distinct categories, or classes. The Kohonen layer has a number of processing elements grouped for each of these classes. The number of processing elements per class depends upon the complexity of the input-output relationship. It is the Kohonen layer that learns and performs relational classifications with the aid of a training set. This network uses supervised learning rules.

In the training mode, this supervised network uses the Kohonen layer such that the distance of a training vector to each processing element is computed and the nearest processing element is declared the winner. There is only one winner for the whole layer. The winner will enable only one output processing element to fire, announcing the class or category the input vector belonged to. If the winning element is in the expected class of the training vector, it is reinforced toward the training vector. If the winning element is not in the class of the training vector, the connection weights entering the processing element are moved away from the training vector. During the training process, individual processing elements assigned to a particular class migrate to the region associated with their specific class.

During the recall mode, the distance of an input vector to each processing element is computed and again the nearest element is declared the winner. That in turn generates one output, signifying a particular class found by the network.

### 4.2.3 Support Vector Machine (SVM)

The support vector machine (SVM) is a classifier derived from statistical learning theory by

Vapnik [5, 6]. SVM have been introduced for solving pattern recognition and nonlinear function estimation problems. In this method one maps the data into a high dimensional input space in which one constructs an optimal separating hyperplane. While classical neural network techniques suffer from the existence of many local minima [7], SVM solutions are obtained from quadratic programming problems possessing a global solution. Being based on the structural risk minimization principle, the quality and complexity of the SVM solution does not depend directly on the dimensionality of the input space. SVM approach is considered a good classifier because of its high generalization performance, even when the dimension of the input space is very high. Intuitively, given a set of points which belongs to either one of two classes, a linear SVM finds the hyperplane leaving the largest possible fraction of points of the same class on the same side, while maximizing the distance of either class from the hyperplane. According to [5], this hyperplane minimizes the risk of misclassifying examples of the test set.

Let  $(x_i, y_i)_{1 < i < N}$  be a set of training examples, where  $x_i \in \mathbb{R}^d$ , is the  $i^{\text{th}}$  input pattern and  $d$  being the dimension of the input space, and  $y_i$  the  $i^{\text{th}}$  output pattern belongs to a class labeled by  $y_i \in \{-1, 1\}$ . The aim is to define a hyperplane which divides the set of examples such that all the points with the same label are on the same side of the hyperplane. The general form of the linear classification function is

$$g(x) = w \cdot x + b \quad (4.1)$$

which corresponds to a separating hyperplane  $w \cdot x + b = 0$ .

We can normalize  $g(x)$  to satisfy  $|g(x)| \geq 1$  for all  $x_i$ , so that the distance from the closest point to the hyperplane is  $\frac{1}{\|w\|}$ . Among the separating hyperplanes, the one for which the distance to the closest point is maximal is called optimal separating hyperplane (OSH). Since the distance to the closest point is  $\frac{1}{\|w\|}$ , finding the OSH amounts to minimizing  $\|w\|$  and the objective function is:

$$\min \phi(w) = \frac{1}{2} \|w\|^2 = \frac{1}{2} (w \cdot w) \quad \text{subject to}$$

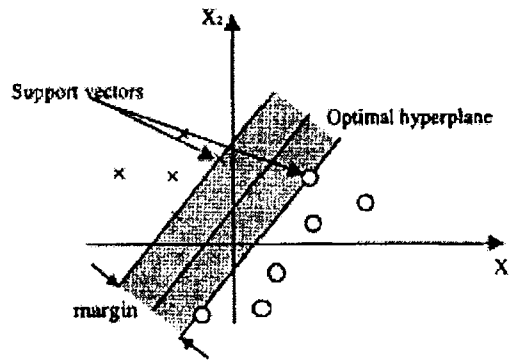


Figure 4.3 SVM classification with a hyperplane that maximizes the separating margin between the two classes (indicated by data points marked by “x”s and “O”s). Support vectors are elements of the training set that lie on the boundary hyperplanes of the two classes.

$$y_i(w \cdot x_i + b) \geq 1, \quad i = 1, \dots, N \quad (4.2)$$

Figure 4.3 shows the OSH, margin and support vectors of an SVM classifier.

If we denote by  $(\alpha_1, \dots, \alpha_N)$  the  $N$  non-negative Lagrange multipliers associated with constraints in Equation (4.2), we can uniquely construct the OSH by solving a constrained quadratic programming problem. The solution  $w$  has an expansion  $w = \sum_i \alpha_i y_i x_i$  in terms of a subset of training patterns, called support vectors, which lie on the margin. The classification function can thus be written as

$$f(x) = \text{sgn}\left(\sum_{i=1}^N \alpha_i y_i x_i \cdot x + b\right) \quad (4.3)$$

When the data is not linearly separable, on the one hand, SVM introduces slack variables and a penalty factor such that the objective function can be modified as

$$\phi(w, \xi) = \frac{1}{2}(w \cdot w) + C\left(\sum_1^N \xi_i\right) \quad (4.4)$$

The parameter  $C$  is chosen by the user, a larger  $C$  corresponding to assigning a higher penalty to errors.

On the other hand, the input data can be mapped through some nonlinear mapping into a higher dimensional feature space in which OSH is constructed. Thus the dot product can be

represented by  $K(x, y) = (\phi(x) \cdot \phi(y))$  when the kernel satisfies the Mercer's condition [5].

Finally we obtain the classification function

$$f(x) = \text{sgn}\left(\sum_{i=1}^N \alpha_i y_i K(x_i, x) + b\right) \quad (4.5)$$

For the kernel function,  $K(., .)$  one typically has the following choices:

1. Linear,

$$K(x, x_i) = x_i^T x \quad (4.6)$$

2. Polynomial,

$$K(x, x_i) = (x_i^T x + t)^d \quad (4.7)$$

where  $t$  the intercept and  $d$  the degree of the polynomial.

3. Radial basis function,

$$K(x, x_i) = \exp(-\|x - x_i\|^2 / \sigma^2); \quad (4.8)$$

where  $\sigma^2$  is the variance of the Gaussian kernel.

4. Multilayer perceptron,

$$K(x, x_i) = \tanh[\kappa x_i^T x + \theta]; \quad (4.9)$$

where  $\kappa$  is the scale parameter and  $\theta$  the bias.

These kernel parameters need to be fine tuned for optimum performance at the time of training.

#### 4.2.4 Least Squares Support Vector Machine (Ls-SVM)

In least squares (LS) version of SVM's, one finds the solution by solving a linear system instead of quadratic programming [8]. This is due to the use of equality instead of inequality constraints in the problem formulation.

The support vector method aims at constructing a classifier of the form Equation (4.5),

which can be re-represented as  $f(x) = \text{sgn}\left(\sum_{j=1}^h w_j \varphi_j(x) + b\right) = \text{sgn}(w^T \varphi(x) + b)$ ,

where  $w^T = [w_1, w_2, \dots, w_h]^T$ ,  $\varphi = [\varphi_1, \varphi_2, \dots, \varphi_h]^T$ . The classifier is constructed as follows.

For the case of two classes, one assumes

$$w^T \varphi(x_i) + b \geq 1, \text{ if } y_i = +1 \quad (4.10)$$

$$w^T \varphi(x_i) + b \leq -1, \text{ if } y_i = -1 \quad (4.11)$$

which is equivalent to

$$w^T [\varphi(x_i) + b] \geq 1, i = 1, \dots, N \quad (4.12)$$

where  $\varphi(\cdot)$  is a nonlinear function which maps the input space into a higher dimensional space.

Ls-SVM classifiers are obtained as solution to the following optimization problem:

$$\min_{w, b, e} J_{LS}(w, b, e) = \frac{1}{2} w^T w + C \frac{1}{2} \sum_{k=1}^N e_k^2 \quad (4.13)$$

subject to the equality constraints

$$y_k [w^T \psi(x_k) + b] = 1 - e_k, k = 1, \dots, N. \quad (4.14)$$

In Equation (4.13), the first term is responsible to find a smooth solution, while the second one minimizes the training errors ( $C$  is the trade-off parameter between the terms).

From this, the following Lagrangian can be formed:

$$L(w, b, e; \alpha) = J_{LS} - \sum_{k=1}^N \alpha_k \{y_k [w^T \psi(x_k) + b] - 1 + e_k\} \quad (4.15)$$

where  $\alpha_k$  are Lagrange multipliers, which can be either positive or negative due to the equality

constraints as follows from the Kuhn-Tucker conditions [9]. The support values  $\alpha_k$  are proportional to the errors at the data points in the Ls-SVM case, while in the standard SVM case many support values are typically equal to zero. The kernel functions are same as that of SVM (Equation 4.6 to 4.9).

#### 4.2.5 $k$ -Means Clustering

$k$ -means [10] is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume  $k$  clusters) fixed a priori. The main idea is to define  $k$  centroids, one for each cluster. These centroids should be placed in a cunning way, because different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early groupage is done. At this point we need to re-calculate  $k$  new centroids as barycenters of the clusters resulting from the previous step. After we have these  $k$  new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop we may notice that the  $k$  centroids change their location step by step until no more changes happen. In other words centroids do not move any more.

Finally, this algorithm aims at minimizing an objective function, in this case a squared error function. The objective function

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2 \quad (4.16)$$

where  $\|x_i^{(j)} - c_j\|^2$  is a chosen distance measure between a data point  $x_i^{(j)}$  and the cluster centre  $c_j$ , is an indicator of the distance of the  $n$  data points from their respective cluster centers. The algorithm is composed of the following steps [11]:

- (1) Choose  $k$  cluster centers to coincide with  $k$  randomly-chosen patterns or  $k$  randomly defined points inside the hypervolume containing the pattern set.



- (2) Assign each pattern to the closest cluster center.
- (3) Re-compute the cluster centers using the current cluster memberships.
- (4) If a convergence criterion is not met, go to step 2. Typical convergence criteria are: no (or minimal) reassignment of pattern to new cluster centers, or minimal decrease in squared error.

Although it can be proved that the procedure will always terminate, the  $k$ -means algorithm does not necessarily find the most optimal configuration, corresponding to the global objective function minimum. The algorithm is also significantly sensitive to the initial randomly selected cluster centers. The  $k$ -means algorithm can be run multiple number of times to reduce this effect.

### 4.3 Error-Correcting Output Coding (ECOC)

SVM's are designed for binary classification. Multiclass support for SVM is provided in our implementation using ECOC technique. Error-correcting output coding (ECOC) proposed by Dietterich and Bakiri [12] is a method for combining binary classifiers to solve multi-class problems. Each class is given a unique binary string of length  $L$  called codewords.  $L$  binary functions are learned to produce  $L$  binary classifiers. Table 4.1 shows a set of 10-bit codewords assigned to five classes, with the columns  $f_0, \dots, f_9$  representing the classifiers to be learned. All columns are distinct and each classifier learned is a disjunction of the classes. For example,  $f_1(z) = 1$  if  $z$  is in class 1 or 5.

ECOC is robust to changes in the size of the training dataset and to the assignment of codewords to classes. ECOC claims to have good error-correction capability, which is measured by the minimum Hamming distance  $d$  between any pair of codewords. Each codeword is able to

**Table 4.1** Error-correcting output codes for five-class problem with minimum Hamming distance of five.

Class	$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$	$f_8$	$f_9$
1	1	1	1	1	1	1	1	1	1	1
2	0	0	0	0	0	0	0	0	1	1
3	0	0	0	0	1	1	1	1	0	0
4	0	0	1	1	0	0	1	1	0	0
5	0	1	0	1	0	1	0	1	0	1

correct at least  $\frac{d-1}{2}$  single-bit errors. For example, the set of codewords in Table 4.1 has a Hamming distance of five, allowing it to correct up to two single-bit errors. This error-correction capability is dependent on the fact that the errors committed by each bit of the output are uncorrelated.

In order to have good error-correction capability, the codewords should possess the following properties:

- Row separation. Each codeword should be well separated from each other in terms of Hamming distance.
- Column separation. Each bit in a column  $f_i$  should be uncorrelated with another column  $f_j$ , where  $j \neq i$ . There should be a large Hamming distance between different columns, and also between a column and the complement of other columns. This ensures that the columns are neither identical nor complementary.

Given  $K \leq 7$  classes, an exhaustive method can be used to select the optimal codewords. For larger  $K$ , a randomized hill-climbing algorithm is used. First,  $K$  random strings of codewords with length  $L$  are generated. Next, the algorithm attempts to improve the minimum Hamming distance by searching for pairs of rows closest to each other and columns that are either too close or too far apart. The four codeword bits that the two rows and columns intersect are located and changed to improve their separation.

## 4.4 Performance Measurements

To evaluate different algorithms for image retrieval, an effective performance measurement is necessary. Different performance measurements have been proposed [13, 14]. The performance measurements are usually based on statistics of the subjective tests. Different performance measurements often use different subjective tests, resulting in different definitions of retrieval accuracy. The different performance measurements are discussed in this section.

### 4.4.1 Recall and Precision Pair (RPP)

RPP is the most widely used retrieval performance measurement in literature. It is basically based on categorical matching. In this method, dataset is converted to binary set according to

relevance or irrelevance to the query, based on subjective test. In the subjective test, each subject selects items relevant to the query from the dataset. Items selected for each query by more than the predefined number of subjects are considered relevant to the query; otherwise, they are treated as irrelevant. The precision and recall are then defined as

$$P = \frac{r}{n_1} = \frac{\text{number of relevant retrieved images}}{\text{number of retrieved images}}$$

$$R = \frac{r}{n_2} = \frac{\text{number of relevant retrieved images}}{\text{total number of relevant images in the database}} \quad (4.17)$$

Precision measures the retrieval accuracy while recall measures the ability of retrieving relevant items from the database. Precision and recall are inversely related, i.e., precision normally degenerates as recall increases.

#### 4.4.2 Average Normalized Modified Retrieval Rank (ANMRR)

In order to be objective in comparisons, a quantitative measure called ANMRR was developed on the basis of the specification of a data set, a query set and the corresponding ground-truth data [15]. The ground-truth data is a set of visually similar images for a given query image.

Consider a query  $q$  with a ground-truth size of  $NG(q)$ . The rank of the  $k^{\text{th}}$  retrieved image for query  $q$  is defined as

$$\text{Rank}^*(k) = \begin{cases} \text{Rank}(k), & \text{if } \text{Rank}(k) \leq K(q) \\ 1.25K, & \text{if } \text{Rank}(k) > K(q) \end{cases}$$

where  $K$ , the tolerance of the system, is defined with respect to the ground truth size  $NG(q)$  as follows.

$$K = 2 * NG(q); \text{ if } NG(q) \geq 20$$

$$K = 3 * NG(q); \text{ if } NG(q) \geq 10$$

$$K = 4 * NG(q); \text{ otherwise.}$$

The average rank (AVR) for query  $q$ ,  $AVR(q) = \frac{1}{NG(q)} \sum_{k=1}^{NG(q)} Rank^*(k)$  and modified retrieval rank (MRR) is  $MRR(q) = AVR(q) - 0.5 * [1 + NG(q)]$ . Normalizing with respect to  $NG(q)$  leads to normalized modified retrieval rank (NMRR)

$$NMRR(q) = \frac{MRR(q)}{1.25 * K - 0.5 * [1 + NG(q)]}. \quad (4.18)$$

$NMRR(q)$  can take on values between 0 (indicating whole ground truth found) and 1 (indicating none from ground truth found), irrespective of ground truth size. The average normalized modified retrieval rank (ANMRR) indicates the retrieval quality over all queries and is defined as

$$ANMRR = \frac{1}{NQ} \sum_{q=1}^{NQ} NMRR(q) \quad (4.19)$$

where NQ is the total number of queries.

## 4.5 Statistical Tools

Various statistical tools used for data analysis in this research are outlined in this section.

### 4.5.1 The McNemar Test

The McNemar test, first suggested for use in benchmark tests of speech recognition systems by Gillick and Cox [16], applies to an evaluation consisting of discrete items which are either correct or incorrect and are independent of one another. It is used to calculate the statistical significance of differences in classification accuracy for related samples. The McNemar test is useful for comparing the classification accuracy of various classifiers.

The discrete items in an evaluation may be classified according to whether each of the two systems under consideration attained correct classification of the item. The resulting counts may be represented by a 2x2 error matrix as shown in Table 4.2. It can be noted from the table that  $f_{12}$  is the number of cases correctly classified by classifier 1 and wrongly classified

**Table 4.2 2x2 error matrix to calculate the statistical significance of differences in classification accuracy based on McNemar test for related samples.**

	Classifier 2 ↓		
	Allocation	Correct	Incorrect
Classifier 1 →	Correct	$f_{11}$	$f_{12}$
	Incorrect	$f_{21}$	$f_{22}$

(incorrect) by classifier 2, and  $f_{21}$  is the number of cases correctly classified by classifier 2 and wrongly classified (incorrect) by classifier 1 .

The McNemar test score  $Z$  is given by:

$$Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \quad (4.20)$$

A positive value  $Z$  indicates classifier 1 is more accurate than classifier 2 and negative value of  $Z$  indicate classifier 2 is more accurate. A positive score of more than 1.96 indicates that first method is a better than the second one at 95% confidence level.

### 4.5.2 Kappa Statistics

To assess the accuracy of any measurement, it is usual to distinguish between the reliability of the data collected and their validity. Reliability is essentially the extent of agreement between repeated measurements and validity is the extent to which a method of measurement provides a true assessment of that which it purports to measure. When studying the variability of observer *categorical* ratings, an important component of possible lack of accuracy is observer disagreement. This is indicated by how observers classify individual subjects into the same category on the measurement scale. Kappa coefficient is one of the most common approaches in this regard [17].

Kappa statistics is an index which compares the agreement against that which might be expected by chance. Kappa can be thought of as chance-corrected proportional agreement, and possible values range from +1 (perfect agreement) via 0 (no agreement above that expected by

**Table 4.3 Kappa statistics computation matrix**

		Measurement A		Total
		Incorrect	Correct	
Measurement B	Incorrect	$f_{11}$	$f_{12}$	$P_{21}$
	Correct	$f_{21}$	$f_{22}$	$P_{22}$
Total		$P_{11}$	$P_{12}$	$T$

Note:

$f_{11}$  – number of cases in which measurement A and measurement B is incorrect;

$f_{12}$  – number of cases in which measurement A is correct and measurement B is incorrect;

$f_{21}$  – number of cases in which measurement A is incorrect and measurement B is correct;

$f_{22}$  – number of cases in which measurement A and measurement B is correct;

$$T = f_{11} + f_{12} + f_{21} + f_{22};$$

$$P_{21} = (f_{11} + f_{12}) / T;$$

$$P_{22} = (f_{21} + f_{22}) / T;$$

$$P_{11} = (f_{11} + f_{21}) / T;$$

$$P_{12} = (f_{12} + f_{22}) / T.$$

chance) to -1 (complete disagreement). A Kappa statistics computation matrix is shown in Table 4.3.

$$Kappa = \frac{\text{Observed Agreement} - \text{Chance Agreement}}{1 - \text{Chance Agreement}} \quad (4.21)$$

where Observed Agreement =  $(f_{11} + f_{22}) / T$ ;

$$\text{Chance Agreement} = (P_{21} * P_{11}) + (P_{22} * P_{12});$$

### 4.5.3 Correlation Coefficient

Two variables are said to be "correlated" or "associated", if knowledge of scores for one of them helps to predict scores for the other. Capacity to predict is measured by a correlation coefficient that can indicate some amount of relationship, no relationship, or some amount of inverse relationship between the variables.

Pearson's correlation coefficient,  $r$ , is the most commonly used measure of association. It summarizes the linear relationship between two variables having ranked categories. That is, the variables may be continuous, ordinal, interval or ratio. Its formula is as follows:

$$r = \frac{1}{n-1} \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{S_x} \right) \left( \frac{Y_i - \bar{Y}}{S_y} \right) \quad (4.22)$$

In this equation,  $n$  is the sample size,  $\bar{X}$  is the observed sample mean for variable  $x$ ,  $\bar{Y}$  is the observed sample mean for variable  $y$ ,  $S_x$  is the standard deviation for variable  $x$  and  $S_y$  is the standard deviation for variable  $y$ .  $X_i$  and  $Y_i$  represent the values of variables  $x$  and  $y$  for the  $i$ <sup>th</sup> individual in the sample. The values for  $r$  generated by this equation will range from -1 to 1. A value of -1 indicates a perfect inverse correlation between the variables (that is, when the value for one variable is high, the value for the other variable is low). A value of 0 indicates no relationship between the variables. A value of 1 indicates perfect correlation between the variables.

## References:

- [1] D.F. Specht and P.D. Shapiro, "Generalization accuracy of probabilistic neural network compared with backpropagation network," in *Proc. of JCNN'91*, pp. 458-461, 1991.
- [2] D. F. Specht, "Probabilistic Neural Networks," *Neural Networks*, vol. 3, pp.109-118, 1990.
- [3] P.D. Wasserman, *Neural Computing: Theory and Practice*, Van Nostrand Reinhold, New York, 1989.
- [4] Kohonen T., *Self-Organization and Associative Memory*, 2nd Edition, Springer-Verlag, Berlin 1987.
- [5] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.
- [6] C. Cortes and V. Vapnik, "Support vector networks," *Machine Learning*, vol. 20, pp. 1-25, 1995.
- [7] Bishop C.M., *Neural networks for pattern recognition*, Oxford University Press, 1995.
- [8] Suykens J.A.K., and Vandewalle J., "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, 1999.
- [9] Fletcher R., *Practical methods of optimization*, 2<sup>nd</sup> Edition, John Wiley and Sons, 1987.
- [10] J. B. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations," in *Proc. of 5-th Berkeley Symp. on Mathematical Statistics and Probability*, Berkeley, University of California Press, vol. 1, pp. 281-297, 1967.
- [11] A.K. Jain, M. N. Murty and P. J. Flynn, "Data clustering: A review", *ACM Computing Surveys*, vol. 31, no. 3, pp. 264-323, 1999.
- [12] Dietterich T. G. and Bakiri G., "Solving multiclass learning problems via Error-Correcting Output Codes," *J. Artificial Intelligence Research*, vol. 2, pp 263-286, 1995.
- [13] G. Lu and A. Sajjanhar, "On Performance Measurement of Multimedia Information Retrieval Systems," in *Proc. of Int'l Conf. on Computational Intelligence and Multimedia Applications*, Australia, pp.781-787, 1998.
- [14] A. Del Bimbo, *Visual Information Retrieval*, Morgan Kaufmann Publishers, Inc. San Francisco, USA, pp. 56-57, 1999.
- [15] *Introduction to MPEG-7: Multimedia Content Description Interface*, B. S. Manjunath, P. Salembier, and T. Sikora, Eds., John Wiley & Sons Ltd., pp. 183-184, 2002.
- [16] L. Gillick and S. Cox, "Some Statistical Issues in the Comparison of Speech Recognition Algorithms", in *Proc. of ICASSP*, vol. 1, pp. 532-535, 1989.
- [17] Cohen J, "A Coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, pp 37-46, 1960.

---

## Chapter 5

# Improved Legendre Moment Descriptor

---

### 5.1 Introduction

Various region-based and contour-based shape descriptors are proposed in the literature as described in Section 2.3. Among them, MPEG-7 proposed Angular Radial Transformation Descriptor (ARTD) as region-based shape descriptor and Curvature Scale Space Descriptor (CSSD) as contour-based shape descriptor [1]. The Zernike Moment Descriptor (ZMD) also has been a popular choice as region-based shape descriptor [2]. In this chapter novel region-based and contour-based shape descriptors with a generic name, Improved Legendre Moment Descriptor (ILMD), which is more accurate than MPEG-7 descriptors for shape description and retrieval is proposed. The implementation details of the ZMD, ARTD and CSSD are provided. The motivations for proposing a new shape descriptor based orthogonal Legendre polynomials are discussed. Theory, implementation details and transformation invariance properties of the proposed ILMD are discussed. Comparison of the retrieval accuracy of the state-of-the-art descriptors with ILMD using MPEG-7 region-shape database and contour-shape database is then detailed. A scheme for affine transformation invariance and experimental results are reported. Finally, a method for CBIR using  $k$ -means clustering of the shape descriptors is presented.

### 5.2 MPEG-7 Shape Descriptors

This section gives an introduction to the theory and implementation details of MPEG-7



shape descriptors. They are implemented and used for performance comparison with the proposed shape descriptor.

### 5.2.1 Zernike Moment Descriptor (ZMD)

Zernike moment descriptor was a major candidate as region-based descriptor in early MPEG-7 experiments [2]. Based on the theory of orthogonal polynomials, Teague proposed the use of orthogonal moments to recover the image from their moment representation and introduced Zernike moments [3]. This allowed independent moment invariants to be constructed to an arbitrarily high order. The complex Zernike moments are derived from Zernike polynomials:

$$V_{nm}(x, y) = V_{nm}(\rho \cos \theta, \rho \sin \theta) = R_{nm}(\rho) \exp(jm \theta) \quad (5.1)$$

where  $\rho$  is the length of vector from  $(x, y)$  to the shape centroid and  $\theta$  is the angle between vector  $\rho$  and  $x$  axis taken in counter-clockwise direction.

The Zernike polynomials are a complete set of complex-valued functions orthogonal on the unit disc, i.e.,  $(x^2 + y^2) \leq 1$ . The radial polynomial  $R_{nm}(\rho)$  is defined as

$$R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} \rho^{n-2s}. \quad (5.2)$$

where  $n = 0, 1, 2, \dots, \infty$  and  $m$  takes on positive and negative integer values subject to the conditions  $n - |m| = \text{even}$ ,  $|m| \leq n$ .

The complex Zernike moments for a discrete image intensity function  $F(x, y)$  of order  $n$  with repetition  $m$  are defined as:

$$A_{nm} = \frac{n+1}{\pi} \sum_x \sum_y F(x, y) V_{nm}^*(x, y), \quad x^2 + y^2 \leq 1 \quad (5.3)$$

where symbol  $*$  denotes the complex conjugate. Due to the constraint of  $n - |m| = \text{even}$  and  $m < n$ , there are  $n/2$  repetition of moments in each order  $n$ .

Since Zernike basis functions take the unit disc as their domain, this disc must be specified before moments can be calculated. In our implementation, the distance from centroid to the farthest pixel of the image is taken as the radius of the disc [4]. This provides scale invariance. The translation invariance is achieved by using the centroid as the origin of the unit disc. The magnitudes of Zernike moment coefficients are then normalized by dividing them by the mass of the shape, further enhancing scale invariance. The rotation of the shape around the unit disc is expressed as a phase change. If  $\theta$  is the angle of rotation,  $A_{mn}^R$  is the Zernike moment of the rotated image and  $A_{mn}$  is the Zernike moment of the original image, then

$$A_{mn}^R = A_{mn} \exp(-jn\theta) \quad (5.4)$$

Hence magnitude of ZMD provides rotational invariance. The precision of shape representation depends upon the number of moments truncated from the expansion. For efficient retrieval, the first 35 moments of order  $1 \dots 10$  are used as ZMD. The similarity between two shapes indexed with Zernike moments descriptors is determined by the city block distance between the two Zernike moments vectors. The block diagram of the whole process of computing ZMD is shown in Figure 5.1.

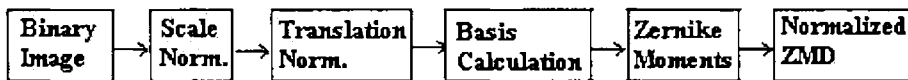


Figure 5.1 Block diagram of computing ZMD

### 5.2.2 Angular Radial Transformation Descriptor (ARTD)

The 2-D Angular Radial Transformation (ART) [5] belongs to the broad Zernike Moment family and provides a compact and efficient way to express pixel distribution within a 2-D object region. The set of orthogonal moment basis is defined on a unit disc in polar coordinates  $(\rho, \theta)$ .

From each shape, a set of ART coefficients  $F_{nm}$  is extracted, using the following formula.

$$F_{nm} = \langle V_{nm}(\rho, \theta), f(\rho, \theta) \rangle$$

$$= \int_0^{2\pi} \int_0^1 V_{nm}^*(\rho, \theta) f(\rho, \theta) \rho \, d\rho \, d\theta \quad (5.5)$$

where  $f(\rho, \theta)$  is an image intensity function in polar coordinates and  $V_{nm}(\rho, \theta)$  is the ART basis function of order  $n$  and  $m$ . The basis functions are separable along the angular and radial directions, and are defined as follows:

$$V_{nm}(\rho, \theta) = \frac{1}{2\pi} \exp(jm\theta) R_n(\rho), \quad (5.6)$$

where  $R_n(\rho) = \begin{cases} 1, & n = 0 \\ 2 \cos(\pi n \rho), & n \neq 0 \end{cases}$

The radial direction  $m$ , is an integer subject to the constraints:  $n - |m|$  is even and  $|m| \leq n$ . The 2-D ART basis functions,  $V_{nm}(\rho, \theta)$ , are complex functions.

Since ART basis function takes the unit disc as their domain, all shapes are normalized into a unit disc of fixed radius of 128 pixels. As the ART and Zernike moments belong to the same family, they have similar transformation invariance properties. The ART descriptor is defined as a set of normalized magnitudes of complex ART coefficients, which guarantee rotation invariance. For scale normalization ART coefficients are divided by the magnitude of the ART coefficient of order  $n = 0, m = 0$ , which is not used as a descriptor element. The first 35 moments coefficients of order  $n = 10$  and  $m = 10$  are used as the ART shape descriptor. The descriptor extraction process is similar to that of ZMD except that all image pixels are mapped to a unit disc of radius 128. The similarity measure used is city block distance.

### 5.2.3 Curvature Scale Space Descriptor (CSSD)

Scale space method has been shown to be very useful in shape analysis since Asada and Brady used it to derive a primal sketch from a shape [6]. Its use for shape retrieval as a contour-based shape descriptor has been proposed by Mokhtarian *et al* [7]. To improve the representation accuracy, an enhanced CSSD was proposed by Abbasi *et al* [8].

To create a CSS description of a contour shape,  $N$  equidistant points are selected on the contour, starting from an arbitrary point on the contour and following the contour clockwise. The  $x$ -coordinates and  $y$ -coordinates of the selected  $N$  points are grouped together into two

series  $X$  and  $Y$  respectively. The contour is then gradually smoothed by the repetitive application of a low-pass filter with the kernel  $(0.25, 0.5, 0.25)$  to  $X$  and  $Y$  coordinates of the selected  $N$  contour points. As a result of the smoothing, the contour evolves and its concave parts gradually flatten-out, until the contour becomes convex. A so-called CSS image can be associated with the contour evolution process. The horizontal coordinates of the CSS image correspond to the indices of the contour points selected to represent the contour  $(1, \dots, N)$ , and vertical coordinates of the CSS image correspond to the amount of filtering applied, defined as the number of passes of the filter. Each horizontal line in the CSS image corresponds to the smoothed contour resulting from  $k$  passes of the filter. For each smoothed contour, the zero-crossings of its curvature function are computed. Curvature zero-crossing points separate concave and convex parts of the contour. Each zero-crossing is marked on the horizontal line corresponding to the smoothed contour and at the location corresponding to the position of this zero-crossing along the contour. The CSS image has characteristic peaks. The coordinate values of the prominent peaks  $(x_{css}, y_{css})$  in the CSS image are extracted. The peaks are ordered based on decreasing values  $y_{css}$ , transformed using a nonlinear transformation and quantized. In addition, the eccentricity and circularity of the contour are also calculated, quantized and stored.

Figure 5.2 shows the shape evolution during filtering (left) and the corresponding CSS image (right). The contour curvature zero crossings  $(A, B, \dots, H)$  and the corresponding points on the CSS image are marked.

The CSSD descriptor consists of the eccentricity and circularity values of the original and filtered contour, the index indicating the number of peaks in the CSS image, the height of the

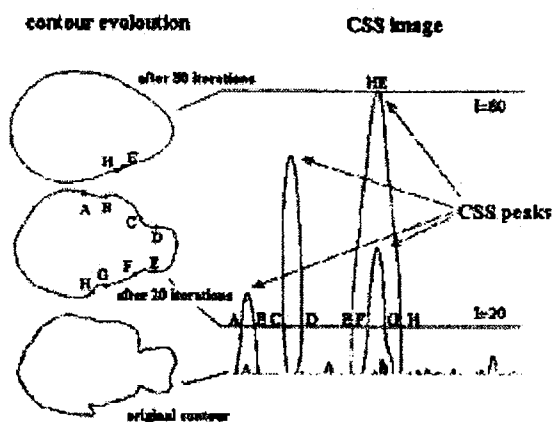


Figure 5.2 Example of contour evolution and corresponding CSS [1]

highest peak and the  $x$  and  $y$  positions on the remaining peaks.

CSS descriptor is translation invariant. Scale invariance is achieved by normalizing all the shapes into fixed number of boundary points. In our implementation, number of boundary points is 128 points. Since rotation of shape causes circular shifting of CSS peaks, the rotation invariance is achieved by circular shifting the highest peak (*primary peak*) to the origin of the CSS map. The similarity between two shapes is then measured by the summation of the peak differences between all the matched peaks and the peak values of all the unmatched peaks [7].

### 5.3 Motivations of Proposing ILMD

This section details how the idea of the proposed descriptor is formulated. It is believed that human beings discriminate shapes mainly by their contour features. The majority of real world objects have clear contours which are readily available. Thus object contour is an important parameter in describing a shape. The powerfulness of CSSD owes to its ability to capture key local features such as the locations and the degrees of convexity (or concavity) of curve segments on the shape boundary. But, CSSD has matching difficulties and it only captures the local shape features. The global features which are also important to shape representation are missed out from the representation. As such, global features such as eccentricity, circularity and number of CSS peaks should be combined to form more practical descriptors. There may be no CSS descriptors for smooth convex shapes such as polygon composed of straight lines. All these factors affect the retrieval result.

The contour-based shape descriptors are usually suitable for describing contour shape without sophisticated boundary. Region-based shape descriptors can be applied to more general applications and are more robust in describing shape with sophisticated boundary compared with contour-based shape descriptors, because they exploit more shape information to derive shape features. Dengsheng Zhang and Guojun Lu conducted a study in which accuracy of region-based descriptor ZMD was evaluated against contour-based descriptors using a contour shape database [9]. The study concludes that ZMD has better contour description ability than CSSD. This is due to the fact that the entire contour pixels of a shape contribute to the ZMD where as few locations of the contour with a convexity account for CSSD. Thus ZMD can be used as a region and contour based descriptor. The ARTD also belongs to the Zernike moment family and hence has similar properties.

Recon Error:	50	31	25	6	7	3	1
	Using Legendre Moments						
Recon Error:	73	64	68	52	42	21	2
	Using Zernike Moments						
Maximum Order of Moments:	7	8	9	10	11	12	13

Figure 5.3 Image Reconstruction of letter 'E' using Legendre moments and Zernike moments [10].

Mukundan *et. al.* reports that Legendre moments based on orthogonal Legendre polynomial have better image representation capability than Zernike moments [10]. The reconstruction error using Legendre moments was less than that using Zernike moments of the same order. Figure 5.3 demonstrates the error in reconstruction of an image using Legendre moments and Zernike moments of various orders. Thus Legendre moments better represent an image with fewer coefficients in comparison with Zernike moments. Liao and Pawlak in [11] and Hakim El Fadili *et. al.* in [12] also reports the use of Legendre moments for image representation. Legendre moments were used in several pattern recognition applications [13 - 18]. These results were the motivation for exploring the Legendre moments for use as a shape descriptor. But, Legendre moments were not popular as a shape descriptor for CBIR. The lack of popularity of Legendre moments as a shape descriptor is due to its lack of inherent invariance to geometric transformations. Hence, this research investigated on improving invariance of the Legendre moments and shape descriptors based on Legendre polynomials.

## 5.4 Legendre Moments

A brief discussion on the theory of Legendre moments is presented in this section. Moments with an orthogonal basis can be used to attain a zero value of redundancy measure in a set of moment functions, so that they correspond to independent characteristics of the image [16]. The image intensity distribution can also be analytically reconstructed from its orthogonal moments. The Legendre polynomials are defined along rectangular image coordinate axes inside a unit circle. Moments with the Legendre polynomial as kernel functions were first introduced by Teague [3]. A detailed overview of Legendre moments is presented in [19].

The  $p^{\text{th}}$  order Legendre polynomial is defined as follows.

$$P_p(x) = \frac{1}{2^p p!} \frac{d^p}{dx^p} (x^2 - 1)^p, \quad x \in [-1, 1]$$

$$= \sum_{j=0}^p a_{pj} x^j \quad (5.9)$$

The values of  $a_{pj}$  can be expressed as:

$$a_{pj} = (-1)^{(p-j)/2} \frac{(p+j)!}{2^p ((p-j)/2)! ((p+j)/2)! j!}, \text{ and } (p-j) \text{ is even}$$

Another expression for  $P_p(x)$  is as follows

$$P_p(x) = \frac{1}{2^p} \sum_{k=0}^{p/2} (-1)^k \frac{(2p-2k)!}{k! (p-k)! (p-2k)!} x^{p-2k}, \text{ } (p-k) \text{ is even} \quad (5.10)$$

The Legendre polynomials form a complete orthogonal set inside the unit rectangle, and hence

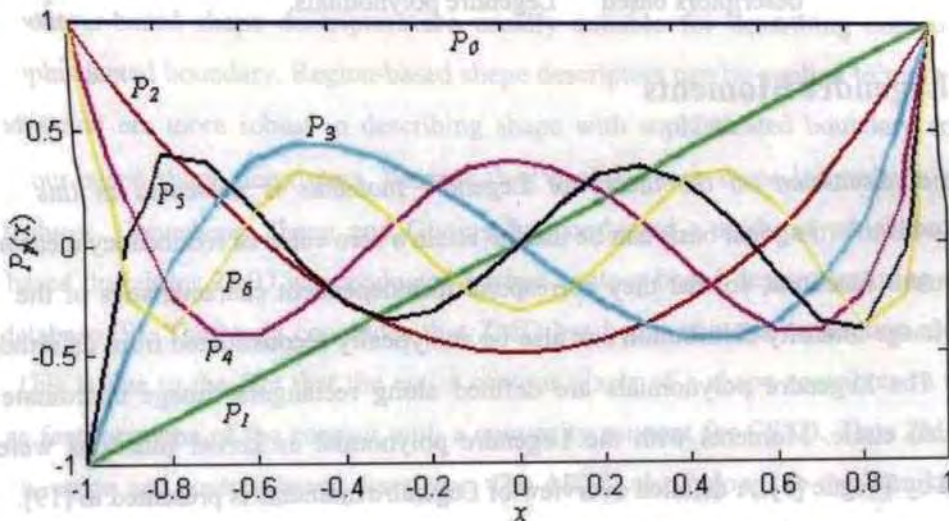


Figure 5.4 Plot of Legendre polynomial function values

$$\int_{-1}^1 P_p(x) P_q(x) dx = \frac{2}{2p+1} \delta_{pq} \quad (5.11)$$

The plots of the functions  $P_p(x)$ ,  $p = 0, 1, \dots, 6$  is given in Figure 5.4. The kernel of Legendre moments are products of Legendre polynomials defined along rectangular image coordinate axes. The Legendre moments of order  $(p+q)$  for an image function  $f(x, y)$  are defined as

$$L_{pq} = \frac{(2p+1)(2q+1)}{4} \int_{-1}^1 \int_{-1}^1 P_p(x) P_q(y) f(x, y) dx dy \quad (5.12)$$

In order to evaluate the Legendre moments, the image coordinate space has to be necessarily scaled so that their representative magnitudes are less than 1. For a digital image  $F(i, j)$  with dimension along the  $x$  and  $y$  axis  $M$  and  $N$  respectively, discrete version of Legendre moments can be written as

$$\lambda_{pq} = \frac{(2p+1)(2q+1)}{(M-1)(N-1)} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} P_p(x_i) P_q(y_j) F(i, j) \quad (5.13)$$

where  $x_i$  and  $y_j$  denote the normalized pixel coordinates in the range  $[-1, 1]$ , given by

$$x_i = \frac{2i - M + 1}{M - 1}; \quad y_j = \frac{2j - N + 1}{N - 1}.$$

The inverse moment transform which follows from orthogonality of Legendre polynomials can be similarly expressed as

$$F(i, j) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \lambda_{mn} P_m \left( \frac{2i - M + 1}{M - 1} \right) P_n \left( \frac{2j - N + 1}{N - 1} \right), \quad i = 0, \dots, M - 1, j = 0, \dots, N - 1 \quad (5.14)$$

The reconstruction of an image from a set of moments from order 0 to order  $n_{max}$  uses the truncated form of the series in Equation (5.14), to get a polynomial approximation of  $F(i, j)$ .



**Table 5.1 List of Legendre moments up to order 6**

Order	Legendre moment of order $p + q$	Number of moments in each order	Total number of moments up to order 6
0	$\lambda_{0,0}$	1	28
1	$\lambda_{0,1}, \lambda_{1,0}$	2	
2	$\lambda_{0,2}, \lambda_{1,1}, \lambda_{2,0}$	3	
3	$\lambda_{0,3}, \lambda_{1,2}, \lambda_{2,1}, \lambda_{3,0}$	4	
4	$\lambda_{0,4}, \lambda_{1,3}, \lambda_{2,2}, \lambda_{3,1}, \lambda_{4,0}$	5	
5	$\lambda_{0,5}, \lambda_{1,4}, \lambda_{2,3}, \lambda_{3,2}, \lambda_{4,1}, \lambda_{5,0}$	6	
6	$\lambda_{0,6}, \lambda_{1,5}, \lambda_{2,4}, \lambda_{3,3}, \lambda_{4,2}, \lambda_{5,1}, \lambda_{6,0}$	7	

The Legendre moment coefficients of a given order representing an image are referred in this thesis as Legendre Moment Descriptor (LMD). As an example, the first 28 Legendre moments of up to order 6 are given in Table 5.1.

## 5.5 Development of ILMD

According to MPEG-7 specification a shape descriptor should be invariant to geometric transformations like translation, rotation and scaling. The LMD does not have these properties. The following procedure is developed to improve invariance to transformations. ILMD is used as a generic name to refer the proposed descriptor.

### 5.5.1 Translation

The object shape need not be always centered to the image. The translation invariance is achieved by selecting the centroid of the shape as the center of the Legendre polynomial basis function. This is achieved by defining a ‘bounding box’ to contain only the shape, with shape centroid as ‘bounding box’ center. Figure 5.5 demonstrates various translated shapes and the new region for computing Legendre moments (‘bounding box’). The image dimensions are shown in gray rectangle and ‘bounding box’ is shown in red rectangle. Table 5.2 lists the first 10 coefficients (truncated to decimal point precision 2) of LMD and translation invariant ILMD of the three shapes shown in Figure 5.5. The last column shows the city block distance between moment coefficients of images with that of case ‘1’ for both the descriptors. It can be seen that the coefficients of the three shapes for ILMD are identical showing the translational invariance.

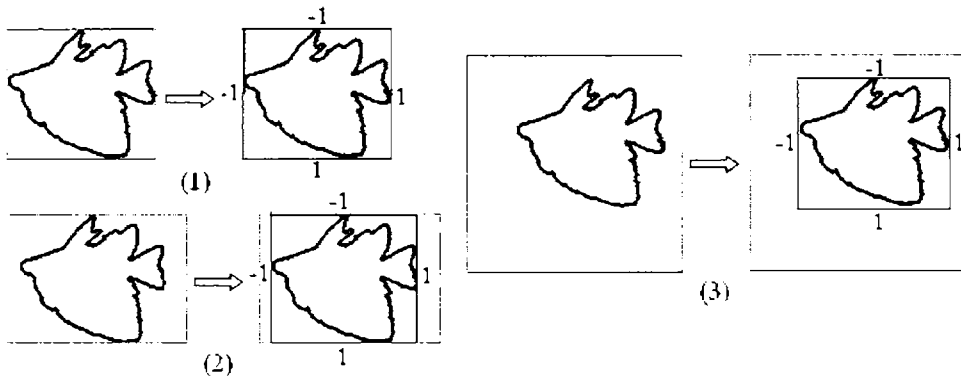


Figure 5.5 Various translated shapes and their proposed region of Legendre moment computation.

