

# **A New Approach to Automatic Saliency Identification in Images Based on Irregularity of Regions**

Mohammad Al-Azawi

A thesis submitted for the degree of Doctor of Philosophy  
Centre for Computational Intelligence  
Faculty of Technology  
De Montfort University

January - 2015

## **Preface**

The work outlined in this dissertation was carried out at the Centre for Computational Intelligence, Faculty of Technology, De Montfort University, over a three-year period from October 2011 to Jan 2015. This dissertation is the result of my work and primarily discusses the outcomes of work done in collaboration, except for a few instances, which are stated in the text and cited.

The material included in this thesis has not been submitted for a degree or diploma or any other qualification at any other university. Furthermore, no part of my dissertation has already been or is currently submitted for any such degree, diploma or other qualification.

## **Acknowledgements**

I would like to express my deep sense of gratitude to my supervisors, Dr. Yingjie Yang and Dr. Howell Istance for their patience, continuous support, encouragement and being available whenever I needed their opinion.

## **Abstract**

This research introduces an image retrieval system which is, in different ways, inspired by the human vision system. The main problems with existing machine vision systems and image understanding are studied and identified, in order to design a system that relies on human image understanding. The main improvement of the developed system is that it uses the human attention principles in the process of image contents identification. Human attention shall be represented by saliency extraction algorithms, which extract the salient regions or in other words, the regions of interest. This work presents a new approach for the saliency identification which relies on the irregularity of the region. Irregularity is clearly defined and measuring tools developed. These measures are derived from the formality and variation of the region with respect to the surrounding regions. Both local and global saliency have been studied and appropriate algorithms were developed based on the local and global irregularity defined in this work.

The need for suitable automatic clustering techniques motivate us to study the available clustering techniques and to development of a technique that is suitable for salient points clustering. Based on the fact that humans usually look at the surrounding region of the gaze point, an agglomerative clustering technique is developed utilising the principles of blobs extraction and intersection. Automatic thresholding was needed in different stages of the system development. Therefore, a Fuzzy thresholding technique was developed.

Evaluation methods of saliency region extraction have been studied and analysed; subsequently we have developed evaluation techniques based on the extracted regions (or points) and compared them with the ground truth data.

The proposed algorithms were tested against standard datasets and compared with the existing state-of-the-art algorithms. Both quantitative and qualitative benchmarking are presented in this thesis and a detailed discussion for the results has been included. The benchmarking showed promising results in different algorithms. The developed algorithms have been utilised in designing an integrated saliency-based image retrieval system which uses the salient regions to give a description for the scene. The system auto-labels the objects in the image by identifying the salient objects and gives labels based on the knowledge database contents. In addition, the system identifies the unimportant part of the image (background) to give a full description for the scene.

# Table of Contents

Preface .....	i
Acknowledgements .....	ii
Abstract.....	iii
Table of Contents .....	iv
List of Figures.....	viii
List of Abbreviations .....	xiii
List of Symbols.....	xv
Chapter 1. Introduction .....	1
1.1 Machine Vision and Content-Based Image Retrieval.....	1
1.2 Saliency Based Image Retrieval Systems (SBIR) .....	3
1.3 Problems with MV and CBIR.....	4
1.3.1 The Semantic Problem .....	4
1.3.2 Whole versus Parts .....	6
1.4 Feasibility of Computational Intelligence.....	7
1.4.1 Uncertainty and Incompleteness of Images.....	7
1.4.2 Adaptivity and Learning.....	7
1.5 The Motivation and the Goal .....	7
1.6 Contributions .....	8
1.7 Organization.....	9
1.8 Publications.....	10
1.9 Conclusions.....	11
Chapter 2. Visual Perception.....	12
2.1 Introduction.....	12
2.2 Visual Perception .....	12
2.2.1 Human Vision System (HVS) and Machine Vision Systems (MVS) .....	13
2.2.2 Human Colour Perception .....	14
2.3 Human Attention.....	15
2.3.1 Attention .....	16
2.3.2 Pre-attentive Phase .....	18
2.3.3 Attentive Phase .....	18
2.3.4 Top-Down and Bottom-Up Attention .....	19
2.3.5 Fixations and Saccades.....	19
2.3.6 Saliency .....	20