Table 5.2 First 10 Legendre moments of images shown in Figure 5.5

Case	Descriptor *	$\lambda_{00}$	$\lambda_{01}$	$\lambda_{10}$	$\lambda_{02}$	$\lambda_{11}$	$\lambda_{20}$	$\lambda_{03}$	$\lambda_{12}$	$\lambda_{21}$	$\lambda_{30}$	distance
1	I	0.91	-0.04	0.01	0.05	0.00	0.05	0.03	-0.01	0.00	-0.05	0.00
	II	0.91	-0.04	0.01	0.05	0.00	0.05	0.03	-0.01	0.00	-0.05	0.00
2	I	0.93	-0.01	0.01	0.09	0.00	0.04	0.04	-0.01	0.00	-0.05	0.11
	II	0.91	-0.04	0.01	0.05	0.00	0.05	0.03	-0.01	0.00	-0.05	0.00
3	I	0.94	-0.04	0.03	0.05	0.02	0.07	0.02	-0.02	0.04	-0.03	0.17
	II	0.91	-0.04	0.01	0.05	0.00	0.05	0.03	-0.01	0.00	-0.05	0.00

\* Descriptor I – LMD; II - ILMD

### 5.5.2 Rotation

As Legendre moment coefficients are defined in the unit rectangle, they are not invariant to rotation. Rotation invariance can be achieved by aligning the major axis of the shape parallel to the x-axis. After aligning the major axis, the Legendre moment is computed in the region of ‘bounding box’ as specified in Section 5.5.1. A shape flipped or rotated to any degree can be finally assigned to one of the four cases after major axis alignment as shown in Figure 5.6 ((2) – (5)). As the basis set is defined in the interval [-1 1], the rectangular region can be divided into four quadrants. An image portion, irrespective of the quadrant in which it is present due to rotation, will have identical LMD with a possible sign change after major axis alignment, as each quadrant can be assumed to be defined over the basis set [0 1] or [0 -1]. This is due to the

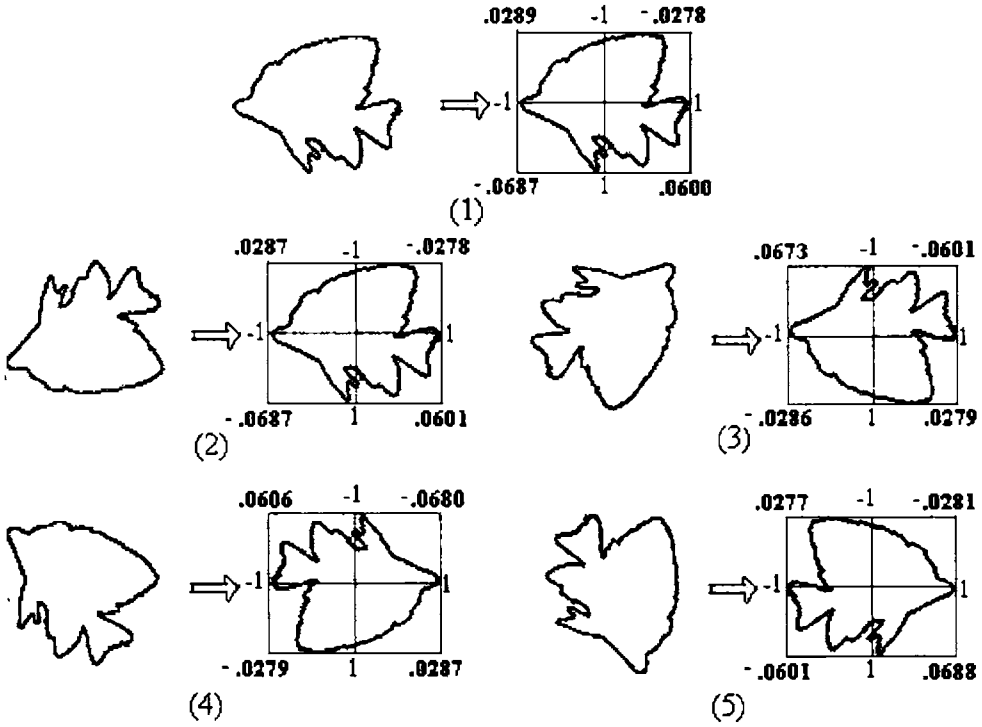


Figure 5.6 An example demonstrating four possible cases of major axis alignment

symmetry property of Legendre polynomial function with respect to origin.

The Legendre moment of order  $(p + q)$  for an  $M \times N$  image can be represented as

$$\lambda_{pq} = \lambda_{pq}^1 + \lambda_{pq}^2 + \lambda_{pq}^3 + \lambda_{pq}^4 \quad (5.15)$$

where  $\lambda_{pq}^n$  represent the Legendre moment of quadrant  $n = 1, \dots, 4$ .

These coefficients can be computed as follows:

$$\lambda_{pq}^1 = \frac{(2p+1)(2q+1)}{(M-1)(N-1)} \sum_{i=0}^{(M/2)-1} \sum_{j=0}^{(N/2)-1} P_p\left(\frac{2i-M+1}{M-1}\right) \times P_q\left(\frac{2j-N+1}{N-1}\right) \times F(i, j)$$

$$\lambda_{pq}^2 = \frac{(2p+1)(2q+1)}{(M-1)(N-1)} \sum_{i=0}^{(M/2)-1} \sum_{j=(N/2)}^{N-1} P_p\left(\frac{2i-M+1}{M-1}\right) \times P_q\left(\frac{2j-N+1}{N-1}\right) \times F(i, j)$$

$$\lambda_{pq}^3 = \frac{(2p+1)(2q+1)}{(M-1)(N-1)} \sum_{i=M/2}^{M-1} \sum_{j=0}^{(N/2)-1} P_p\left(\frac{2i-M+1}{M-1}\right) \times P_q\left(\frac{2j-N+1}{N-1}\right) \times F(i, j)$$

$$\lambda_{pq}^4 = \frac{(2p+1)(2q+1)}{(M-1)(N-1)} \sum_{i=M/2}^{M-1} \sum_{j=N/2}^{N-1} P_p\left(\frac{2i-M+1}{M-1}\right) \times P_q\left(\frac{2j-N+1}{N-1}\right) \times F(i, j)$$
(5.16)

Figure 5.6 (1) shows a shape whose major axis is originally aligned parallel to the x-axis. Figure 5.6 ((2) – (5)) show the original shapes with various degree of rotation and the resulting shape after major axis alignment. The ‘bounding box’ and quadrants are shown in red color. The Legendre moments of order  $p = 1$  and  $q = 1$  for each quadrant is marked on the corner. It can be noted that, a portion of shape irrespective of the quadrant in which it is present has similar LM coefficient, excepting a sign change.

Although theoretically coefficients should be exactly same, the slight variation observed in practice is due to the mapping error introduced due to image rotation at the time of generation of candidate shapes and major axis alignment. The proposed rotation invariant descriptor is computed by taking the sum of the absolute value of each quadrant as shown below.

$$\lambda'_{pq} = \left| \lambda_{pq}^1 \right| + \left| \lambda_{pq}^2 \right| + \left| \lambda_{pq}^3 \right| + \left| \lambda_{pq}^4 \right|$$
(5.17)

**Table 5.3 First 10 Legendre moments of images shown in Figure 5.6**

Case	Descriptor *	$\lambda_{00}$	$\lambda_{01}$	$\lambda_{10}$	$\lambda_{02}$	$\lambda_{11}$	$\lambda_{20}$	$\lambda_{03}$	$\lambda_{12}$	$\lambda_{21}$	$\lambda_{30}$	distance
1	I	0.08	0.04	0.01	-0.02	0.00	-0.02	-0.03	-0.01	0.00	-0.05	0.00
	II	0.10	0.15	0.15	0.03	0.18	0.03	0.10	0.13	0.11	0.11	0.00
2	I	0.48	-0.01	0.01	-0.52	-0.12	-0.57	0.01	-0.02	0.01	-0.02	1.71
	II	0.10	0.15	0.15	0.03	0.18	0.03	0.10	0.13	0.11	0.11	0.00
3	I	0.04	-0.01	0.00	-0.05	0.00	-0.06	0.02	0.00	0.03	0.00	0.31
	II	0.10	0.15	0.15	0.03	0.18	0.03	0.10	0.13	0.10	0.11	0.01
4	I	0.04	0.00	0.01	-0.06	0.00	-0.05	0.00	-0.03	0.00	-0.02	0.23
	II	0.10	0.15	0.15	0.03	0.18	0.03	0.10	0.13	0.10	0.11	0.01
5	I	0.50	0.01	0.01	-0.57	0.11	-0.51	-0.02	-0.01	-0.02	-0.02	1.66
	II	0.10	0.15	0.15	0.03	0.18	0.03	0.10	0.13	0.10	0.11	0.01

\* Descriptor I – LMD; II – ILMD

Table 5.3 lists the first 10 coefficients of Legendre moments (Descriptor I) and rotation invariant ILMD (Descriptor II) of shapes shown in Figure 5.6. The last column shows the corresponding city block distance between coefficients of each descriptor with respect to case 1. It can be noted that rotation invariance improved descriptor for the rotated shapes is closely matching in comparison with LMD.

### 5.5.2.1 Major Axis Algorithm

Usual major axis determination algorithms are applicable for contour shapes. Here an algorithm applicable to both contour and region shape is developed. Normally, the major axis is found by traversing all the points on the shape boundary and determining the line joining the two boundary points with the farthest distance. However, for region shape, boundary information is not known a priori. So, the major axis for a shape is computed by searching the outer border point pairs on the shape boundary in a number of directions. The algorithm involves three major steps. (i) finding the bounding box of the shape; (ii) finding the pair of boundary points in the range of 0 to 179 degree (with respect to x-axis); (iii) finding the two points of the farthest apart in the found boundary points.

The bounding box is determined by computing the left and top most shape point and width and breadth of the object in the image. Then the image is cropped to an enclosing rectangle to contain only the shape. The pair of boundary points is determined by traversing the line passing through the centroid of the shape at various orientations. The search limits are bounding box borders. The first shape point which comes across on the line is the starting point and last shape point is the end point. A list of pairs of boundary point is maintained by scanning the shape for various orientations. A scan in the range of 0 to 179 degree incrementing by 1 degree will cover the entire object shape. Now, the distance between start and end points of each boundary pair is computed. The major axis orientation will correspond to orientation of the boundary pairs which are farthest apart. This information is used to align the major axis of the shape parallel to the x-axis.

### 5.5.3 Scaling

Images normally appear in various scales and it is important for the shape descriptor to be

invariant to scaling, for efficient image retrieval. Other invariance properties should be maintained while providing scale invariance. A new scheme is proposed and an already existing scheme is evaluated for scale invariance to Legendre moments.

### 5.5.3.1 Method I

It is not necessary that the image dimension to be proportionally changed as per scale change of the object shape it contains. The basis function for Legendre polynomial is defined in the unit range  $[-1, 1]$  in both dimensions. So, using the ‘bounding box’ of the object shape, instead of entire image area, for Legendre moment computation will provide scale invariance. The zeroth order moment corresponds to the mass of the image. Further improvement in scale invariance can be achieved by normalizing each Legendre moment coefficient with the zeroth order moment. The developed descriptor combining the previously defined invariance properties (Equation (5.17)) is denoted as ILMD1 and is defined as

$$ILMD1 = \left[ \frac{\lambda'_{01}}{\lambda'_{00}}, \frac{\lambda'_{10}}{\lambda'_{00}}, \frac{\lambda'_{02}}{\lambda'_{00}}, \frac{\lambda'_{11}}{\lambda'_{00}}, \frac{\lambda'_{20}}{\lambda'_{00}}, \dots \right] \quad (5.18)$$

### 5.5.3.2 Method II

Chee-Way Chong *et al.* proposed a set of scale invariants of Legendre moments, which is shown to be better than invariants generated using corresponding geometric moments [20]. Here, the descriptors remain unchanged for elongated, contracted and reflected non-symmetrical as well as symmetrical images. The scale invariants proposed in [20] is given below.

Assuming that the original object is scaled non-uniformly with different factors,  $a$  and  $b$ , along  $x$  and  $y$ -axes, respectively, the scale Legendre moments can be defined by modifying Equation (5.12) as follows:

$$L_{pq} = \frac{(2p+1)(2q+1)}{4} \int_{-1}^1 \int_{-1}^1 P_p(ax) P_q(ay) f(x, y) dx dy, \quad (a \neq b) \in (R - \{0\}) \quad (5.19)$$

The scaled Legendre polynomials along  $x$ -axis can be expressed as a series of decreasing powers of  $x$  as follows:

$$P_p(ax) = B_{pp}a^p x^p + B_{p(p-2)}a^{p-2} x^{p-2} + B_{p(p-4)}a^{p-4} x^{p-4} + \dots + B_{pk}a^k x^k \quad (5.20)$$

$$\text{where } B_{pk} = (-1)^{(p-k)/2} \frac{1}{2^p} \frac{(p+k)!}{((p-k)/2)!((p+k)/2)!k!}$$

The relation between the original and scaled Legendre polynomials is then formed by rearranging Equation (5.20) as shown below:

$$\sum_{k=0}^p \delta_{pk} P_k(ax) = a^p \sum_{k=0}^p \delta_{pk} P_k(x); \quad (p-k) \text{ is even} \quad (5.21)$$

where  $\delta_{pp} = 1$  and

$$\delta_{pk} = \sum_{r=0}^{\dot{p}-2} \frac{-B_{(p-r)k} \delta_{p(p-r)}}{B_{kk}}; \quad \dot{p} - r \text{ is even} \quad (5.22)$$

$$\dot{p} = (p - k) \geq 2$$

By employing Equations (5.21) and (5.22), the scaled Legendre polynomials along  $y$ -direction can then be deduced as follows:

$$\sum_{d=0}^q \delta_{qd} P_d(by) = b^q \sum_{d=0}^q \delta_{qd} P_d(y); \quad (q-d) \text{ is even} \quad (5.23)$$

Equations (5.21) and (5.23) form the kernel of the scale invariants of Legendre moments. The invariants are denoted as  $\psi_{pq}$ . They are expressed as a series of  $(p+q)^{\text{th}}$ ,  $(p+q-2)^{\text{th}}$ ,  $(p+q-4)^{\text{th}}$ , etc. orders of original or scaled Legendre moments as follows:

$$\psi_{pq} = \sum_{k=0}^p \sum_{d=0}^q \left[ \frac{C_{pq}}{C_{kd}} \delta_{pk} \delta_{qd} \dot{L}_{kd} \right] = a^{p+1} b^{q+1} \sum_{k=0}^p \sum_{d=0}^q \left[ \frac{C_{pq}}{C_{kd}} \delta_{pk} \delta_{qd} L_{kd} \right] \quad (5.24)$$

$$\text{where } C_{pq} = \frac{(2p+1)(2q+1)}{(M-1)(N-1)}$$

Using the following relations:

1.  $\psi_{00} = abL_{00}$ ,
2.  $\psi_{p0} = a^{p+1}b \sum_{k=0}^p \frac{C_{p0}}{C_{k0}} \delta_{pk} L_{k0}$ ,
3.  $\psi_{0q} = ab^{q+1} \sum_{d=0}^q \frac{C_{0q}}{C_{0d}} \delta_{qd} L_{0d}$ ,

the scale factor  $a$  and  $b$  contained in Equation (5.24) can be cancelled out. The normalized scale invariants of Legendre moments,  $\omega_{pq}$ , are subsequently derived as follows:

$$\omega_{pq} = \frac{\psi_{pq} \psi_{00}^{\xi+1}}{\psi_{(p+\xi)0} \psi_{0(q+\xi)}}; \quad p, q \text{ and } \xi = 0, 1, 2, 3, \dots \quad (5.25)$$

The descriptors derived in (5.25) are denoted as aspect ratio invariants. They are applicable to images with uniform as well as non-uniform scaling. Combining translation and rotation invariance properties to the scale invariance can be done by aligning the major axis of the shape to the  $x$ -axis and taking Legendre basis function domain as the shape's 'bounding box'. Thus, using Equation (5.17), the new invariant descriptor denoted as ILMD2 could be expressed as

$$ILMD2 = \frac{\sum_{k=0}^p \sum_{d=0}^q \left[ \frac{C_{pq}}{C_{kd}} \delta_{pk} \delta_{qd} \lambda'_{kd} \right] * (\lambda'_{00})^{\xi+1}}{\sum_{k=0}^{p+\xi} \frac{C_{(p+\xi)0}}{C_{k0}} \delta_{(p+\xi)k} \lambda'_{k0} * \sum_{d=0}^{q+\xi} \frac{C_{0(q+\xi)}}{C_{0d}} \delta_{(q+\xi)d} \lambda'_{0d}}; \quad (5.26)$$

where  $p, q$  and  $\xi = 0, 1, 2, 3, \dots$

The process of deriving ILMD1 and ILMD2 are summarized in a flowchart shown in Figure 5.7.

### 5.5.4 Comparison of LMD, ILMD1 and ILMD2

The retrieval effectiveness of the LMD and newly developed descriptors ILMD1 and



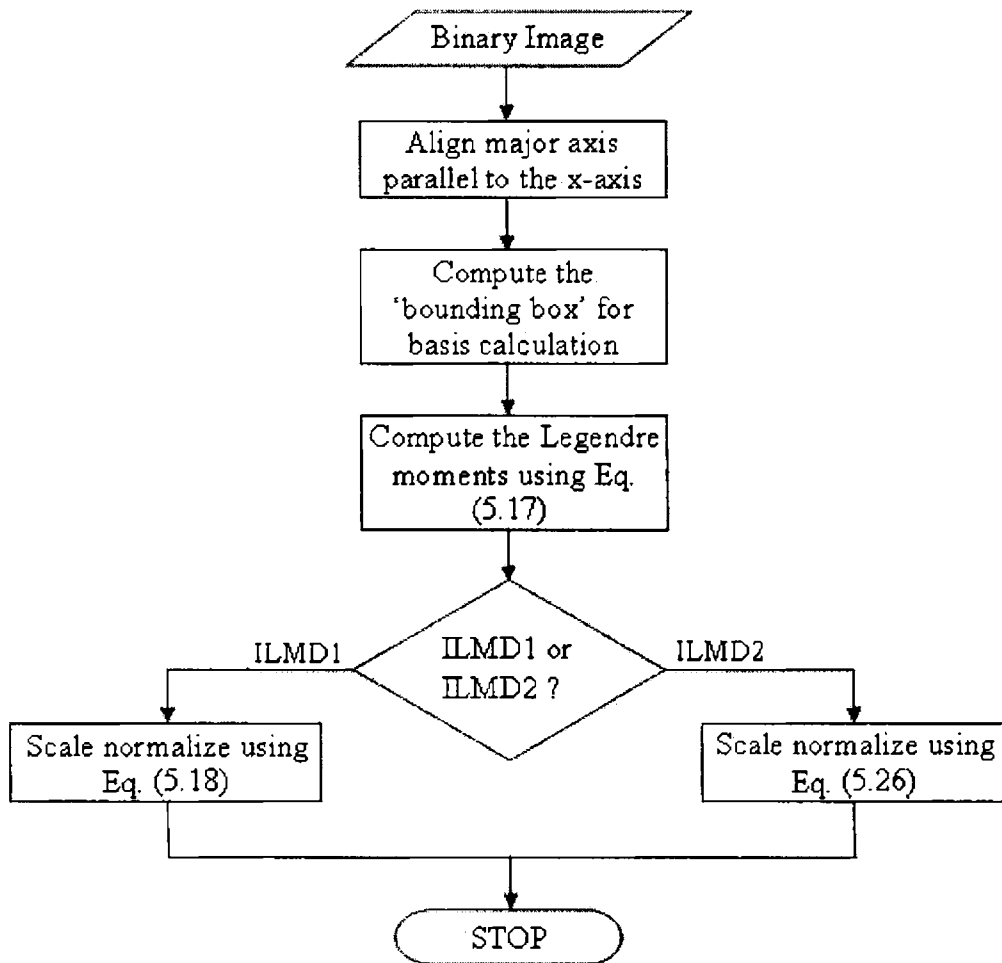


Figure 5.7 Flowchart of development of ILMD1 and ILMD2

ILMD2 are tested on the region-based shape database MPEG-7 CE-2 whole. This database is chosen for the selection of a better descriptor, as it contains a large number of images with small to large shape variations. MPEG-7 region shape whole database CE-2 consists of 3621 shapes of mainly trademarks. 651 shapes from 31 classes of shapes are selected as queries. Each class has 21 members generated through scaling, rotation and perspective transformation. The performance evaluation is carried out using recall and precision pair (RPP) method outlined in Section 4.4.1. In all cases Legendre moments of order up to  $p + q = 6$  are chosen as descriptor coefficients. The primary aim of this experiment is to choose the best descriptor under similar test conditions. Tuning for the optimum number of coefficients and similarity measure is deferred until the best descriptor is selected. Figure 5.8 shows precision – recall plot of the LMD, ILMD1 and ILMD2 using MPEG-7 CE-2 whole database.

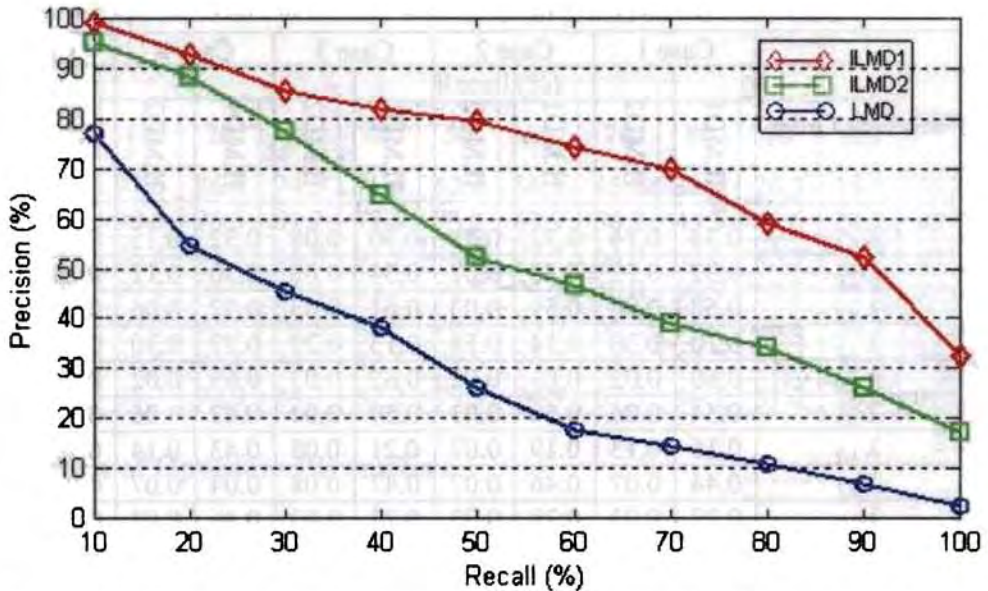


Figure 5.8 Average precision-recall of 651 retrievals using LMD, ILMD1 and ILMD2 on MPEG-7 region shape whole database CE-2.

It is observed that LMD has the lowest performance due to its lack of invariance to transformations like translation, rotation and scaling. The performance of ILMD1 is found to be better than that of ILMD2. Inferior performance of ILMD2 is due to its lack of discriminative ability. Figure 5.9 shows some images from MPEG-7 CE-2 database scaled at 40% of actual size. First three shapes are the same but with different scales. Last two shapes are of different category from the first set. Table 5.4 lists their ILMD1 and ILMD2 descriptors up to order 6 truncated to decimal point precision 2. The last row lists the city block distance of each shape with the first shape. For ILMD2, distance of the scaled shapes (shapes (2) and (3)) is less in comparison with that of ILMD1. But, distance of dissimilar shapes (shapes (4) and (5)) is less



Figure 5.9 Some examples of MPEG-7 region shape whole database CE-2

Table 5.4 ILMD1 and ILMD2 descriptors up to order 6 for shapes shown in Figure 5.9

Moment Order	Case 1		Case 2		Case 3		Case 4		Case 5	
	ILMD1	ILMD2	ILMD1	ILMD2	ILMD1	ILMD2	ILMD1	ILMD2	ILMD1	ILMD2
$\lambda_{0,1}$	0.34	0.14	0.32	0.08	0.30	0.08	0.32	0.15	0.31	0.15
$\lambda_{1,0}$	0.32	0.12	0.28	0.07	0.34	0.08	0.30	0.12	0.29	0.11
$\lambda_{0,2}$	0.58	0.06	0.59	0.03	0.61	0.03	0.52	0.06	0.51	0.06
$\lambda_{1,1}$	0.10	0.30	0.14	0.18	0.13	0.20	0.27	0.30	0.38	0.30
$\lambda_{2,0}$	0.60	0.02	0.61	0.02	0.62	0.03	0.65	0.02	0.65	0.02
$\lambda_{0,3}$	0.54	0.06	0.58	0.03	0.59	0.04	0.42	0.06	0.42	0.06
$\lambda_{1,2}$	0.16	0.13	0.19	0.07	0.21	0.08	0.43	0.14	0.46	0.14
$\lambda_{2,1}$	0.44	0.07	0.46	0.07	0.47	0.08	0.04	0.07	0.06	0.08
$\lambda_{3,0}$	0.27	0.03	0.28	0.03	0.27	0.03	0.46	0.03	0.47	0.03
$\lambda_{0,4}$	0.38	0.05	0.41	0.03	0.38	0.03	0.37	0.05	0.44	0.05
$\lambda_{1,3}$	0.59	0.13	0.61	0.08	0.59	0.09	0.88	0.13	1.01	0.13
$\lambda_{2,2}$	0.65	0.03	0.59	0.02	0.64	0.02	1.37	0.02	1.34	0.03
$\lambda_{3,1}$	0.47	0.10	0.42	0.08	0.41	0.08	0.33	0.10	0.38	0.10
$\lambda_{4,0}$	0.20	0.02	0.16	0.02	0.16	0.02	0.12	0.02	0.10	0.02
$\lambda_{0,5}$	0.08	0.05	0.11	0.03	0.10	0.03	0.38	0.05	0.38	0.05
$\lambda_{1,4}$	0.31	0.10	0.21	0.06	0.29	0.07	0.22	0.11	0.29	0.11
$\lambda_{2,3}$	0.71	0.02	0.80	0.03	0.76	0.03	0.40	0.02	0.30	0.02
$\lambda_{3,2}$	0.37	0.03	0.44	0.03	0.42	0.03	0.56	0.03	0.57	0.04
$\lambda_{4,1}$	0.41	0.07	0.49	0.07	0.45	0.07	0.38	0.07	0.34	0.07
$\lambda_{5,0}$	0.11	0.02	0.14	0.02	0.15	0.02	0.19	0.02	0.20	0.02
$\lambda_{0,6}$	0.27	0.04	0.30	0.02	0.38	0.02	0.07	0.05	0.08	0.05
$\lambda_{1,5}$	1.43	0.11	1.42	0.07	1.36	0.07	0.83	0.11	0.85	0.11
$\lambda_{2,4}$	0.23	0.01	0.28	0.01	0.29	0.01	1.84	0.01	1.81	0.01
$\lambda_{3,3}$	0.34	0.02	0.47	0.03	0.44	0.04	1.08	0.02	0.93	0.03
$\lambda_{4,2}$	1.63	0.03	1.62	0.03	1.69	0.03	1.35	0.02	1.36	0.02
$\lambda_{5,1}$	0.56	0.08	0.40	0.07	0.49	0.08	0.10	0.08	0.09	0.08
$\lambda_{6,0}$	0.69	0.02	0.65	0.02	0.74	0.02	0.24	0.02	0.23	0.02
Distance	0.00	0.00	1.25	0.60	1.12	0.55	7.88	0.06	8.05	0.09

than the distance to the similar shapes. For ILMD1, scaled shapes are closer than the dissimilar shapes. This accounts for the inferior retrieval performance of the ILMD2. It has less discriminative ability between shapes. Moreover, ILMD2 is computationally intensive than ILMD1. So, ILMD1 is selected as the optimum descriptor among the three candidate descriptors.

**Table 5.5 Retrieval performance of ILMD1 with different orders on MPEG-7 CE-2**

Order $p+q$	No. of coefs.	Recall (%)										Mean
		10	20	30	40	50	60	70	80	90	100	
1 to 3	9	97.6	86.4	74.1	71.4	62.4	56.1	45.8	32.9	26.3	14.3	56.73
1 to 4	14	98.6	90.5	81.2	77.8	73.2	67.9	56.2	45.5	41.6	24.7	65.72
1 to 5	20	98.8	91.5	84.8	81.9	76.1	72.7	64.0	52.6	45.9	28.9	69.72
1 to 6	27	99.1	92.7	85.4	82.0	79.5	74.2	70.0	58.9	52.2	32.6	72.66
1 to 7	35	99.0	92.2	84.6	82.9	78.7	73.0	66.2	56.7	50.6	29.3	71.32
1 to 8	44	98.8	92.1	84.7	82.1	79.5	74.7	68.0	57.2	51.4	32.3	72.08
1 to 9	54	98.7	92.0	84.7	82.2	79.5	74.9	67.3	56.3	51.0	29.2	71.58

### 5.5.5 Selection of Optimum Parameters

According to MPEG-7, it is not necessary that image could be reconstructed from its shape descriptor. The number of moment coefficients used is an important parameter in deciding the retrieval accuracy and compactness of the descriptor. For ILMD1, the zeroth order moment will be unity after scale normalization, so it is not used in the descriptor. To select the optimum number of moments to be used, retrieval accuracy experiments are conducted using ILMD1 and MPEG-7 CE-2 database with starting order 1 and ending order in the range from 3 to 9. Table 5.5 lists the precision for various recall rates using different order of ILMD1. The last column shows the overall performance for a given order. As we trade off retrieval precision with number of coefficients, ILMD1 with order 1 to 6 is found to be best suitable. It is also compact with only 27 numbers of coefficients in the descriptor.

Another parameter to be fixed is the similarity measure. Both Euclidean distance and city block distance can be used to define the distance between the feature vectors. To decide best similarity measure, retrieval is conducted using both distance measures and compared. Figure 5.10 illustrates the retrieval accuracy of ILMD1 with order 1 to 6 using Euclidean and city block distance measure on MPEG-7 CE-2 database. It can be noted that the performance of city block distance measure is slightly superior to Euclidean distance. Hence, city block distance measure is adopted as the similarity measure for ILMD1.

To conclude, ILMD1 with coefficients in the order 1 to 6 is found to be a suitable candidate as a shape descriptor. The optimum similarity measure is city block distance.

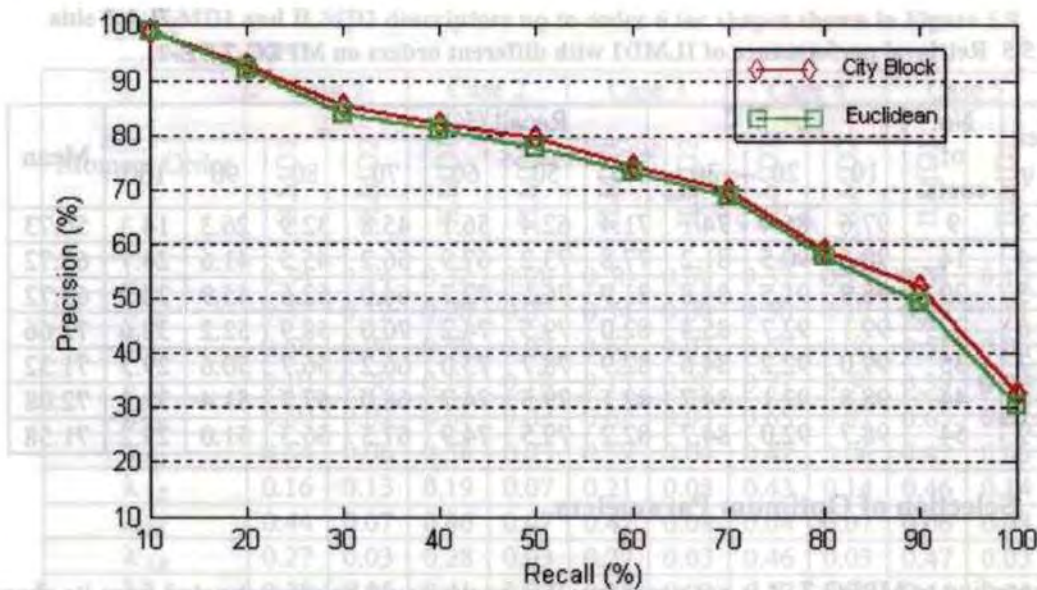


Figure 5.10 Retrieval accuracy of ILMD1 using Euclidean and City block distance measures on MPEG-7 CE-2 database

### 5.5.6 Modifications to ILMD

To characterize the descriptor as contour-based or region-based descriptor, the properties of Legendre polynomials is analyzed. Kuijlaars and Martinez-Finkelshtein reported that Legendre polynomials have asymptotic zero distribution [21]. The number of zeros of the orthogonal polynomials play a vital part in describing high-spatial-frequency components of the image. It can be noted from the plot of Legendre polynomials (Figure 5.4) that the edge region of  $[-1, 1]$  distributes more zeros than the central region of  $[-1, 1]$ . Thus, the following conjecture comes into being: if a target image covers two edge areas for Legendre polynomials in  $[-1, 1]$ , it would have better image representation quality than any other case. In the case of ILMD as Legendre basis function region is restricted to the 'bounding box' of the shape, object contour appear in the edge region of the Legendre polynomial. Thus, outer region of shape, i.e. contour, are better represented than the inner region. This is the requirement for a contour-based shape descriptor and hence ILMD1 should have better contour description capability. Thus, the newly proposed descriptor ILMD1 is a contour-based descriptor.

Bo Fu *et. al.* proposed a modification to the Legendre polynomial to distribute more zeroes in the central region for efficient representation of 'small images' [22]. The actual object region

in the 'small image' is small in comparison with the total image dimension. In original Legendre polynomial, since zeroes are located in the edge region actual object points are weakly covered. In the modified Legendre polynomial, since more zeroes are in the central region, object region in 'small images' are strongly covered. The modification to the Legendre polynomial is as follows. The domain of the Legendre polynomial is expanded to the set of all real numbers with a period of 2:

$$V_p(2m+x) = P_p(x), \quad (5.27)$$

where  $x \in [-1, 1]$ ,  $m \in Z$ ,  $Z$  means the set of all integers and  $P_p(x)$  is the Legendre polynomial.

After translating the expanded Legendre polynomials by a random real  $x_*$  (assume  $x_* > 0$ ), polynomials are defined in the range  $[-1, 1]$ . The modified Legendre polynomials with translation  $x_*$  can be expressed as follows:

$$V_p^{x_*}(x) = \begin{cases} P_p(x - x_*), & x \geq -1 + x_* \\ P_p(x + 2 - x_*), & x < -1 + x_* \end{cases} \quad (5.28)$$

where  $x \in [-1, 1]$ . This modified Legendre polynomials with random translation  $x_* \in [-1, 1]$  have the same orthogonality property as the original Legendre polynomials. Figure 5.11 shows

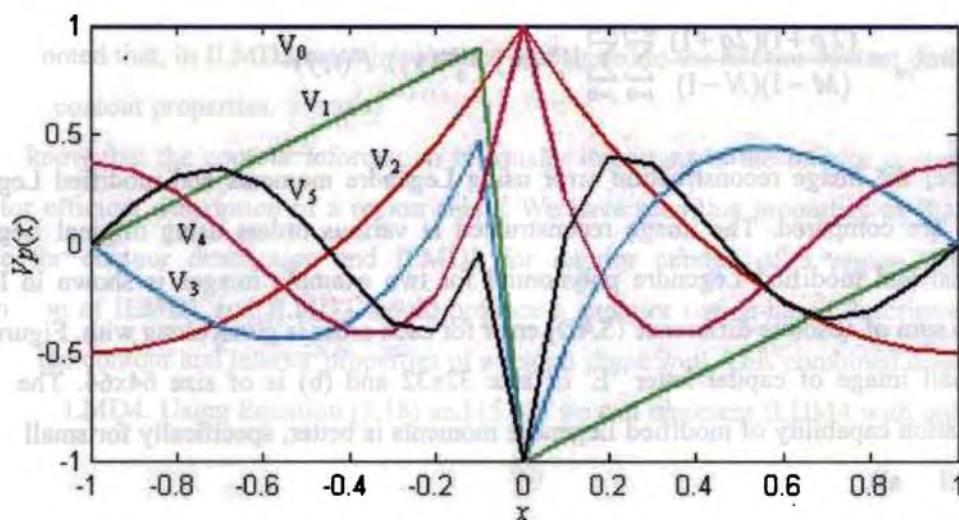


Figure 5.11 Plot of modified Legendre polynomial function values

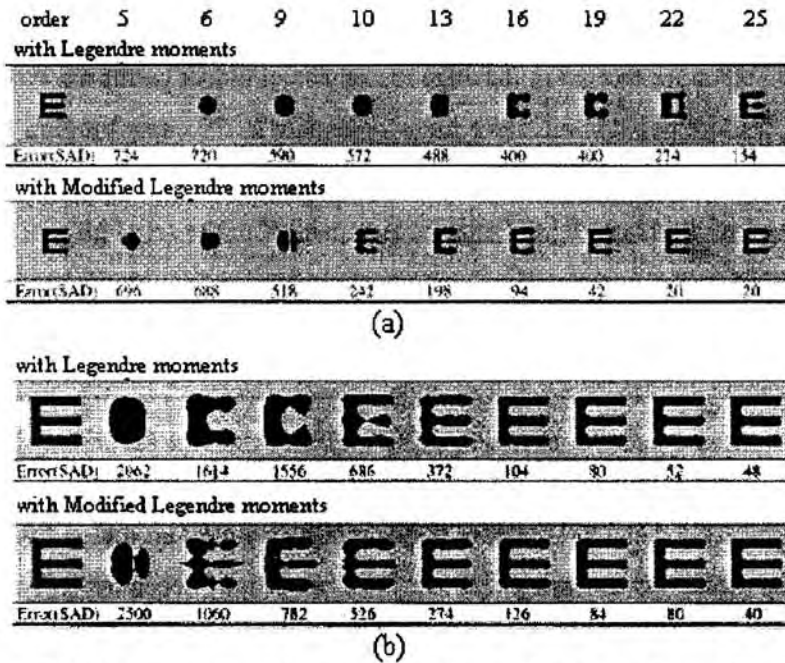


Figure 5.12 Reconstructed images at various orders using Legendre moments and modified Legendre moments for (a) a small image (b) a larger image [22]

the modified Legendre polynomials of 0-6<sup>th</sup> orders with a translation  $x_0 = 1$ . It can be noted that the two edge intervals of  $[-1, 1]$  with the maximal zero distribution for Legendre polynomials are combined into the central region of  $[-1, 1]$  for the new Legendre polynomials. The Legendre moments for an image  $F(i, j)$  of order  $p + q$  with the new polynomial can be defined as follows

$$m\lambda_{pq} = \frac{(2p+1)(2q+1)}{(M-1)(N-1)} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} V_p^x(x_i) V_q^x(y_j) F(i, j) \quad (5.29)$$

In [22] the image reconstruction error using Legendre moments and modified Legendre moments are compared. The image reconstructed at various orders using original Legendre polynomial and modified Legendre polynomial for two example images is shown in Figure 5.12. The sum of absolute difference (SAD) error for each order is given along with. Figure 5.12 (a) is small image of capital letter 'E' of size 32x32 and (b) is of size 64x64. The image representation capability of modified Legendre moments is better, specifically for small images (Figure 5.12 (a)).

### 5.5.6.1 As a Shape Descriptor

A region-based descriptor should be able to describe the interior content of the shape also. So, an efficient region-based descriptor should have more distribution of zeroes in the central region. The modified Legendre polynomials have this property and hence is a suitable candidate as a region-based shape descriptor. But as in the case of Legendre moments, modified Legendre moments do not have invariance to geometric transformations.

The modified Legendre polynomial has same properties such as orthogonality, symmetry with respect to origin etc. as in the case of original Legendre polynomials. So, the technique adopted in computation of ILMD1 could be applied here also to improve invariance to transformations. From Equation (5.17), a new equation for modified Legendre moments for invariance improvement can be written as

$$\lambda_{pq}^n = |m\lambda_{pq}^1| + |m\lambda_{pq}^2| + |m\lambda_{pq}^3| + |m\lambda_{pq}^4| \quad (5.30)$$

where  $m\lambda_{pq}^n$  is the modified Legendre moment for the quadrant  $n = 1, 2, \dots, 4$ . After scale normalization the new improved Legendre moment descriptor can be represented as

$$ILMD3 = \frac{\lambda_{01}^n}{\lambda_{00}^n}, \frac{\lambda_{10}^n}{\lambda_{00}^n}, \frac{\lambda_{20}^n}{\lambda_{00}^n}, \dots \quad (5.31)$$

It is to be noted that, in ILMD3 more importance is assigned to the interior content of the shape rather than contour properties.

We know that the contour information is equally important as the interior content of the shape for efficient description of a region shape. We have seen that properties of ILMD1 are suitable for contour description and ILMD3 for interior content of a region shape. So, combination of ILMD1 and ILMD3 would produce a superior region-based descriptor, which describes the contour and interior properties of a region shape well. This combined descriptor is denoted as ILMD4. Using Equation (5.18) and (5.31) we can represent ILDM4 with order  $p + q$  as follows:



**Table 5.6 Retrieval performance of ILMD4 with different orders on MPEG-7 CE-2**

Order $P + q$	No. of coefs.	Recall (%)										Mean
		10	20	30	40	50	60	70	80	90	100	
1 to 3	18	98.5	91.7	86.6	81.5	76.5	68.0	54.7	52.9	38.9	23.2	67.25
1 to 4	28	99.0	94.8	88.1	86.6	83.6	77.8	66.2	65.4	54.0	31.6	74.71
1 to 5	40	99.1	95.4	91.0	89.2	87.5	83.0	73.0	67.6	61.0	33.5	78.03
1 to 6	54	99.0	95.5	92.5	91.0	89.1	86.3	79.9	73.1	67.5	39.2	81.31
1 to 7	70	99.1	95.4	92.2	91.0	88.2	85.1	77.6	72.8	65.0	38.4	80.48
1 to 8	88	99.0	95.3	92.4	91.0	88.4	84.9	78.9	72.1	66.6	38.9	80.75
1 to 9	108	99.2	95.4	92.2	91.1	88.1	84.1	76.0	72.1	65.1	37.3	80.06

$$ILMD4 = \left[ \frac{\lambda'_{01}}{\lambda'_{00}}, \frac{\lambda'_{10}}{\lambda'_{00}}, \dots, \frac{\lambda'_{pq}}{\lambda'_{00}}, \frac{\lambda''_{01}}{\lambda''_{00}}, \frac{\lambda''_{10}}{\lambda''_{00}}, \dots, \frac{\lambda''_{pq}}{\lambda''_{00}} \right] \quad (5.32)$$

To determine the optimum number of coefficients required as a region-based shape descriptor experiments were conducted to test the retrieval effectiveness at various orders (order of ILMD1 and ILMD3) on MPEG-7 CE-2 database. Table 5.6 lists the precision for various recall rates using different orders of moments. ILMD4 with orders 1 to 6 is found to have optimum performance as we trade off performance against the number of coefficients. Thus, newly proposed region-based shape descriptor is ILMD4 with orders 1 to 6 having 54 coefficients.

## 5.6 Experimental Setup and Results

In order to evaluate the retrieval effectiveness of the proposed descriptors two sets of experiments are conducted. First the contour representation accuracy of the descriptors are evaluated and then the region representation accuracy. The performance measure used is recall precision pair (RPP). Also, a quantitative evaluation of the performance is reported using ANMRR outlined in Section 4.4.2.

### 5.6.1 Comparison of Contour Representation Accuracy

The contour representation accuracy of the ILMD1 is compared with ZMD, ARTD and CSSD. Although, ILMD1, ZMD and ARTD describe the entire region they can also be used to represent the contour alone. The CSSD is a strict contour descriptor. The MPEG-7 contour

shape database CE-1 was used for the experiments. Two pixel wide contours of the shapes in the database is extracted and used for all the descriptors to test their accuracy for contour shape.

### 5.6.1.1 Test Database

The CE-1 database consists of shapes acquired from real world objects. It takes into consideration the common shape distortions in nature and the inaccurate nature of shape boundaries from segmented shapes. It is designed to test contour shape descriptor's behavior under different shape distortions. The database consists of three parts, Set A, Set B and Set C. Set A has two subsets, Set A1 and Set A2.

Set A1 is for testing of scale invariance. It consists of 420 shapes of 70 classes. Each class has 6 similar member shapes. All the 420 shapes in the database are used as queries to test the retrieval accuracy.

Set A2 is for testing of rotation invariance. It consists of 420 shapes of 70 classes. Each class has 6 similar member shapes. All the 420 shapes in the database are used as queries to test the retrieval accuracy.

Set B is for testing of similarity-based retrieval or for testing shape descriptors' robustness to various arbitrary shape distortions including rotation, scaling, arbitrary skew and stretching, defection, indentation and other variations. It has 1400 shapes which have been classified to 70 classes. Each class in Set B has 20 similar member shapes. All the 1400 shapes in the database are used as queries to test the retrieval accuracy.

Set C is for testing shape descriptors' robustness to non-rigid object distortions. Set C consists of 200 affine transformed bream fishes and 1100 marine fishes which are unclassified. The 200 bream fishes are used as queries.

### 5.6.1.2 Retrieval Results and Analysis

In the plot of average precision for various recall rates for Set A1 shown in Figure 5.13, ILMD1 is observed to be better than other descriptors. Performance of ZMD and ARTD are comparable although ARTD is slightly better. CSSD has the least performance. In the First 12 retrieved results shown in Figure 5.14, ILMD1 and ARTD retrieves all the 6 butterflies while ZMD misses 2 and CSSD misses 4 cases. In each example results, the top left shape is the query

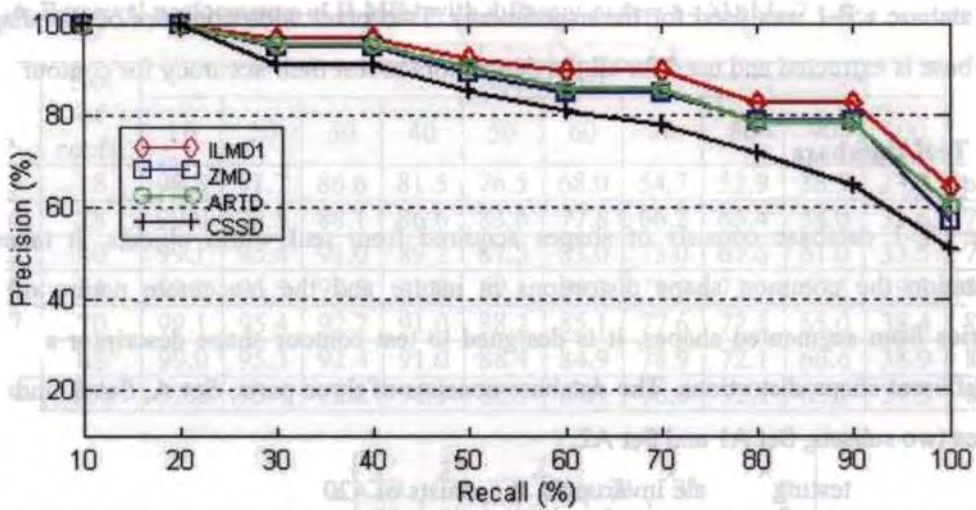


Figure 5.13 Average precision-recall of 420 retrievals using ILMD1, ZMD, ARTD and CSSD on Set A1 of MPEG-7 contour shape database CE-1

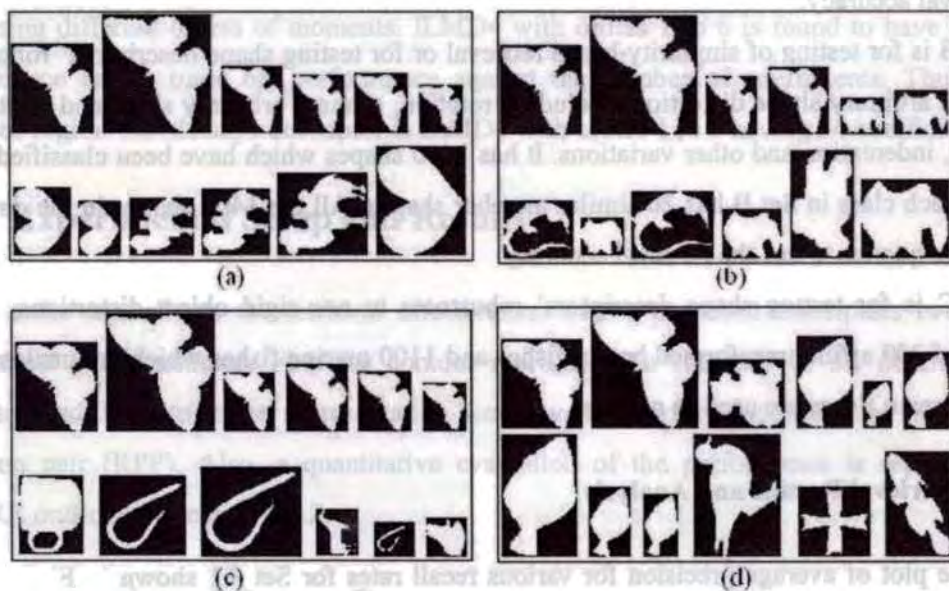


Figure 5.14 Retrieval results of query image 'butterfly-1' on CE-1 Set A1 of CE-1 using (a) ILMD1 (b) ZMD (c) ARTD (d) CSSD.

shape. The retrieved shapes are ranked in descending order of similarity to the query shape, and are arranged in left to right and top to bottom order.

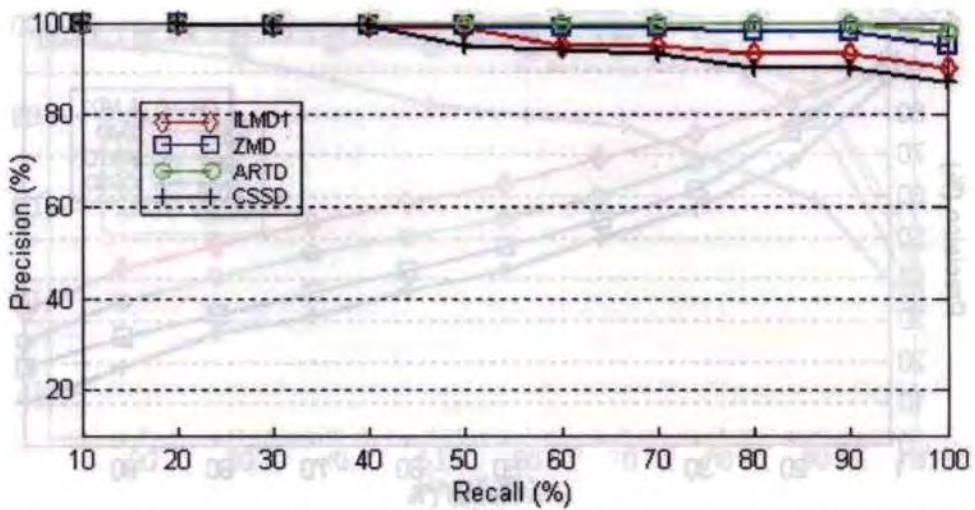


Figure 5.15 Average precision-recall of 420 retrievals using ILMD1, ZMD, ARTD and CSSD on Set A2 of MPEG-7 contour shape database CE-1

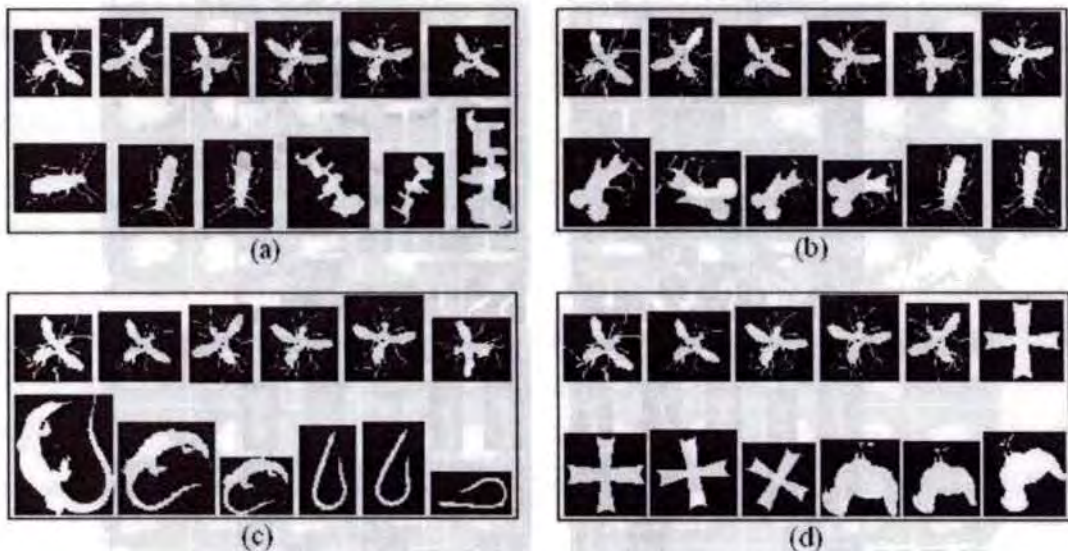


Figure 5.16 Retrieval results of query image 'fly-1' on CE-1 Set A2 of CE-1 using (a) ILMD1 (b) ZMD (c) ARTD (d) CSSD

In Figure 5.15, the RPP plot for Set A2, ARTD shows superior performance while performance of ZMD is comparable. The performance of ILMD1 is slightly inferior to ZMD and CSSD has the least performance. The superior performance of ARTD and ZMD is due to their inherent invariance to rotation. For ILMD1, some mapping error is introduced due to rotation for major axis alignment. Still, for maximum recall rate ILMD1 is only about 3% lower than ZMD. In the first 12 retrieved results shown in Figure 5.16, ILMD1, ZMD and ARTD

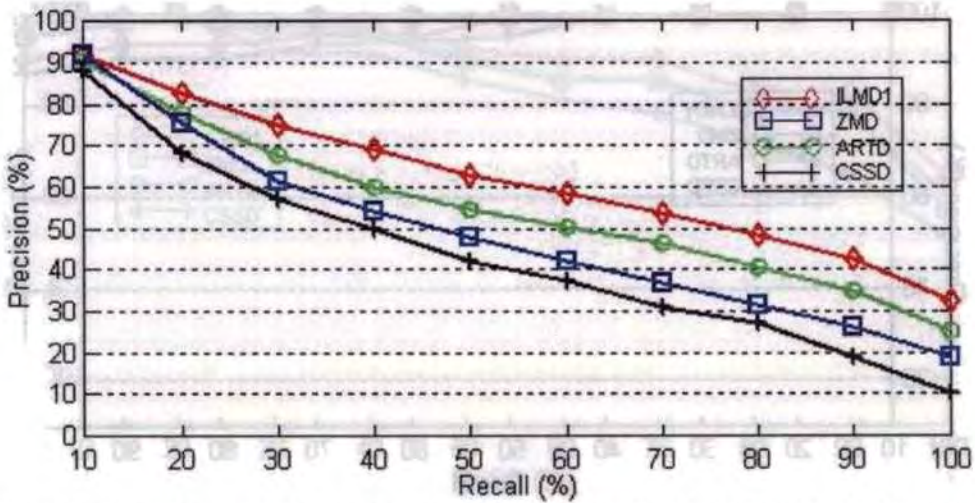


Figure 5.17 Average precision-recall of 1400 retrievals using ILMD1, ZMD, ARTD and CSSD on Set B of MPEG-7 contour shape database CE-1

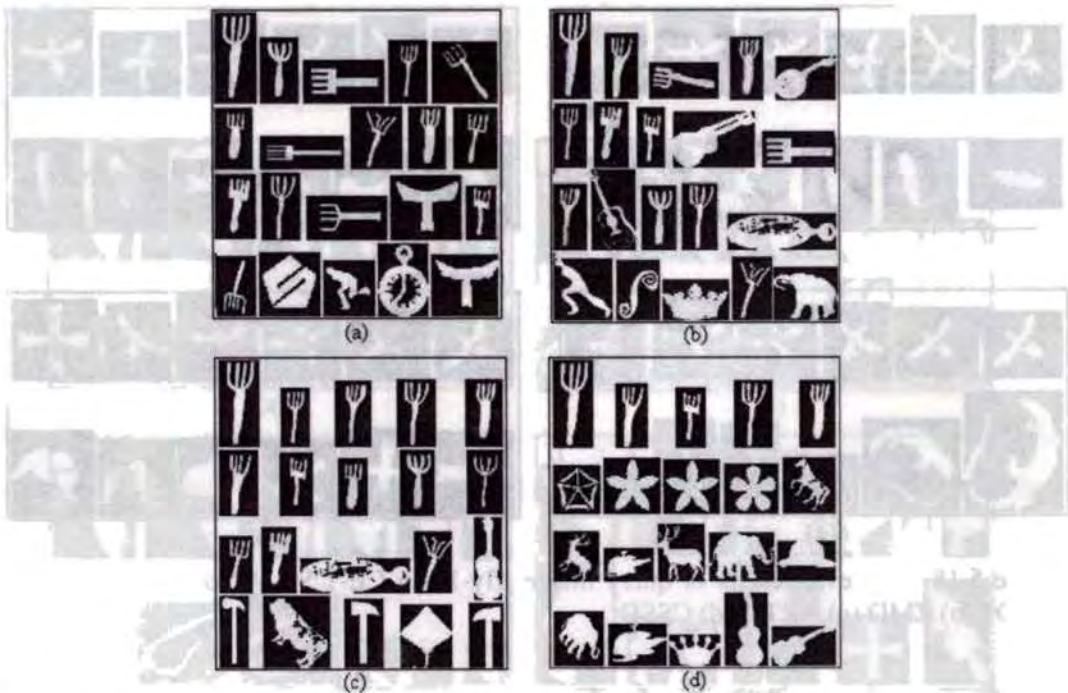


Figure 5.18 Retrieval results of query image 'fork-1' on CE-1 Set B using (a) ILMD1 (b) ZMD (c) ARTD (d) CSSD

retrieved all the 6 fly shapes matching to the query image while CSSD missed 1. It can be noted that among the first 12 samples, ILMD1 retrieved 3 beetle shapes which are closer to the query fly shape. Thus ILMD1 retrieves more perceptually similar shapes than other descriptors.

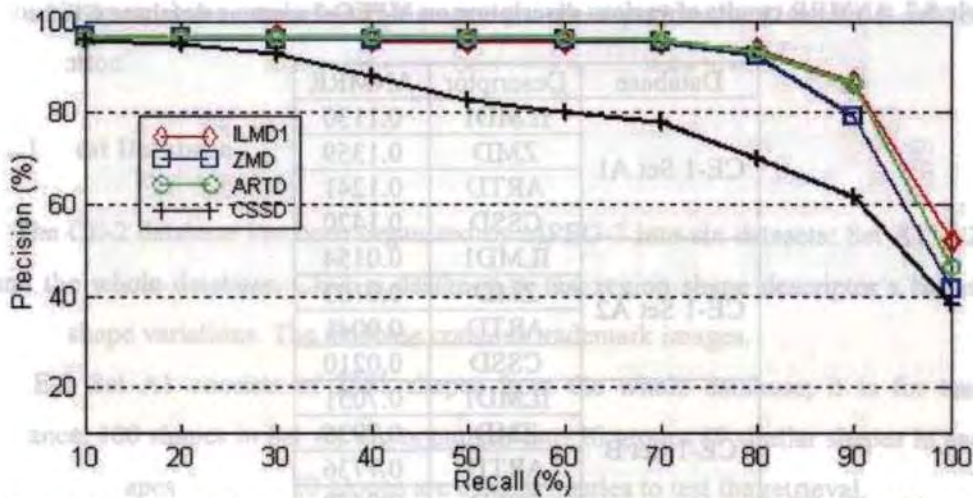


Figure 5.19 Average precision-recall of 200 retrievals using ILMD1 ZMD, ARTD and CSSD on Set C of MPEG-7 contour shape database CE-1

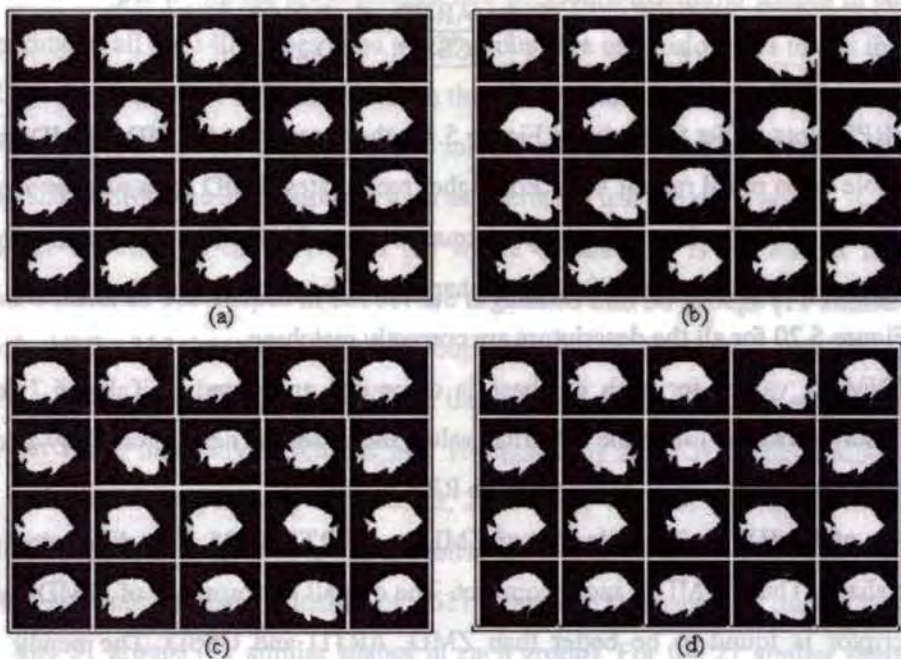


Figure 5.20 Retrieval results of query image 'bream-0' on CE-1 Set C using (a) ILMD1 (b) ZMD (c) ARTD (d) CSSD

Figure 5.17, precision - recall plot for Set B, indicates that ILMD1 has better precision than other descriptors. The retrieval accuracy of ARTD is better than ZMD and CSSD has the least performance of all. In the sample query result shown in Figure 5.18, ILMD1 retrieved 15 fork shapes among 20 in the top 20 retrievals while ARTD retrieved 13, ZMD 11 and CSSD 5.

**Table 5.7 ANMRR results of various descriptors on MPEG-7 contour database CE-1**

Database	Descriptor	ANMRR
CE-1 Set A1	ILMD1	0.1130
	ZMD	0.1359
	ARTD	0.1241
	CSSD	0.1420
CE-1 Set A2	ILMD1	0.0154
	ZMD	0.0103
	ARTD	0.0041
	CSSD	0.0210
CE-1 Set B	ILMD1	0.7051
	ZMD	0.7936
	ARTD	0.7736
	CSSD	0.8472
CE-1 Set C	ILMD1	0.2049
	ZMD	0.2182
	ARTD	0.2166
	CSSD	0.2324

In the RPP plots of Set C shown in Figure 5.19, the precision of ILMD1, ZMD and ARTD are comparable up to recall rate of 80%. For higher recall rates ILMD shows better performance than ARTD and ARTD shows better performance than ZMD. The CSSD has the lowest retrieval accuracy. As there are 200 similar shapes to the query shape, the first 20 retrievals shown in Figure 5.20 for all the descriptors are correctly matching.

The ANMRR values for each database is computed and listed in Table 5.7 for all the candidate descriptors. The lower the ANMRR value, the better the descriptor. They also indicate superior performance of ILMD1 supporting the RPP results.

The size of ILMD1 is 27 while that of ZMD and ARTD is 35. The dimension of CSSD varies with shape. Thus, ILMD is more compact. The overall performance of ILMD1 as contour shape descriptor is found to be better than ZMD, ARTD and CSSD. The newly proposed descriptor ILMD1 is found to be a better contour based descriptor.

### 5.6.2 Comparison of Region Representation Accuracy

The retrieval accuracy of the newly developed region descriptor ILMD4 is compared with standard region based descriptors ZMD and ARTD using a region shape database. The MPEG-7

proposed region shape database CE-2 is used for the experiments. Also, Set B of CE-1 is used for evaluation.

### 5.6.2.1 Test Database

The CE-2 database has been organized by MPEG-7 into six datasets: Set A1, A2, A3, A4, B and the whole database. CE-2 is designed to test region shape descriptor's behavior under different shape variations. The database contains trademark images.

CE-2 Set A1 consists of 2881 shapes from the whole database; it is for test of scale invariance. 100 shapes in Set A1 are organized into 20 groups (5 similar shapes in each group). All the 100 shapes from the 20 groups are used as queries to test the retrieval.

CE-2 Set A2 consists of 2921 shapes from the whole database; it is for test of rotation invariance. 140 shapes in Set A2 are organized into 20 groups (7 similar shapes in each group). In our experiment, all the 140 shapes from the 20 groups are used as queries to test the retrieval.

CE-2 Set A3 consists of 3101 shapes from the whole database; it is for test of rotation/scale invariance. 330 shapes in Set A3 are organized into 30 groups (11 similar shapes in each group). All the 330 shapes from the 30 groups are used as queries to test the retrieval.

CE-2 Set A4 consists of 3101 from the whole database; it is for test of robustness to perspective transform. 330 shapes in Set A4 are organized into 30 groups (11 similar shapes in each group). All the 330 shapes from the 30 groups are used as queries to test the retrieval.

CE-2 Set B consists of 2811 shapes from the whole database, it is for subjective test. 682 shapes in Set B are manually sorted out into 10 classes by MPEG-7. The number of similar shapes in each class is respectively 68, 248, 22, 28, 17, 22, 45, 145, 45, 42. All the 682 shapes from 10 classes are used as queries to test the retrieval.

The CE-2 whole database consists of 3621 shapes, 651 shapes of the 3621 shapes are organized into 31 groups (21 similar shapes in each group). For the 21 similar shapes in each group, there are 10 perspective transformed shapes, 5 rotated shapes and 5 scaled shapes. The 31 groups of shapes reflect overall shape operations, and they test the overall robustness of a shape descriptor. The whole database is 17-29% larger in size than the individual sets. All the 651 shapes are used as queries to test the retrieval accuracy.

To test the retrieval accuracy of the descriptors in the presence of occlusion a new database was constructed using shapes from CE-1 Set B. The database consists of 10 classes with 5



shapes in each class. Each member of the class has undergone mild to severe occlusion. All the 50 shapes were used as queries to test the retrieval accuracy.

Details of CE-1 Set B database was given in the previous section. Even though, this database is intended for testing contour description accuracy, the entire image region (not just contour) can be used to test the accuracy of region based descriptors also. All the 1400 shapes in the database are used as queries to test the retrieval accuracy.

### 5.6.2.2 Retrieval Results and Analysis

In the plot of average precision for various recall rates for CE-2 Set A1 shown in Figure 5.21, ILMD4 has the highest overall precision and accuracy of ARTD is better than that of ZMD. For low recall rates the performance of the three descriptors are comparable. Figure 5.22 shows first 12 retrievals of a sample query shape. Out of 5 similar shapes, 1 is missing for ILMD4 and 2 for ZMD and ARTD.

Precision – recall plot Figure 5.23 for CE-2 Set A2 shows overall high performance for all the descriptors. Performance of ZMD and ARTD are comparable and is slightly better than (by about 3% for 100% recall) ILMD4. This is due to inherent rotation invariance of ZMD and ARTD. In the first 12 retrieval results shown in Figure 5.24, all 7 members of the query class are retrieved by the three descriptors.

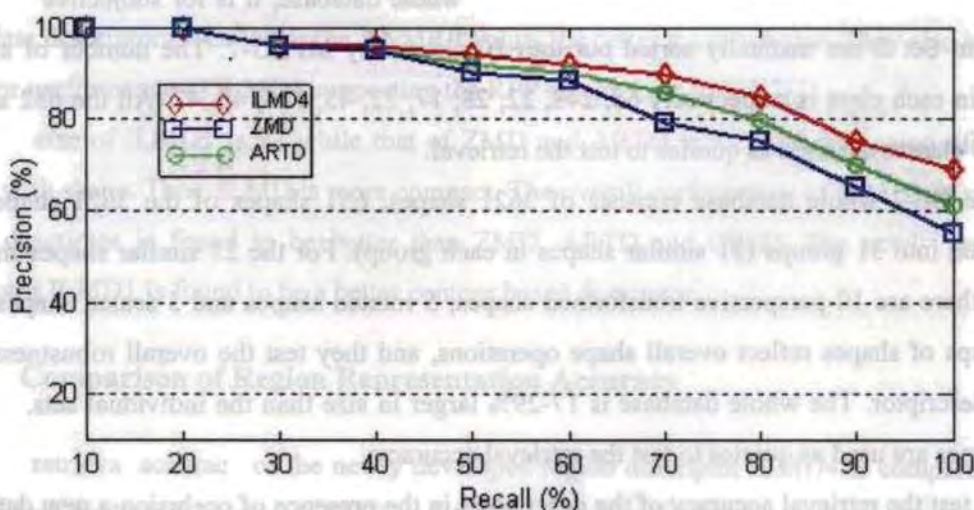


Figure 5.21 Average precision-recall of 100 retrievals using ILMD4, ZMD, and ARTD on Set A1 of MPEG-7 region shape database CE-2



Figure 5.22 Retrieval results of query image '513' on CE-2 Set A1 using (a) ILMD4 (b) ZMD (c) ARTD

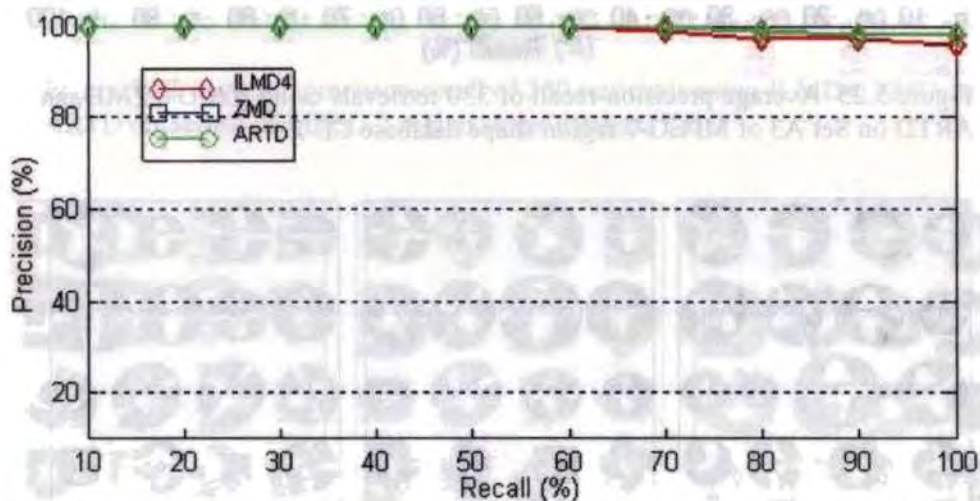


Figure 5.23 Average precision-recall of 140 retrievals using ILMD4, ZMD, and ARTD on Set A2 of MPEG-7 region shape database CE-2

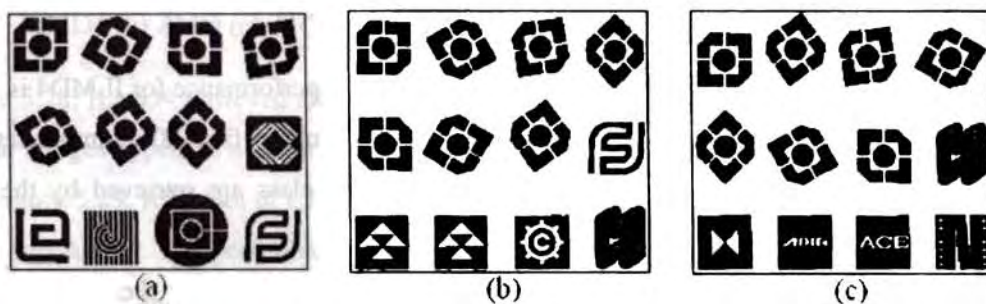


Figure 5.24 Retrieval results of query image '277' on CE-2 Set A2 using (a) ILMD4 (b) ZMD (c) ARTD

Figure 5.25, RPP plot of CE-2 Set A3, shows overall high performance for all the descriptors. Performance of ZMD and ARTD are comparable and is slightly better than (by

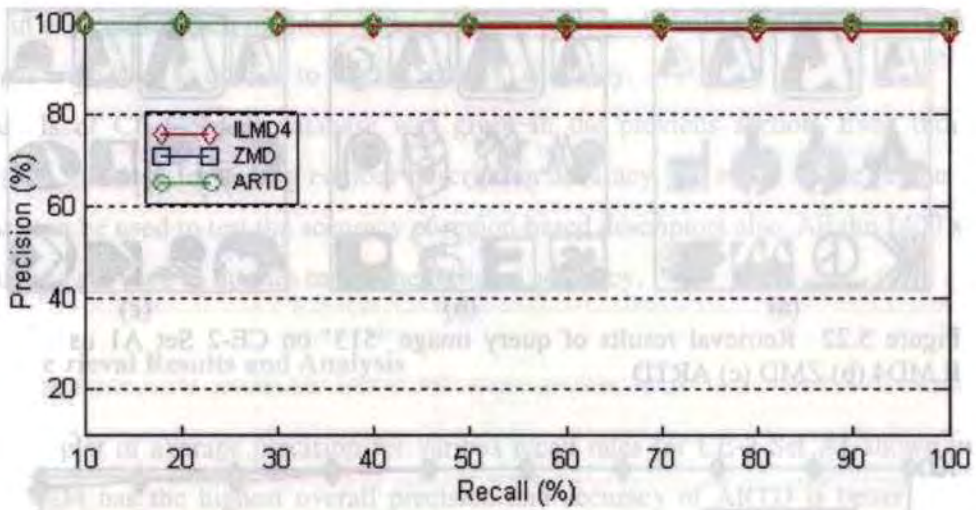


Figure 5.25 Average precision-recall of 330 retrievals using ILMD4, ZMD, and ARTD on Set A3 of MPEG-7 region shape database CE-2

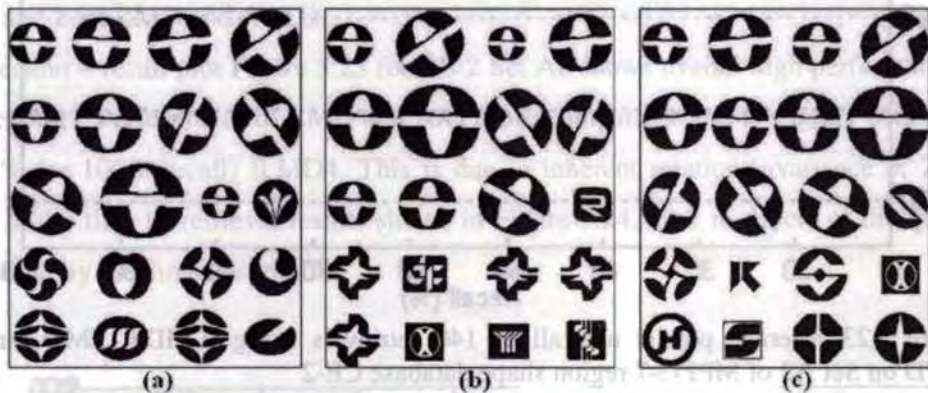


Figure 5.26 Retrieval results of query image '1484' on CE-2 Set A3 using (a) ILMD4 (b) ZMD (c) ARTD

about 2% for 100% recall) than ILMD4. Minor degradation in performance for ILMD4 is due to the re-sampling error introduced in major axis alignment. In the first 20 example retrieval results shown in Figure 5.26, all 11 members of the query class are retrieved by the three descriptors.

In precision – recall plot Figure 5.27 for CE-2 Set A4, ILMD4 has the highest overall precision and the accuracy of ARTD is better than that of ZMD. Set A4 is for perspective transformed images and for some highly skewed images in which the major axis may be different with respect to the original shape. This affects overall accuracy of ILMD4. This issue is addressed in the next section. Top 20 retrieval results for a query shape are shown in

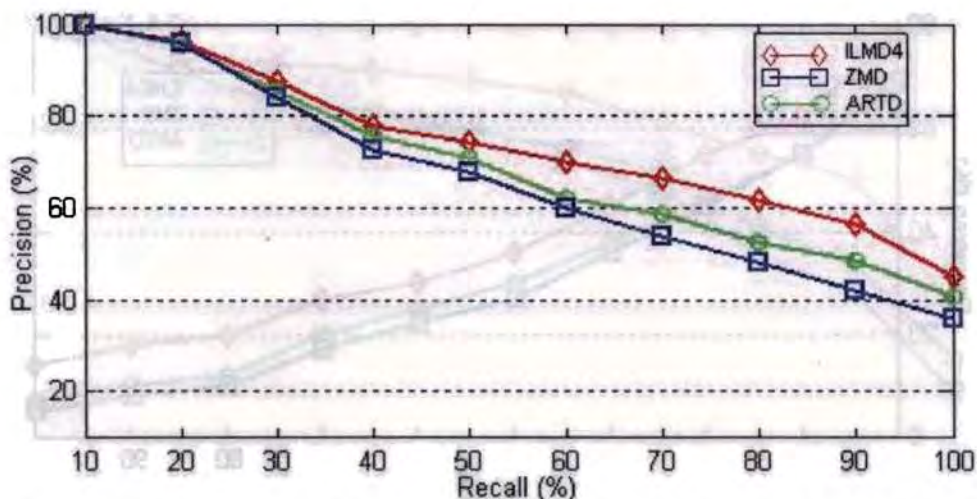


Figure 5.27 Average precision-recall of 330 retrievals using ILMD4, ZMD, and ARTD on Set A4 of MPEG-7 region shape database CE-2

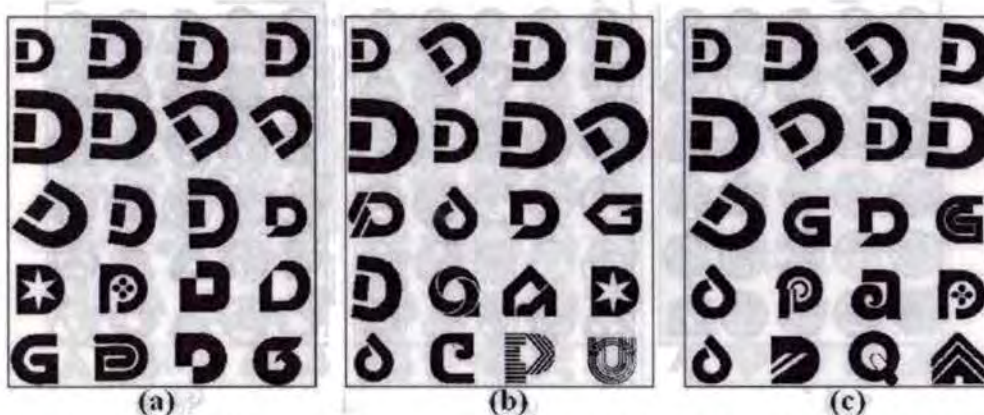


Figure 5.28 Retrieval results of query image '1470' on CE-2 Set A4 using (a) ILMD4 (b) ZMD (c) ARTD

Figure 5.28. ILMD4 retrieved all the 11, while ZMD and ARTD retrieved 9 similar shapes. The similar images appear at a closer distance for ILMD4 than for other descriptors.

Figure 5.29, RPP plot of CE-2 Set B, shows comparatively better accuracy for ILMD4 than ZMD and ARTD. The performance of ARTD is better than that of ZMD. The comparative low retrieval performance on Set B for all the descriptors is due to the fact that the grouping within the set is too rough. Top 30 example retrieval results are shown in Figure 5.30 for a query shape. It can be noted that for ILMD4 more matching shapes appear closer to the query image.

In Figure 5.31 for CE-2 whole database, ILMD4 has the highest overall precision and the accuracy of ARTD is better than that of ZMD. In Figure 5.32, top 30 retrieval results, all the

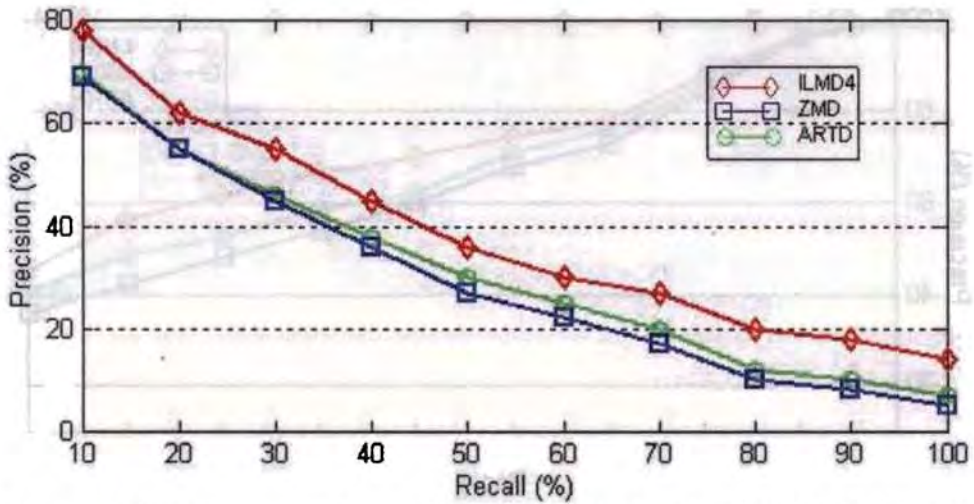


Figure 5.29 Average precision-recall of 682 retrievals using ILMD4, ZMD, and ARTD on Set B of MPEG-7 region shape database CE-2

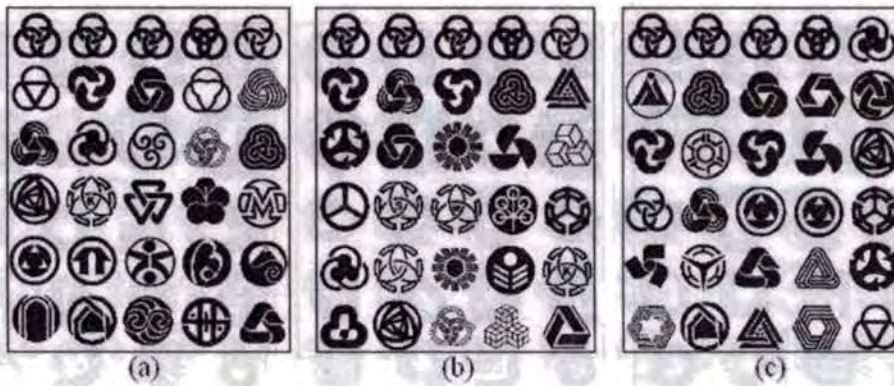


Figure 5.30 Retrieval results of query image '394' on CF-2 Set B using (a) ILMD4 (b) ZMD (c) ARTD

similar images are retrieved by ILMD4 and ARTD while some are missing for ZMD. The distance at which the matching image appears varies for the descriptors, and for ILMD4 they are closer. This database also contains some perspective transformed images and hence limitations to ILMD4 as specified for Set A4 exist, which will be addressed in the next section.

Some example classes of the occlusion database is shown in Figure 5.33. The RPP plot for occlusion database shown in Figure 5.34 indicates comparable performance of ILMD4 and ARTD although performance of ILMD4 is slightly better. The performance of ZMD is slightly inferior to ARTD. In the example top 12 retrieval results shown in Figure 5.35, it can be noted that similar image appears closer to query image for ILMD4 and ARTD than for ZMD.

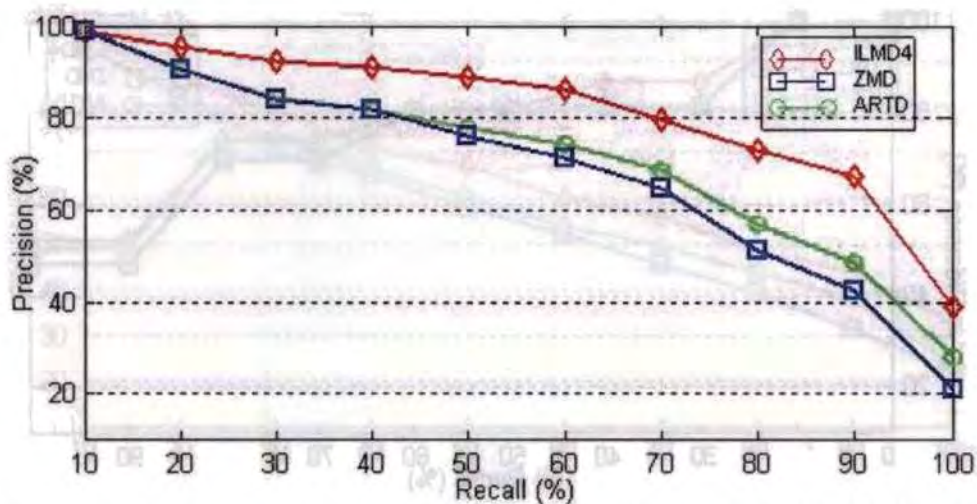


Figure 5.31 Average precision-recall of 651 retrievals using ILMD4, ZMD, and ARTD on MPEG-7 whole region shape database CE-2

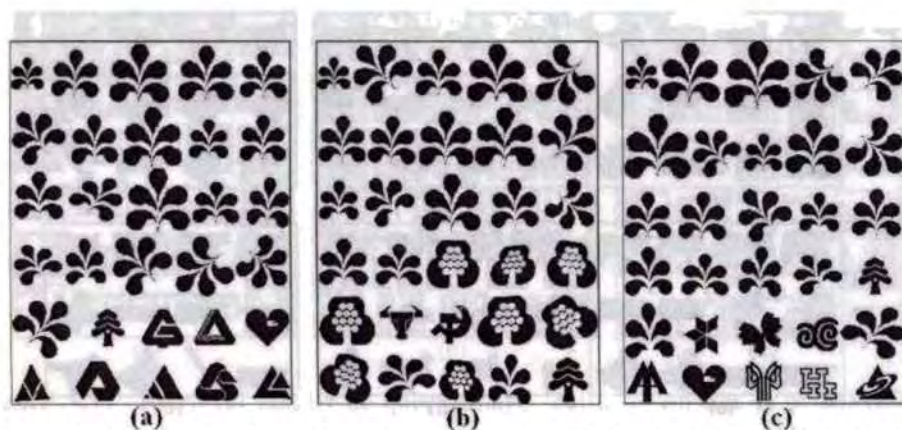


Figure 5.32 Retrieval results of query image '2024' on CE-2 whole database using (a) ILMD4 (b) ZMD (c) ARTD

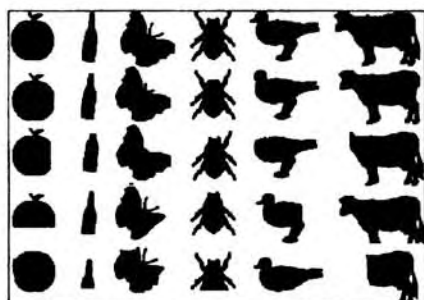


Figure 5.33 Some examples of occlusion database

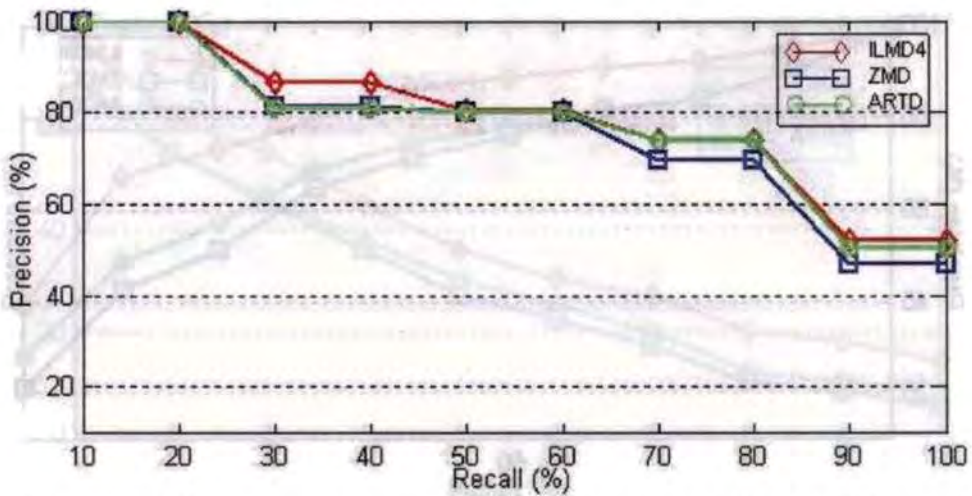


Figure 5.34 Average precision-recall of 50 retrievals using ILMD4, ZMD, and ARTD on occlusion database

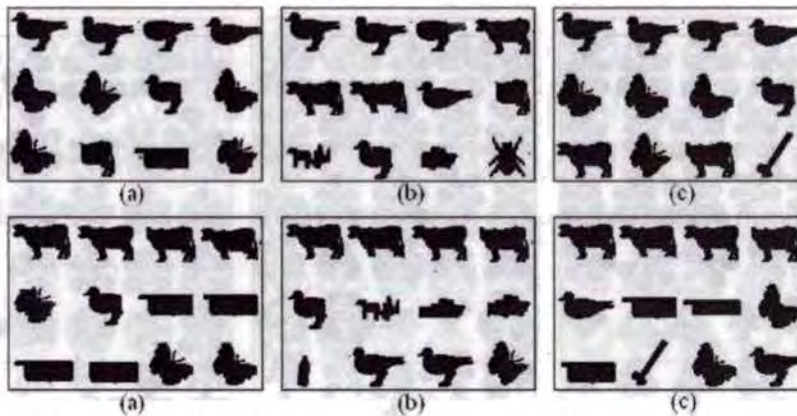


Figure 5.35 Retrieval results of two queries on occlusion database using (a) ILMD4 (b) ZMD (c) ARTD

ILMD4 shows better precision than other descriptors in RPP plot of CF-1 Set B using entire shape region, as plotted in Figure 5.36. Performance of ARTD is better than ZMD. The first 30 retrieval results for an example query shape is shown in Figure 5.37. ILMD4 retrieves all the 20 cattle shapes, while ZMD retrieved 8 and ARTD 13. ILMD4 retrieved shapes of horses, which are similar to cattle, as additional images in the first 30 retrieval results.