2.4	Content – Based Image Retrieval .....	21
2.5	Image Processing and Machine Vision.....	24
2.6	Image Feature Extraction.....	25
2.6.1	Low-Level Features .....	25
2.6.2	High Level Features .....	30
2.7	Image Segmentation .....	31
2.7.1	Image Thresholding.....	31
2.7.2	Colour Image Segmentation .....	33
2.8	Clustering.....	34
2.8.1	Clustering Types.....	35
2.8.2	Basic Models of Clustering .....	36
2.8.3	Sequential Clustering Technique (SCT).....	40
2.8.4	Parallel Clustering Technique (PCT) .....	43
2.8.5	Iterations Stopping Criteria .....	44
2.8.6	Application on Points Clustering.....	46
2.9	Computational Intelligence.....	46
2.9.1	Machine Learning Techniques .....	47
2.9.2	Artificial Neural Networks .....	48
2.9.3	Uncertainty and Fuzzy Intelligence.....	51
2.10	Conclusions.....	53
Chapter 3.	Data Collection and Analysis .....	54
3.1	Introduction.....	54
3.2	Datasets .....	54
3.2.1	Saliency Datasets .....	54
3.2.2	Image Retrieval Dataset .....	55
3.3	Saliency Ground Truth Data .....	57
3.3.1	Manual Ground Truth Data (MGTD).....	57
3.3.2	Eye Tracking Ground Truth Data (ETGTD) .....	60
3.4	Benchmarking Data .....	62
3.5	Saccade Points Extraction.....	63
3.6	Conclusions.....	64
Chapter 4.	Saliency Identification and Evaluation.....	65
4.1	Introduction.....	65
4.2	Salient Points and Regions Extraction.....	67

4.2.1	Wavelet-Based Techniques .....	67
4.2.2	Location-Based Saliency .....	68
4.2.3	Background Suppression .....	68
4.2.4	Corner-Based Techniques .....	69
4.2.5	Feature Maps-Based Saliency .....	69
4.2.6	Frequency Spectra-Based Saliency .....	70
4.2.7	Colour-Based Saliency .....	70
4.3	Evaluation of Saliency .....	71
4.3.1	Previous Work .....	72
4.3.2	Proposed Evaluation Measures.....	75
4.4	Salient and Gaze Points Clustering.....	91
4.5	Blobs – Based Clustering (BBC) .....	91
4.6	Iterative Blobs-Based Clustering (IBBC) .....	95
4.7	Irregularity-Based Saliency Identification (IBSI).....	104
4.7.1	Statistical Measures as Descriptors .....	105
4.7.2	Local Saliency Identification (LSI) .....	110
4.7.3	Global Saliency Identification (GSI).....	135
4.7.4	Two Steps Saliency Identification (TSSI) .....	138
4.7.5	Improving the TSSI .....	139
4.7.6	Results .....	143
4.7.7	Benchmarking.....	147
4.8	Neural Network Based Salient Points Clustering .....	149
4.9	Conclusions.....	153
Chapter 5.	Saliency Based Image Contents Identification (SBICI) .....	154
5.1	Introduction.....	154
5.1	Image Contents Identification.....	155
5.2	Colour – Based CBIR .....	156
5.2.1	Bin-by-Bin Distance (BBD) .....	156
5.2.2	Cross-Bin Distance (CBD) .....	158
5.2.3	Global vs. Local Histogram.....	159
5.3	Texture – Based CBIR .....	160
5.4	Saliency Based Image Retrieval (SBIR).....	165
5.5	Distance Measures .....	170
5.6	Evaluation of Image Retrieval Techniques.....	177