It can be noted that images retrieved in addition to those belonging to the query class are perceptually more similar to the query image for ILMD4 than for other descriptors. The ANMRR values of retrieval for these sets of database are computed and listed in Table 5.8. The ANMRR results also agree with the RPP results.

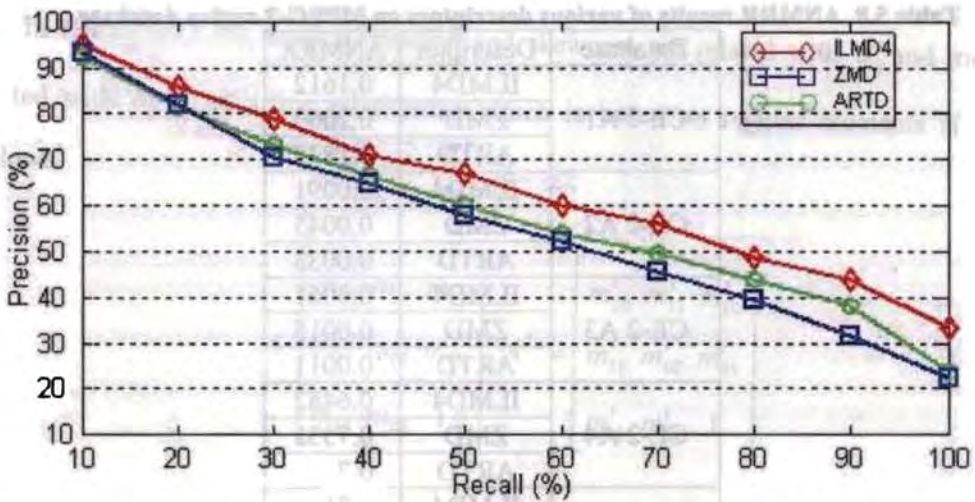


Figure 5.36 Average precision-recall of 1400 retrievals using ILMD4, ZMD, and ARTD on Set B of MPEG-7 shape database CE-1 as region shape

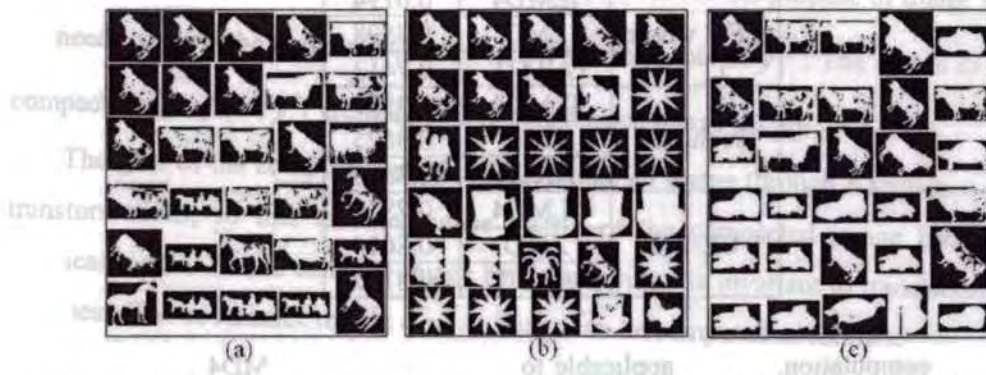


Figure 5.37 Retrieval results of query image 'cattle-1' on Set B of MPEG-7 shape database CE-1 as region shape using (a) ILMD4 (b) ZMD (c) ARTD

The dimension of ILMD4 is 54 and that of ZMD and ARTD are 35. But the retrieval performance of ILMD4 is significantly better than other descriptors. Hence, ILMD4 is a suitable region based descriptor.

## 5.7 Enhancement to ILMD for Affine Transformations

Affine or perspective transformed shapes are seen common in nature. The major axis of the shape may change due to high degree of skew. As the ILMD computation relies on the major axis of a shape, this will contribute to degradation in retrieval performance. To address this problem, a preprocessing step is proposed. The shape is normalized through compacting



Table 5.8 ANMRR results of various descriptors on MPEG-7 region database

Database	Descriptor	ANMRR
CE-2 A1	ILMD4	0.1612
	ZMD	0.2060
	ARTD	0.1833
CE-2 A2	ILMD4	0.0091
	ZMD	0.0045
	ARTD	0.0033
CE-2 A3	ILMD4	0.0061
	ZMD	0.0015
	ARTD	0.0011
CE-2 A4	ILMD4	0.6482
	ZMD	0.7531
	ARTD	0.7136
CE-2 B	ILMD4	0.8134
	ZMD	0.8691
	ARTD	0.8508
CE-2 whole	ILMD4	0.6194
	ZMD	0.7598
	ARTD	0.7413
Occlusion	ILMD4	0.3585
	ZMD	0.3625
	ARTD	0.3620
CE-1 B	ILMD4	0.7844
	ZMD	0.8614
	ARTD	0.8160

prior to ILMD computation. This is applicable to both ILMD1 and ILMD4. Leu proposed an algorithm for shape normalization in [23]. The shape normalization process is summarized here.

Let  $f(u, v)$  be the affine transformed image and  $f(x, y)$  the standard normalized image. The affine relationship between  $f(u, v)$  and  $f(x, y)$  are as follows.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = A \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, A = \begin{bmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ 0 & 0 & 1 \end{bmatrix}, \quad (5.33)$$

where  $\begin{bmatrix} u \\ v \end{bmatrix}^T$  is the affine transformed position corresponding to point  $\begin{bmatrix} x \\ y \end{bmatrix}^T$ , and  $A$  is the affine coefficient matrix.

In Equation (5.33), scaling, rotation and skewness are related with  $a_{ij}$  and translation is related with  $b_{ij}$ . The affine relationship which is based on regular moments is defined as follows.

$$A \begin{bmatrix} m_{20} & m_{11} & m_{10} \\ m_{11} & m_{02} & m_{01} \\ m_{10} & m_{01} & 1 \end{bmatrix} A^T = \begin{bmatrix} m'_{20} & m'_{11} & m'_{10} \\ m'_{11} & m'_{02} & m'_{01} \\ m'_{10} & m'_{01} & 1 \end{bmatrix} \quad (5.34)$$

where  $m_{pq}$  is the  $pq$ -order regular moment of the compact image  $f(x, y)$  and  $m'_{pq}$  the  $pq$ -order regular moment of the affine transformed image  $f(u, v)$ . For the purpose of image normalization we need to know the mapping from point  $[u \ v]^T$  to point  $[x \ y]^T$ . The aim is to construct the compact image  $f(x', y')$  from the affine transformed image  $f(u, v)$ .

The goal of the compact algorithm is to adjust an image through a sequence of two linear transformations, so that the covariance matrix of the compacted image becomes a scaled identical matrix. So the compact image by the algorithm is invariant to translation, scaling and skewness. The covariance matrix  $C$  of an image can be computed as follows.

$$C = \begin{bmatrix} m'_{20} - m'_{10}{}^2 & m'_{11} - m'_{10}m'_{01} \\ m'_{11} - m'_{10}m'_{01} & m'_{20} - m'_{01}{}^2 \end{bmatrix} \quad (5.35)$$

Let  $E$  be a matrix whose rows are formed from the eigenvectors of  $C$ , ordered so that the last row is the eigenvector corresponding to the smallest eigenvalue. Eigenvectors and eigenvalues of  $E$  are needed to adjust the covariance matrix to a scaled identical matrix. The matrix  $C$  can be uncorrelated by multiplying with the matrix  $E$ .

$$C' = ECE^T = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \quad (5.36)$$

where  $\lambda_1$  and  $\lambda_2$  are the two eigenvalues of the matrix  $C$  and

$$E = \begin{bmatrix} e_{1x} & e_{1y} \\ e_{2x} & e_{2y} \end{bmatrix} \quad (5.37)$$

where  $e_{ix}$  and  $e_{iy}$  are corresponding eigenvectors of eigenvalues  $\lambda_i$ .

Multiplying by scaling matrix  $S$ , the covariance matrix of  $C$  becomes scaled identity matrix.

$$SC'S^T = c^2 I \quad (5.38)$$

where  $c$  is a scaling constant and

$$S = \begin{bmatrix} \frac{c}{\sqrt{\lambda_1}} & 0 \\ 0 & \frac{c}{\sqrt{\lambda_2}} \end{bmatrix} \quad (5.39)$$

Combining above two equations,

$$SECE^T S^T = c^2 I \quad (5.40)$$

Now we can get the compact image from the affine transformed image by the Equations (5.34) and (5.40).

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = SE \begin{bmatrix} u - m'_{10} \\ v - m'_{01} \end{bmatrix} \quad (5.41)$$

Two examples for image normalization through compacting is given in Figure 5.38. It can be noted that affine transformed shape is transformed to normal shape after compacting. The enhanced descriptor ILMD1 is denoted as EILMD1 and ILMD4 as EILMD4.

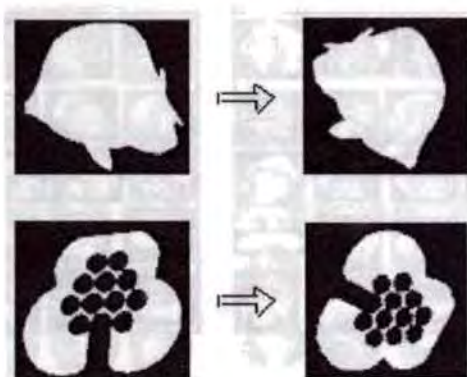


Figure 5.38 Two examples of image normalization through compacting

### 5.7.1 Retrieval Experiment and Results

The improvement in retrieval accuracy due to the affine transformation invariance of the descriptors EILMD1 and EILMD4 is experimentally evaluated. Database sets containing affine transformed shapes are chosen as the test set. EILMD1 is evaluated using CE-1 Set C and EILMD4 using CE-2 Set A4 and CE-2 whole database.

Figure 5.39 demonstrates the average precision-recall of 200 retrievals using ILMD1 and EILMD1 on Set C of MPEG-7 contour shape database CE-1. The performance of ILMD1 is better than ZMD, ARTD and CSSD as detailed earlier. It can be noted that compaction improves the performance of the ILMD1. First 30 retrievals of an example query result for

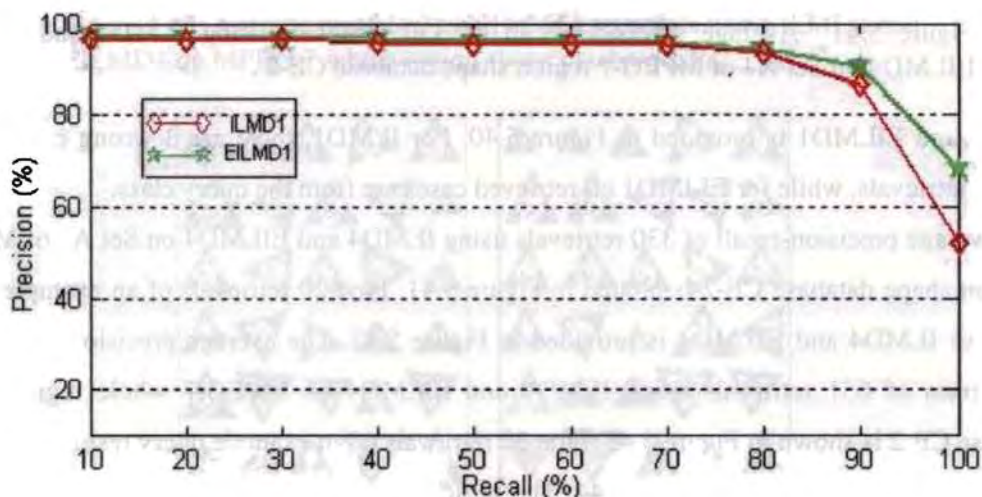


Figure 5.39 Average precision-recall of 200 retrievals using ILMD1 and EILMD1 on Set C of MPEG-7 contour shape database CE-1

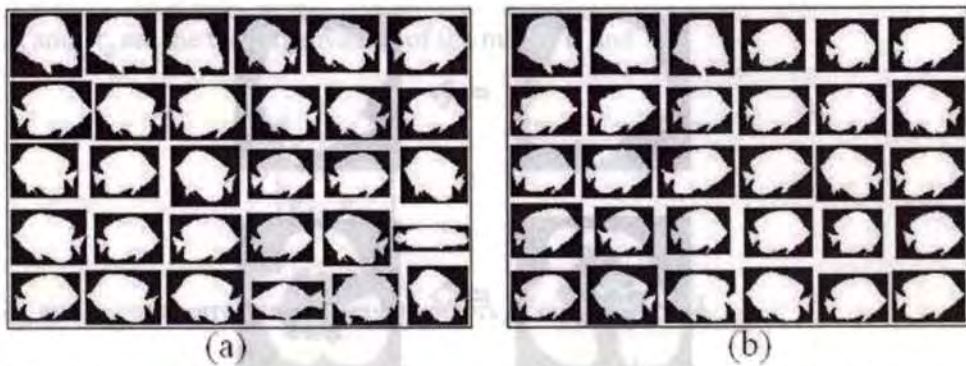


Figure 5.40 Retrieval results of query image 'bream-124' on CE-1 Set C using (a) ILMD1 (b) EILMD1.

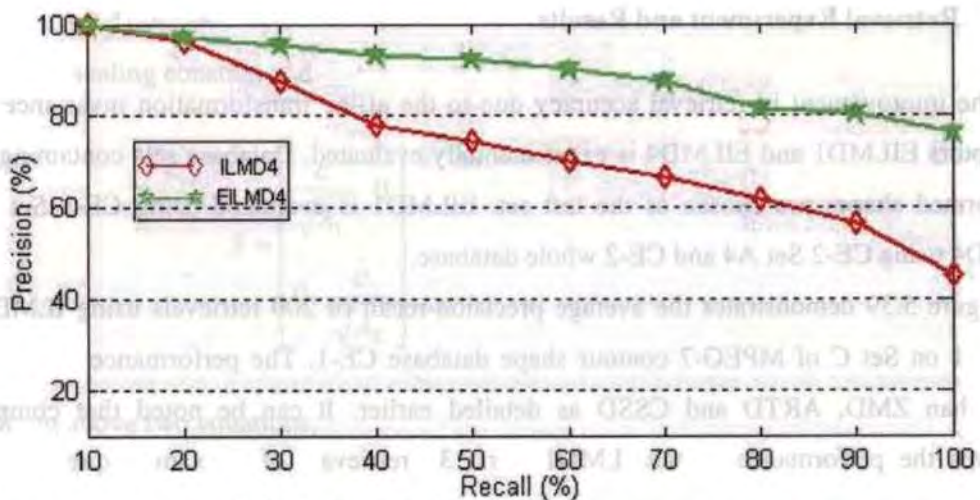


Figure 5.41 Average precision-recall of 330 retrievals using ILMD4 and EILMD4 on Set A4 of MPEG-7 region shape database CE-2 .

ILMD1 and EILMD1 is provided in Figure 5.40. For ILMD1, there are 3 wrong cases in the first 30 retrievals, while for EILMD1 all retrieved cases are from the query class.

Average precision-recall of 330 retrievals using ILMD4 and EILMD4 on Set A4 of MPEG-7 region shape database CE-2 is plotted in Figure 5.41. First 20 retrievals of an example query result for ILMD4 and EILMD4 is provided in Figure 5.42. The average precision for various recall rates of 651 retrievals using ILMD4 and EILMD4 on MPEG-7 whole region shape database CE-2 is shown in Figure 5.43. First 30 retrievals of an example query result for ILMD4 and EILMD4 is provided in Figure 5.44. The retrieval results show that for EILMD4 member images of the query class are retrieved correctly at a shorter distance than for ILMD4. It can be noted that the overall precision of EILMD4 is better than ILMD4 in both cases. The

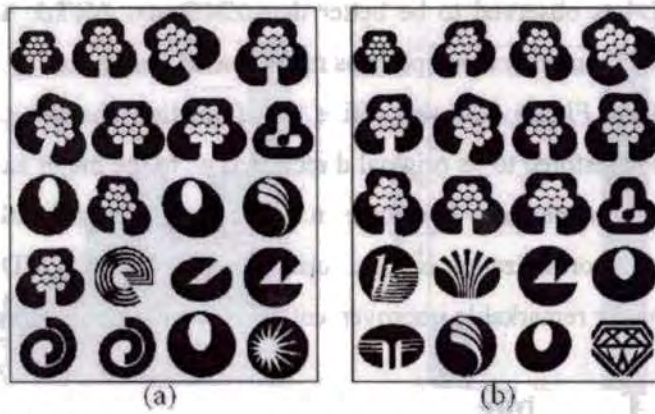


Figure 5.42 Retrieval results of query image '1970' on CE-2 Set A4 using (a) ILM D4 (b) EILMD4

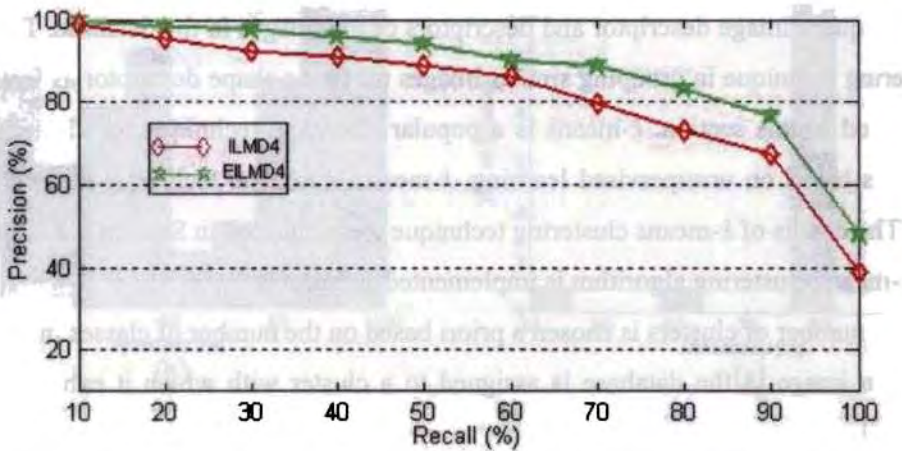


Figure 5.43 Average precision-recall of 651 retrievals using ILM D4 and EILMD4 on MPEG-7 whole region shape database CE-2

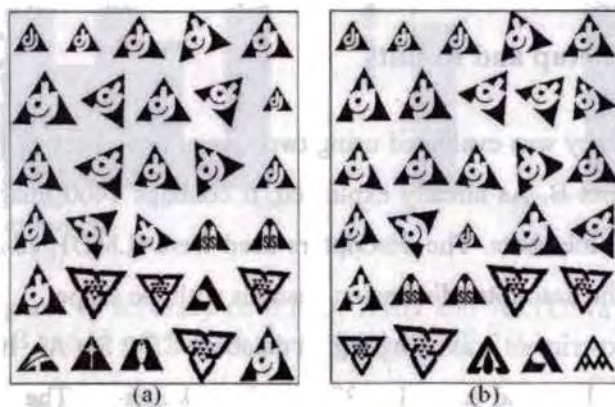


Figure 5.44 Retrieval results of query image '2842' on CE-2 whole database using (a) ILM D4 (b) EILMD4

performance of ILMD4 is observed to be better than ZMD and ARTD. Thus, EILMD4 has better retrieval accuracy than other descriptors as region based descriptor.

It can be noted from Figure 5.38 that affine transformation changes the major axis of a shape. The major axis is restored to its original direction (i.e., to the major axis of the non affine transformed shape), by compacting. A change in major axis causes the shape to be aligned differently while rotation normalization is done during computation of ILMD. Correction of this deficiency is the reason for remarkable improvement in retrieval accuracy of enhanced ILMD.

## 5.8 Clustering Technique for CBIR

The similarity measure is usually carried out by computing the distance, for e.g. city block, between the query image descriptor and descriptors of the images in the database. The accuracy of a clustering technique in grouping similar images using the shape descriptor as feature vector is investigated in this section.  $k$ -means is a popular statistical technique for clustering pattern vectors. It is based on unsupervised learning.  $k$ -means clustering technique is used for image retrieval. The details of  $k$ -means clustering technique were outlined in Section 4.2.5.

The  $k$ -means clustering algorithm is implemented as follows.

1. The number of clusters is chosen a priori based on the number of classes in the database
2. Each image in the database is assigned to a cluster with which it exhibits maximum similarity.
3. New cluster centers are computed as the centroids of the clusters.
4. Steps 2 and 3 are repeated until there are no changes in the cluster centers.

### 5.8.1 Experimental Setup and Results

The clustering accuracy was evaluated using two sets of experiments. First set is using the contour database CE-1 Set B. As already explained, it contains 1400 images with 70 classes, each class containing 20 members. The descriptors used were ILMD1, ZMD and ARTD. The CSSD was not used, as the descriptor dimension changes with the shape.

The second set of experiment was using region database CE-2 Set A1. It contains 20 classes with 5 shapes in each class and a total of 2881 trademark images. The candidate descriptors

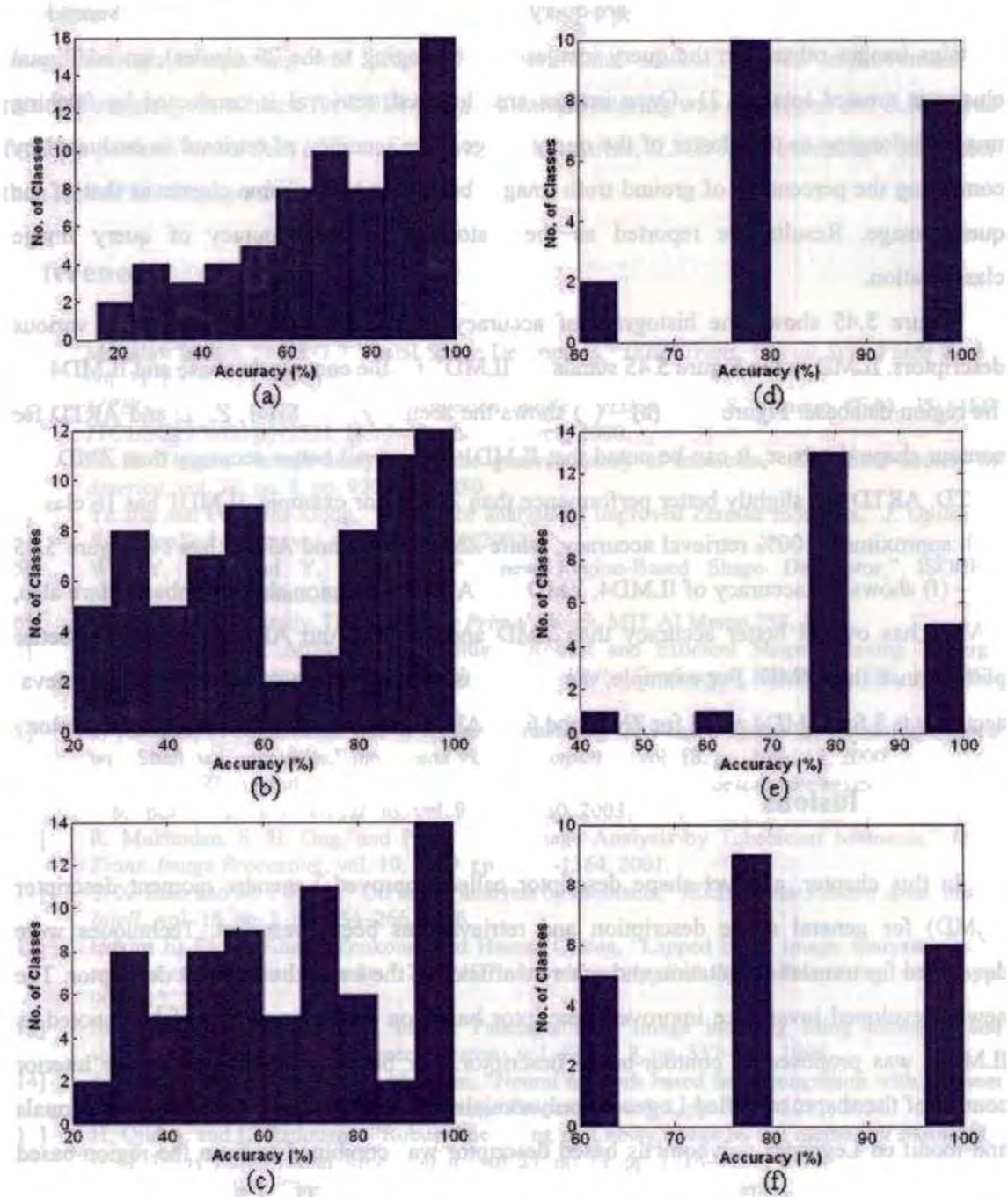


Figure 5.45 Histogram of accuracy of ILMD, ZMD and ARTD. (a – c) contour shape database CE-1 Set B and (d – f) Region shape database CE-2 A1

were ILMD, ZMD and ARTD. All 1400 images in the first set and 100 images in the second set are used as query images.



In the first set as all the images are query images, 70 clusters are created. As the second set contains images other than the query images (not belonging to the 20 classes), an additional cluster is created totaling 21. Once images are clustered, retrieval is conducted by fetching images belonging to the cluster of the query image. The accuracy of retrieval is evaluated by computing the percentage of ground truth images belonging to the same cluster as that of the query image. Results are reported as the histogram of the accuracy of query image classification.

Figure 5.45 shows the histogram of accuracy of various query classes using various descriptors. ILMD in the Figure 5.45 stands for ILMD1 for the contour database and ILMD4 for the region database. Figure 5.45 (a) – (c) shows the accuracy of ILMD1, ZMD and ARTD for contour shape database. It can be noted that ILMD1 has overall better accuracy than ZMD and ARTD. ARTD has slightly better performance than ZMD. For example, ILMD1 has 16 classes with approximate 100% retrieval accuracy, while ZMD has 12 and ARTD has 14. Figure 5.45 (d) – (f) shows the accuracy of ILMD4, ZMD and ARTD for region shape database. Here also, ILMD4 has overall better accuracy than ZMD and ARTD. And ARTD has slightly better performance than ZMD. For example, the number of classes with approximate 100% retrieval accuracy is 8 for ILMD4 and 5 for ZMD and 6 for ARTD. Overall, ILMD is a better descriptor.

## 5.9 Conclusions

In this chapter, a novel shape descriptor called improved Legendre moment descriptor (ILMD) for general shape description and retrieval has been presented. Techniques were developed for translation, rotation and scale invariance of the Legendre moment descriptor. The newly developed invariance improved descriptor based on Legendre polynomials, denoted as ILMD1, was proposed as contour-based descriptor. For better representation of the interior content of the shape, modified Legendre polynomials was used. Original Legendre polynomials and modified Legendre polynomials based descriptor was combined to form the region-based descriptor, denoted as ILMD4. MPEG-7 shape descriptors ARTD, CSSD and ZMD were implemented and compared with the performance of the proposed descriptor. The retrieval effectiveness of ILMD1 was evaluated using MPEG-7 contour shape database CE-1 and found to be better than ZMD, ARTD and CSSD. The retrieval effectiveness of ILMD4 was evaluated using MPEG-7 region shape database CE-2 and found to be better than ZMD and ARTD. For

providing ILMD invariance to affine transformations, a preprocessing step to normalize the image through compacting was proposed. Experimental studies showed improvement in retrieval accuracy. A scheme for CBIR using  $k$ -means clustering was developed and ILMD was found to perform better than other descriptors. To summarize, ILMD is a better shape descriptor than the MPEG-7 shape descriptors.

## References:

- [1] Miroslaw Bober, "MPEG-7 Visual Shape Descriptors," *IEEE Trans. Circuit Syst. Video Tech.*, vol. 11, no. 6, pp. 716-719, 2001.
- [2] *MPEG-7 Visual part of experimentation model version 5.0*, S. Jeannin (Ed.), ISO/IEC JTC1/SC29/WG11/N3321, Nordwijkerhout, March, 2000.
- [3] M. R. Teague, "Image analysis via the general theory of moments," *J. Optical Society of America*, vol. 70, no. 8, pp. 920-930, 1980.
- [4] Ye Bin and Peng Jia-Xiong, "Invariance analysis of improved Zernike moments," *J. Optics A: Pure Applied Optics*, vol. 4, pp. 606-614, 2002.
- [5] W. -Y. Kim and Y. -S. Kim, "A new Region-Based Shape Descriptor," ISO/IEC MPEG99/M5472, Maui, Hawaii, 1999.
- [6] H. Asada and M. Brady, *The Curvature Primal Sketch*, MIT AI Memo 758, 1984.
- [7] F. Mokhtarian, S. Abbasi and J. Kittler. "Robust and Efficient Shape Indexing Through Curvature Scale Space," in *Proc. British Machine Vision Conference*, Edinburgh, UK, pp.53-62, 1996.
- [8] S. Abbasi, F. Mokhtarian and J. Kittler. "Enhancing CSS-based Shape Retrieval for Objects with Shallow Concavities," *Image and Vision Computing*, vol. 18, pp. 199-211, 2000.
- [9] Dengsheng Zhang and Guojun Lu, "Evaluation of MPEG-7 shape descriptors against other shape descriptors," *Multimedia Systems*, vol. 9, pp. 15-30, 2003.
- [10] R. Mukundan, S. H. Ong, and P. A. Lee, "Image Analysis by Tchebichef Moments," *IEEE Trans. Image Processing*, vol. 10, no. 9, pp. 1357-1364, 2001.
- [11] S. X. Liao and M. Pawlak, "On image analysis by moments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 3, pp. 254-266, 1996.
- [12] Hakim El Fadili, Khalid Zenkour and Hassan Qjidaa, "Lapped block image analysis via the method of Legendre moments," *EURASIP Journal on Applied Signal Processing*, vol. 9, pp. 902-913, 2003.
- [13] M.K. Mandal, T. Aboulnasr and S. Panchanathan, "Image indexing using moments and wavelets," *IEEE Trans. Consumer Electron.*, vol. 42, no. 3, pp. 557-565, 1996.
- [14] J. Haddadnia, K. Faez, and P. Moallem, "Neural network based face recognition with moment invariants," in *Proc. of Int'l Conf. on Image Processing*, vol. 1, pp. 1018-1021, 2001.
- [15] H. Qjidaa, and L. Radouane, "Robust line fitting in a noisy image by the method of moments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 11, pp. 1216-1223, 1999.
- [16] C.H. Teh, and R.T. Chin, "On image analysis by the method of moments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 4, pp. 496-513, 1988.
- [17] Simon X. Liao and Miroslaw Pawlak, "Chinese character recognition with Legendre moments", in *Proc. of the third international conference on Young computer scientists*, Beijing, China, pp. 881 - 883, 1993.
- [18] Simon X. Liao and Miroslaw Pawlak, "Chinese character recognition via orthogonal moments", in *Information Theory and Applications II (Lecture Notes in Computer Science)*, vol. 1133, Springer, Berlin, Heidelberg, 1996.

- [19] R Mukundan and K. R. Ramakrishnan, *Moment Functions in Image Analysis — Theory and Applications*, World Scientific Publication Co., Singapore, pp. 49-55,1998.
- [20] Chee-Way Chong, P. Raveendran and R. Mukundan, "Translation and scale invariants of Legendre moments," *Pattern Recognition*, vol. 37, pp. 119–129, 2004.
- [21] A. B. J. Kuijlaars and A. Martinez-Finkelshtein, "Strong asymptotics for Jacobi polynomials with varying nonstandard parameters," *J. Anal. Math.*, vol. 94, pp. 195–234, 2004.
- [22] Bo Fu, Jianzhong Zhou, Yuhong Li, Guojun Zhang and ChengWang, "Image analysis by modified Legendre moments," *Pattern Recognition*, vol. 40, pp. 691–704, 2007.
- [23] Leu J. -G, "Shape Normalization Through Compacting," *Pattern Recognition Letters*, vol. 10, pp. 243-25, 1989.

---

## Chapter 6

# Development of a Relevance Feedback Technique for CBIR

---

### 6.1 Introduction

A novel shape descriptor called improved Legendre moment descriptor (ILMD) has been developed in the previous chapter. The retrieval accuracy of ILMD was compared with MPEG-7 shape descriptors such as ARTD, ZMD and CSSD using MPEG-7 contour and region shape database. It was found that ILMD has superior performance as a shape descriptor. Techniques for further improvement of the retrieval accuracy using relevance feedback are discussed in this chapter. In the relevance feedback scheme, the user selects a set of positive and/or negative examples to a query image and the system refines the query based on these user feedbacks and retrieves a new set of images. Relevance feedback technique was reviewed in Section 2.4.

Classifying images automatically and accurately is important for both effective image organization and retrieval. Achieving high image classification accuracy is quite challenging because semantically related images may not be linearly separable in the feature space. So, the accuracy of applications like object recognition relies on 1) feature set chosen to represent the object and 2) the ability of the classifier to separate feature sets of multiple classes. Object shape features provide a powerful clue to object identity and functionality and can be used for object recognition [1]. So, shape descriptors such as ILMD, ZMD and ARTD are used as feature descriptors. CSSD is not used as it has variable number of coefficients for various shapes. The

similarity evaluation based purely on distance measure between feature descriptors of the images is often not sufficient to cater to the broader user needs of a general purpose CBIR system. The query aim and thus the notion of ‘similarity’ may differ from user to user for the same query image, and hence the capability to incorporate user feedback to further fine tune and influence the result set will be of much value to the CBIR system. The relevance feedback can be treated as a classification problem in which sample images provided by the user are employed to train a classifier, which then classifies the database into images that are similar and dissimilar to the query image. Such a relevance feedback system demands a general and accurate classification of the images in a database using very few training samples and a system is proposed keeping this in mind. The classification accuracy of statistical and neural network based classifiers such as Support Vector Machine (SVM) [2], Least-Square Support Vector Machine (Ls – SVM) [3], Probabilistic Neural Network (PNN) [4], Learning Vector Quantization (LVQ) network [5] are investigated in this chapter.

The Support Vector Machine (SVM) approach is considered as a good candidate for relevance feedback because of its high generalization performance, even when the dimension of input space is very high. For several pattern classification applications, SVMs have already been shown to provide better generalization performance than traditional techniques, such as neural networks [6, 7]. In least squares (LS) version of SVM’s, one finds the solution by solving a linear system instead of quadratic programming. This is due to the use of equality instead of inequality constraints in the problem formulation. The PNN and LVQ are also supervised neural network that is widely used in the area of pattern recognition, nonlinear mapping, and estimation of the probability of class membership and likelihood ratios [8]. Details of these classifiers were outlined in Section 4.2.

Among the various descriptors and classifiers, the best descriptor – classifier combination need to be selected. For that, classification accuracy of neural network based classifiers with shape descriptors as feature vector at various training set size is computed. Statistical analysis is carried out to choose the combination of ideal classifier and shape descriptor which gives maximum performance. A relevance feedback scheme using this descriptor – classifier combination is then developed.

## 6.2 Selection of Optimum Shape Descriptor and Classifier

An ideal shape descriptor should have features of similar shapes clustered and dissimilar shapes separable in pattern vector space. Also, an ideal classifier should be able to classify patterns of similar shapes to same class. An optimum descriptor-classifier combination is the one which has maximum classification accuracy with a few numbers of training samples. Experiments are conducted to identify the best classifier – shape descriptor combination for various training set size, both for contour and region shape database.

SVM and Ls-SVM are binary classifiers. The multi-class support is achieved using a combination of binary classifiers based on Error Correcting Output Coding (ECOC) [9] approach. Details of ECOC approach were provided in Section 4.3.

## 6.3 Experimental Setup and Results

For the contour shape description, candidate shape descriptors are ILMD1, ZMD and ARTD, detailed in Chapter 5. The database used for evaluation of classification accuracy is MPEG-7 CE-1 Set B. It has 70 classes with 20 members in each class. Thus, based on the ground truth, shapes are to be assigned to one of the 70 classes by the classifier.

For the region shape description, candidate shape descriptors are ILMD4, ZMD and ARTD. The database used for evaluation of classification accuracy is MPEG-7 CE-2 whole database. It has 31 classes with 21 members in each class. Thus, based on the ground truth, shapes are to be assigned to one of the 31 classes by the classifier.

### 6.3.1 Comparison of Classification Accuracy

Experiments were conducted using the candidate descriptors for all classifiers at various training set size on contour and region shape database. A suitable percentage of the images from each class was used for training. Testing was initially done using the training set and validation was done using the rest of the members in each class. The aim of the experiment was to determine a suitable combination of a classifier and shape descriptor, which gives better classification results.

Firstly, experiment was conducted to choose an appropriate kernel function for optimum

**Table 6.1** Classification accuracy of SVM and Ls-SVM for various kernel functions using CE1-B database

Training set size (%)	Classification Accuracy (%)																	
	ILMD1						ZMD						ARTD					
	Linear		RBF		Polynomial		Linear		RBF		Polynomial		Linear		RBF		Polynomial	
	Svm	Ls-svm	svm	Ls-svm	svm	Ls-svm	svm	Ls-svm	svm	Ls-svm	svm	Ls-svm	svm	Ls-svm	svm	Ls-svm	svm	Ls-svm
10	14	2	65	26	3	4	10	3	55	22	3	3	23	8	59	24	5	1
20	3	2	71	38	3	4	2	7	65	31	2	6	20	3	68	33	5	1
30	5	4	76	45	2	5	3	3	70	35	2	1	20	3	71	38	4	2
40	5	5	80	50	3	7	4	4	74	43	2	2	19	2	75	47	5	2
50	5	7	83	55	3	7	3	3	77	47	2	8	21	5	77	54	5	1

classification performance of the SVM and Ls-SVM classifiers using the candidate shape descriptors. The classification accuracy using RBF, linear and polynomial kernel functions were evaluated. Details of the kernel functions were given in Section 4.2. The accuracy was measured for 5 different training sizes for each class in the database, starting at 10% and incrementing up to 50% of the class size as the training size. The remaining data set was used as the test set for validation. The results are presented as classification accuracy; the ratio of the number of cases correctly classified to the test set size expressed as a percentage, for different training sizes. Table 6.1 lists the average classification accuracy for contour shape database and Table 6.2 for region shape database. The RBF kernel function was found to have better performance for both SVM and Ls-SVM. So, for further studies with SVM and Ls-SVM, RBF kernel function is used.

Experiments were conducted for PNN and LVQ classifiers also. To validate the accuracy of classifiers in classifying patterns used for training, the entire test patterns were used for training and testing. Classification accuracy of 100% was obtained for all cases as expected. For validation purpose, the training set size was increased from 10% to 50% for each class in the

**Table 6.2 Classification accuracy of SVM and Ls-SVM for various kernel functions using CE2 whole database**

Training set size (%)	Classification Accuracy (%)																	
	ILMD4						ZMD						ARTD					
	Linear		RBF		Polynomial		Linear		RBF		Polynomial		Linear		RBF		Polynomial	
	svm	Ls-svm	svm	Ls-svm	svm	Ls-svm	svm	Ls-svm	svm	Ls-svm	svm	Ls-svm	svm	Ls-svm	svm	Ls-svm	svm	Ls-svm
10	46	4	78	62	53	11	41	13	70	54	48	9	45	3	72	56	50	14
20	51	25	81	68	61	14	46	36	74	59	56	11	47	11	76	62	59	10
30	60	24	85	74	70	18	53	16	78	63	63	12	55	21	81	69	65	14
40	61	14	90	80	77	12	56	17	83	68	73	13	58	27	85	75	74	9
50	64	39	92	83	81	19	60	35	88	77	75	12	61	23	88	78	77	8

**Table 6.3 Classification accuracy of various classifiers using CE1-B database**

Training set size (%)	Classification Accuracy (%)											
	ILMD1				ZMD				ARTD			
	SVM	PNN	LVQ	Ls-SVM	SVM	PNN	LVQ	Ls-SVM	SVM	PNN	LVQ	Ls-SVM
10	65	65	32	26	55	63	21	22	59	65	31	24
20	71	71	39	38	65	69	29	31	68	70	43	33
30	76	75	44	45	70	73	34	35	71	73	47	38
40	80	79	52	50	74	77	28	43	75	76	45	47
50	83	80	48	55	77	79	28	47	77	79	40	54

database. The patterns in each class other than training set are used as the test patterns. The classification accuracy of various classifiers for a given training set size for the candidate descriptors were computed. The results for SVM and Ls-SVM are given for RBF kernel function.

Table 6.3 consolidates the average classification accuracy using contour shape database for



**Table 6.4 Classification accuracy of various classifiers using CE2 whole database**

Training set size (%)	Classification Accuracy (%)											
	ILMD4				ZMD				ARTD			
	SVM	PNN	LVQ	Ls-SVM	SVM	PNN	LVQ	Ls-SVM	SVM	PNN	LVQ	Ls-SVM
10	78	74	73	62	70	70	69	54	72	74	69	56
20	81	77	76	68	74	75	70	59	76	78	69	62
30	85	81	76	74	78	80	73	63	81	82	70	69
40	90	87	76	80	83	84	73	68	85	88	70	75
50	92	92	78	83	88	88	68	77	88	92	70	78

all the classifiers. For ILMD1, SVM has overall better performance. The performance of PNN is comparable to SVM, but latter is slightly superior. Performance of LVQ and Ls-SVM are not satisfactory. For ZMD and ARTD, PNN has overall better performance. Performance of SVM is comparable to PNN, but PNN has better performance for lower training set size. Here also, accuracy of LVQ and Ls-SVM are comparatively low. Comparing the best performing classifier of each descriptor, SVM with ILMD1 has overall higher classification accuracy.

Table 6.4 consolidates the average classification accuracy using region shape database. For ILMD4, SVM has overall better classification accuracy than PNN, LVQ and Ls-SVM. Even though, PNN has comparable performance at high training set size, performance of SVM for low training set size is superior. Performance of LVQ and Ls-SVM are comparatively low. For ZMD and ARTD, performance of SVM and PNN are comparable, while PNN is slightly better. LVQ and Ls-SVM have low performance for both the descriptors. Comparing the best classifier for the three descriptors, ILMD4 with SVM shows better performance. This is more evident at low training set sizes.

### 6.3.2 Statistical Analysis

The comparison of classification performance was also undertaken in a statistically rigorous fashion. Here, the statistical significance of differences in the accuracy of classifications derived using different methods was assessed using McNemar test, without correction for continuity, for related samples. This is a nonparametric test that may be applied to confusion matrices that are  $2 \times 2$  in dimension [10]. Details of the McNemar Test were outlined in Section 4.5.1. This test is

based upon the standardized normal test statistic. The McNemar test result,  $Z$  – score is calculated as

$$Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \quad (6.1)$$

where  $f_{12}$  and  $f_{21}$  represent the off-diagonal entries of the error matrix.

The McNemar test was conducted to choose best descriptor for a given classifier, best classifier for a given descriptor and best descriptor – classifier combination. To choose best descriptor for a given classifier, for example, the error matrix was formed for one descriptor against another with an entry for each test case at a particular training size. The final confusion matrix for any two descriptors was formed with entries  $f_{12}$  detailing how many cases first descriptor could classify correctly and the second descriptor wrongly and  $f_{21}$  denoting for how many cases first descriptor was wrong and second descriptor correct. This test is done for all the three descriptors, one against another at different training sizes for all the four classifiers. The error matrix is similarly formed for test cases also. A positive score indicates that first method is a better than the second one and score more than 1.96 shows results at 95% confidence level.