5.7	Object Identification Experimental Results .....	179
5.8	Background Identification .....	181
5.8.1	Artificial Neural Network Based Background Identification.....	182
5.8.2	Background Identification Features .....	184
5.8.3	Neural-Fuzzy-Based Background Identification .....	188
5.9	Background Identification Experimental Results .....	192
5.10	Conclusions.....	194
Chapter 6.	Discussion and Conclusions .....	195
6.1	Introduction.....	195
6.2	General Review.....	195
6.3	Academic Contributions .....	198
6.4	Conclusions.....	200
6.5	Limitation and Further Possible Improvement .....	202
References	.....	204



## List of Figures

Figure 1-1: Results obtained from Google.com, (a) query image, (b) the retrieved images. ....	5
Figure 1-2: The effect of rotation in the query image on the results obtained from Google.com, (a) query image, (b) the retrieved images, (c) query Image rotated by 90 degree, (d) the retrieved images. ....	6
Figure 2-1: Different importance may be given for an object. ....	16
Figure 2-2: Attention phases (Aspects). ....	17
Figure 2-3: Pre-attentive visual pop-out features. ....	19
Figure 2-4: Number of CBIR researches vs. year in (a) IEEE, (b) Google. ....	22
Figure 2-5: Image Processing Chain [57]. ....	25
Figure 2-6: GLCM extraction, (a) intensity values, (b) $GLCM(x, y, 1, 0^\circ)$ , (c) $GLCM(x, y, 1, 180^\circ)$ . ....	29
Figure 2-7: Thresholding technique using histogram analysis. (a) Original grey image; (b) histogram of the grey image; (c) selecting arbitrary threshold (100), (d) selecting the threshold value equal to the valley (200). ....	32
Figure 2-8: Typewritten text thresholding with bimodal technique, (a) original grey image, (b) histogram of the grey image, (c) selecting the threshold value equal to the valley (200). ....	33
Figure 2-9: Converting colour image into one band image: (a) original image; (b) RGB averaging conversion; (c) histogram of (b). ....	34
Figure 2-10: Clustering techniques. ....	36
Figure 2-11: Hard-clustering example. ....	37
Figure 2-12: Distance Matrices, (a) $D$ before applying thresholding, (b) $S$ after thresholding ....	38
Figure 2-13: Soft-clustering example. ....	39
Figure 2-14: Mapping of objects to clusters using soft (Fuzzy) clustering. ....	40
Figure 2-15: SCT clustering results. ....	41
Figure 2-16: Clustering process, (a) points to be clustered, (b) results of applying sequential clustering. ....	41
Figure 2-17: RSCT process. ....	42
Figure 2-18: Clustering process, results of applying RSCT, (a) no threshold is used, (b) $T=200$ , (c) $T=100$ . ....	42
Figure 2-19: Results of applying PCT with different number of clusters $nT$ , (a) $nT = 1$ , (b) $nT = 2$ , (c) $nT = 3$ (d) $nT = 6$ . ....	45
Figure 2-20: Points clustering example, (a) gaze points, (b) gaze points after removing the background, (c) SCT, (d) RSCT $T = \infty$ , (e) RSCT $T = 100$ , (f) RSCT $T = 50$ , (g) PCT $nT = 1$ , (h) PCT $nT = 4$ . ....	47
Figure 3-1: High retrieval rate of WANG dataset. ....	56
Figure 3-2: Manual points extraction, (a) software interface, (b) points clicked by the user, (c) clustered points. ....	58

Figure 3-3: The distribution of the participants based on (a) gender, (b) age, (c) nationality, and (d) education level. ....	59
Figure 3-4: The data obtained from OGAMA, (a) the gazing points, (b) the fixation points, (c) the regions of interest, (d) the gazing raw data. ....	60
Figure 3-5: The distribution of the participants based on (a) gender, (b) age, (c) nationality, (d) education level. ....	62
Figure 3-6: Extracting the saccade points $\gamma = 0.5$ .....	64
Figure 4-1: Two different images with similar colour distribution and statistical measures. ....	66
Figure 4-2: Colour-based image segmentation (a) original image, and its reduced colour image, (b) segments obtained by applying colour based segmentation. ....	66
Figure 4-3: The extraction of the centroid from a set of points, (a) the points on the original image, (b) the same points on a white background, (c) the centroid, (the yellow circle), (d) the centroid and points on a white background. ....	78
Figure 4-4: Comparing two points sets, (a) automatically extracted points, (b) manually labelled points, (c) the distance between centroids, (d) the same distance in (c), (d) the distance between the centroid magnified. ....	79
Figure 4-5: The distance between the centroid of the computationally extracted data and the eye-tracking data, (a) fixations obtained from the eye tracking data, (b) the distance between the centroids. ....	79
Figure 4-6: The effect of the weighting of the centroids on the overall centroid (a) without weighting, (b) with weighting, (c) weighted using equation 4-14 ....	83
Figure 4-7: Comparing two different regions using XOR, (a) first saliency map, (b) second saliency map, (c) binary saliency map of (a) , (d) binary saliency map of (b), (e) the result of applying the XOR between the two maps in (c) and (d). ....	85
Figure 4-8: Sample of the falsely detected salient points that were used in the experiment, (a) points, (b) regions. ....	86
Figure 4-9: The effect of the number (quantity) of the falsely detected salient points FDSP.....	88
Figure 4-10: Sample of the falsely detected salient points that were used in the experiment 2, (a) points, (b) regions. ....	89
Figure 4-11: The effect of the distribution of the falsely detected salient points FDSP. ....	91
Figure 4-12: The selection of the size of the blob, (a) points, (b) $\beta = 0.02$ , (c) $\beta = 0.10$ , (d) $\beta = 0.20$ . ....	93
Figure 4-13: Union of overlapped blobs to form an interesting region, (a) overlapped blobs, (b) merging the blobs to form a region, (c) extracting the dimensions of the region, (d) the interesting region. ....	94
Figure 4-14: Converting gazing points to blobs and constructing clouds from the overlapped blobs, (a) $\beta = 0.02$ , (b) $\beta = 0.10$ , (c) $\beta = 0.15$ .....	95
Figure 4-15: Converting gaze points to blobs, (a) $\beta = 0.02$ , (b) $\beta = 0.10$ , (c) $\beta = 0.20$ . ....	96