**Table 6.5 Z - scores of one descriptor against another for various classifiers using CE1-B**

Training Set Size (%)	Z score											
	LVQ			PNN			Ls-SVM			SVM		
	ILMD1 vs ZMD	ILMD1 vs ARTD	ARTD vs ZMD	ILMD1 vs ZMD	ILMD1 vs ARTD	ARTD vs ZMD	ILMD1 vs ZMD	ILMD1 vs ARTD	ARTD vs ZMD	ILMD1 vs ZMD	ILMD1 vs ARTD	ARTD vs ZMD
10	5.79	0.30	5.28	0.68	0	0.69	2.12	0.59	0.71	5.31	3.22	2.08
20	5.33	-2.01	7.30	0.71	0.29	0.27	3.80	2.66	0.79	3.20	1.33	1.40
30	5.39	-1.30	6.87	0.70	0.76	0	5.31	3.85	1.41	3.18	2.76	0.26
40	12.54	3.80	8.74	0.72	1.32	-0.26	3.89	1.26	2.19	3.23	2.78	0.31
50	10.43	4.25	6.35	0.29	0.25	0	4.25	0.29	3.79	3.20	3.26	0

**Table 6.6 Z-scores of one classifier against another for various descriptors using CE1-B**

Training set size (%)		10	20	30	40	50
ILMD1	svm vs pnn	0	0	0.34	0.28	1.37
	svm vs lvq	17.21	16.82	16.59	14.68	18.42
	svm vs Ls-SVM	20.20	17.31	16.21	15.82	14.69
	pnn vs lvq	17.24	16.79	16.25	14.37	16.80
	pnn vs Ls-SVM	20.12	17.37	15.76	15.32	13.09
	lvq vs Ls-SVM	3.21	0.35	-0.29	0.71	-3.81
ZMD	svm vs pnn	-4.23	-2.03	-1.32	-1.37	-0.69
	svm vs lvq	17.74	18.82	18.81	24.09	25.70
	svm vs Ls-SVM	17.32	17.89	18.37	16.29	15.79
	pnn vs lvq	22.10	21.00	21.20	25.56	26.81
	pnn vs Ls-SVM	21.62	19.69	19.71	17.78	16.82
	lvq vs Ls-SVM	-0.29	-0.76	-0.36	-7.80	-9.69
ARTD	svm vs pnn	-3.17	-0.79	-0.81	-0.37	-0.73
	svm vs lvq	9.21	13.15	12.57	15.83	19.28
	svm vs Ls-SVM	18.24	18.31	17.29	14.69	11.81
	pnn vs lvq	17.82	14.31	13.65	16.34	20.23
	pnn vs Ls-SVM	21.31	19.38	18.47	15.36	18.40
	lvq vs Ls-SVM	3.85	5.33	4.82	-0.79	-7.36

It is to be noted that McNemar test was conducted on the individual classification result of each query image in the database, not on the classification accuracy results.

Table 6.5 lists the Z scores of McNemar test conducted to choose best descriptor for a given classifier using contour shape database. It can be noted that ILMD1 against ZMD and ARTD gives positive values in most of the cases for all the four classifiers. This indicates that ILMD1 is better than other descriptors. For PNN classifier, Z-scores of ILMD1 against ZMD and ARTD are small positive values. This shows that, their performance is comparable but ILMD1 is slightly superior. For SVM, Z-scores of ILMD1 are greater than 1.96 (95% confidence level) indicating superior performance. Also, ARTD against ZMD gives positive values for most of the cases indicating that overall performance of ARTD is better than ZMD. In order to determine the best classifier for a given descriptor using contour shape database,

**Table 6.7 Z-scores of descriptor-classifier combination using CE1-B**

Descriptor-classifier	Training set size (%)				
	10	20	30	40	50
ILMD1 with SVM Vs ZMD with PNN	0.72	0.78	1.31	1.28	2.04
ILMD1 with SVM Vs ARTD with PNN	0	0.29	1.30	2.01	2.07
ARTD with PNN Vs ZMD with PNN	0.69	0.27	0	-0.26	0

McNemar test was conducted. Table 6.6 lists the Z scores of all the three descriptors with entries for one classifier against another for contour shape database. For ILMD1, SVM shows better performance than other classifiers as the Z-scores are positive values. The performance of SVM and PNN are comparable, but SVM is slightly superior to PNN. Comparing LVQ and Ls-SVM, LVQ has better performance. For ZMD and ARTD, PNN is the best classifier as performance of PNN is marginally better than SVM. For ZMD, Ls-SVM is a better classifier than LVQ. LVQ shows better performance than Ls-SVM for ARTD at low training set sizes.

To identify the best descriptor – classifier combination, performance of SVM classifier for ILMD1 and PNN classifier for ZMD and ARTD using contour shape database need to be compared. The Z-scores of these combinations are given in Table 6.7. The positive values indicate that ILMD1 with SVM classifier is better than ARTD and ZMD with PNN. Also, the overall performance of ARTD with PNN is slightly better than ZMD with PNN.

McNemar test was conducted using the classification results for the region shape database also. The McNemar test results to choose best descriptor for a given classifier using region shape database CE2 are listed in Table 6.8. For LVQ, ILMD4 against ZMD and ARTD gives positive values indicating that ILMD4 has better performance. ARTD against ZMD gives negative values showing that ZMD is a better descriptor. For PNN, ILMD4 against ZMD gives positive values and against ARTD low negative values. Thus ILMD4 has better classification

**Table 6.8 Z - scores of one descriptor against another for various classifiers using CE2 whole database**

Training Set Size (%)	Z score											
	LVQ			PNN			Ls-SVM			SVM		
	ILMD4 vs ZMD	ILMD4 vs ARTD	ARTD vs ZMD	ILMD4 vs ZMD	ILMD4 vs ARTD	ARTD vs ZMD	ILMD4 vs ZMD	ILMD4 vs ARTD	ARTD vs ZMD	ILMD4 vs ZMD	ILMD4 vs ARTD	ARTD vs ZMD
10	2.02	2.09	0	2.07	0	2.14	4.28	3.27	0.81	4.22	3.18	0.66
20	3.21	3.89	-0.28	0.76	-0.30	1.21	4.83	3.32	1.30	3.81	2.73	0.70
30	1.33	3.30	-1.35	0.39	-0.26	0.63	5.87	2.71	3.26	3.79	2.10	1.24
40	1.28	3.26	-1.27	1.40	-0.29	2.10	6.30	2.76	3.81	3.61	2.79	0.71
50	5.33	4.15	0.73	2.15	0	2.01	3.17	2.73	0.29	2.05	2.17	0

results than ZMD, but comparatively slightly inferior to ARTD using PNN classifier. Also, ARTD is better than ZMD. In the case of Ls-SVM classifier, ILMD4 is better than ZMD and ARTD, and ARTD is better than ZMD. ILMD4 against ZMD and ARTD gives positive values for SVM classifier, indicating its higher accuracy. Also, ARTD is slightly superior to ZMD for SVM classifier. In order to determine the best classifier for the three descriptors using region shape database, McNemar test was conducted by comparing the results of one classifier against another for each descriptor. Table 6.9 lists the Z score of all the three descriptors with entries for one classifier against another for region shape database. For ILMD4, SVM gives positive Z-scores against other classifiers showing its superior classification accuracy. Most of the values are above 1.96, indicating the superiority of SVM. PNN is better than LVQ and Ls-SVM. LVQ shows better classification results at lower training set size than Ls-SVM. In the case of ZMD, except for training set size 10% and 50%, PNN is found to be slightly superior to SVM. Positive Z-scores indicate that PNN is better than other classifiers. SVM is better than LVQ and Ls-SVM. LVQ is better than Ls-SVM except at high training set size. PNN shows better classification accuracy than all other classifiers for ARTD. SVM is better than LVQ and Ls-SVM. Also, LVQ is better than Ls-SVM except for training set size of 40% and 50%

**Table 6.9 Z - scores of one classifier against another for various descriptors using CE2 whole database**

Training set size (%)		10	20	30	40	50
ILMD4	Svm vs pnn	2.09	2.02	2.11	1.38	0
	Svm vs lvq	2.70	2.71	4.85	7.32	7.27
	svm vs Ls-SVM	8.32	6.85	5.82	5.33	4.86
	pnns vs lvq	0.31	0.32	2.74	5.85	7.33
	pnns vs Ls-SVM	6.35	4.90	3.89	3.81	4.86
	lvq vs Ls-SVM	8.83	4.22	0.75	-2.04	-2.71
ZMD	Svm vs pnn	0	-0.28	-0.76	-0.25	0
	Svm vs lvq	0.29	2.03	2.73	5.36	10.44
	svm vs Ls-SVM	8.33	7.84	7.78	7.87	5.81
	pnns vs lvq	0.26	2.73	3.81	5.86	10.38
	pnns vs Ls-SVM	8.40	8.33	8.87	8.33	5.89
	lvq vs Ls-SVM	7.80	5.75	5.24	2.73	-4.84
ARTD	Svm vs pnn	-0.71	-0.75	-0.32	-1.23	-2.08
	Svm vs lvq	1.36	3.77	5.88	7.71	9.32
	svm vs Ls-SVM	8.39	7.37	6.33	5.37	5.27
	pnns vs lvq	2.76	4.88	6.31	9.28	11.27
	pnns vs Ls-SVM	9.26	8.30	6.90	6.88	7.31
	lvq vs Ls-SVM	6.82	3.84	0.30	-2.75	-4.18

for ARTD.

To identify the best descriptor – classifier combination, performance of SVM classifier for ILMD4, and PNN classifier for ZMD and ARTD using region shape database need to be compared. The Z-scores of these combinations are given in Table 6.10. ILMD4 with SVM gives positive Z-score against other descriptor-classifier combinations indicating that it is the best combination. Also, ARTD with PNN is better than ZMD with PNN.

The results obtained in the statistical analysis using McNemar test agrees with the result obtained using classification accuracy statements for both contour shape and region shape database.

**Table 6.10 Z-scores of descriptor-classifier combination using CE2 whole database**

Descriptor-classifier	Training set size (%)				
	10	20	30	40	50
ILMD4 with SVM Vs ZMD with PNN	4.21	3.23	2.76	3.28	2.08
ILMD4 with SVM Vs ARTD with PNN	2.12	1.32	1.45	0.74	0
ARTD with PNN Vs ZMD with PNN	2.14	1.21	0.63	2.10	2.01

## 6.4 Retrieval Using Relevance Feedback

### 6.4.1 Method

From the experimental results obtained in the previous section, it can be seen that ILMD (ILMD1 for contour shape database and ILMD4 for region shape database) with SVM classifier (RBF kernel function) gives optimum classification results. So, it is selected for relevance feedback implementation. During the process of relevance feedback, user marks a set of images as relevant or irrelevant (positive and negative samples). The SVM classifier is used as a binary classifier to separate the images in the database as relevant class or irrelevant class. Training images can be selected from the top few images retrieved using a standard technique such as city block distance measure, outlined in Chapter 5. The classification of images is done according to the distance from each image to the separating hyperplane. By sorting images based on their distance to the hyperplane, the retrieval results can be improved. The steps for relevance feedback are described below.

Mark relevant and irrelevant images, from the top few images retrieved using city block distance measure. The SVM classifier is then trained by using a set of training examples  $(x_i, y_i)_{1 < i < N}$ , where  $x_i \in R^d$ , is the  $i^{\text{th}}$  input pattern and  $d$  being the dimension of the input

space, and  $y_i$  the  $i^{\text{th}}$  output pattern belongs to a class labeled by  $y_i \in \{-1, 1\}$ . The classification function of SVM (Equation (4.5)) is given as

$$f(x) = \text{sgn}\left(\sum_{i=1}^N \alpha_i y_i K(x_i, x) + b\right) \quad (6.2)$$

where  $K(\cdot)$  is the kernel function,  $\alpha_i$  the Lagrange multipliers and  $b$  a constant.

Now the images in the database are classified into two classes: those belonging to the query class and those not belonging to the query class. In the first learning iteration, the images marked as relevant or irrelevant are close to the query and lie on the separating margin. These are support vectors and contribute more to classification accuracy. Samples far away from hyperplane do not contribute to classification accuracy. The advantage of using an initial retrieval result in selecting relevant and irrelevant samples is that support vectors can be identified. From the results of first iteration, more samples can be marked as relevant or irrelevant to further refine the classifier. As SVM has the power to learn with few training samples, satisfactory results can be obtained with low training set size.

For sorting the images based on similarity to a query image using SVM classifier, the  $\text{sgn}()$  in the classification function (Equation (6.2)) is ignored to get the similarity distance to the query as shown below.

$$f(x) = \left(\sum_{i=1}^N \alpha_i y_i K(x_i, x) + b\right) \quad (6.3)$$

$f(x)$  is computed for all the images in the database. Now sort the images based on the increasing value of  $f(x)$  to get the new query result. Thus, query result can be displayed in decreasing order of similarity.

## 6.4.2 Experimental Setup and Results

ILMD1 with SVM was tested using contour database CE1- B and ILMD4 with SVM using region database CE-2 whole. All the images in the database are used as test pattern to the classifier. The images are classified into query class or otherwise.



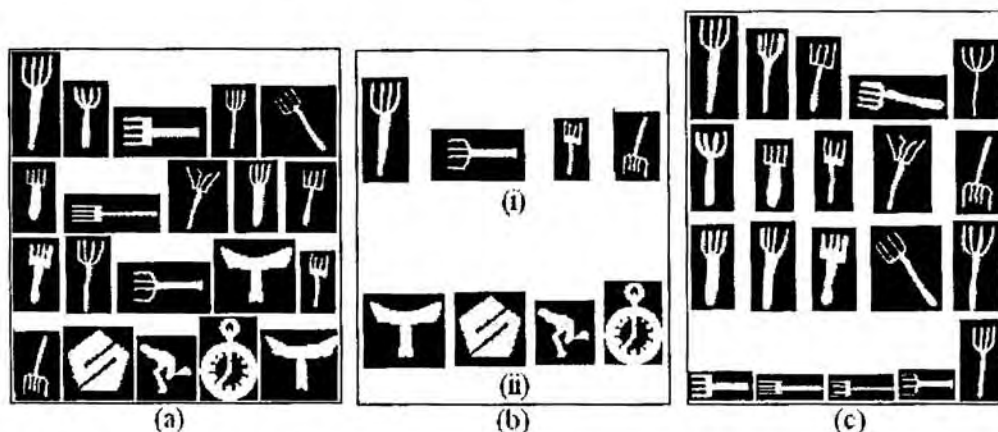


Figure 6.1 Query results of a shape for ILMD1 with SVM classifier with relevance feedback using CE1-B database.(a) retrieval result using city block distance, (b) Feed back samples (i) relevant (ii) irrelevant, (c) Images from query class correctly classified

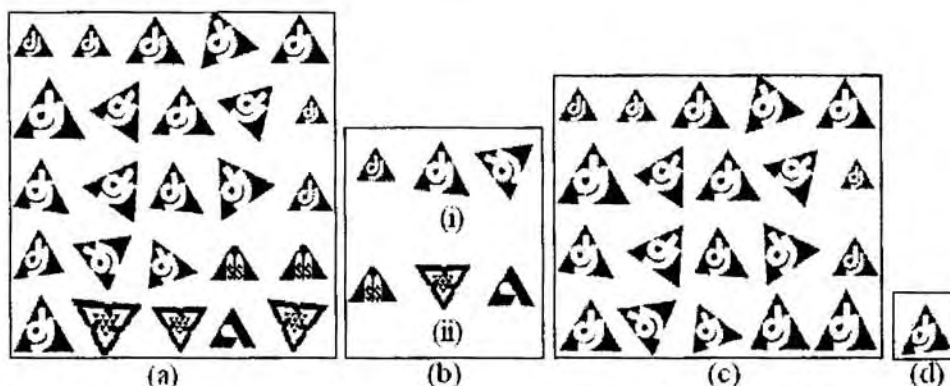


Figure 6.2 Query results of a shape for ILMD4 with SVM classifier with relevance feedback using CE2 whole database.(a) retrieval result using city block distance, (b) Feed back samples (i) relevant (ii) irrelevant, (c) Images from query class correctly classified (d) Images from query class misclassified

The results of a query shape for single training iteration using contour shape database is shown in Figure 6.1. First 20 retrieval results using distance measure are shown in Figure 6.1 (a). Top – left image is the query image. In the top 20 images, 5 images are dissimilar to the query image. For relevance feedback, 4 samples each are selected as relevant and irrelevant from the top 20 query results for training the SVM classifier as shown in Figure 6.1 (b). All the 20 similar images in the query class were correctly classified. Figure 6.1 (c) shows the correctly classified images in the query class. The results for a query shape for single training iteration using region shape database is shown in Figure 6.2. First 25 retrieval results using distance

measure are shown in Figure 6.2 (a). In the top 25 images, 2 images are missing from the query class. For relevance feedback, 3 samples each are selected as relevant and irrelevant from the top 25 query results for training the SVM classifier as shown in Figure 6.2 (b). Out of 21 similar images in the query class, 20 images were correctly classified and 1 image was misclassified. Figure 6.2 (c) shows the correctly classified images in the query class and Figure 6.2 (d) misclassified image in the query class. The result can be improved by increasing the training set size and number of iterations.

Retrieval accuracy is related to the relevant images selected for each query class. This may vary and is user subjective. Hence, a consolidated score of improvement in retrieval accuracy cannot be provided.

## 6.5 Conclusions

A new relevance feedback technique to improve the retrieval accuracy in CBIR using shape descriptors was developed in this chapter. The classifiers considered were SVM, Ls-SVM, PNN and LVQ. ILMD (ILMD1 for contour shape and ILMD4 for region shape), ZMD and ARTD were the shape descriptors used as feature vectors to classifiers. Experiments were conducted to select the best kernel function for SVM and Ls-SVM and RBF was found to give optimum performance for both classifiers. The best descriptor – classifier combination, for contour and region shape database, were determined by comparison of classification accuracy and statistical analysis using McNemar test for various training set size. Experimental studies conducted have shown that ILMD1 with SVM (for contour shape database) and ILMD4 with SVM (for region shape database) give maximum classification performance. A new relevance feedback technique was developed by training an SVM classifier with ILMD features using relevant and irrelevant samples to improve the retrieval accuracy of a CBIR system.

## References:

- [1] Mirosław Bober, "MPEG-7 Visual Shape Descriptors," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 11, no. 6, pp. 716–719, 2001.
- [2] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, 1995.
- [3] Suykens J.A.K. and Vandewalle J., "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, 1999.

- [4] Wasserman P.D., *Advanced Methods in Neural Computing*, Van Nostrand Reinhold, New York, 1993.
- [5] Kohonen T., *Self-Organization and Associative Memory*, 2nd Edition, Springer-Verlag, Berlin, 1987.
- [6] Schölkopf B., Sung K., Burges C., Girosi F., Niyogi P., Poggio T. and Vapnik V., "Comparing Support Vector Machines with Gaussian Kernels to Radial Basis Function Classifiers," *IEEE Trans. on Signal Processing*, vol. 45, no. 11, pp. 2758–2765, 1997.
- [7] Henning Muller, Nicolas Michoux, David Bandon and Antoine Geissbuhler, "A Review of Content-Based Image Retrieval Systems in Medical Applications - Clinical Benefits and Future Directions," *Int'l J. Med Inform.*, vol. 73, no. 1, pp. 1-23, 2004.
- [8] Specht D. F. and Romsdahl H., "Experience with adaptive -probabilistic neural network and adaptive general regression neural network," in *Proc. of ICNN'94*, pp. 1203-1208, 1994.
- [9] Dietterich T. G. and Bakiri G., "Solving multiclass learning problems via Error-Correcting Output Codes," *J. of Artificial Intelligence Research*, vol. 2, pp. 263-286, 1995.
- [10] Foody G. M., "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," *J. of Photogrammetric Engineering and Remote Sensing*, vol. 70, no. 5, pp. 627–633, 2004.

---

## Chapter 7

# Content-Based Retrieval of Scoliosis Images

---

### 7.1 Introduction

In this chapter, a novel algorithm for the CBIR of spine images with scoliosis, which retrieves images which are clinically similar to a query image, based on automatically derived features is proposed. Scoliosis is an abnormality due to the lateral curvature of the spine, the magnitude of which is measured using Cobb Technique [1]. Details of the pathology scoliosis and Cobb technique were outlined in Chapter 3. Initially, the motivations for proposing a technique for CBIR of scoliosis images are discussed. A new algorithm developed for automatic estimation of position and orientation of the spine is outlined. A rule based algorithm developed for strategic vertebrae selection and Cobb angle measurement is then presented. A set of automatically derived features necessary for indexing the scoliosis image for CBIR was formulated and is described in the next section. The software system 'SpineSearch' designed based on the developed algorithm is detailed next. Then, a discussion on the applications of the developed technique and concluding remarks are outlined.

### 7.2 Motivations for proposing CBIR of scoliosis Images

Nowadays, in hospitals 100's of x-ray images are generated per day [2]. Due to the difficulty in storage and tracking of these films, x-rays are either digitized using a high resolution x-ray scanner or digitally generated using direct digital radiography. With the advent

of picture archiving and communication systems (PACS), digital images are stored in them for efficient search and retrieval. In PACS, images are usually stored in Digital Imaging and Communications in Medicine (DICOM) [3] format. Image retrieval with PACS and Electronic Medical Records (EMR) systems is based on textual descriptions.

Modern communication standards use non-image data (textual) for standardized description of information such as technical parameters related to the imaging modality, patient information, body region examined, study, etc. In order to provide comprehensive, detailed coverage for multi-specialty biomedical imaging, the College of American Pathologists (CAP), secretariat of the Systematized Nomenclature of Human and Veterinary Medicine (SNOMED), have entered into partnership with the DICOM Standards Committee and other professional organizations to develop a nomenclature for diagnostic imaging applications [4]. The SNOMED DICOM microglossary was developed to provide context-dependent value sets for DICOM coded-entry data elements, and semantic content specifications for reports and other structures composed of multiple data elements [5]. Although, the capability of storing explicitly-labeled coded descriptors in DICOM images and reports, improves the potential for selective retrieval of images and related information, the controlled terminology within the DICOM tables has been found to be insufficiently detailed for order entry systems [6, 7].

Information such as filenames, keywords and patient identification numbers are typically stored in the DICOM header and are used to retrieve images. Correct retrieval from such systems typically requires exact keywords. Doctors, however, may use different keywords to describe the same image because of the difficulty in interpreting some images and the lack of a unified lexicon. So, queries based on such information are proven to contain a fairly high rate of errors. For example, for the field anatomical region, error rates of 16% have been reported [7]. Clinical decision support techniques such as case-based reasoning [8] or evidence-based medicine [9] demands an even stronger need to retrieve images that can be valuable for supporting certain diagnoses and for planning therapeutic strategies. Therefore, there is a need for adding CBIR capability in the PACS systems in order to provide image indexing and retrieval based on content [10]. CBIR, based on visual features, not only allow the retrieval of cases of patients having similar diagnoses but also retrieve cases with visual similarity but different diagnoses.

Indexing of scoliosis images requires reliable measurement of the curve. The poor

reproducibility of Cobb angle measurement commonly raises uncertainty about the validity of identified changes or progression in curvature [11]. The scoliosis measurement is carried out either manually using a protractor or using a computer assisted technique. Even in computer assisted techniques, there will be an error in Cobb measurement due to the manual selection of end vertebrae levels. The average intraobserver standard deviation (SD) has been reported as  $3.5^\circ$  and interobserver SD range from  $2.8^\circ$  to  $7.2^\circ$  [12 – 16]. Regarding the type of radiographs, a recent study conducted by Wills et al. [17] demonstrates that both digitally and traditionally acquired scoliosis radiographs have similar error rates.

For scoliosis images, textual annotations have several limitations. Variability exists in the selection of strategic vertebrae, measurement of Cobb angle and in the assignment of curve type in a classification scheme. For the Cobb angle measurement, the definition of end vertebrae introduces the main source of error [18]. Moreover, since many classification schemes are in use today, ambiguity is present in selecting a classification scheme. Thus, manual annotation of scoliosis images will vary from surgeon to surgeon. A perfect image search engine should make use of the textual descriptions (such as the DICOM header) as well as the features derived from the content of the image to form a hybrid CBIR system.

In the literature, no techniques have been so far reported for CBIR of scoliosis images. The user subjectivity in scoliosis measurement can be eliminated by using an automated method for strategic vertebra selection and Cobb angle computation. Selecting image features as descriptors instead of a classification scheme eliminates the ambiguity in selecting a classification scheme. Keeping this in mind, a technique for efficient indexing and retrieval of scoliosis images is developed for CBIR.

### **7.3 Automatic Estimation of Position and Orientation of Spine**

In this section, a method to automatically determine the location and orientation of spine in a digital radiograph is developed. A detailed review of literature related to spine localization was presented in Section 3.10.1. The state-of-the-art technique for vertebrae segmentation is based on the Deformation Model using active shape modeling (ASM) [19, 20]. The knowledge of position and orientation of spine are important for the success of the template matching based method. So, a method is developed for spine localization based on mathematical morphology making use of the vertebral morphometry. It computes the orientation of the spine and a point

on the right end plate of the vertebral faces.

### 7.3.1 Methods

The gray-scale characteristics of the vertebrae are complex. The spine exhibits visually recognizable ridge characteristics, usually extending throughout the length of the spine, but particularly visible in the region of the lower vertebrae. Figure 7.1(a) shows the original gray-scale profile of a cervical spinal column. The gray-scale profile of the vertebra, marked in Figure 7.1(a) using a rectangle, is demonstrated by the surface characteristic plot in Figure 7.1(b). The gray-scale image is represented as a topographic model of a terrain where the pixel value represents the height at the pixel coordinates. The surface is rendered using the top-view direction. The ridge point cluster on the posterior boundary of the vertebra is encircled in Figure 7.1(b). The boundary of a vertebra is usually marked using six standard morphometric points as shown in Figure 7.1(c). These six points have the semantic relevance as follows:

- Points 1 and 4 marks the upper and lower posterior corners of the vertebra, respectively;
- Points 3 and 6 marks the upper and lower anterior corners of the vertebra, respectively;
- Points 2 and 5 marks the median along the upper and lower vertebra edge in the sagittal view, respectively.

The curvature of a curve that fit into the midpoints of the top and bottom (points 2 and 5) of each vertebral body for all the vertebrae of the same type (for e.g. cervical), can be taken as the curvature of the spine in that region [21]. These midpoints do not have any prominent associated

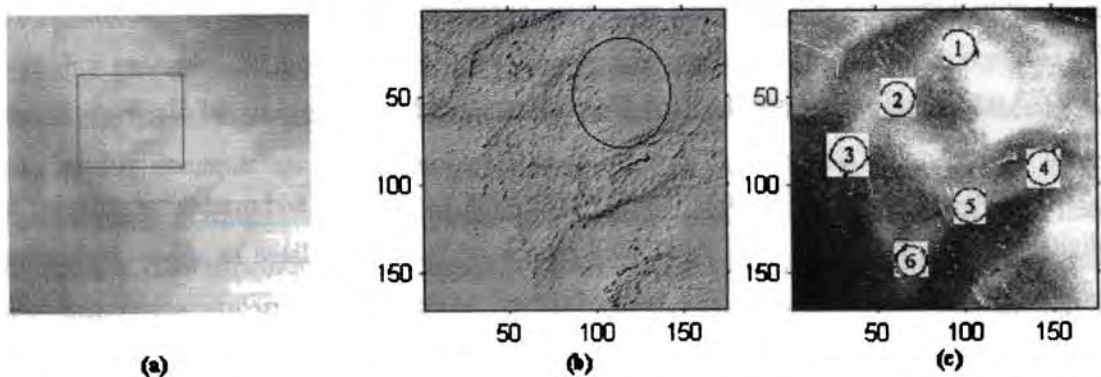


Figure 7.1 (a) Original gray-scale image of a cervical spinal column. (b) Surface characteristic plot of the vertebra. (c) Six standard morphometric points marked on the histogram equalized image of the vertebra body.

visual characteristics, either in gray-scale or in shape. So, they are very poor candidates for detection until the spine anatomy is already known after a finer-grained segmentation. However, the gray-scale ridge points appear near the right edges of the vertebral faces. Hence a line joining these must be conjectured to give a reasonable approximation to the spine location and orientation. These ridge points have bright gray-scale values and can be visualized as being on the local higher regions in elevation space of the spine area. Considering these facts, a method based on mathematical morphology was introduced to detect these points as local maxima.

The detection of ridge point clusters in the outer boundary of the vertebrae is carried out using mathematical morphological operation [22]. A gray-scale image is defined as a 3-D set and two images  $f$  and  $b$  can be represented as follows.

$$f = \{(x, y, z) \mid z \leq g(x, y)\}, b = \{(x, y, z) \mid z \leq h(x, y)\} \quad (7.1)$$

where  $g(x, y)$  and  $h(x, y)$  are the gray-scale of the pixel at location  $(x, y)$  of  $f$  and  $b$  respectively. In mathematical morphology, the image  $b$  that is used for processing image  $f$  is called a structuring element (SE). For a flat SE (a binary), the erosion is simplified to find the minimum gray level and the dilation to find the maximum during the process, in the neighborhood defined by the SE.

The dilation of  $f$  by  $b$  can be represented as

$$f \oplus b = \{(x, y, z) \mid z \leq \max [g(x-x_1, y-y_1) + h(x_1, y_1)], \forall_{x_1, y_1}\} \quad (7.2)$$

and erosion of  $f$  by  $b$  as

$$f \ominus b = \{(x, y, z) \mid z \leq \min [g(x+x_1, y+y_1) - h(x_1, y_1)], \forall_{x_1, y_1}\} \quad (7.3)$$

where  $(x \pm x_1), (y \pm y_1) \in D_f$  and  $x_1, y_1 \in D_b$ ;  $D_f$  and  $D_b$  are the domains of  $f$  and  $b$ , respectively.

A 2-Dimensional flat SE,  $b$ , which is 'disc' shaped in the Euclidean metric and centered at the origin is constructed as follows.

$$b(x, y) = \left\{ (x, y) \in Z^2 : \sqrt{x^2 + y^2} \leq (r + 0.5) \right\} \quad (7.4)$$



where  $r$  is the radius of the disc, chosen as 20.

For each of the test image, a margin area is used on all four image sides (left, right, top, and bottom) and pixels that were outside this margin limit were not processed. This was done to avoid the frequent problems encountered by the presence of very bright pixels due to light leakage near the edges of the images. As a pre-processing step, histogram equalization is performed on the image for enhancing the contrast between the vertebrae and the surrounding regions. The contrast-enhanced image is morphologically opened with the SE. The opening operation includes two procedures; erosion followed by dilation.

$$f \circ b = (f \ominus b) \oplus b \quad (7.5)$$

The Top-Hat transformation [22] is then performed by subtracting the opened image from the original image to detect objects having the size of the SE and high gray-scale profile.

$$h = f - (f \circ b) \quad (7.6)$$

All the 8-connected structures that are lighter than their surroundings and connected to the image border are then removed by an erosion operation. This operation removes some of the non-spine edges in the image. Any grain (i.e., connected component) with area less than 100 in the neighborhood defined by a 'cross' shaped flat SE (Equation (7.7)) are removed to eliminate isolated clutters in the image.

$$c(x, y) = \left\{ (x, y) \in Z^2 : |x| + |y| \leq 1 \right\} \quad (7.7)$$

The resultant image is a coarsely segmented one with ridges on the vertebral faces enhanced. The ridge point clusters located in each vertebra face need to be connected together to mark the outer boundary of the spine. The Radon Transform [23] of the segmented image is computed for locating the strongest line in the image. The projection of the image matrix onto  $x$ -axis in the range of 0-100 degree is found. A projection of a 2-Dimensional function at a particular angle is the line integral in that direction. The Radon transform of the image function  $f(x, y)$  parallel to the  $y'$  axis can be computed as follows:

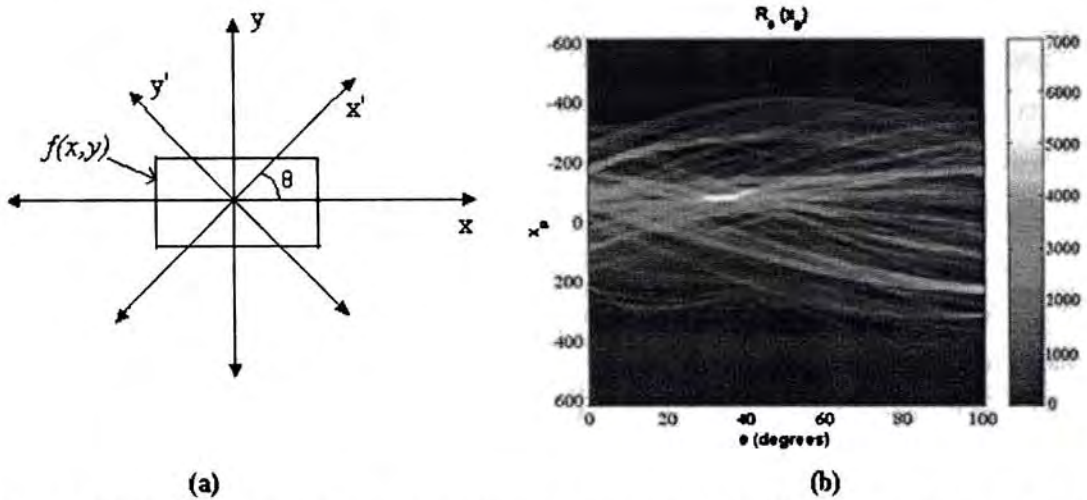


Figure 7.2 (a) The Geometry of the Radon Transform. (b) The Radon Transform of a test image using 100 projections

$$R_\theta(x') = \int_{-\infty}^{+\infty} f(x' \cos \theta - y' \sin \theta, x' \sin \theta + y' \cos \theta) dy' \quad (7.8)$$

where 
$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

The geometry of the Radon transform is shown in Figure 7.2(a) and the Radon coefficient values computed for various angles in a test image is shown in Figure 7.2(b). The locations of strong peaks in the Radon transform matrix correspond to the high line integral values in the image. As the segmented image has the ridge point clusters boosted, a line joining these points will correspond to the strongest peak in the Radon transform matrix. From the orientation ( $\theta$ ) and coordinate location on the  $x'$  axis ( $x_p$ ) of the largest absolute value of radon coefficient, corresponding  $y'$  coordinate ( $y_p$ ) can be determined. Now, we know a point ( $x_p, y_p$ ) in the posterior boundary and the orientation of the spine. A line drawn passing through ( $x_p, y_p$ ) and perpendicular to  $\theta$ , marks the posterior boundary of the spine in the original x-ray image. The coarse level segmentation convergence is said to be achieved, if the error in computed orientation and location of the posterior boundary with respect to the ground truth information is within tolerable limits.

In some cases, where the image contains metallic objects like ornaments, orthodontic treatment accessories etc, failure of segmentation convergence occur. This is because of the

higher gray-scale value of these objects than the spine ridge points. In such cases, the proposed method will fail as the Radon Transform joins the metallic object points instead of spine ridges due to its high gray scale value. This can be corrected by another pass, in which the metallic object is detected and removed from the Top-Hat segmented image. The removal of the external objects can be done by using H-maxima transformation [24], which suppresses all maxima in a given image whose height is lower than a given threshold value. The regional maxima are connected components of pixels with an intensity value above a specified threshold and whose external pixels have a lower intensity. The threshold is fixed as the 90% of the maximum gray-scale value of the coarsely segmented image. As the metallic object has higher gray-scale value, it will be identified as the regional maxima, and is subtracted from the coarsely segmented image before Radon transform computation. The H-maxima transform is computed in a 2-

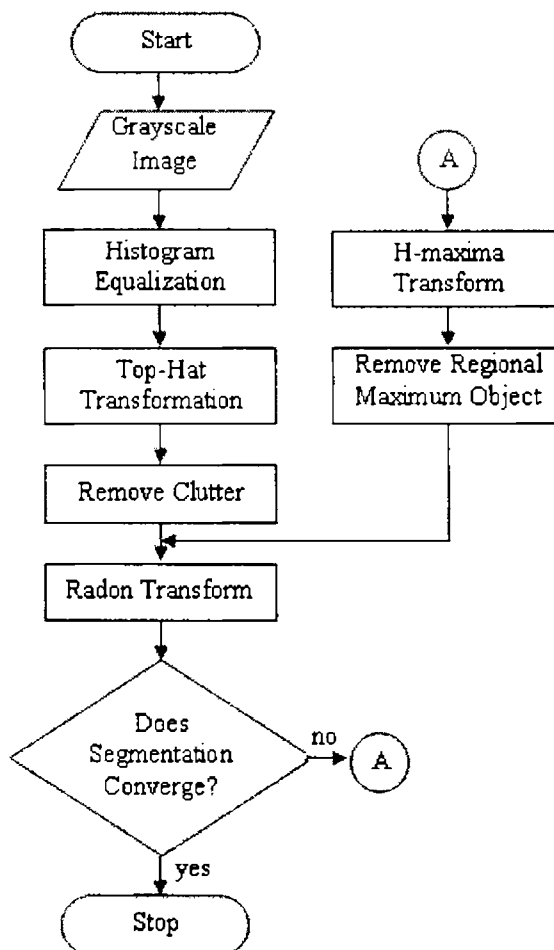


Figure 7.3 The flowchart of the spine localization process

Dimensional eight – connected neighborhood. The foreign object removed and ridge points boosted coarsely segmented image is subjected to Radon Transform computation as already explained. A flow chart representation of the complete process is given in Figure 7.3.

### 7.3.2 Experimental Setup and Results

The developed algorithm is tested on an ensemble of 100 images chosen from the archive maintained by National Library of Medicine (NLM) as part of the second National Health and Nutrition Examination Survey (NHANES II) [25]. The digitized x-ray images were originally generated from cervical radiographs using a Lumisys laser scanner at a resolution of 146 dpi. The candidate images having various image size and spine orientations are randomly chosen from the database. The lateral view cervical spine images had pixel depth of 8-bits. The seven vertebrae present in a cervical image are usually referred to as C1 to C7.

For each image in the test set we used the expert-collected  $(x, y)$  coordinate (provided by NLM) of the posterior bottom point (point 4 in Figure 7.1(b)) for C2 and C6 (or C5) vertebra. The slope of the straight line fit to the spine joining these points gives the spine orientation for each image in the test set. The orientation of the line joining the points 4 or points 5 (middle point of bottom face of vertebra body) of each vertebrae will be the same as they are parallel. As our computed location points are on the posterior boundary, we chose to connect point 4 of each vertebra as the ground truth (which is on the posterior boundary) so that the location accuracy also can be evaluated. The error in orientation was calculated by taking the absolute difference of the spine orientation computed using the proposed algorithm and that obtained as the ground truth for each of the image in the test set. From the literature it is observed that a tolerance up to 15 degree is permissible in orientation estimation [26]. The absolute error using the developed algorithm was found to be within the tolerable limit for all the 100 images in the test set. For 8 cases, the algorithm had to undergo a second pass after removal of the metallic objects present

Table 7.1 Computed spine orientation accuracy

Serial Number	Orientation Error range (degrees)	% of images in the range
1	0 - 3	57
2	3 - 6	25
3	6 - 9	12
4	9 - 12	3
5	12 - 15	3

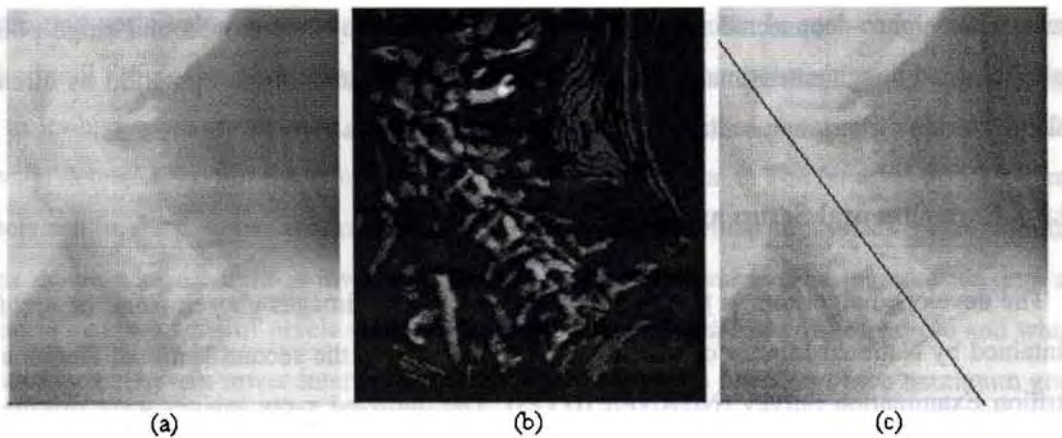


Figure 7.4 (a) Cervical spine image. (b) Intermediate result of Segmentation using mathematical morphology. (c) A marker drawn using computed spine location and orientation

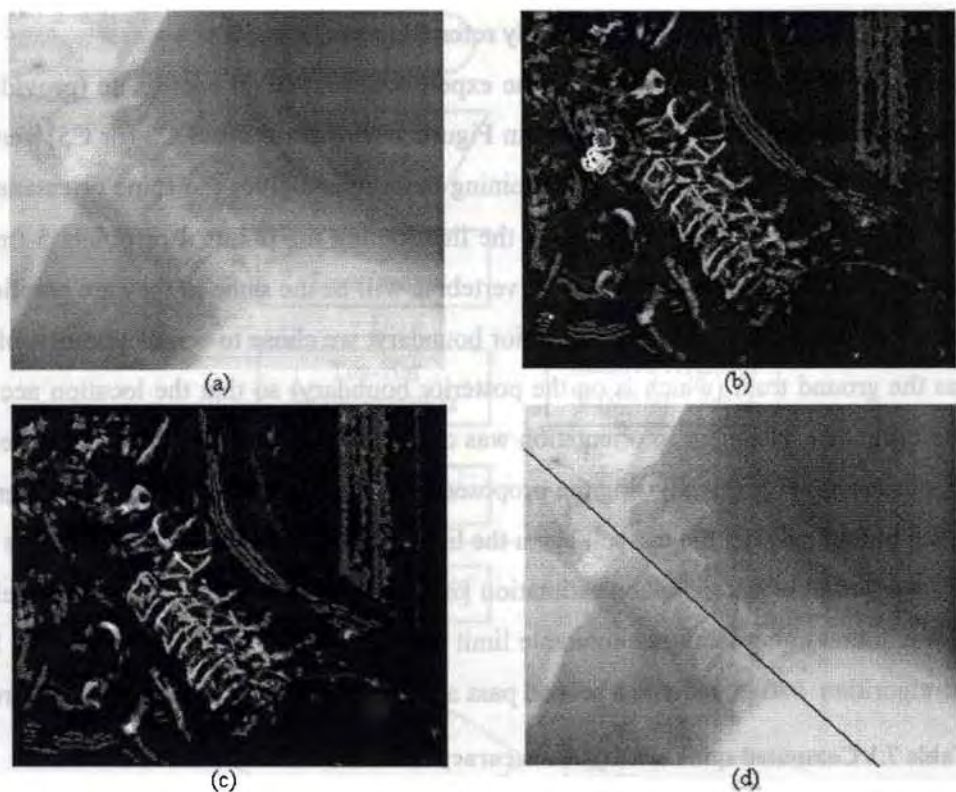


Figure 7.5 (a) Cervical spine image containing a metallic object. (b) Intermediate result of segmentation using mathematical morphology. (c) Image after removal of the metallic object using H-Maxima Transform. (d) A marker drawn using computed spine location and orientation.

in them. Table 7.1 tabulates the percentage of the number of images that fall in different ranges

of the orientation error. It is seen from the table that for more than 80 percent of the cases the error is less than 6 degrees. The average orientation error for 100 cases was found to be 4.6 degrees.

A point computed on the posterior boundary of the vertebra identifies the location of the vertebra. The accuracy of the location of the point identified can be calculated by computing the perpendicular distance from the point to the straight line approximated (by joining point 4) from the ground truth data using analytical geometry. The maximum deviation was found to be  $\pm 5.2$  mm. The computed location point of C2 and C6 is joined to draw a marker on the posterior boundary of the spine for display purpose. Figure 7.4 demonstrates the images at various stages of the algorithm and shows a marker drawn for the computed spine location. Figure 7.5 shows the case of an image in which a foreign object is present.