Figure 4-16: Changes of variables with iterations, (a) points to be clustered, (b) clustered points, (c) internal distance, (d) external Distance, (e) internal standard deviation, (f) external standard deviation, (g) $\omega x n$ vs. $\omega c n$ and $\epsilon n$ , (h) $\Delta \epsilon(n)/\Delta n$ . ....	101
Figure 4-17: Numerical example on formal and irregular regions, (a) formal region, (b) irregular region, (c) difference from mean for (a), (d) difference from mean for (b), (e) $Jx, y - \mu R \times \sigma R$ of (a), (f) $Jx, y - \mu S \times \sigma S$ of (b). ....	113
Figure 4-18 one dimensional representation of (a) R, (b) S, (c) R and S, (d) 1D $Jx, y - \mu R \times \sigma R$ of (a) and $Jx, y - \mu S \times \sigma S$ of (b). ....	114
Figure 4-19: Image and space representation, (a) original image, (b) intensity image, (c) mean image, (d) Cartesian representation of the mean image, (e) difference between intensity and mean images, (f) Cartesian representation of the difference image. ....	115
Figure 4-20: The results of applying the irregularity measures on an image, (a) standard deviation image, (b) Cartesian representation of the standard deviation image, (c) variation values , (d) Cartesian representation for the variation. ....	116
Figure 4-21: Examples of applying the irregularity on images to extract the salient object. ....	117
Figure 4-22: Average distance measured using XORD, (a) vs. sub-image size, (b) vs. overlapping. ....	120
Figure 4-23: Bimodal thresholding technique, (a) normal technique, (b) the application of fuzzy membership function in this technique. ....	124
Figure 4-24: The application of saliency enhancement, (a) original image, (b) the saliency enhanced image ISM. ....	126
Figure 4-25: Fuzzy membership function representation of the histogram. ....	127
Figure 4-26: Applying Fuzzy bimodal histogram thresholding, (a) NS region, (b) LS regions, (c) S regions, (d) VS regions. ....	128
Figure 4-27: Applying the thresholding technique on the saliency enhanced image, (a) important regions, (b) important object. ....	128
Figure 4-28: Illustration of a case where there is a regular region inside the object, (a) irregularity map, (b) points clustering and merging using IBBC after 4 iterations, (c) same as (b) but when IBBC stops after 8 iterations, and (d) the extracted objects. ....	129
Figure 4-29: Comparison with edge detector, (a) irregularity map, (b) edge map obtained by Canny edge detector, (c) edge map obtained from Sobel edge detector. ....	129
Figure 4-30: LSI benchmarking with state-of-the-art algorithms, (a) F-Measure, (b) different measures comparison. ....	131
Figure 4-31: XORS measure for different saliency extraction methods and for different images classes. ....	132
Figure 4-32: Histogram Smoothing, (a) Original Histogram, (b) Smoothed histogram. ....	137
Figure 4-33: Global Salient Region Extraction, (a) original image, (b) saliency enhanced image, (c) the salient object after thresholding. ....	137
Figure 4-34: TSSI application on an image. ....	138
Figure 4-35: Numerical example of applying BMF. ....	139

Figure 4-36: Example of applying BMF on real image, (a) binary image with FDSP, (b) the resultant image after applying BMF. ....	140
Figure 4-37: Extracting the object from the image, (a) using the region extracted from LSI, (b) using TSSI, (c) mask, (d) using RTSSI. ....	142
Figure 4-38: Comparison of the three methods of salient objects extraction. ....	143
Figure 4-39: The distance between different saliency identification methods and the ground truth data. ....	143
Figure 4-40: Stages and processes of the refined two stage saliency identification (RTSSI) block diagram. ....	144
Figure 4-41: Example of applying the RTSSI algorithm, (a) original image, (b) local saliency identification, (c) local masks, (d) global saliency enhancement, (e) binary saliency map after thresholding, (f) ANDing the masks in (c) and the binary image in (e), (g) applying BMF, (h) extracting the masks, (i) the obtained objects. ....	145
Figure 4-42: Quantitative comparison with different saliency extraction techniques, (a) F-Measure curves, (b) Exclusive OR curves, (c) average measures comparison. ....	149
Figure 4-43: The application of neural network on salient regions extraction, (a) original image, (b) important points (red) and saccade points (blue), (c) saliency map, (d) the extracted objects. ....	152
Figure 4-44: The application of the trained neural network on salient regions extraction on images similar to the image in Figure 4-43. ....	152
Figure 5-1: CBIR basic diagram. ....	155
Figure 5-2: Experimental results of CBIR based on histogram comparison using Minkowski distances between (b) red histogram, (c) green histogram, (d) blue histogram, and (e) grey histogram. ....	158
Figure 5-3: Finding the match for the query image using the formula given in Equation 5-17, (a) grass texture query image, (b) cloud texture query image. ....	164
Figure 5-4: finding the matched for the query image using the formula given in Equation 5-19 , (a) grass texture query image, (b) cloud texture query image. ....	165
Figure 5-5: SBIR System. ....	166
Figure 5-6: Histogram for the bands of an image. ....	168
Figure 5-7: Histogram components. ....	169
Figure 5-8: Example of simple image and its histogram, (a) image pixels' values, (b) histogram for the entire image, (c) histogram of the region surrounding the object. ....	173
Figure 5-9: Separating the histogram into (a) object histogram and (b) background histogram. ....	174
Figure 5-10: Actual histogram for (a) object and (b) background. ....	174
Figure 5-11: Comparing the measures of the histogram cut and object cut, (a) object, (b) background. ....	175
Figure 5-12: Object and background features membership functions. ....	175
Figure 5-13: Distance measures for the three possible cases, RBII, EII, and OBII ....	177
Figure 5-14: Evaluation of the three retrieval ways, (a) Precision, (b) Recall. ....	178
Figure 5-15: Retrieving evaluation using WEEM. ....	179

Figure 5-16: Evaluation and comparison among EII, RBII, and OBI, (a) precision, (b) recall, (c) WEEM, (d) average measures.....	180
Figure 5-17: the variation of the statistical measures with respect to coarseness (a) Measures that change directly with coarseness, (b) measures that do not change directly with coarseness. ....	186
Figure 5-18: The effect of irregularity on the GLCM, (a) original image, (b) 1D GLCM 16 grey levels, (c) 2D GLCM 16 grey levels, (d) 1D GLCM 256 grey levels. ....	187
Figure 5-19: The variation of the statistical measures with respect to regularity (a) measures that change directly with regularity, (b) measures that do not change directly with regularity. ....	187
Figure 5-20: GLCM, (a) original image, (b) 16 grey level GLCM, (c) 256 grey level GLCM, (d) 16 grey level 2-D GLCM histogram, (e) 256 grey level 2-D GLCM histogram, (f) 16 grey level 3-D GLCM contour, (g) 16 grey level 3-D GLCM contour, (h) 16 grey level 2-D GLCM histogram, (i) 256 grey level 2-D GLCM histogram. ...	188
Figure 5-21: The proposed algorithm block diagram. ....	190
Figure 5-22. FSC vs. Hue, Saturation=0.5 and Value = 0.5.....	191

## List of Abbreviations

Abbreviation	Definition
AB	Appearance-Based Methods Saliency
ANN	Artificial Neural Network
ANNs	Artificial Neural Networks
AUC	Area Under the Curve
BBC	Blobs-Based Clustering
BBD	Bin-by-Bin Distance
BMF	Binary majority filter
BMHT	Bimodal Histogram Thresholding
BPANN	Back Propagation Artificial Neural Network
BUFB	Bottom-Up Feature-Based Approaches Saliency
CBD	Cross-Bin Distance
CBIR	Contents-Based Image Retrieval
CC	Correlation Coefficient
CCG	Cluster Centre of Gravity
CED	Computationally Extracted Data
CI	Computational Intelligence
DSSO	Different scenes with similar objects
ECL	Error Correction Learning
EEM	Efficiency Evaluation Measure
EII	Entire Image Identification
EMD	Earth Mover's Distance
ETD	Eye Tracking Data (interest points extracted using eye tracker)
ETGT	Eye Tracking Ground Truth
FBMT	Fuzzy-Based Bimodal thresholding
FDSP	falsely detected salient point
FIS	Fuzzy Interfacing Systems
FLV	Fuzzy Linguistic variables
FPR	false positive rate
FPSP	First Point Selection Problem
GCH	Global Colour Histogram
GLCMs	Grey-Level Co-occurrence Matrices
GSI	Global Saliency Identification
GSM	Global Saliency Mask
HANN	Hopfield Artificial Neural Network
HLF	High-level features
HOG	Histograms of Oriented Gradients
HSI	Hue, Saturation, and Intensity Colour Model
HSL	Hue, Saturation, and Lightness Colour Model
HSV	Hue, Saturation, and Value Colour Model
HT	Hough Transform
HVS	Human Vision System
IBBC	Iterative Blobs-Based Clustering
IBH	Intersection Between the Histograms
IBSI	Irregularity-Based Saliency Identification
IOP	Integration of Parts Saliency
IPC	Image Processing Chain
IRS	Image Retrieval System
KLD	Kullback-Leibler Divergence
LCH	Local Colour Histogram
LLF	Low-level features
LSI	Local Saliency Identification
LSM	Local Saliency Mask
MBL	Memory – Based Learning
MD	Manual Data (interest points extracted manually)