### 7.3.3 Discussion

In ASM based vertebrae segmentation algorithms, a template of the spine is placed at different orientations and locations for maximum matching before applying deformation model to fit to the spine. The developed algorithm is useful as an initialization step for template positioning in ASM based vertebrae segmentation algorithms to save computational overhead. The developed algorithm is based on the vertebra morphometry. The location of the posterior boundary of the spine and its orientation information can also be used to define a region of interest (ROI) containing the spine. The image could be rotated as per the orientation of the spine, so that spine will align parallel to the vertical axis. The width of the ROI should be chosen, based on the resolution of the image, in such a way to include the entire vertebra body in the box. For the 146 dpi images, a width of 250 pixels (from posterior to anterior boundary) is found to contain the spinal column, even in cases of spines with large curvature. The length of the ROI can be taken as the length of the image itself.

## 7.4 Measurement and Retrieval of Scoliosis Images

This section details the acquisition of land mark points for curve digitization, rule based algorithm proposed for automatic selection of strategic vertebrae and measurement of the

various parameters of a scoliosis image. Also, feature descriptor for scoliosis images is formulated and a query structure for CBIR is designed.

#### 7.4.1 Acquisition of Land Mark Points

For the measurement of scoliosis, end plate tilt angles of the involved vertebrae need to be computed. This can be done from the corner points (points 1, 3, 4 and 6) shown in Figure 7.1 (c), of a vertebra. As the spine in the case of scoliosis has large unpredictable curvature and the vertebrae show significant shape variations (often wedged) from image to image, template matching based automatic vertebra segmentation method cannot be applied to extract corner points of the vertebra. Moreover, accuracy of the points selected is an important factor in Cobb angle measurement, which cannot be ensured by automated methods as x-ray images are highly smeared. So, the landmark points are manually selected by the user. The approximate center of the sacrum, the top of the iliac crests and the four corner points of the vertebrae (or that of pedicles) are the landmark points of interest. The points marked on the iliac crests will aid in defining the inter cristal line to draw the central sacral line (CSL). The landmark points of the image were selected using a mouse pointing device on an image displayed on a computer screen. The landmark points marked on an example image are shown in Figure 7.6.

The superior – inferior (S - I) tilt angle is defined as the average of the superior and inferior



Figure 7.6 An example scoliosis image with landmark points marked.

end plate tilt angle of a vertebra with respect to the horizontal axis, and average tilt of the left and right end plate tilt angle is the left – right (L - R) tilt angle of the vertebra with respect to the vertical axis. Given two end points of an end plate, the tilt angle of an end plate is computed as the arc tangent of the ratio of the difference in  $y$ -coordinates to  $x$ -coordinates. The sign convention followed for processing is such that the tilt to the right with respect to its corresponding geometrical axis is taken as positive and otherwise negative. A disc space tilt is the absolute difference of the average S – I tilt angle of the vertebra just above and below the disc.

#### 7.4.2 Rule Based Algorithm for Strategic Vertebra Selection

For measuring different parameters of scoliosis curve, identification of strategic vertebrae is important. The manual selection of strategic vertebra is the major source of error in scoliosis measurement, even in computer assisted methods [18]. A rule based algorithm is developed for the automatic selection of strategic vertebrae of a scoliosis image. The algorithm is based on the following observations.

- A plot of the left, right, superior and inferior end plate tilt angle of the vertebrae will give rise to a biphasic curve (Figure 7.7 (a) – (d)). So, average left –right end plate tilt angle (with respect to vertical axis of the image) and superior – inferior end plate tilt angle (with respect to the horizontal axis of the image) will also give rise to a biphasic curve denoted as L – R curve (Figure 7.8 (a)) and S – I curve (Figure 7.8 (b)) respectively. A biphasic curve is a curve with positive and negative part.
- The zero crossing vertebra (ZCV) - a vertebra where a sign change over takes place, of the above curves corresponds to the apical vertebra (AV) (Figure 7.8 (a) and (b)).
- The immediate neighboring crest (maximum) and trough (minimum) of the AV in the L – R or S – I curve will correspond to the upper end vertebra (UEV) and lower end vertebra (LEV) of the curve respectively (Figure 7.8 (a) and (b)).
- A plot of the distance to central sacral line (CSL) of the mid point of each vertebra will give rise to a biphasic curve and AV will correspond to the maximum distances in each phase (Figure 7.8 (c)).
- A plot of the disc space tilt angle of the vertebrae will give rise to a biphasic curve and its zero crossing vertebra corresponds to the end vertebrae (EV) (Figure 7.8 (d)).



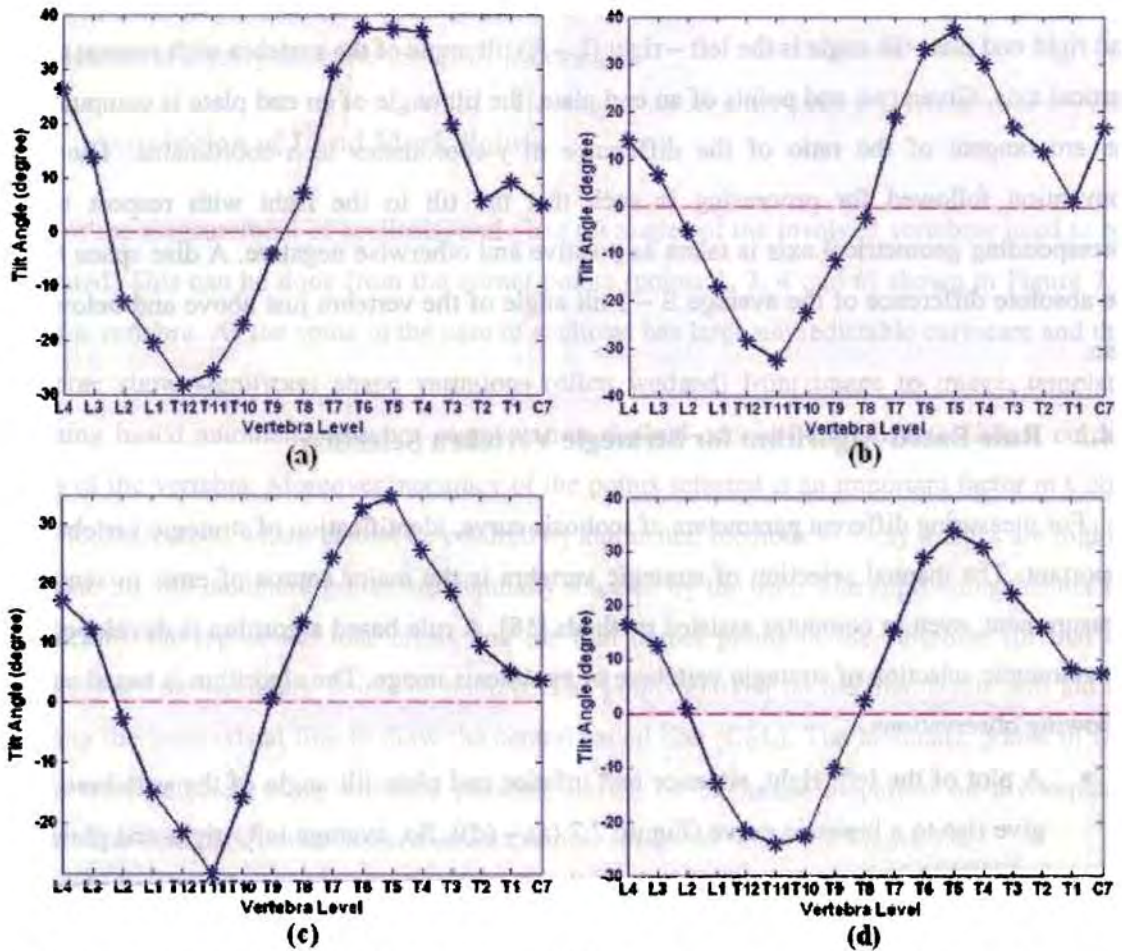


Figure 7.7 (a) Left (b) Right (c) Superior (d) Inferior end plate tilt angle plot of each vertebra of the image shown in Figure 7.6

- Absolute sum of the tilt angle of the end vertebrae is the Cobb angle.

Let us take an example image shown in Figure 7.6 and validate our hypothesis. Figure 7.8 (a) – (d) shows the L-R curve, S-I curve, distance to the CSL and disc space slope of the vertebrae levels from L4 to C7 respectively of a typical scoliosis image. The ground truth information (as plotted by an expert spine surgeon) about this image is that the spine has two major curves – a lumbar curve with apical vertebra at level L2 and end vertebrae T12 & L4, and a thoracic curve with apical vertebra as T8-T9 disc and end vertebrae T5 & T11. From Figure 7.8 it can be noted that the vertebrae of interest derived from observations, matches with that of the ground truth. Our rules for identifying the vertebrae of interest is defined based on the S-I curve (Figure 7.8 (b)) and disc space tilt angle curve (Figure 7.8 (d)). The algorithm for

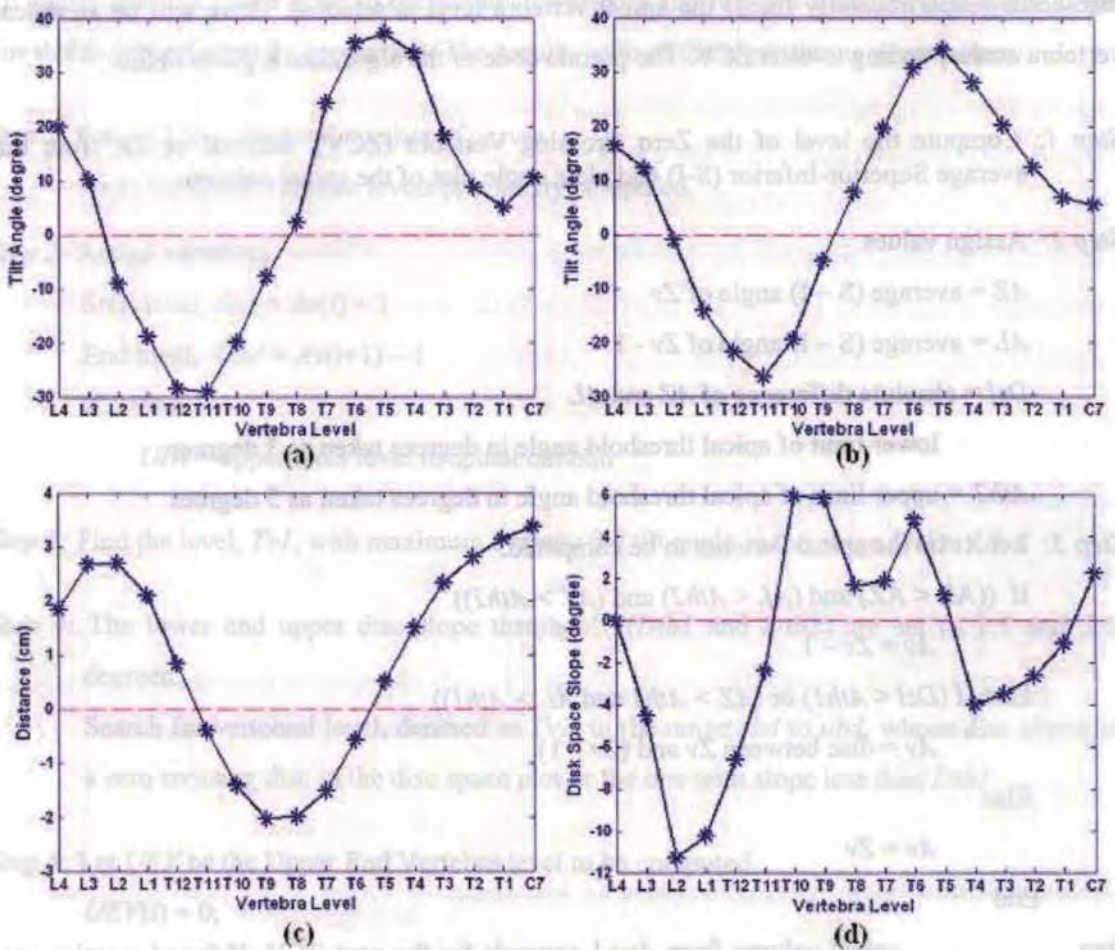


Figure 7.8 (a) Average left - right (L - R) end plate tilt angle plot (b) superior - inferior (S - I) end plate tilt angle plot. (c) Plot of the distance to the CSL of each vertebra (d) Plot of the Disk slope of each vertebra of the image shown in Figure 7.6.

determining the apical vertebra (AV), lower end vertebra (LEV), upper end vertebra (UEV) and other parameters are outlined below.

### 7.4.2.1 Apical Vertebra

The apical vertebra is the apex of a curve. A scoliosis patient can have one or more curves in the spine. The candidate vertebra levels for apical vertebra in the S - I plot is zero crossing vertebra (ZCV) level, one level below ZCV or the disc between these two levels. Based on the relative value of the S - I tilt angle of ZCV and ZCV - 1 level vertebra and two pre-specified

thresholds (experimentally fixed) the apical vertebra level is selected. There will be an apical vertebra corresponding to each ZCV. The pseudo code of the algorithm is given below.

*Step 1:* Compute the level of the Zero Crossing Vertebra (ZCV), denoted as  $Z_v$ , from the average Superior-Inferior (S-I) end plate angle plot of the spinal column.

*Step 2:* Assign values.

$AZ$  = average (S – I) angle of  $Z_v$

$AL$  = average (S – I) angle of  $Z_v - 1$

$DzI$  = absolute difference of  $AZ$  and  $AL$

$Ath1$  = lower limit of apical threshold angle in degrees taken as 3 degrees

$Ath2$  = upper limit of apical threshold angle in degrees taken as 5 degrees

*Step 3:* Let  $A_v$  be the apical vertebra to be computed.

If (( $AL < AZ$ ) and ( $AL < Ath2$ ) and ( $AZ > Ath2$ ))

$A_v = Z_v - 1$

Else if ( $DzI < Ath1$ ) or ( $AZ > Ath1$  and  $AL > Ath1$ )

$A_v$  = disc between  $Z_v$  and ( $Z_v - 1$ )

Else

$A_v = Z_v$

End

*Step 4:* Search the spinal column from  $A_v+1$  onwards for the next ZCV. If found, continue to Step 3 until all ZCV are processed.

#### 7.4.2.2 Upper End Vertebra

Corresponding to each apical vertebra (AV) of a curve, there will be an upper end vertebra (UEV) and lower end vertebra (LEV). First, let us select the upper end vertebra of all curves present in an image. The search for UEV is performed in the range of one level above the AV of a curve to one level below of AV of the immediately above curve. For the top most curve, uppermost visible vertebra is the search limit. The candidate levels for UEV are the level with maximum average tilt angle in the S –I curve and / or a vertebra level whose disc above is parallel (zero crossing disc in the disc space slope plot). The UEV is selected based on the S – I

tilt angles and disc space slope of the candidate vertebra and two pre-specified disc space slope thresholds (experimentally computed). The pseudo code of the algorithm is given below.

*Step 1:* For  $i = 1$  to  $n$ , the total number of curves

$A_v$  is the apical vertebra levels previously computed.

*Step 2:* Assign variables.

Start level,  $Slvl = A_v(i) + 1$

End level,  $Ulvl = A_v(i+1) - 1$

If  $i == n$

$Ulvl =$  upper most level in spinal column

*Step 3:* Find the level,  $Tv1$ , with maximum average S-I tilt angle in the range  $Slvl$  to  $Ulvl$ .

*Step 4:* The lower and upper disc slope threshold ( $Dth1$  and  $Dth2$ ) are set to 1.5 and 3.5 degrees.

Search for vertebral level, denoted as  $Tv2$ , in the range  $slvl$  to  $ulvl$ , whose disc above is a zero crossing disc in the disc space plot or the one with slope less than  $Dth1$ .

*Step 5:* Let  $UEV$  be the Upper End Vertebra level to be computed.

$UEV(i) = 0$ ;

If ( $Tv1 == Tv2$ )

$UEV(i) = Tv1$

Else

$Dsl1 =$  Absolute disc slope of  $Tv1$

$Dsl2 =$  Absolute disc slope of  $Tv2$

If  $\text{abs}(Tv1 - Tv2) == 1$ )

If ( $Tv2 < Tv1$ )

If ( $Dsl1 < Dth2$ )

$UEV(i) = Tv1$ ;

Else

$UEV(i) = Tv2$ ;

End

```

    Else
         $UEV(i) = Tv1;$ 
    End
Else If  $\text{abs}(Tv1 - Tv2) == 2$ 
    If  $(Tv1 > Tv2)$ 
        If  $(Dsl1 < Dth2)$ 
             $UEV(i) = Tv1;$ 
        Else
             $UEV(i) = Tv2;$ 
        End
    Else
         $UEV(i) = Tv2;$ 
    End
Else
     $UEV(i) = Tv1;$ 
End
End

```

*Step 6:* Continue to Step 1 until all  $n$  curves are processed

#### 7.4.2.3 Lower End Vertebra

The search for LEV of curves is carried from top to bottom in the range one level below the apical vertebra to one level above the apical vertebra of the immediately lower curve. The rules are similar to that of UEV. For the lowermost curve the search starts from the first visible vertebra. The disc space slope is computed as the absolute difference in S – I tilt angle of the candidate vertebra and that of the vertebra immediately below.

#### 7.4.3 Measurement

Parameters of interest to a scoliosis image are Cobb angle, apical distance, spinal balance, and pelvic inclination. The definitions of these parameters were given in Section 3.6. Once strategic vertebrae are identified these parameters can be automatically computed.

The Central Sacral Line (CSL) is the line passing through the center of the sacrum and perpendicular to the line joining the tangents of the iliac crests [27]. This is based on the assumption that the leg lengths are equal and the lateral edge of the radiograph was used to define the vertical axis. Let us formulate the equation of the CSL which can be later used for computing the distance of the center of each vertebra to the CSL.

Let  $(x_1, y_1)$  be the top-mid of the left iliac crest and  $(x_2, y_2)$  the corresponding point of the right iliac crest. Figure 3.4 details the standard terminology. The center of sacrum is represented by  $(x_3, y_3)$ . So, CSL will be the line passing through  $(x_3, y_3)$  and perpendicular to the line joining  $(x_1, y_1)$  and  $(x_2, y_2)$ .

Let  $m_p$  be the slope of the line joining  $(x_1, y_1)$  and  $(x_2, y_2)$ ,

$$m_p = \frac{y_2 - y_1}{x_2 - x_1} \quad (7.9)$$

Any point  $(x, y)$  on the line perpendicular to (7.9) and passing through  $(x_3, y_3)$  is defined as

$$\frac{y - y_3}{x - x_3} = \frac{-1}{m_p} \quad (7.10)$$

Equation (7.10) can be rearranged as

$$x + m_p * y - m_p y_3 - x_3 = 0 \quad (7.11)$$

The apical distance is calculated as the perpendicular distance of the centroid (intersecting of the lines connecting the superior lateral corners of the vertebral body to the contralateral inferior lateral corners) of the apical vertebra (AV) to the CSL [27]. The center of the AV,  $(x_4, y_4)$ , is taken as the average of the corner points of the AV. The distance of  $(x_4, y_4)$  to CSL is defined as

$$d_a = \frac{x_4 + m_p * y_4 - m_p * y_3 - x_3}{\sqrt{1 + m_p^2}} \quad (7.12)$$

Spinal balance (or coronal balance) is the horizontal distance between the sagittal C7 plumbline dropped vertically from C7 centroid to the central sacral line [27], which is computed in a similar way. The number of pixels is converted to distance in cm based on the resolution of the image. The pelvic inclination is the measure of angle made by the line joining the top of the iliac crests to the horizontal axis [27]. The Cobb angle of a curve can be calculated as the absolute sum of the tilt angle of the inferior end plate of lower end vertebra and superior end plate of the upper end vertebra.

#### 7.4.4 A New Classification Scheme

A classification scheme is important to determine the prognosis and treatment of scoliosis. Important scoliosis classification schemes using AP view images are Ponsetty and Friedman scheme [28] and King *et. al.* scheme [29]. The new classification scheme is based on Ponsetty and Friedman scheme and details of which was outlined in Section 3.8.1. Adolescent idiopathic scoliosis is classified into five different curve types based on the convexity, location and curve magnitude [28]. But, the spinal balance is also found to be an important factor in assigning therapeutic strategy. To include all other curve types like infantile, juvenile, non idiopathic curves etc and to consider spinal balance, two additional curve types are defined. The newly assigned different curve types are

- Type I: Single Right Thoracic;
- Type II: Right Thoraco-Lumbar,
- Type III: Left Lumbar,
- Type IV: Double Major with Right Thoracic,
- Type V: Double Thoracic,
- Type VI: Others, triple curves, and cervicothoracic curves (with low de-compensation; < 2.5 cm),
- Type VII: All Others (with moderate or high de-compensation; > 2.5 cm).

Curve classification is implemented using a set of rules based on parameters such as curve convexity (left or right), number of curves, location of curve (lumbar, thoracic etc) and curve magnitude.

### 7.4.5 Feature Descriptor and Distance Measure

The feature descriptor designed to represent the scoliosis image for CBIR is described here. The set of features considered important for a scoliosis image are the curve type, number of curves, convexity of curve, Cobb angle, spinal balance, AV levels, apical distances, levels of UEV and LEV, average Superior - Inferior end plate tilt angle of T1, average Superior - Inferior end plate tilt angle of L4 and pelvic inclination. In a large database, reducing the result set according to clinical interest of a query image is necessary. This filtering can be achieved by using either curve type or number of curves feature, depending on whether the user is interested in the sorted images based on a classification scheme or not. The number of entries for the features Cobb angle, AV levels, apical distances and levels of UEV and LEV will correspond to the number of curves in an image. There will be a total of 12 features for a single curve and 17 features for a double curve.

The Euclidean distance is used as the distance measure to compute distance between the query image and images in the database. The Euclidean distance is defined as

$$L_2(f_1, f_2) = \sqrt{\sum (f_1(i) - f_2(i))^2}; \quad \forall i \quad (7.13)$$

where  $f_1$  is feature vector of query image and  $f_2$  that of images in the database,  $i$  is the number of elements of the feature vector.

### 7.4.6 Query Design

The most frequent query paradigm for CBIR is Query-By-Example, in which the user supplies a query image and the system returns images in the database that are similar to the query image based on visual characteristics. For scoliosis images clinically, images similar to a query image are those images that have similar treatment strategy. The similarity (or dissimilarity) is quantified as the distance of the feature descriptor of the query image to that of the images in the database. The indexing process automatically extracts a set of features (feature descriptor) from each stored image in the database and stores in a table of a relational database.

Two types of query schemes are designed for the CBIR of scoliosis images. First scheme is for retrieving images of a specific curve type in a sorted order. All records matching the curve



type are initially fetched and then ordered according to the convexity of curves and number of curves. The results are further ordered based on the Euclidean distance of the other 14 features (for a double curve) of query image with that of database images. This method is intended for users who are particular in using a classification scheme, even though this method has the limitation of ambiguity in selecting a classification scheme. The second scheme is for retrieving images matching the number of curves instead of curve type. The initial filtering is based on the number of curves feature and ordered by the convexity of the curve. Further ordering is done based on the Euclidean distance of the other 14 features. This method eliminates the variability due to the ambiguity in selecting a classification scheme.

The database schema consists of a single table with the following attributes. For the case of images with double curve, the table attributes are

<u>Field Name</u>	<u>Type</u>	<u>Comments</u>
m_PID	INTEGER	- unique patient identifier (primary key)
m_Type	SMALLINT	- curve type
m_NoCurves	SMALLINT	- Number of curves
m_conv1	NUMBER(1)	- convexity of first curve (left or right)
m_Av1	FLOAT	- level of AV of first curve
m_Av2	FLOAT	- level of AV of second curve
m_Ad1	FLOAT	- apical distance of first curve
m_Ad2	FLOAT	- apical distance of second curve
m_Lev1	SMALLINT	- level of LEV of first curve
m_Lev2	SMALLINT	- level of LEV of second curve
m_Uev1	SMALLINT	- level of UEV of first curve
m_Uev2	SMALLINT	- level of LEV of second curve
m_Cobb1	FLOAT	- Cobb angle of first curve
m_Cobb2	FLOAT	- Cobb angle of second curve
m_Simbalance	FLOAT	- spinal imbalance
m_T1tilt	FLOAT	- tilt of T1 vertebra
m_L4tilt	FLOAT	- tilt of L4 vertebra
m_Apelvis	FLOAT	- Pelvic inclination
m_Image	BLOB	- Image file

The SQL syntax for both the query schemes is given below.

The query using a classification scheme for the table 'apview' is given as,

```
SELECT * FROM apview T1,
  (SELECT * FROM apview WHERE m_PID = 306) T2
  WHERE T1.m_Type = T2.m_Type
  ORDER BY abs(T2.m_cur1 - T1.m_cur1),
           abs(T2.m_NoCurves - T1.m_NoCurves),
           SQRT(POWER(abs(T2.m_cobb1 - T1.m_cobb1),2) +
                POWER(abs(T2.m_cobb2 - T1.m_cobb2),2)+
                POWER(abs(T2.m_simbalance - T1.m_simbalance),2) +
                POWER(abs(T2.m_ad1 - T1.m_ad1), 2) +
                POWER(abs(T2.m_ad2 - T1.m_ad2), 2) +
                POWER(abs(T2.m_apelvis - T1.m_apelvis), 2) +
                POWER(abs(T2.m_t1tilt - T1.m_t1tilt), 2) +
                POWER(abs(T2.m_l4tilt - T1.m_l4tilt), 2) +
                POWER(abs(T2.m_av1 - T1.m_av1), 2) +
                POWER(abs(T2.m_av2 - T1.m_av2), 2) +
                POWER(abs(T2.m_lev1 - T1.m_lev1), 2) +
                POWER(abs(T2.m_lev2 - T1.m_lev2), 2) +
                POWER(abs(T2.m_uev1 - T1.m_uev1), 2) +
                POWER(abs(T2.m_uev2 - T1.m_uev2), 2));
```

The query without using a classification scheme is given as,

```
SELECT * FROM apview T1,
  (SELECT * FROM apview WHERE m_PID = 306) T2
  WHERE T1.m_NoCurves = T2.m_NoCurves
  ORDER BY abs(T2.m_cur1 - T1.m_cur1),
           SQRT(POWER(abs(T2.m_cobb1 - T1.m_cobb1),2) +
                POWER(abs(T2.m_cobb2 - T1.m_cobb2),2)+
```

```

POWER(abs(T2.m_s imbalance - T1.m_s imbalance),2) +
POWER(abs(T2.m_ad1 - T1.m_ad1), 2) +
POWER(abs(T2.m_ad2 - T1.m_ad2), 2) +
POWER(abs(T2.m_apelvis - T1.m_apelvis), 2) +
POWER(abs(T2.m_t1tilt - T1.m_t1tilt), 2) +
POWER(abs(T2.m_l4tilt - T1.m_l4tilt), 2) +
POWER(abs(T2.m_av1 - T1.m_av1), 2) +
POWER(abs(T2.m_av2 - T1.m_av2), 2) +
POWER(abs(T2.m_lev1 - T1.m_lev1), 2) +
POWER(abs(T2.m_lev2 - T1.m_lev2), 2) +
POWER(abs(T2.m_uev1 - T1.m_uev1), 2) +
POWER(abs(T2.m_uev2 - T1.m_uev2), 2));

```

The query plan generated for the second query is given below.

---

```

SQL OPTIMIZER OUTPUT:
SELECT COMMAND.
PROJECT [337] (
| SORT [336] IN TEMP2(
| | PROJECT [295] (
| | | JOIN [2][NESTED_LOOP-JOIN](
| | | | PROJECT [255] (
| | | | | PROJECT [341] (
| | | | | | dharma.apview. [342](
| | | | | | | INDEX SCAN OF (
| | | | | | | | sys_001_000001006,
| | | | | | | | | (apview.m_pid) = (306)
| | | | | | | )
| | | | | | )
| | | | | , apview.m_pid
| | | | | , apview.m_type
| | | | | , apview.m_nocurves
| | | | | , apview.m_curl
| | | | | , apview.m_av1
| | | | | , apview.m_av2
| | | | | , apview.m_ad1
| | | | | , apview.m_ad2
| | | | | , apview.m_lev1
| | | | | , apview.m_lev2
| | | | | , apview.m_uev1
| | | | | , apview.m_uev2

```

```

| | | | | , apview.m_cobb1
| | | | | , apview.m_cobb2
| | | | | , apview.m_simbalance
| | | | | , apview.m_t1tilt
| | | | | , apview.m_l4tilt
| | | | | , apview.m_apelvis
| | | | | )
| | | | | , PEXPR1
| | | | | , PEXPR2
| | | | | , PEXPR3
| | | | | , PEXPR4
| | | | | , PEXPR5
| | | | | , PEXPR6
| | | | | , PEXPR7
| | | | | , PEXPR8
| | | | | , PEXPR9
| | | | | , PEXPR10
| | | | | , PEXPR11
| | | | | , PEXPR12
| | | | | , PEXPR13
| | | | | , PEXPR14
| | | | | , PEXPR15
| | | | | , PEXPR16
| | | | | , PEXPR17
| | | | | , PEXPR18
| | | | | , PEXPR18
| | | | | )
| | | | | ,
| | | | | (PEXPR3) = (PEXPR22)
| | | | | ,
| | | | | PROJECT [235] (
| | | | | | dharma.t1. [0](
| | | | | | | TABLE SCAN
| | | | | | )
| | | | | , apview.m_pid
| | | | | , apview.m_type
| | | | | , apview.m_nocurves
| | | | | , apview.m_curl
| | | | | , apview.m_av1
| | | | | , apview.m_av2
| | | | | , apview.m_ad1
| | | | | , apview.m_ad2
| | | | | , apview.m_lev1
| | | | | , apview.m_lev2
| | | | | , apview.m_uev1
| | | | | , apview.m_uev2
| | | | | , apview.m_cobb1

```

```

| | | | , apview.m_cobb2
| | | | , apview.m_simbalance
| | | | , apview.m_t1tilt
| | | | , apview.m_l4tilt
| | | | , apview.m_apelvis
| | | | , apview.rowid
| | | | )
| | | | )
| | | | , PEXPR20
| | | | , PEXPR21
| | | | , PEXPR22
| | | | , PEXPR23
| | | | , PEXPR24
| | | | , PEXPR25
| | | | , PEXPR26
| | | | , PEXPR27
| | | | , PEXPR28
| | | | , PEXPR29
| | | | , PEXPR30
| | | | , PEXPR31
| | | | , PEXPR32
| | | | , PEXPR33
| | | | , PEXPR34
| | | | , PEXPR35
| | | | , PEXPR36
| | | | , PEXPR37
| | | | , PEXPR1
| | | | , PEXPR2
| | | | , PEXPR3
| | | | , PEXPR4
| | | | , PEXPR5
| | | | , PEXPR6
| | | | , PEXPR7
| | | | , PEXPR8
| | | | , PEXPR9
| | | | , PEXPR10
| | | | , PEXPR11
| | | | , PEXPR12
| | | | , PEXPR13
| | | | , PEXPR14
| | | | , PEXPR15
| | | | , PEXPR16
| | | | , PEXPR17
| | | | , PEXPR18
| | | | , PEXPR38
| | | | , PEXPR19
| | | | , abs ((PEXPR4) - (PEXPR23))

```

```

| | , sqrt (((((((((((((power (abs ((PEXPR13) - (PEXPR32)),2)) + (power (abs ((PEXPR14) -
| (PEXPR33)),2))) + (power (abs ((PEXPR15) - (PEXPR34)),2))) + (power (abs ((PEXPR7) -
| (PEXPR26)),2))) + (power (abs ((PEXPR8) - (PEXPR27)),2))) + (power (abs ((PEXPR18) -
| (PEXPR37)),2))) + (power (abs ((PEXPR16) - (PEXPR35)),2))) + (power (abs ((PEXPR17) -
| (PEXPR36)),2))) + (power (abs ((PEXPR5) - (PEXPR24)),2))) + (power (abs ((PEXPR6) -
| (PEXPR25)),2))) + (power (abs ((PEXPR9) - (PEXPR28)),2))) + (power (abs ((PEXPR10) -
| (PEXPR29)),2))) + (power (abs ((PEXPR11) - (PEXPR30)),2))) + (power (abs ((PEXPR12) -
| (PEXPR31)),2)))
| | )
| SORT BY (
| | , PEXPR39
| | , PEXPR40
| ),
| , PEXPR1
| , PEXPR2
| , PEXPR3
| , PEXPR4
| , PEXPR5
| , PEXPR6
| , PEXPR7
| , PEXPR8
| , PEXPR9
| , PEXPR10
| , PEXPR11
| , PEXPR12
| , PEXPR13
| , PEXPR14
| , PEXPR15
| , PEXPR16
| , PEXPR17
| , PEXPR18
| , PEXPR19
| , PEXPR20
| , PEXPR21
| , PEXPR22
| , PEXPR23
| , PEXPR24
| , PEXPR25
| , PEXPR26
| , PEXPR27
| , PEXPR28
| , PEXPR29
| , PEXPR30
| , PEXPR31
| , PEXPR32
| , PEXPR33
| , PEXPR34

```

```
| , PEXPR35  
| , PEXPR36  
| , PEXPR37  
| , PEXPR38  
| , PEXPR39  
| , PEXPR40  
| )  
, PEXPR1  
, PEXPR2  
, PEXPR3  
, PEXPR4  
, PEXPR5  
, PEXPR6  
, PEXPR7  
, PEXPR8  
, PEXPR9  
, PEXPR10  
, PEXPR11  
, PEXPR12  
, PEXPR13  
, PEXPR14  
, PEXPR15  
, PEXPR16  
, PEXPR17  
, PEXPR18  
, PEXPR19  
, PEXPR20  
, PEXPR21  
, PEXPR22  
, PEXPR23  
, PEXPR24  
, PEXPR25  
, PEXPR26  
, PEXPR27  
, PEXPR28  
, PEXPR29  
, PEXPR30  
, PEXPR31  
, PEXPR32  
, PEXPR33  
, PEXPR34  
, PEXPR35  
, PEXPR36  
)
```

---

The query plan for the first query is similar, except for the filter predicate. It can be noted that the designed query is optimized for performance as the join between ‘apview’ table and its derived table is Nested – Loop Join.

#### 7.4.7 Experimental Setup

The experiment is carried out using 20 standing idiopathic A-P scoliosis images (including both DICOM and TIFF format). The ground truth information (correct curve apex, end vertebrae and Cobb angle) is generated by an expert spine surgeon analyzing the images using a computer-assisted measurement tool called Amrita MedVision [<http://www.amritatech.com/medvisonradiology.htm>]. The Cobb angle is measured by placing lines along the superior end plate of the superior vertebral body and along the inferior end plate of the inferior vertebral body of the manually identified end vertebrae. The 20 images used for evaluation contains 30 curves (10 double curve images). The curve pattern classification in the ground truth data is based on the newly proposed classification technique. Among these patients, there are 10 cases with Type I curve pattern, 1 with Type III, 4 with Type IV, 1 with Type V, 3 with Type VI and 1 with Type VII. The Cobb angle value ranged between 33 ° and 92 °. The tolerance range of error for Cobb angle is taken to be 3 ° for the measured angle with respect to the ground truth, as this is well within the range of reported intraobserver measurement error range.

The landmark points are acquired by marking twice (at an interval of seven days) each image by three observers (two orthopedic surgeons and one non-clinician researcher) taking the images in random order. The feature descriptor is computed and inserted into the table of a relational database management system (RDBMS) with the patient ID (for our test images same as image name) as a primary key using an open database connectivity (ODBC) application.

The aims of the experiment are given below.

- Determine vertebrae of interest and measure Cobb angle automatically using the landmark points input by the three users at two different times independently.
- Compute the average error and standard deviation of these six sets of measurement data with respect to the ground truth data.
- Assign a curve type based on new classification scheme.
- Compute the inter and intra observer reliability in measurement.



- Index the images using the feature descriptor detailed and conduct Query-By-Example test to evaluate the accuracy of CBIR system.

#### 7.4.7.1 Statistical Analysis

The measurement error in Cobb angle is computed as the average of the difference and standard deviation of the measured data with the ground truth data. The interobserver and intraobserver error analysis is carried out using statistical tools called Kappa Statistics [30] and Pearson correlation. Kappa is an index which compares the agreement against that which might be expected by chance, with possible value range from +1 (perfect agreement) via 0 (no agreement above that expected by chance) to -1 (complete disagreement). In the case of correlation coefficient, a value of -1 indicates a perfect inverse correlation between the variables (that is, when the value for one variable is high, the value for the other variable is low). A value of 0 indicates no relationship between the variables. A value of 1 indicates a perfect correlation between the variables. The details of Kappa Statistics and Pearson correlation were outlined in Section 4.5. The cases of disagreement for Cobb angle is selected as the ones whose difference with ground truth was more than a tolerance range of 3°. For AV and EV a difference in level is treated as disagreement. Depending on the resolution of the image used there is an intrinsic error associated with Cobb angle measurement. In this study, this intrinsic error associated with angular measurement is taken as 0.5 degrees. So, all angle measurements are rounded off to account for this measurement error.

#### 7.4.8 Results

Table 7.2 lists the number of erroneous measurements, average error and maximum error of the apical vertebra level, end vertebrae level and Cobb angle of the measurement data with the ground truth data of the curves under investigation. Out of 30 apical vertebrae, 4 cases differed in all the measurements. The difference in level is half level for all the 4 cases (the disc adjacent to the vertebra in question). Out of the 60 end vertebrae (total count of LEV and UEV), a maximum of 8 cases showed a difference in level with average and maximum difference of 1 level. The number of curves for which the Cobb angle differed more than the tolerance range of

**Table 7.2. Measurement error in apex level, end vertebra level and Cobb angle**

Parameter	Observer	Trial	AV	EV	Cobb
No. of Erroneous Measurements	Observer 1	Trial 1	4	8	6
		Trial 2	4	8	7
	Observer 2	Trial 1	4	7	7
		Trial 2	4	8	8
	Observer 3	Trial 1	4	6	8
		Trial 2	4	7	6
Average Error	Observer 1	Trial 1	Adjacent Disc (.5)	1	2 degree
		Trial 2	Adjacent Disc (.5)	1	2 degree
	Observer 2	Trial 1	Adjacent Disc (.5)	1	2 degree
		Trial 2	Adjacent Disc (.5)	1	2 degree
	Observer 3	Trial 1	Adjacent Disc (.5)	1	2 degree
		Trial 2	Adjacent Disc (.5)	1	2 degree
Maximum Difference	Observer 1	Trial 1	Adjacent Disc (.5)	1	5 degree
		Trial 2	Adjacent Disc (.5)	1	6 degree
	Observer 2	Trial 1	Adjacent Disc (.5)	1	6 degree
		Trial 2	Adjacent Disc (.5)	1	6 degree
	Observer 3	Trial 1	Adjacent Disc (.5)	1	5 degree
		Trial 2	Adjacent Disc (.5)	1	6 degree

3 degrees from the ground truth varied from 6 to 8 for various measurements and the maximum difference in Cobb angle is 6 degrees. The average error in Cobb angle is 2.0 ° with a standard deviation of 3.0 °. The Cobb angle difference is within tolerable limit for some cases even though their end vertebra levels differed from ground truth. The average Cobb angle error for whose end vertebra level is different from the ground truth is 5 degrees. The curve types of the images automatically identified by the rule based program are matching to that plot ground truth data by an expert spine surgeon manually. An average sample standard deviation of 2.0 ° is obtained between intraobserver measurements and a value of 4.0 ° for interobserver

Table 7.3 Correlation coefficient results

		Observer 1		Observer 2		Observer 3		
		Trial 1	Trial 2	Trial 1	Trial 2	Trial 1	Trial 2	
GT	AV	0.998	0.998	0.998	0.998	0.998	0.998	
	EV	0.997	0.996	0.995	0.997	0.996	0.995	
	Cobb	0.980	0.982	0.982	0.985	0.984	0.982	
Observer1	Trial 1	AV	1	1	0.999	0.999	0.998	0.998
		EV	1	0.998	0.995	0.997	0.995	0.996
		Cobb	1	0.990	0.973	0.980	0.975	0.974
	Trial 2	AV	1	1	0.999	0.999	0.999	0.999
		EV	0.998	1	0.995	0.996	0.994	0.995
		Cobb	0.989	1	0.974	0.979	0.977	0.973
Observer2	Trial 1	AV	0.999	0.999	1	1	0.998	0.999
		EV	0.995	0.995	1	0.995	0.994	0.994
		Cobb	0.975	0.975	1	0.984	0.977	0.975
	Trial 2	AV	0.999	0.999	1	1	0.998	0.998
		EV	0.997	0.996	0.995	1	0.991	0.993
		Cobb	0.977	0.977	0.979	1	0.969	0.970
Observer3	Trial 1	AV	0.998	0.999	0.998	0.998	1	1
		EV	0.995	0.994	0.994	0.991	1	0.994
		Cobb	0.972	0.975	0.972	0.969	1	0.975

repeated observations. The average intraobserver standard deviation (SD) has been reported as 3.5 ° and interobserver SD range from 2.8 ° to 7.2 ° in the literature [12 – 16]. The accuracy of the developed technique is better compared to the reported intraobserver measurement error ranges.

Table 7.4 Kappa statistics results

		Observer 1		Observer 2		Observer 3		
		Trial 1	Trial 2	Trial 1	Trial 2	Trial 1	Trial 2	
GT		AV	0.87	0.88	0.88	0.85	0.78	0.81
		EV	0.69	0.73	0.65	0.71	0.66	0.68
		Cobb	0.73	0.75	0.72	0.76	0.71	0.72
Observer1	Trial 1	AV	1	1	0.91	0.90	0.90	0.92
		EV	1	0.85	0.78	0.77	0.79	0.76
		Cobb	1	0.89	0.83	0.80	0.75	0.74
	Trial 2	AV	1	1	0.90	0.93	0.91	0.93
		EV	0.85	1	0.75	0.76	0.76	0.78
		Cobb	0.89	1	0.74	0.79	0.77	0.78
Observer2	Trial 1	AV	0.91	0.90	1	1	0.88	0.89
		EV	0.79	0.82	1	0.85	0.78	0.79
		Cobb	0.75	0.71	1	0.84	0.77	0.75
	Trial 2	AV	0.89	0.89	1	1	0.88	0.82
		EV	0.77	0.76	0.85	1	0.76	0.73
		Cobb	0.77	0.79	0.83	1	0.79	0.80
Observer3	Trial 1	AV	0.88	0.89	0.90	0.89	1	1
		EV	0.75	0.74	0.73	0.71	1	0.84
		Cobb	0.72	0.75	0.72	0.79	1	0.85

Table 7.3 details the correlation coefficient values obtained for various measurements against ground truth data and also between each other. A high value of correlation coefficient obtained shows good correlation of the measured data with the ground truth data. High correlation coefficient values obtained for measurement of an observer against another observer

indicate good interobserver reliability of the system. Regarding curve pattern type, all the curves are classified into correct curve type in all measurements with a perfect agreement score of 1 (not shown in Table 7.3). Similarly, the correlation coefficient shows high value for repeated measurements of the same observer indicating good intraobserver repeatability also.

The overall reliability and repeatability of the developed Cobb angle measurement scheme is quantified by empirical study based on Kappa statistics. The results are listed in Table 7.4. Comparative high positive value obtained for the measurements of AV level, EV level and Cobb angle by the three observers indicates that measurement values closely agree. Significant positive values obtained for measurement of an observer against another observer indicate good interobserver reliability of the system. Regarding curve pattern type, all the curves are classified into correct curve type with a perfect agreement score of +1 (not shown in Table 7.4). Similarly, the Kappa coefficient shows positive value for repeated measurements of the same observer indicating good intraobserver repeatability.

The CBIR system is evaluated by Query-By-Example test using two types of queries: one is filtering the result based on curve type and the other is based on the number of curves in an image. Five images are used as query image. In the first category of query, the system retrieved correct set of images of the query image type in the decreasing order of similarity (more similar first), for all the query cases. Figure 7.9 (a) shows an example result of a query using a Type IV image. Only four Type IV images are present in the database and all of them are returned correctly in the increasing order of distance to the query image. The same query images are used for retrieval in the second category of query scheme. The system retrieved images similar to query image with the same number of curves in a sorted manner. Figure 7.9 (b) shows an example query result with the first 10 images retrieved. It can be noted from Figure 7.9 (b) that, the first four images are of same type (Type IV) as the query image. Rest of the images have similar number of curves, and closer apex and end vertebra levels, Cobb angle etc (from ground truth information) to the query image in the increasing order of distance. The first image in each set (Figure 7.9 (a) and Figure 7.9 (b)) is the query image.

#### **7.4.9 Discussion**

General purpose medical CBIR systems such as MedGIFT uses global image features like general shape, color and texture [31]. Usage of such features will aid only in a general

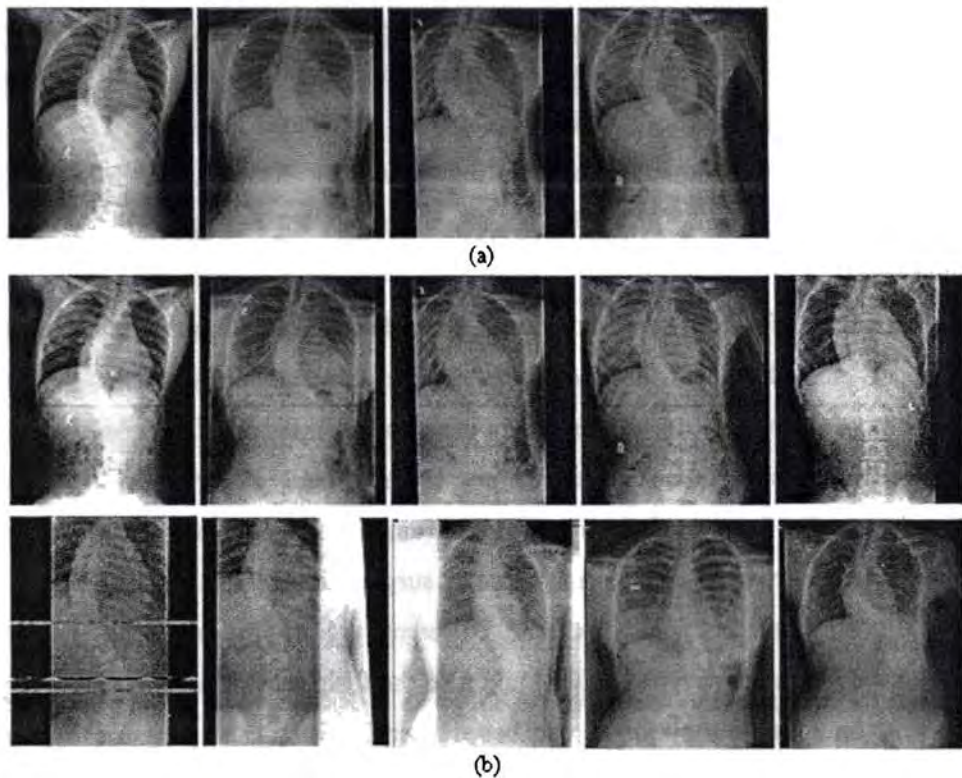


Figure 7.9 CBIR test results. First image in each set is the query image. (a) Retrieval results based on curve type (b) Retrieval results based on number of curves

classification, for example, to group x-ray images of spine and skull separately. But, the real application of CBIR lies in identifying similar images based on the pathology of interest. For that, local features specific to the pathology need to be identified and used as feature descriptors. In the developed method, a set of features specific to the pathology scoliosis is identified and used as feature descriptors. They were automatically derived from a given set of landmark points.

Nachiappan *et al.* [32] proposed a scoliosis measurement system in which the observer marks the lateral and medial intersecting points and the maximum angle is taken as the Cobb angle. The system does not identify the end vertebrae and multiple curves are not taken care of. In comparison with the scoliosis measurement system proposed by Stokes *et al.* [13, 33], the newly developed system is more accurate as the CSL is computed with respect to the spinal coordinates (level of iliac crests) than the global coordinates (lateral and horizontal edges of radiographs). Also, the rules for identifying the vertebrae of interest were formulated based on various parameters instead of just maximum end plate tilt angle or apical distance.

The proposed algorithmic approach could be applied to the sagittal plane deformities such as lordosis and kyphosis which are also measured using Cobb technique with lateral view radiographs. Pathologies of lordosis and kyphosis were introduced in Section 3.9. The system can support other classification schemes like King *et al.* [29], Peking Union Medical College (PUMC) method [34], Lenke *et al.* [35], SRS classification [36] etc by incorporating sagittal plane profile also.

Since many classification schemes are in use today, there is ambiguity in selecting a classification scheme. They are also not interoperable. Classification systems are not fully reliable because of the variability that exists in curve classification [37]. Large variations are present among scoliosis images. A classification scheme broadly categorizes the images into five or six classes. So, large variations are present among images identified as similar by a classification scheme. Now in scoliosis treatment, surgical strategy is decided based on the curve type assigned by a classification scheme. Different surgical approaches for correction are in use today for a given curve type [34, 38] and also the fusion levels selected may vary from surgeon to surgeon. The only option in text based querying is to use a classification scheme, which has the before mentioned disadvantages. To resolve them, the developed system has a query scheme to retrieve similar images without using any classification scheme at all (second query scheme). In the proposed hybrid CBIR system (text and visual information), treatment procedure adopted and results obtained can be recorded, along with preoperative and postoperative images. Thus, a surgeon can search for similar previous cases in the database and assess the outcomes of various techniques for a given case in hand to decide surgical strategy, making the classification schemes obsolete. If a large database of pre and post-operative scoliosis images could be maintained, better diagnosis and treatment suggestions can be provided by an expert CBIR system. For those who are particular in using a classification scheme, a query scheme (first query scheme) is provided which helps in sorting images within a class type based on similarity, addressing the issue of large variations of images. The developed system provides additional important information for instrumentation planning such as distance to CSL of each vertebra, end plate tilt angle, pelvic inclination, vertebral body height, intervertebral disc space height etc. The proposed rule-based measurement system has the advantage that the scoliosis measurement can be carried out by any person who is trained in identifying the landmark points. The flexibility of the user makes it useful for applications like

large scale scoliosis screening, in which the most prohibitive part is the cost of a spine expert [39]. The developed system has also got applications in research and education for navigating large image databases.

## 7.5 SpineSearch – The CBIR Software System

A software system called ‘SpineSearch’ is developed based on the algorithm developed for CBIR of scoliosis images. The developed system provides graphical user interface (GUI) to measure the Cobb angle and facilitates content based search for similar images in the database, based on Query-By-Example paradigm. The design and graphical user interface of the software are discussed in this section.

### 7.5.1 Design

The ‘SpineSearch’ was designed using object oriented programming system (OOPS) approach. The GUI interface and functional modules of the system was developed using Microsoft Visual C++ (<http://www.microsoft.com>) and Dharma SQL Server (<http://www.dharma.com>) is used as the relational database management system (RDBMS) engine in Windows XP platform. The database connectivity is implemented using open database

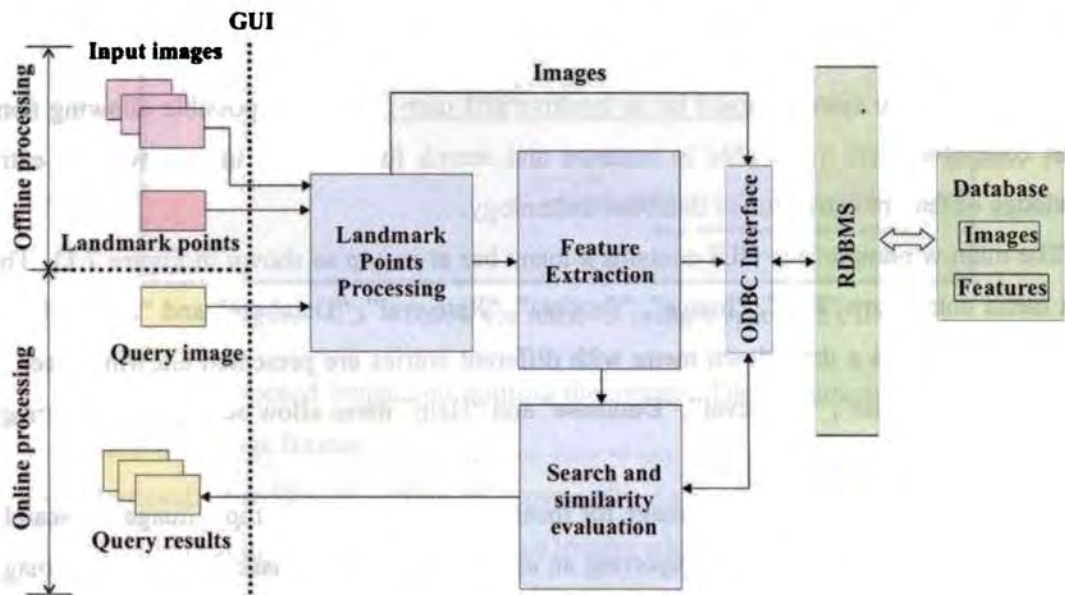


Figure 7.10 Block diagram of SpineSearch CBIR system



connectivity (ODBC) interface; hence the system could support any SQL engine like Oracle as the back end.

The block diagram of the ‘SpineSearch’ CBIR system is given in Figure 7.10. A basic module “Landmark Points Processing” works on the landmark points input by the user, by marking them on the scoliosis image. It computes the tilt angle of each end plate of the vertebra and other distance and angle parameters. This module is used while offline processing for database population and online processing for search of similar images (if the query image is a new one which is not present in the database). The ‘Feature Extraction’ block forms the feature descriptor from the attributes generated by the “Landmark Points Processing” block. The feature descriptor used were outlined in Section 7.4.4 The “Search and Similarity Evaluation” block uses the feature descriptors and does the distance calculation between the query image and images in the database. This block projects the query result images to the user. All the three blocks interact with DBMS engine to store and retrieve images and features using an ODBC interface block. As the DBMS engine is interfaced through ODBC, the architecture of ‘SpineSearch’ can be easily extended to client-server model and a centralized multi-center database can be maintained to improve the powerfulness of the CBIR system. Proper message handling is ensured to propagate messages or error information by the SQL engine, to be displayed to the user, to facilitate tracking.

### **7.5.2 Graphical User Interface**

The GUI of the system should be as intuitive and user friendly as possible allowing non-expert computer users to be able to measure and search for scoliosis images without extra knowledge of feature extraction or database technology.

The main window of the GUI contains a menu bar at its top as shown in Figure 7.11. The main menu entries are “File”, “Image”, “Process”, “Retrieval”, “Database” and “Help”. From each of these entries a drop down menu with different entries are presented allowing a set of actions. But only “File”, “Retrieval”, “Database” and “Help” items allow actions until an image is opened.

- **File:** Allows opening of an image for measurement or as an example image for search. This menu has options for importing an already marked landmark points on the image or exporting landmark points to a file for later use. Also, it has provision for closing and



Figure 7.11 Main window of the GUI of the SpineSearch

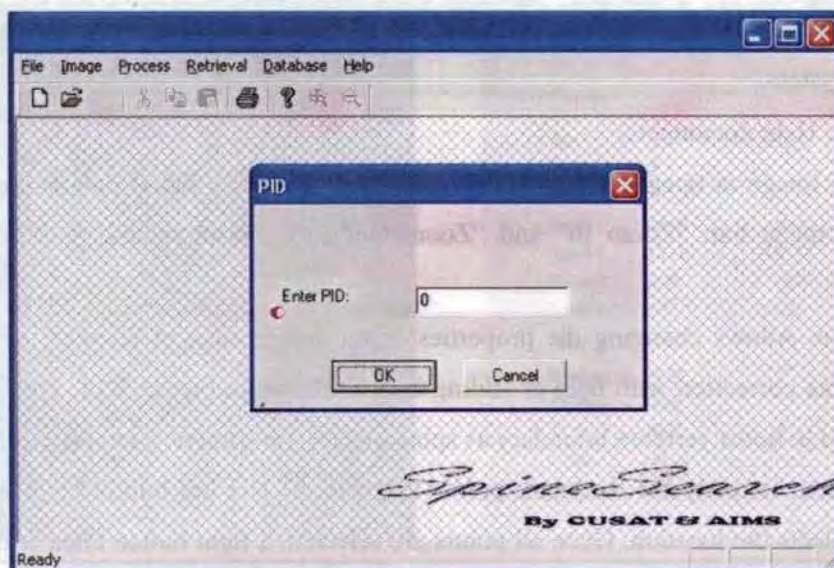


Figure 7.12 Window for retrieval using a specified PID

printing the opened image and quitting the system. The system supports both DICOM and TIFF image format.

- Retrieval: Provides an option to browse the database by specifying a PID (patient identifier) unique to image and retrieve images whose content is matching to the content of the image specified by PID. A window for the same is shown in Figure 7.12.



Figure 7.13 Interface to adjust image properties

- Database: Initiates database schema creation. Need to be done only once at first use of the system.
- Help: Help documents.

Once an image is opened using “File” menu, other already listed options and two more icons in the menu bar, “Zoom In” and “Zoom Out”, to control scaling gets activated. The option.

- Image: Allows changing the properties of the image such as contrast, brightness and gamma correction with help of sliding bars as shown in Figure 7.13. These options are useful to boost vertebra boundary as spine images are sometimes highly smeared.

Now the system is ready for landmark points acquisition. The left button click of the mouse marks and selects the location. Once all points are selected, a right button click will acquire the points. Figure 7.14 shows the interface with landmark points marked as red spots.

The “Process” option in the menu bar is now ready to use. The drop down list have the “Cobb”, “Insert”, “Update”, “Show” and “Chart” options.

- Cobb: Computes the Cobb angle and feature descriptor of the curve. The result is displayed in a window as shown in Figure 7.15. A miniature version of the image is displayed on the left pane and all curve parameters are listed on the right pane.
- Insert: Allows inserting the feature descriptor and image in to the database.



Figure 7.14 Interface to mark landmark points

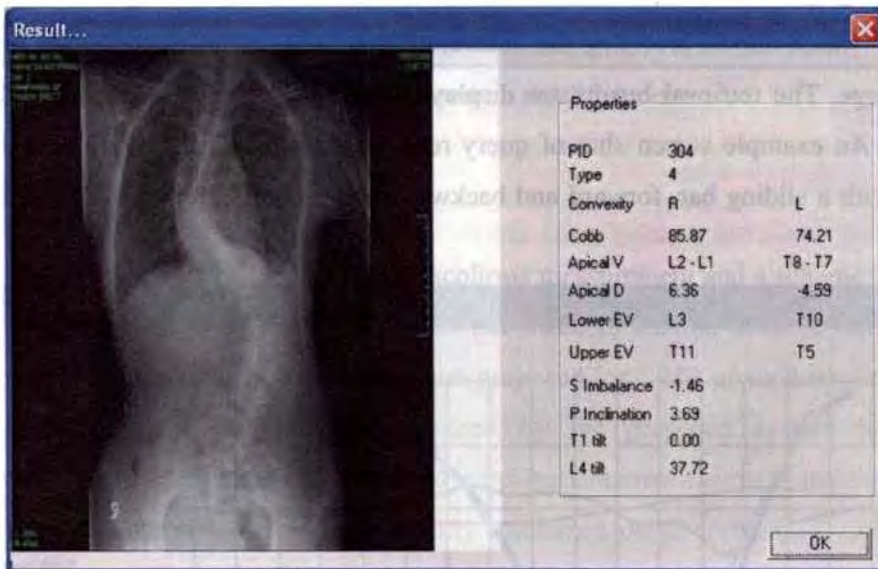


Figure 7.15 Result window of Cobb angle measurement

- Update: Allows updating the feature descriptor of the image already present in the database.
- Show: To display the result window obtained during Cobb computation.
- Chart: Allows viewing different tilt angles of various end plates of each vertebra. Also, plots the distance to CSL and disc space slope of each vertebra. Various check boxes to select are

Inferior:	Inferior end plate tilt angle
Superior:	Superior end plate tilt angle
Left:	Left end plate tilt angle
Right:	Right end plate tilt angle
Dcsl:	Distance to CSL for each vertebra
S-I:	Average Superior – Inferior tilt angle
L-R:	Average Left – Right tilt angle
Dslope:	Disc space slope of each vertebra

This is a quite useful interface for spine surgeons to examine the scoliosis curve for instrumentation planning. Figure 7.16 shows the user interface with some options for plotting selected.

The “Retrieval” menu item has two options in the drop down menu. One is the “Select” option already discussed to specify the PID and retrieve images matching to it. Another option is “Search” which allows browsing the database for images similar in content to the currently opened image. The retrieval results are displayed one image per screen along with its curve properties. An example screen shot of query result is shown in Figure 7.17. The interface is provided with a sliding bar, forward and backward buttons and edit box to navigate the result set.

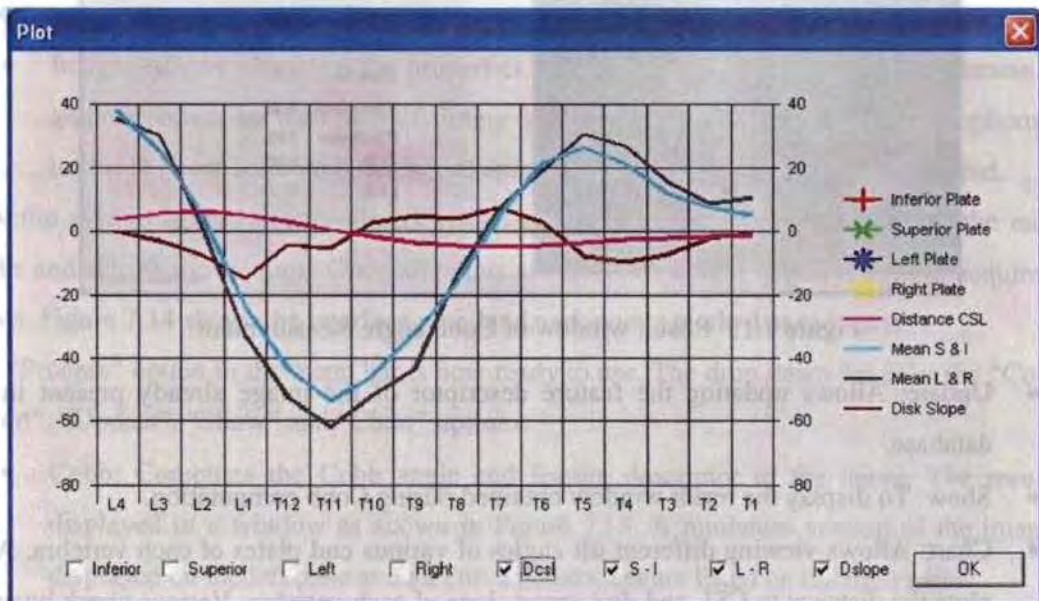


Figure 7.16 User interface for plotting scoliosis curve parameters

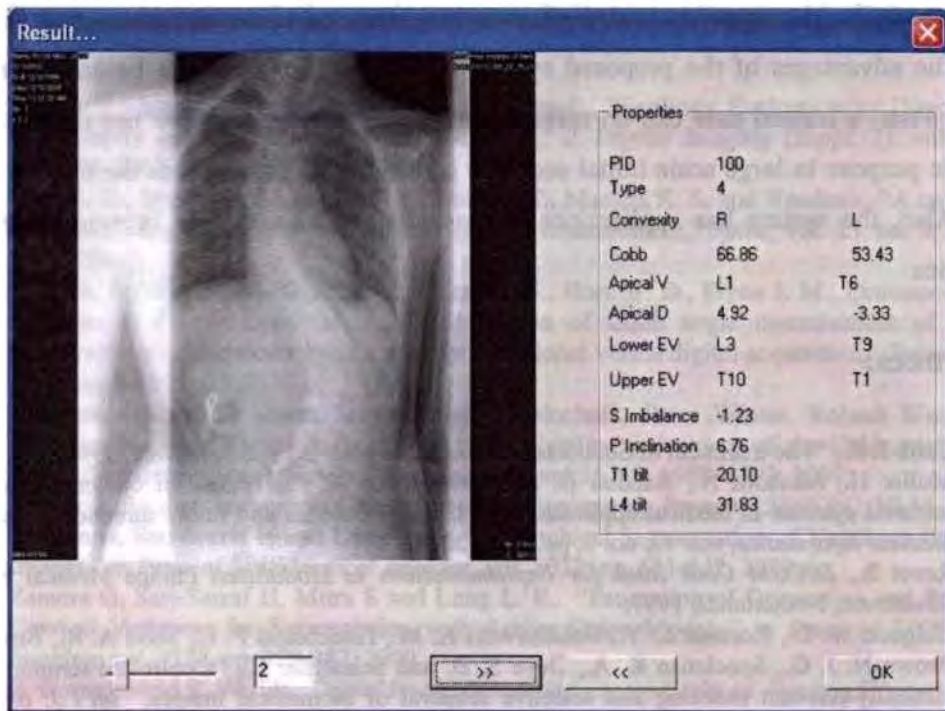


Figure 7.17 CBIR search result window

## 7.6 Conclusions

A novel technique with better accuracy in scoliosis measurement and a scheme for CBIR of scoliosis images are developed. An algorithm for automatic estimation of spine orientation and location is also developed. This is the first system proposed for CBIR of scoliosis images to the best of our knowledge. The accuracy obtained by the proposed system for scoliosis measurement is better, compared to the reported measurement error ranges in the literature. The inter observer and intra observer reliability study conducted using correlation coefficient and Kappa statistics proved the reliability and repeatability of the system. The developed system has dual applications: one as an accurate scoliosis measurement system and the other as a CBIR system to aid diagnosis and treatment. The new system has applications in diagnosis for case-based reasoning and evidence-based medicine. The system has eliminated the variability in end vertebrae selection and ambiguity in choosing a classification scheme, to aid in decision making of surgical approach. The system is quite useful for a spine surgeon in instrumentation planning by outcome analysis of the similar previous cases. Also, the system provides important parameters like distance to CSL of each vertebra, end plate tilt angle, pelvic inclination,

vertebral body height, intervertebral disc space height etc to assist instrumentation planning. One of the advantages of the proposed system is that the user need not be an expert spine surgeon. Thus, a trained user can do offline database population and also use the system for diagnostic purpose in large scale initial scoliosis screening, so as to reduce the cost of a spine expert. Also, the system has applications in research and education to browse large image repositories.

## References:

- [1] Cobb J. R., "The treatment of Scoliosis," *Connecticut Med. J.*, vol. 7, pp. 467, 1943.
- [2] Muller H., Michous N., Bandon D. and Geissbuhler A., "A review of content-based image retrieval systems in medical applications — Clinical benefits and future directions," *Int'l J. of Medical Informatics*, vol. 73, no. 1, pp. 1-23, 2004.
- [3] Revet B., *DICOM Cook Book for Implementations in Modalities*, Philips Medical Systems, Eindhoven, Netherlands, 1997.
- [4] Bidgood W. D., Korman L. Y., Golichowski A. M., Hildebrand P. L., Mori A. R., Bray B. B., Brown N. J. G., Spackman K. A., Dove S. B. and Schoeffler K., "Controlled terminology for clinically-relevant indexing and selective retrieval of biomedical images," *Int'l J. on Digital Libraries*, vol. 3, no. 1, pp. 278–287, 1997.
- [5] Bidgood W. D., "The SNOMED DICOM Microglossary – Controlled terminology resource for data interchange in biomedical imaging," *Methods of Information in Medicine*, vol. 37, no. 4-5, pp. 404–414, 1998.
- [6] Kimura M., Kuranishi M., Sukenobu Y., Watanabe H., Nakajima T., Morimura S. and Kabata S., "JJ1017 image examination order codes - standardized codes supplementary to DICOM for image modality, region, and direction," in *Proc. SPIE*, vol. 4685, pp. 307–315, 2002.
- [7] Gld M. O., Kohnen M., Keyzers D., Schubert H., Wein B. B., Bredno J. and Lehmann T. M., "Quality of DICOM header information for image categorization, in *Proc. SPIE*, vol. 4685, pp. 280–287, 2002.
- [8] Bozec C. L., Jaulent M. C., Zapletal E and Degoulet P., "Unified modeling language and design of a case-based retrieval system in medical imaging," In *Proc. of the Annual Symposium of the American Society for Medical Informatics (AMIA)*. Nashville, TN, USA, 1998.
- [9] Bui A. A. T., Taira R. K. , Dionision J. D. N., Aberle D. R., El-Saden S and Kangarloo H, "Evidence-based radiology," *Academic Radiology*, vol. 9, no. 6, pp. 662-669, 2002.
- [10] Lehmann T. M., Wein B. B. and Greenspan H., "Integration of Content-based Image Retrieval to Picture Archiving and Communication Systems," in *Proc. of Medical Informatics Europe (MIE 2003)*, Amsterdam, IOS Press, 2003.
- [11] Carman D. L., Browne R. H. and Birch J. G., "Measurement of scoliosis and kyphosis radiographs: intraobserver and interobserver variation," *J. Bone and Joint Surg.*, vol. 72-A, no. 3, pp. 328–333, 1990.
- [12] Morrissy R. T., Goldsmith G. S., Hall E. C, Kehl D and Cowie G. H., "Measurement of the Cobb angle on radiographs of patients who have scoliosis," *J. Bone and Joint Surg.*, vol. 72-A, no. 3, pp. 320–327, 1990.
- [13] Stokes I. A. F and Aronsson D. D., "Computer-assisted algorithms improve reliability of King classification and Cobb angle measurement of scoliosis," *Spine*, vol. 31, no. 6, pp. 665–670, 2006.

- [14] Beauchamp M, Labelle H, Grimard G, Stanciu C, Poitras B and Dansereau J., "Diurnal variation of Cobb angle measurement in adolescent idiopathic scoliosis," *Spine*, vol. 18, no. 12, pp.1581–1583, 1993.
- [15] Berliner L, Somchai Kreang-Arekul and Kaufman L., "Scoliosis Evaluation by Direct Digital Radiography and Computerized Post-Processing," *J. Digital Imaging (Suppl. 1)*, vol. 15, pp. 270–274, 2002.
- [16] Shea K. G., Stevens P. M., Nelson M., Smith J. T., Masters K. S. and Yandrow, "A comparison of manual versus computer-assisted radiographic measurement," *Spine*, vol. 23, no. 5, pp. 551–555, 1998.
- [17] Wills B. P., Auerbach J. D., Zhu X., Caird M. S., Horn B. D., Flynn J. M., Drummond D. S., Dormans J. P. and Ecker M. L., "Comparison of Cobb angle measurement of scoliosis radiographs with preselected end vertebrae: traditional versus digital acquisition," *Spine*, vol. 32, no. 1, pp. 98-105, 2007.
- [18] Michaela Gstoettner, Katrin Sekyra, Nadja Walochnik, Peter Winter, Roland Wachter and Christian M. Bach, "Inter- and intraobserver reliability assessment of the Cobb angle: manual versus digital measurement tools," *Eur. Spine J.*, vol.16, pp.1587–1592, 2007.
- [19] Thoma G. R., *Annual Report - Communications Engineering Branch*, LHNCBC, NLM, 2002.
- [20] Zamora G, Sari-Sarraf H and Long L. R., "Hierarchical Segmentation of Vertebrae from X-ray Images," in *Proc. of SPIE Medical Imaging*, vol. 5032, pp. 631-642, 2003.
- [21] Zamora G, Sari-Sarraf H, Mitra S and Long L. R., "Estimation of Orientation and Position of Cervical Vertebrae for Segmentation with Active Shape Models," in *Proc. of SPIE Medical Imaging*, vol. 4322, pp. 378-387, 2000.
- [22] R. Gonzalez and R. Woods, *Digital Image Processing - second edition*, Pearson Education, India, pp. 519-561, 2002.
- [23] Bracewell and N. Ronald, *Two-Dimensional Imaging*, Englewood Cliffs, NJ: Prentice Hall, pp. 505-537, 1995.
- [24] Pierre Soille, *Morphological Image Analysis: Principles and Applications*, Springer- Verlag, New York, 1999.
- [25] <http://archive.nlm.nih.gov/proj/ftp/ftp.php>, National Library of Medicine, NHANES II X-ray images.
- [26] Long L. R. and Thoma G. R., "Landmarking and feature localization in spine x-rays," *J. of Electronic Imaging*, vol. 10, pp. 939-956, 2001.
- [27] Bridwell K. H. and DeWald R. L (Eds), *The Text Book of Spinal Surgery (Second Edition)*, Lippincott-Raven Publishers, Philadelphia, 1997.
- [28] Ponsetti I. V. and Friedman B., "Prognosis in Idiopathic Scoliosis," *J Bone Joint Surg. Am*, vol. 32, pp. 381-395, 1950.
- [29] King H. A., Moe J. H., Bradford D. S. and Winter R. B., "The selection of fusion levels in thoracic idiopathic scoliosis," *J. Bone Joint Surg. Am*, vol. 65, no. 9, pp. 1302–1313, 1983.
- [30] Cohen J., "A Coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, pp 37-46, 1960.
- [31] H. Müller, A. Rosset, J. P. Vallée and A. Geissbuhler, "Comparing feature sets for content-based medical information retrieval," in *Proc. SPIE Med. Imaging*, San Diego, CA, vol. 5351, pp. 99–109, 2004.
- [32] Nachiappan C, Dangerfield P. H., Giakas G, Cochrane T and Dorgan J. C., "Computer-assisted Cobb measurement of scoliosis. *Eur Spine J.*, vol. 11, pp. 353–357, 2002.
- [33] Stokes I. A. F. and Aronsson D. D., "Identifying Sources of Variability in Scoliosis Classification Using a Rule-Based Automated Algorithm," *Spine*, vol. 27, no. 24, pp. 2801–2805, 2002.
- [34] Qiu G, Zhang J, Wang Y, Xu H, Zhang J, Weng X, Lin J, Zhao Y, Shen J, Yang X, Luk K. D, Lu D and Lu W. W., "A new operative classification of idiopathic scoliosis: A Peking Union Medical College Method," *Spine*, vol. 30, no. 12, pp. 1419-1426, 2005.



- [35] Lenke L. G., Betz R. R., Harms J., Bridwell K. H., Clements D. H., Lowe T. G. and Blanke K., "Adolescent idiopathic scoliosis: a new classification to determine extent of spinal arthrodesis," *J. Bone and Joint Surg.*, vol. 83-A, no. 8, pp. 1169-81, 2001.
- [36] Lowe T, Berven S. H., Schwab F. J. and Bridwell K. H., "The SRS classification for adult spinal deformity: building on the King/Moe and Lenke classification systems," *Spine*, vol. 31, no. 19S, pp. 119-125, 2006.
- [37] Richards B. S., Sucato D. J, Konigsberg D. E and Ouellet J. A., "Comparison of reliability between the Lenke and King classification systems for adolescent idiopathic scoliosis using radiographs that were not premeasured," *Spine*, vol. 28, no. 11, pp. 1148-56, 2003.
- [38] Rhee J. M., Bridwell K. H., Won D. S. , Lenke L. G., Chotigavanichaya C and Hanson D. S., "Sagittal plane analysis of adolescent idiopathic scoliosis: The effect of anterior versus posterior instrumentation," *Spine*, vol. 27, no. 21, pp. 2350-2356, 2002.
- [39] Panayotis N Soucacos, Panayotis Konstantinos Soucacos, Konstantinos C Zacharis, Alexandros E. Beris and Theodore A. Xenakis, "School-Screening for Scoliosis - A Prospective Epidemiological Study in Northwestern and Central Greece," *J. Bone and Joint Surg.*, vol. 79A, pp. 1498-1503, 1997.

---

## Chapter 8

# Conclusions and Future Work

---

### 8.1 Conclusions

In this thesis, techniques for content-based image retrieval of general and biomedical images have been investigated. For general images shape descriptors, and for biomedical images scoliosis images have been the focus of the research. State of the art shape representation techniques and CBIR systems were reviewed. The main contributions were the development of novel shape descriptors and a system for CBIR of scoliosis images.

MPEG-7 has adopted two notions of shape descriptors: region-based and contour based. The ARTD is the region-based and CSSD the contour-based shape descriptor adopted by MPEG-7. ZMD is also widely used as region shape descriptor. A previous study reports that ZMD has better contour description ability than CSSD. The ARTD also belongs to the Zernike moment family and hence has similar properties. Also there are reports that Legendre moments based on orthogonal Legendre polynomial have better image representation capability than Zernike moments. But, they lack invariance to geometric transformations. Improving invariance to transformations, a new shape descriptor called Improved Legendre Moment Descriptor (ILMD) based on Legendre polynomials has been developed. ARTD, ZMD and CSSD were implemented and used for comparison of the retrieval performance of the proposed descriptor.

Techniques were developed for the improvement of invariance of the Legendre moments to geometric transformations like translation, rotation and scaling. Legendre polynomials have asymptotic zero distribution. Hence, invariance improved Legendre moment descriptor was used as a contour-based descriptor. To better represent the interior content of a shape as region-

based descriptor, a modified Legendre polynomial to distribute more zeroes in the central region was selected. A new region shape descriptor was formed by a combination of invariance improved Legendre polynomial and modified Legendre polynomial based moments. Retrieval performance of the proposed descriptors were compared using recall – precision pair (RPP) and average normalized modified retrieval rank (ANMRR) with CSSD, ZMD and ARTD using MPEG-7 contour shape database CE-1 and region shape database CE-2. Experiments showed that retrieval performances of the proposed descriptors are better than CSSD, ZMD and ARTD. To improve the performance of ILMD in the case of affine transformations, a pre-processing step was proposed. Shapes were normalized by compacting to improve invariance to affine transformations. A scheme for CBIR was developed using  $k$ -means clustering of shape descriptors. ILMD found to have better clustering accuracy than other descriptors.

User interaction is an important aspect in improving retrieval accuracy of CBIR. Hence, a new relevance feedback scheme was developed using neural networks, as a classification problem. The classification accuracy of ILMD, ZMD and ARTD shape descriptors using SVM, Ls-SVM, PNN and LVQ classifiers were investigated. ILMD with SVM was found to be the optimum shape descriptor – classifier combination, by statistical analysis using McNemar Test. So, a new relevance feedback scheme was designed using ILMD and SVM. Relevant feed back samples were selected by the user for training the SVM classifier from an initial retrieval results using city block distance measure. Then, SVM classifies the image into query class or otherwise. Results indicated improvement in retrieval accuracy.

The present technique for querying scoliosis images is based on the curve type assigned using a classification scheme. The manual annotated text-based querying and retrieval of scoliosis images is not accurate due to the following reasons of user subjectivity: variability in the selection of strategic vertebrae, ambiguity due to the existence of many classification schemes and large intra class variation of scoliosis images within a curve type. Moreover, different classification schemes are not interoperable. Hence, a novel scheme for CBIR of scoliosis images was developed. In the literature, no technique is so far reported for CBIR of scoliosis images.

An algorithm based on the mathematical morphology was developed for automatic determination of location and orientation of spine in digital radiographs. A rule based algorithm was formulated for automatic selection of strategic vertebrae using some landmark points input

by the user. It is based on the features such as maxima and minima, and zero – crossings of the end plate tilt angle of each vertebrae. Cobb angle and other parameters of scoliosis curve were automatically computed using the strategic vertebra levels and end plate tilt angles. Experiments were conducted by three independent observers marking the images twice, at an interval of seven days. Accuracy was evaluated using average error and standard deviation. Inter and intra observer reliability of the system was measured using statistical techniques such as Kappa statistics and correlation coefficient. The results showed that error ranges obtained are low compared to the reported error ranges in the literature. Also, statistical studies showed interobserver and intraobserver reliability. Hence, the developed system is accurate, reliable and repeatable.

A feature descriptor for scoliosis images was designed using the curve type, number of curves, convexity of curve, Cobb angle, spinal balance, AV levels, apical distances, levels of UEV and LEV, average superior-inferior end plate tilt angle of T1, average superior - inferior end plate tilt angle of L4 and pelvic inclination. Two query schemes were designed: one using a classification scheme (for those who are particular in using a classification scheme) and another without using a classification scheme. Query-By-Example tests were conducted using both query schemes. It was found that the system retrieved images clinically similar to the query image, in both schemes. Although, the first query scheme has the limitation of using a classification scheme it resolves the issue of large intra class variation. The query scheme without using a classification scheme retrieved all the images belonging to the query image class, in addition to other similar images, in increasing order of distance. Thus, query scheme without a classification scheme is superior as it eliminates ambiguity of a classification scheme.

The pre and post operative images and treatment procedure adopted have been incorporated into the CBIR system. The CBIR of scoliosis images has special importance, since different treatment strategies are adopted by various surgeons worldwide for a similar case, based on the classification scheme they use. Traditional, querying using a classification scheme will retrieve only images classified using specified classification scheme. But, using the developed CBIR system outcome analysis of various techniques can be carried out for similar previous cases, as it does not use a classification scheme. A software system called ‘SpineSearch’ was developed based on the proposed algorithm for CBIR of scoliosis images. Since, the user of the system need not be a spine surgeon; the cost involved in large scale scoliosis screening is reduced. For

spine surgeons, the system is useful in diagnostic purpose for case-based reasoning and evidence-based medicine and as an accurate measurement and instrumentation planning tool. The developed system has also applications in research and education, for navigation of large image repositories.

## 8.2 Future Work

The ILMD has been applied to object shapes of binary image. Extensions to gray scale image are a possible improvement of the scope of the descriptor. Also, combination of ILMD with other features such as color or texture could be investigated for improving accuracy of CBIR.

In the case of CBIR of scoliosis images, the proposed system is based on the association of coronal plane profile to surgical strategy. Currently, the system does not consider the sagittal profile, axial plane rotation and side bending radiographs. Future research could be to integrate this information also to aid the surgeon. Incorporating the sagittal and axial plane profile would improve the utility of the system in assigning surgical protocol. Also, it would help supporting many new classification schemes like Lenke *et.al.*, PUMC, SRS etc. By incorporating a large database across multi centers worldwide, the efficiency of the CBIR system can be improved.

---

## Appendices

*During initial part of the research, wavelet transform based techniques were studied by the author. Algorithms developed for wavelet transform computation as part of the research are included as appendices in this thesis. Appendix A describes a fast algorithm based on FFT for discrete wavelet transform and Appendix B details a computational structure and algorithm for wavelet packet decomposition on massively parallel processors machine.*

---

## Appendix A

# Development of a Modified FFT-Based Algorithm for DWT

---

### A.1 Introduction

The Discrete Wavelet Transform (DWT) [1], in which both time, scale parameters are discrete, has been recognized as a natural wavelet transform for discrete time signals. The demand for real-time operations in many signal processing tasks with large data sets has necessitated fast and computationally efficient algorithms [2, 3, 4, 5, 6] for wavelet transform. Also, many parallel algorithms [7] are available for a variety of parallel processing architectures.

This appendix primarily focuses on the development of an FFT-based algorithm for real-time computation of the DWT. The computational advantage of the proposed algorithm is compared with the FFT-based Fast Wavelet Transform algorithm proposed by Rioul [5] in terms of number of computations per point, for various wavelet kernel size and decomposition levels.

### A.2 Computational Structure for Fast Wavelet Transform (FWT)

The computational reorganization proposed by Rioul [5] to reduce the computational load of the well known pyramidal algorithm [8] for DWT and the FFT-based algorithm for its implementation is discussed in this section.

According to the pyramidal structure proposed by Mallat [8], the DWT elementary cell (for each level) contains two filtering operations (a highpass filter  $H(z)$  and a lowpass filter  $G(z)$ ),

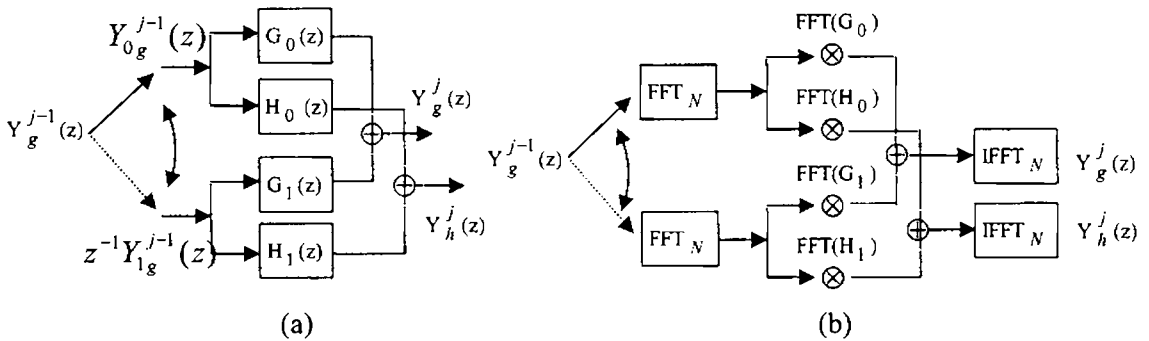


Figure A.1 (a) Elementary DWT cell for FWT and (b) FFT-based implementation of DWT cell

which are followed by dyadic downsampling. The arithmetic complexity of an FIR filter implementation can be reduced by bringing together the computation of several successive outputs [9]. Since the filter outputs are decimated, the filter bank building blocks can be reorganized [6, 10] based on biphase decomposition (separating into odd and even sequences). The reorganized computational structure is shown in Figure A.1 (a). The  $z$ -domain representation of the biphase decomposition of a sequence  $y(n)$  is

$$Y(z) = Y_0(z^2) + z^{-1}Y_1(z^2) \quad (\text{A.1})$$

where  $Y_0(z) = \sum_n y[2n]z^{-n}$ , the even part and  $Y_1(z) = \sum_n y[2n+1]z^{-n}$ , the odd part.

Similarly, the biphase decomposition of the  $L$ -tap filters  $G(z)$  and  $H(z)$  results in their  $L/2$ -tap components  $G(z) = G_0(z) + G_1(z)$  and  $H(z) = H_0(z) + H_1(z)$ . Now, the approximation subband for level  $j$  can be obtained by

$$Y_g^j(z) = G_0(z)Y_{0_g}^{j-1}(z) + z^{-1}G_1(z)Y_{1_g}^{j-1}(z) \quad (\text{A.2a})$$

and the detail subband

$$Y_k^j(z) = H_0(z)Y_{0_g}^{j-1}(z) + z^{-1}H_1(z)Y_{1_g}^{j-1}(z) \quad (\text{A.2b})$$

where  $1 \leq j \leq J$ ;  $J$  is the lowest level of decomposition and  $Y_{0_g}^0(z)$  and  $Y_{1_g}^0(z)$  are the biphase components of the input sequence  $x(n)$ .



An FFT-based implementation of the above structure is described next. The input of the DWT cell is given in blocks of  $B$  samples (each  $L/2$  filter operates on  $B/2$  length samples) and the wrap-around effect [9] due to cyclic-convolution can be avoided by using overlap-add or overlap-save method. For a filter of length  $L/2$  and a sequence of length  $B/2$ , the input block-length without wrap-around effect for  $N$  point DFT is,

$$B = 2N - (L - 2) \quad (\text{A.3})$$

A length- $N$  FFT of the biphase components of the input sequence and the wavelet / scaling filters are computed. Now, four frequency-domain convolutions are performed by multiplying the (Hermitian symmetric) FFT of the input by the (Hermitian symmetric) FFT of the corresponding filter as shown in Figure A.1 (b). The corresponding subsequences are added together for approximation / detail subband and inverse FFT (IFFT) is computed. The approximation subband is used as input for further decomposition. As the data size gets halved at each level due to subsampling, waiting for more blocks from the previous level can be done so that each cell has the same input block length of  $B$  for FFT/IFFT operation. A length- $N$  FFT is most efficient for an optimized value of the block length.

The signal can be reconstructed from the wavelet representation by transposition of the analysis algorithm and by using synthesis filters  $\tilde{H}(z)$  and  $\tilde{G}(z)$ , which are time-reversed versions of the corresponding analysis filters [5].

### A.3 Proposed Modified FFT-Based Algorithm

The proposed algorithm is based on the frequency domain subsampling and makes use of the computational advantage of fast-convolution provided by the structure discussed in Section A.2.

By eliminating the calculation of the approximation subband in levels other than the lowest one of the FWT algorithm, two FFT and one IFFT operations in the intermediate levels can be avoided.

The proposed modified algorithm is explained below. The input for the first level is data samples taken in blocks of length  $B$  (as per Equation (A.3)) and split into even- and odd-

indexed sequences of length  $B/2$  (Equation (A.1)). Now, compute  $N$  point FFT of these sequences and the initial FFT length  $N$  is chosen satisfying the condition,  $N \geq 2^{J-1} * L/2$ ; where  $J$  is the maximum decomposition level (division by 2 factor as we use biphasic components). As the sequence size reduces by half on entering the next level, the FFT length of the biphasic components of the wavelet/scaling filter coefficients are  $N/2^{j-1}$  for level  $j$ . For the first level, the frequency-domain convolutions are performed by multiplying (Hermitian symmetric) FFT of the input by the (Hermitian symmetric) FFT of the wavelet filter and the resulting sequences are added together. Length- $N$  IFFT is applied to obtain the input block's wavelet coefficients for the first level. For the approximation coefficients, the (Hermitian symmetric) FFTs of the scaling filter are used and the resulting sequences are added together.

Without computing IFFT of the approximation coefficients, decomposition for the lower levels can be done in the Fourier space. If  $Y_g^{j-1}[k]$  is a length- $N$  Fourier transform, the length- $N/2$  Fourier transform of its downsampled version is

$$Y_g^j[k] = \sum_{k=1}^{N/2} \frac{1}{2} (Y_g^{j-1}[k] + Y_g^{j-1}[k + N/2]) \tag{A.4}$$

The  $Y_g^j[k]$  corresponds to the Fourier transform of even samples and that of odd ones have a

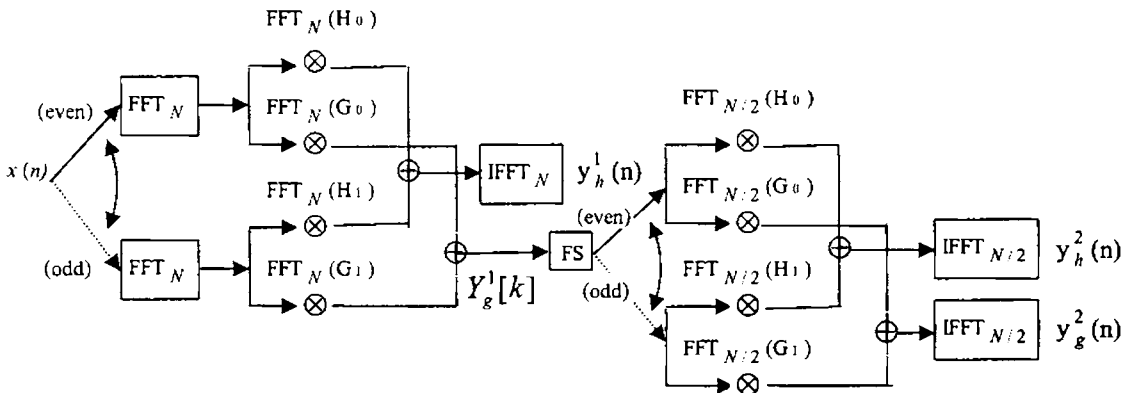


Figure A.2 Proposed FFT-based algorithm. (FS stands for Fourier-domain subsampling)

phase shift [6]. The Fourier-domain approximation coefficients at level  $(j-1)$  are downsampled as per Equation (A.4) to halve the resolution and passed as input for next level  $j$ . The Hermitian symmetric Fourier-coefficients of the resulting odd / even sequences are multiplied with the corresponding Hermitian symmetric filter  $\text{FFT}_{N:2^{j-1}}$  to get the detail / approximation coefficients for level  $2 \leq j \leq J$ . The length- $N/2^{j-1}$  IFFT can be done in each level  $j$  for the wavelet coefficients to get the detail subband and in the last level, length- $N/2^{J-1}$  IFFT is used for approximation subband computation. The coefficients from various blocks are grouped together to get the wavelet / approximation subband. So, every level, except the first and last, has only one IFFT operation (for detail subband) other than complex multiplication and block addition operations. A two level decomposition implementation of the proposed algorithm is shown in Figure (A.2).

Since each block at any stage is processed independently, the algorithm has no inter-block dependency as in the case of FWT. So, no hidden synchronization overhead is involved in the proposed algorithm implementation.

For the synthesis part, the transposed flow graph of the analysis algorithm shown in Figure (A.2) can be used with synthesis filter components which are time-reversed versions of analysis filters.

## A.4 Computational Complexity

The computational complexity in terms of the number of real multiplications and real additions required by the candidate algorithms is calculated in this section. The total number of operations (multiplications + additions) is considered as the appropriate criterion for performance comparison of various algorithms [5].

The “split-radix” FFT [11] used in both algorithms has best known complexity for length  $N = 2^k$ . For real data, the split-radix FFT (or inverse FFT) requires exactly

$$2^{k-1}(k-3) + 2 \text{ (real) multiplications} \quad (\text{A.5})$$

and

$$2^{k-1}(3k-5) + 4 \text{ (real) additions} \quad (\text{A.6})$$

The filters ( $H(z)$  and  $G(z)$ ) in the computation of the DWT usually have equal length. The filter FFT's can be pre-computed and applied as and when needed.

The FWT algorithm has a regular computational structure. The operations required by an elementary cell are counted as follows [5]. The four frequency-domain convolution operations require  $4N/2$  complex multiplications. Assuming that a complex multiplication is done with three real multiplications and three real additions [9], the computational complexity of an elementary cell can be expressed as

$$2 \text{ FFT}_N + 4*3*N/2 \text{ mults} + 4*3*N/2 \text{ adds} + 2*N/2 \text{ adds} + 2 \text{ IFFT}_N$$

This can be simplified as  $k2^{k+1} + 8$  multiplications and  $(3k - 1)2^{k+1} + 16$  additions [5]. The total number of elementary cells required for depth  $J$  decomposition is  $2*(1 - 2^{-J})$ . So, the total number of multiplications required per point is

$$M_1 = ((k2^{k+1} + 8) * 2 * (1 - 2^{-J})) / B \quad (\text{A.7})$$

and the total number of additions per point is

$$A_1 = (((3k - 1)2^{k+1} + 16) * 2 * (1 - 2^{-J})) / B \quad (\text{A.8})$$

In the case of proposed algorithm, the first level has two length- $N$  FFT and one length- $N$  IFFT. The last level has two length- $N/2^{J-1}$  IFFT and all the intermediate levels has one length- $N/2^{j-1}$  IFFT computations per input block. The frequency-domain convolution complexity is  $4*(N/2^j)$  and block addition complexity is  $2*(N/2^j)$ , where  $1 \leq j \leq J$ . The frequency-domain downsampling operation for level  $j$  requires  $N/2^j$  additions. The total operations on an input block for decomposition depth  $J$  can be expressed as

$$(2 \text{ FFT}_N)_1 + (2 \text{ IFFT}_{N/2^{J-1}})^J + (\text{IFFT}_{N/2^{j-1}})_{1 \rightarrow (J-1)} + \left(\frac{N}{2^j}\right)_{1 \rightarrow (J-1)} \text{ adds} + (4*3*\frac{N}{2^j})_{1 \rightarrow J} \text{ mults}$$

$$+ (4*3*\frac{N}{2^j})_{1 \rightarrow J} \text{ adds} + (2*\frac{N}{2^j})_{1 \rightarrow J} \text{ adds} + (2*\frac{N}{2^j})_{1 \rightarrow (J-1)} \text{ adds.}$$

The subscript of parenthesis denotes the level in which the operations are performed. The total number of multiplications per point, using Equation (A.5), is

$$M_2 = ((2^k(k-3) + 4) + (2^{k-J}(k-J-2) + 2) + (\sum_{j=1}^J 2^{k-j}(k-j+10) + 2)) / B \quad (\text{A.9})$$

The total number of additions per point, using Equation (A.6), is

$$A_2 = ((2^k(3k-5) + 8) + (2^{k-J}(3(k-J)-2) + 4) + (\sum_{j=1}^J 3 * 2^{k-j}(k-j+4) + 4) + (\sum_{j=1}^{J-1} 2^{n-j})) / B \quad (\text{A.10})$$

The computational complexity per point of the FWT and that of the proposed algorithm are calculated using the above equations. Appropriate initial FFT length which gives best performance for a given algorithm is chosen for a particular wavelet kernel size. The results are detailed below.

## A.5 Results and Discussions

Table A.1 lists the resulting number of real multiplications per input point required by the candidate algorithms for various wavelet kernel sizes at different decomposition depths. The proposed algorithm has less number of multiplications per point for filter size greater than four and decomposition depth greater than one. Also, the performance improves with an increase in decomposition depth.

Table A.2 lists the number of real additions per point required for both algorithms. The number of real additions is less for the proposed algorithm compared to FWT for filter size greater than two. The same trend as in Table A.1 can be seen regarding the improvement in addition complexity also with an increase in wavelet kernel size and level.

Although both Vetterli's algorithm [6] and the proposed algorithm uses Fourier-domain subsampling, the latter has better performance due to the use of subsampled sequences for initial FFT computation (FFT length being more close to the best performance length) and Hermitian symmetry property.

**Table A.1 FFT-Based DWT algorithms: multiplication complexity per point\***

Filter Length	LEVEL 1		LEVEL 2		LEVEL 3		LEVEL 4		LEVEL 5		FFT LENGTH	
	I	II	I	II	I	II	I	II	I	II	I	II
2	3.0	5.1	4.5	6.4	5.2	7.0	5.6	7.4	5.8	7.5	2	32
4	4.0	5.2	6.0	6.6	7.0	7.2	7.5	7.6	7.7	7.8	4	32
8	5.2	6.3	7.8	7.6	9.1	8.3	9.8	8.7	10.1	8.8	16	64
16	6.5	7.4	9.8	8.7	11.8	9.4	12.3	9.7	12.7	9.9	32	128
32	7.9	8.5	11.8	9.8	13.8	10.5	14.8	10.8	15.3	11.0	64	256
64	9.1	9.5	13.6	10.9	15.9	11.5	17.1	11.9	17.6	12.0	256	512
128	10.2	10.6	15.4	12.0	17.9	12.6	19.2	13.0	19.9	13.1	512	1024

\*Each entry gives the number of real multiplications per input point for various decomposition levels. The notations I and II represents the FWT algorithm and the proposed algorithm respectively. The last column shows the corresponding initial FFT length.

**Table A.2 FFT-Based DWT algorithms: addition complexity per point\***

Filter Length	LEVEL 1		LEVEL 2		LEVEL 3		LEVEL 4		LEVEL 5		FFT LENGTH	
	I	II	I	II	I	II	I	II	I	II	I	II
2	6.0	13.7	9.0	15.0	10.5	15.7	11.2	16.1	11.6	16.3	2	32
4	9.3	14.1	14.0	15.5	16.3	16.2	17.5	16.6	18.0	16.8	4	32
8	14.1	17.4	21.2	18.7	24.7	19.4	26.5	19.8	27.4	20.0	16	64
16	18.2	20.6	27.3	22.0	31.9	22.7	34.2	23.0	35.3	23.2	32	128
32	22.3	23.9	33.5	25.2	39.1	25.9	41.9	26.2	43.3	26.4	64	256
64	26.2	27.1	39.3	28.4	45.8	29.1	49.1	29.5	50.7	29.6	256	512
128	29.6	30.4	44.5	31.7	51.9	32.4	55.6	32.7	57.4	32.9	512	1024

\*Each entry gives the number of real additions per input point for various decomposition levels. The notations I and II represents the FWT algorithm and the proposed algorithm respectively. The last column shows the corresponding initial FFT length.

## A.6 Conclusion

A computationally efficient FFT-based DWT algorithm is presented in this appendix. The FWT algorithm [4] has been proved to be better in performance than pyramidal algorithm by Mallat [8] and FFT-based Vetterli's Algorithm [6]. The computational complexity calculations show that the proposed algorithm provides remarkable savings for wavelet kernel size greater than four (which are widely used), compared to FWT. Also, the performance of the algorithm increases with decomposition depth. The lack of inter-block dependency is a useful feature in parallel processing environment. The proposed algorithm is best suited for computationally intensive applications, such as in image processing.

**References:**

- [1] I. Daubechies, "The wavelet transform, time-frequency localizations and signal analysis," *IEEE Trans. Inf. Theory*, vol. 36, no. 9, pp. 961-1005, 1990.
- [2] Guoan Bi, "On computation of the discrete wavelet transform," *IEEE Trans. Signal Proc.*, vol. 47, no. 5, pp. 1450-1453, 1999.
- [3] M. J. Shensa, "The discrete wavelet transform: wedding the 'a trous and mallat Algorithms,'" *IEEE Trans. Signal Proc*, vol. 40, no. 10, pp. 2464-2482, 1992.
- [4] S. G. Mallat, "A theory for multi-resolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 11, no. 7, pp. 674-693, 1989.
- [5] I. Daubechies and W. Sweldens, "Factoring wavelet transform into lifting steps," *J. Fourier Anal. Appl.*, vol. 4, no. 3, pp. 247-269, 1998.
- [6] J. N. Patel, A. A. Khokhar and Leah H. Jamieson, "Scalability of 2-D wavelet transform algorithms: Analytical and experimental results on MPPs," *IEEE Trans. Signal Proc*, vol. 48, no. 12, pp. 3407-3419, Dec. 2000.
- [7] P. P. Vaidyanathan, *Mutirate Systems and Filterbanks*, A. V. Oppenheim, Ed, Prentice Hall Signal Processing Series, 1993.
- [8] O. Rioul and P. Duhamel, "Fast algorithms for discrete and continuous wavelet transforms," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 569-586, 1992.
- [9] M. Vetterli, "Wavelets and filter banks: Relationships and new results," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, Albuquerque, NM, pp. 1723-1726, 1990.
- [10] H. J. Nussbaumer, *Fast Fourier Transform and Convolution Algorithms*, Berlin: Springer, 1981.
- [11] P. Duhamel, "Implementation of split-radix FFT algorithms for complex, real, and real-symmetric data," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 4, pp. 285-295, 1986.

---

## Appendix B

# Development of a Computational Structure for Fast Computation of Wavelet Packet Transform on MPPs

---

### B.1 Introduction

In this appendix, a Parallel Multiple Subsequence (PMS) structure is developed for wavelet packet (WP) decomposition. In PMS structure, subbands are computed using subsequences obtained directly from the input data, improving parallelism in computation. An algorithm for implementation of PMS on massively parallel processors (MPPs) is also developed.

Wavelet packets, which comprise of the entire family of subband coded decompositions, is an ideal tool in multiresolution analysis. In the wavelet transform [1] computation, the signal is decomposed into coarse scale approximations and the detail signal. This procedure is applied recursively to the coarse scale approximations leading to the well known filter bank tree wavelet decomposition structure. In the WP decomposition the recursive procedure is applied to both coarse scale approximation and detail signals, which leads to a complete binary tree, giving more flexibility in frequency resolution.

Several efficient parallel algorithms [2, 3] proposed for the fast wavelet transform are applicable to WP decomposition also. Some of the works in parallel wavelet packet decomposition includes subband based approaches for performing the best basis selection on



parallel MIMD [4, 5] and SIMD [6] architectures, parallel wavelet packet decomposition in numerics [7] and some of their applications [8].

But, most of these algorithms are based on the filter bank tree structure. The delay associated with the implementation grows exponentially with the number of levels [9]. For instance, the set of basis functions for Short Time Fourier representation of a signal requires the lowest level WP subbands only. With filter bank tree structure, one has to perform unnecessary computations by way of evaluating the higher level subbands. One of the important factors limiting the range of scalability in parallel processing is the sequential component of the algorithm [2].

## B.2 Wavelet Packet Transform Algorithms

This section briefly describes the filter bank tree algorithm by Mallat [10] and proposed PMS structure [11] based algorithm for WP decomposition is then explained.

### B.2.1 The Filter Bank Tree WP Algorithm

The wavelet packet decomposition extends the discrete wavelet transform in a way that each level  $j$  consists of  $2^j$  subbands, generated by a tree of low pass and high pass operations. Consider the analysis filter bank of the 1-D WP scheme shown in Figure B.1. In this figure, the analysis filters  $H(z)$  represents a high pass filter and  $G(z)$  represents a low pass filter. The WP transform of a discrete signal  $x(n)$  can be computed by convolving with filters  $H(z)$  and  $G(z)$  followed by dyadic downsampling. This process is repeated on both sequences until the required

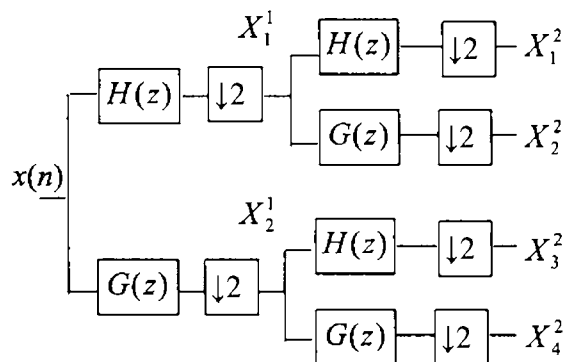


Figure B.1 Two level WP decomposition using filter bank tree algorithm

level of decomposition is reached.

The WP subbands at any level  $j$  are given by

$$X_{2^{i-1}}^j(n) = \sum_k h(k) X_i^{j-1}(2n - k) \quad (\text{B.1a})$$

$$X_{2^i}^j(n) = \sum_k g(k) X_i^{j-1}(2n - k) \quad (\text{B.1b})$$

where  $X_1^0 = x(n)$ ; the input sequence ( $n \in \mathbb{Z}$ ),  $j = 1, 2, \dots, J$ ; denotes different levels and  $1 \leq i \leq 2^{j-1}$  is the subband index within a level.

## B.2.2 Parallel Multiple Subsequence (PMS) Structure Based Algorithm

The PMS structure [11], originally developed for DWT, is based on the principle of polyphase splitting for subband decomposition. Here, an extension to the PMS structure for WP transform is developed. From the wavelet (defined by its filter  $H(z)$ ) and its smoothing function (defined by its filter  $G(z)$ ) we compute the filter coefficients for the subbands at each level by successive convolutions and upsampling. Subbands at various levels are computed directly by convolving the corresponding filter with the original data.

The subbands can be computed based on the PMS structure as follows

$$X_{2^{i-1}}^j(k) = \sum_{p=1}^{2^j} x_{j,p}(k) * h_{i,p}^j(-k) \quad (\text{B.2a})$$

$$X_{2^i}^j(k) = \sum_{p=1}^{2^j} x_{j,p}(k) * g_{i,p}^j(-k) \quad (\text{B.2b})$$

where  $*$  denote convolution and  $1 \leq i \leq 2^{j-1}$  is the subband index,

$$x_{j,p}(k) = x(2^j k + p - 1),$$

$$h_{i,p}^j(k) = h_i^j(2^j k + 2^j - p + 1), \text{ and}$$

$$g_{i,p}^j(k) = g_i^j(2^j k + 2^j - p + 1).$$

The PMS structure for second level WP decomposition is shown in Figure B.2. Being a regular structure, this can be extended to any level. The PMS structure has got parallelism both within and between levels, making it highly suitable in parallel processing environment.

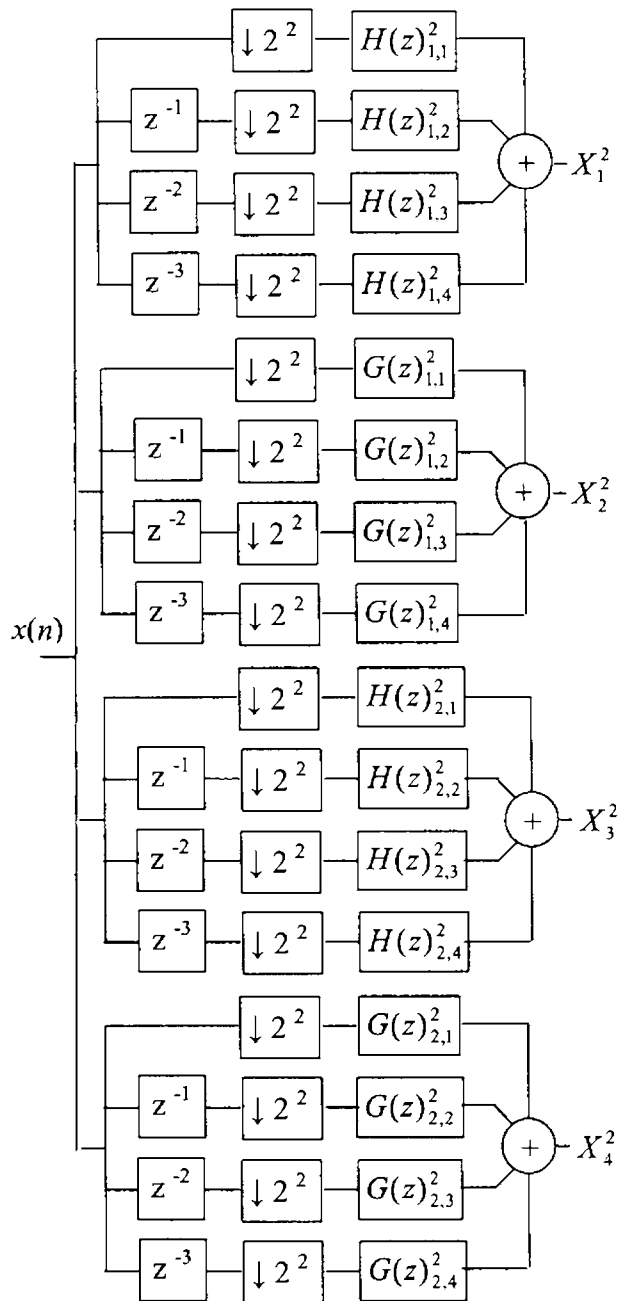


Figure B.2 Second level WP decomposition using PMS structure

### B.3 Algorithm Analysis

The scalability and computational complexity of the algorithms described in Section (B.2) are analyzed on coarse-grained machines. The platform used is a distributed memory

architecture in which each processor has fast access to local memory

### B.3.1 Computational Model and Assumptions

The notion of scalability of an algorithm and parameters of the computational model are defined based on references [12, 13]. Let  $t_f$  be the time required for one floating point operation. The time required for the complete transfer of a message containing  $m$  words between two processors that are  $l$  connections away is given by the  $t_s + (t_w m) * l$ , where  $t_s$  is the startup time, and  $t_w = \text{bytes-per-word} / B$ , where  $B$  is the bandwidth of the communication channel between the processors in bytes per second. So, the total execution time mainly consists of two parts: one corresponding to the computation complexity and the other corresponding to the communication complexity.

Let  $T(n,p)$  be the time taken by an algorithm on a  $p$  processor architecture with input data size  $n$ . The algorithm is considered scalable on the structure if  $T(n,p)$  increases linearly with an increase in the data size or decrease linearly with the increasing number of processors (machine size). We assume that  $p < n$ , as we are interested in large problem sizes generally.

The performance study of the algorithms is done by varying the machine size ( $p$ ), problem size ( $n$ ) and the wavelet filter kernel size ( $L$ ) for different levels of decomposition. For the sake of simplicity in analysis, we assume that the problem size and machine size are powers of 2, i.e.,  $2^n$  and  $2^p$  respectively. The scalability and performance in parallel environment is analyzed for generating subbands at a given level only.

### B.3.2 Data Distribution Strategy

The main problem faced, when dealing with multicomputers, is how to perform an efficient mapping of tasks and data to the processors which raises the questions of load balancing and communication minimization. Both questions are closely connected with the data distribution task. For both the algorithms, two methods of handling border data ( $L$  coefficients of the neighboring Processing Elements (PE) are required to compute a single output coefficient at the border) can be used [14]. They are

- *Data Swapping*: Each PE computes only non-redundant data and then exchanges these results with the appropriate neighboring PEs, in order to get the necessary data for the next

calculation step (i.e., the next decomposition level).

- *Data Overlapping*: In the initialization step, each PE is provided not only its share of the original signal but also the data set which is required to compute the redundant data. This avoids additional communication with neighbor PEs to obtain the border data.

Appropriate data distribution scheme is chosen in the analysis for a given algorithm.

### B.3.3 Analysis of the Filter Bank Tree Algorithm

The parallel implementation algorithm used here is based on WP image decomposition algorithm proposed by Feil and Uhl [5]. It is found that for data distribution, in a filter bank tree algorithm, the data overlapping approach is not competitive at all over a wide range of different architectures [6]. So, the data distribution scheme used here will be Data swapping method.

The most natural way to distribute the computational work of a WP transform can be found on a distributed memory architecture with the number of PEs equal to a power of 2, i.e.,  $p=2^a$ . The input data  $2^n$  for each level (will be approximately the same as the original input data, ignoring the increase in length caused by convolution, as all subbands are retained in WP decomposition) is partitioned into  $2^a$  parts of equal size  $2^{n-a}$ . The partitioning is done in two different ways depending upon whether level  $j$  is smaller or larger than  $a$ . Let  $i$  denote the subband index and  $0 \leq i < 2^j$ . If  $j < a$ , a subband with index  $i$  is not assigned to a single PE but is shared by PEs with processor index in the range  $2^{a-j} * i$  to  $2^{a-j} * (i+1) - 1$ . Therefore, in the initialization step, those  $2^{a-j}$  PEs will exchange their data in order to have the entire shared subband residing on each of them. Then, in the second step, they will calculate their own part of the subband they share at level  $j+1$ . If  $j \geq a$ ,  $2^{j-a}$  subbands and also their two children reside on each PE. Thus no communication among PEs is needed for the subset of subbands residing on each PE at level  $j$ .

The message communication required for level  $j$  is  $L$  (filter length) data units across  $2^{a-j+1}-1$  PEs for subband computation and  $2^{n-a}$  data units across  $2^{a-j}$  PEs for data redistribution on entering a new level. This is required for all the  $2^j$  subbands of the level. Thus, the overall communication amount (ie., the number of datapoints sent) can be expressed as

$$m = \sum_{j=1}^{a-1} (L * (2^{a-j+1} - 1) + 2^{n-a} * 2^{a-j}) * 2^j \quad (\text{B.3})$$

The total number of PEs involved in the message transfer at various stages is

$$k = \sum_{j=1}^{a-1} 3 * 2^a - 2^j.$$

Based on the parameters described in Section B.3.1, the total time required for message communication is  $k * t_s + m * t_w$ . The computation of each output coefficient requires  $2L$  floating point operations (additions and multiplications). As each processor holds  $2^{n-a}$  data units, the total computation time is  $2L * 2^{n-a} * J * t_f$ , where  $J$  is the maximum decomposition level. Thus the total time taken for WP decomposition is given by,

$$T_1 = (2L * 2^{n-a} * J * t_f) + (k * t_s + m * t_w) \quad (\text{B.4})$$

### B.3.4 Analysis of the PMS Structure Based Algorithm

As the PMS structure is tailored for the parallel computation of the subbands of a given level directly from the original input sequence, there is no sequential part in the algorithm. So, the data distribution scheme proposed for PMS structure based algorithm is Data Overlapping approach i.e., all necessary data desired to compute the subbands is sent to the processors initially. The proposed data distribution strategy is outlined below.

The number of subbands in a regular WP decomposition scheme is  $2^j$  for level  $j$ . But, PMS structure splits each subband (and the corresponding filter) again into  $2^j$  subsequences. This results in  $2^j * 2^j$  subsequences for the level  $j$ . The input data is partitioned into  $2^a$  parts of equal size  $2^{n-a}$ . The data partitioning can be done in two different ways depending upon whether  $2j$  is smaller or larger than  $a$ . Let  $i$  denote the subsequence index and  $0 \leq i < 2^{2j}$ . If  $2j < a$ , the number of subsequences is less than the number of available PEs and each subsequence with an index  $i$  is not assigned to a single PE but is shared by PEs with processor index in the range  $2^{a-2j} * i$  to  $2^{a-2j} * (i + 1) - 1$ . The redundant data units to be distributed initially among PEs is  $L_j / 2^j$ , where  $L_j = (L - 1)(2^j - 1) + 1$ ; the filter length for level  $j$ . As each PE is having the entire data units required, no message communication need to be performed in this distribution scheme and the computational work is uniformly distributed. If  $2j \geq a$ ,

$2^{2j-a}$  subsequences can reside on each PE. Then, initial redundant data distribution is also not required.

The computation of each output coefficient requires  $2 * (L_j / 2^j) * 2^j = 2 L_j$  floating point operations. Since each processor holds  $2^{n-a}$  data units, the total computation time is  $2 L_j * 2^{n-a} * t_f$ . As there is no message passing required, the total time taken for WP decomposition is

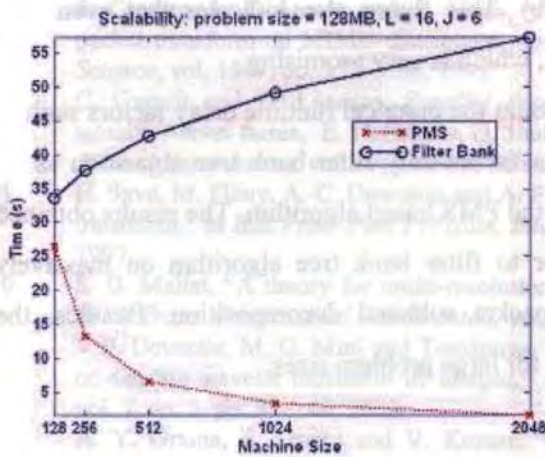
$$T_2 = 2 L_j * 2^{n-a} * t_f \quad (\text{B.5})$$

## B.4 Analytical Results and Discussion

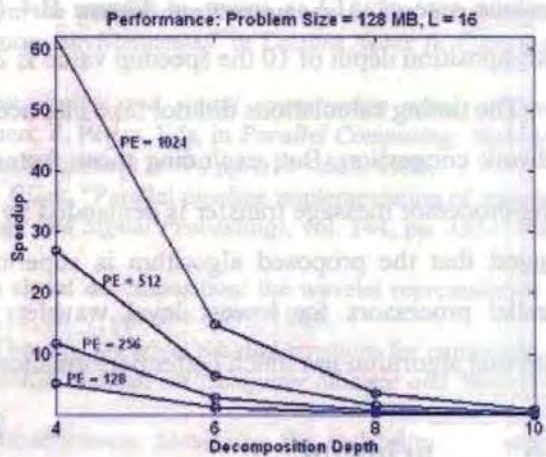
In order to get an approximate figure of the timings, the system parameters of Intel Paragon XP / S machine [2] is used in equation (B.4) and (B.5). The paragon machine has a 2 – D mesh (torus) connection structure with support for number of processors in the range of 64 – 4000. The per node memory capacity is 128 MB. The communication bandwidth of the machine is 200 MB/s. Each processor has a peak performance (64 bits) of 75 Mflop / s and the communication latency is around 100  $\mu$ secs. The performance measurement criterion used here is speedup, which is taken as the ratio of execution time of the filter bank tree algorithm to that of the PMS structure based algorithm, i.e.,

$$\text{speedup} = T_1 / T_2 \quad (\text{B6})$$

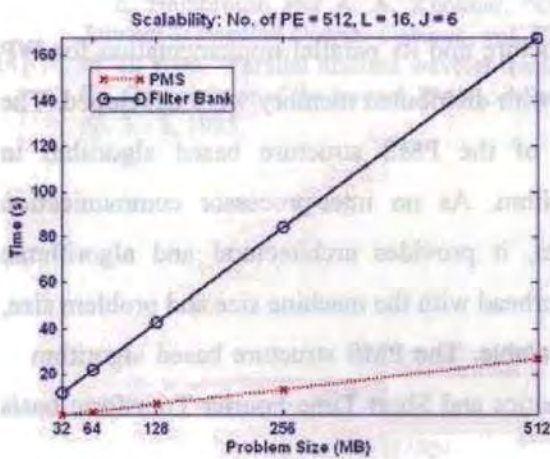
Figure B.3 compares the scalability of the candidate algorithms for increasing machine and problem size. Figure B.3(a) plots the execution time of filter bank tree and PMS based algorithms at a decomposition depth of 6 on various machine size with fixed wavelet kernel size ( $L = 16$ ) for a problem size of 128 MB. The execution time decreases for the PMS based algorithm with an increase in machine size whereas it linearly increases for the filter bank tree algorithm due to the communication overhead. Figure B.3(b) shows the execution time for various problem sizes on 512 processors of the Paragon using a 16-tap wavelet kernel for various problem sizes. It can be noted that the execution time increases linearly with the problem size and hence PMS algorithm perfectly satisfies the scalability criterion. Although the execution time for both the algorithms increases with problem size, due to the communication



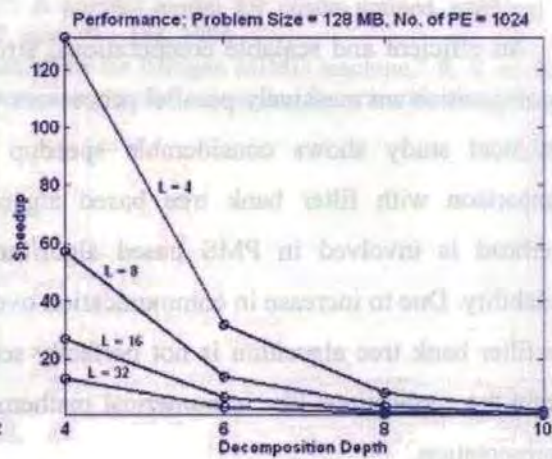
(a)



(a)



(b)



(b)

Figure B.3 Comparison of scalability of filter bank tree and PMS algorithm at decomposition depth 6 and filter kernel size 16 for (a) different machine size and (b) different problem size

Figure B.4 Comparison of performance (speedup) of filter bank tree and PMS algorithm at different decomposition depth and fixed problem size of 128 MB. (a) filter kernel size 16 on different machine size (b) number of PEs 1024 and different filter kernel size.

overhead caused by the data re-distribution between the levels, the rate of increase of filter bank tree algorithm is much faster than that of PMS.

Figure B.4 shows the speedup of the proposed algorithm over the filter bank tree algorithm for various decomposition levels. The speedup increases significantly with machine size up to level 8 as shown in Figure B.4 (a). The plot of speedup for a usual range of filter length, 32, at



machine size of 512 is given in Figure B.4 (b). This figure also indicates that even at a decomposition depth of 10 the speedup value is 2, which is very promising.

The timing calculations did not take into account the practical runtime delay factors such as network congestion. But, excluding these factors favors only filter bank tree algorithm as no inter-processor message transfer is demanded by the PMS based algorithm. The results obtained suggest that the proposed algorithm is superior to filter bank tree algorithm on massively parallel processors for lowest level wavelet packet subband decomposition. Besides, the proposed algorithm has much better performance for large problem sizes.

## B.5 Conclusion

An efficient and scalable computational structure and its parallel implementation for WP decomposition on massively parallel processors with distributed memory were developed. The analytical study shows considerable speedup of the PMS structure based algorithm in comparison with filter bank tree based algorithm. As no inter-processor communication overhead is involved in PMS based algorithm, it provides architectural and algorithmic scalability. Due to increase in communication overhead with the machine size and problem size, the filter bank tree algorithm is not perfectly scalable. The PMS structure based algorithm is useful for applications like in numerical mathematics and Short Time Fourier Transform basis representation.

## References:

- [1] I. Daubechies, "The Wavelet Transform, Time – Frequency localizations and signal analysis," *IEEE Trans. Inf. Theory*, vol. 36, pp. 961 – 1005, 1990.
- [2] J. N. Patel, A. A. Khokhar and Leah H. Jamieson, "Scalability of 2-D Wavelet Transform algorithms: analytical and experimental results on MPPs," *IEEE Trans. Signal Processing*, vol. 48, no. 12, pp. 3407 – 3419, 2000.
- [3] O. Rioul and P. Duhamel, "Fast algorithms for discrete and continues wavelet transforms," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 569 – 586, 1992.
- [4] A. Uhl, "Wavelet packet best basis selection on moderate parallel MIMD architectures," *Parallel Computing*, vol. 22, no. 1, pp. 149 – 158, 1996.
- [5] M. Feil and A. Uhl, "Multicomputer algorithms for wavelet packet image decomposition," in *Proc. 14<sup>th</sup> International Parallel and Distributed Processing Symposium*, pp. 793 – 800, 2000.
- [6] M. Feil and A. Uhl, "Algorithms and programming paradigms for 2-D wavelet packet decomposition on multicomputers and multiprocessors," in P. Zinterhof, M. Vajtersic, and A. Uhl, Eds, *Parallel Computation*, (Proceedings of ACPC'99, Lecture Notes on Computer Science, Springer-Verlag, 1557), pp. 367 – 376, 1999.

- [7] S. Corsaro, L. D'Amore and A. Murli, "On the parallel implementation of the fast wavelet packet transform on MIMD distributed memory environments," in *Lecture Notes In Computer Science*, vol. 1557, pp. 357 - 366, 1999.
- [8] C. Guerrii and D. Lazzaro, *Parallel deconvolution and signal compression using adapted wavelet packet bases*, E. Hollander, G. Joubert, F. Peters, Eds, in *Parallel Computing: State-of-Art and Perspectives*, vol. 11, Elsevier Science Publishers B. V., pp. 617 - 624, 1996.
- [9] H. Sava, M. Fleuy, A. C. Downton and A. F. Clark, "Parallel pipeline implementation of wavelet transform," in *IEE Proc. Part I (Vision, Image and Signal Processing)*, vol. 144, pp. 355 - 359, 1997.
- [10] S. G. Mallat, "A theory for multi-resolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674 - 693, 1989.
- [11] V.P. Devassia, M. G. Mini and Tessamma Thomas, "A novel parallel structure for computation of discrete wavelet transform of images," *AMSE Journal on Computer Science and Statistics*, vol. 7, no. 3, pp. 25 - 40, 2002.
- [12] A. Y. Grama, A. Gupta and V. Kumar, "Isoefficiency: Measuring the scalability of parallel algorithms and architectures," *IEEE Parallel Distrib. Syst.*, vol. 1, pp. 12 - 21, 1993.
- [13] S. E. Hambrusch and A. A. Khokhar, "C<sup>3</sup>: A parallel model for coarse-grained machines," *Journal of Parallel Distrib. Comput.*, vol. 32, pp. 139 - 154, 1996.
- [14] M. L. Woo, "Parallel discrete wavelet transform on the Paragon MIMD machine," R. S. et. al., Ed, *Proceedings of the seventh SIAM conference on parallel processing for scientific computing*, pp. 3 - 8, 1995.

---

# List of Publications

## 1. Journals Papers

1. **V. P. Dinesh Kumar** and Tessamma Thomas, "Performance Study of Improved Legendre Moment Descriptor as Region-Based Shape Descriptor," *Journal of Pattern Recognition and Image Analysis*, vol. 18, no. 1, pp. 23-29, 2008.
2. **V. P. Dinesh Kumar**, Tessamma Thomas and K. V. Menon, "CBIR of Scoliosis Radiographs," *Journal of Spinal Disorders and Techniques*, 2008. (Accepted)
3. **V. P. Dinesh Kumar** and Tessamma Thomas, "A Comparative Study of Classification Accuracy of Orthogonal Moment Based Shape-Descriptors and Neural Networks," *AMSE Journal of Signal Processing and Pattern Recognition (I B)*, 2007. (Accepted)
4. **V. P. Dinesh Kumar** and Tessamma Thomas, "An Improved Legendre Moment Descriptor (ILMD) as Contour-Shape Descriptor," *AMSE Journal, Advances in Modelling, Series A: General Mathematics*, vol. 43, no.1, pp. 71- 84, 2006.
5. **V. P. Dinesh Kumar** and Tessamma Thomas, "Automatic Estimation of Position and Orientation of Spine in Digitized X-rays Using Mathematical Morphology," *Journal of Digital Imaging*, vol.18, no.3, pp.234-241, 2005.
6. P.S. John, James Chacko, Antony Joseph Thoppil, Tessamma Thomas, Ananda Resmi S, and **V. P. Dinesh Kumar**, "A Novel Computer Assisted Technique for Pedicle Screw Insertion," *International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 3, pp. 59-63, March 2007.

## 2. Conference Papers

1. **V. P. Dinesh Kumar**, V.P. Devassia, M.G. Mini and Tessamma Thomas, "A Parallel Multiple Subsequence Structure for Wavelet Packet Transform on MPPs," *Proc. of the International Conference on Emerging Technologies and Applications in Engineering, Technology and Sciences (ICETAETS - 2008)*, Gujrat, India, pp.733-737, January 13-14, 2008.
2. **V.P. Dinesh Kumar** and Tessamma Thomas, "Clustering of Invariance Improved Legendre Moment Descriptor for Content Based Image Retrieval," *Proc. of the*

- IEEE International Conference on Signal Processing, Communications and Networking (ICSCN-2008)*, Madras Institute of Technology, Chennai, India, pp.378-382, January 4-6, 2008.
3. **V. P. Dinesh Kumar**, Tessamma Thomas, and K.V. Menon, "Content Based Image Retrieval of Scoliosis Images," *Proc. of International Conference on Spinal Deformities*, Delhi, India, October 5 – 7, 2007
  4. **V. P. Dinesh Kumar**, Tessamma Thomas, and K.V. Menon, "Development of Software for Automatic Scoliosis Measurement," *National Conference on Biotechnology in Molecular Medicine*, Cochin, India, January 13 - 14, 2007.
  5. **V. P. Dinesh Kumar** and Tessamma Thomas, "Spine Vertebrae localization using Mathematical Morphology," *Proc. of Annual Meeting of the Society for Computer Applications in Radiology (SCAR)*, Vancouver, Canada, May 20-23, 2004
  6. **V. P. Dinesh Kumar**, Linesh J, Roopesh Kattan and Tessamma Thomas, "Invariance improved Legendre Moments as Contour-shape descriptor," *Proc. of Int. Conf. on Modelling and Simulation (MS'2004)*, Belarus, pp. 160-163, April 27-29, 2004.
  7. **V. P. Dinesh Kumar** and Tessamma Thomas, "A modified FFT- based algorithm for real-time computation of Discrete Wavelet Transform," *Proc. of the IEEE TENCON-2003*, Bangalore, pp.1030-1034, October 2003.
  8. Tessamma Thomas, **V. P. Dinesh Kumar**, P.S. John, Antony Joseph Thoppil, and James Chacko, "A Video Based Tracking System for Pedicle Screw Insertion," *Proc. of the 4<sup>th</sup> International Conference on Computer Science and Its Applications (ICCSA-2006)*, San Diego, California, USA, June 27-29, 2006.
  9. Ananda Resmi S, **V. P. Dinesh Kumar**, James Chacko, Antony Joseph Thoppil, P.S.John and Tessamma Thomas, "A Novel Computer Assisted Technique for Pedicle Screw Insertion," *Proc. of International Conference on Advanced Information and Telemedicine Technologies for health (AITTH '05)*, Belarus, October 2005.

---

# Resume

## Personal Profile

Name : Dinesh Kumar V.P.  
Sex : Male  
Date of Birth : 09-05-1974  
Area of Interest : CBIR and DBMS.  
Languages Known : English, Hindi and Malayalam  
Computer Skills : C/C++, Java, SQL, ODBC and JDBC  
SUN SOLARIS and MS WINDOWS  
MATLAB, PURIFY, QUANTIFY etc.  
Professional Experience : 4 years of software industry experience  
Permanent Address : "Sopanam"  
50/1351-D, Ambedkar Road  
North Edappally  
Cochin - 682024  
Kerala, India  
Email Address : dkvp10@yahoo.com



## Academic Profile:

Course	University/Institution	Year of Passing	% of Marks	Remarks
Ph.D	Cochin University of Science and Technology	Doing	N.A	Full Time
M.Sc. (Electronics Science)	Cochin University of Science and Technology	1996	87.4	First Rank
B.Sc. (Physics)	M.G.University	1994	89.7	First Class

## **Major Achievements**

- First Rank Holder in M.Sc Electronic Science degree of Cochin University of Science & Technology in 1996.
- Awarded K.G. Nair Endowment Gold Medal
- Granted senior research fellowship from CSIR, India.
- Member of Academic Council of Cochin University of Science & Technology during 2004-2005.

## **Professional Experience:**

M/s Dharma Systems (P) Ltd  
94, 4<sup>th</sup> B Cross, Industrial Layout,  
Koramangala 5<sup>th</sup> Block, Bangalore – 560 095  
India.  
Ph: 080 25533622  
Position Held: Senior Software Engineer

He was involved in the Design/Development of various software products of Dharma Systems. Integrator is a Distributed Database Management System developed by them to provide a unified SQL access to different datasources. Dharma SQL engine and Applink are components (stand alone) of Integrator which provides SQL interface to legacy file based DBMS (like ISAM) and mainframe databases (like CICS). He was the project leader for various version releases of Integrator and has also worked in the performance improvement projects of many these software products. He also lead a team for the development and release of an SQL engine called c-tree SQL Server ([www.faircom.com](http://www.faircom.com)) based on Dharma SQL engine with FairCom corporation's c-tree as the file manager. This project was done as offshore project with duration of 18 months. The other projects were of short durations of 3/6 months. A list of important projects undertaken is given below.

- *C-tree SQL server project*
- *Design of Integrator Admin tool enhancement*
- *Integrator ODBC gateway enhancement*
- *Performance improvement project for i2 technologies*
- *Integrator release activities*
- *Performance improvement project for Applink and Integrator*
- *Solaris porting of Integrator*