<b>Abbreviation</b>	<b>Definition</b>
MGTM	Manual Ground Truth Map
MLANN	Multi-Layers Neural Net
MSE	Mean Squared Error
MV	Machine Vision
MVS	Machine Vision System
OBII	Object-Based Image Identification
PAR	Precision and Recall
PCT	Parallel Clustering Technique
QD	Quadratic Distance
RBCD	Region-Based Centroid Distance
RBII	Region-Based Image Identification
RGB	Red, Green, and Blue Colour Model
RLT	Reinforcement Learning Technique
ROC	Receiver Operating Characteristics
ROI	Region of Interest
RSCT	Refined Sequential Clustering Technique
RSM	Refined Saliency Mask
RTSSI	Refined Two Steps Saliency Identification
SBICI	Saliency Based Image Contents Identification
SBII	Saliency Based Image Identification
SBIR	Saliency Based Image Retrieval
SCD	Single Centroid Distance
SCT	Sequential Clustering Technique
SIFT	Scale Invariant Feature Transform
SLANN	Single Layer Artificial Neural Net
SLT	Supervised Learning Technique
SOA	singularity of attention
SOMANN	Self-Organizing Map Artificial Neural Net
SSDP	Same Scene with Different Perspective
TDKB	Top-Down Knowledge-Based Saliency
TM	Template Matching Saliency
TPR	True Positive Rate
TSSI	Two Steps Saliency Identification
ULT	Unsupervised Learning Technique
WEEM	Weighted Efficiency Evaluation Measure
WRBCD	Weighted Region-Based Centroid Distance
XORD	Exclusive OR – Based Distance
XORS	Exclusive OR – Based Similarity

## List of Symbols

symbol	Definition
$p_{ij}$	Pixel intensity value at location $(i, j)$
$R$	Image region
$J(x, y)$	Image Function
$\mathbb{I}$	Representing the image as a set of pixels
$\mathcal{B}(x, y)$	Binary Image
$T$	Threshold value
$D(X, Y)$	Distance measure between X and Y
$S(X, Y)$	Similarity measure between X and Y
$P$	Set of extracted salient points
$H$	Set of ground truth salient points
$\mathbb{C}$	Cloud of points (Special set of points)
$\alpha, \beta$	Constants and usually used as tuning factors
$E(\cdot)$	Expected value function
$\mu$	Mean
$v_f$	First order variation measure
$\sigma$	Standard deviation
$v^{norm}$	Normalised variation measure
$\mathbb{P}_{\mathbb{I}}$	Power set of the image set $\mathbb{I}$
$\mathbb{S}$	Subset of the image set $\mathbb{I}$
$\mathbb{r}$	Sub-region in an image
$\mathbb{R}$	Set of all regions in an Image
$\psi$	Feature extraction function
$\mathbb{d}$	Description vector
$\mathbb{D}$	Set of descriptions
$\mathbb{R}_I$	Set of irregular regions
$\mathbb{R}_U$	Set of Regular (unimportant) regions
$\mathcal{H}$	Histogram represented as a set
$\mathcal{O}$	Set of grey levels that belong to object
$\mathcal{B}$	Set of grey levels that belong to background
$\mathcal{A}$	Angular Second Moment (ASM)
$\mathcal{C}$	Contrast
$\mathcal{C}$	Correlation
$\mathcal{M}$	Inverse Difference Moment (IDM)
$\mathcal{E}$	Entropy
$\mathcal{D}(\cdot, \cdot)$	Fuzzy Distance Measure
$H_I$	Important information in an image
$H_U$	Unimportant information in an image
$R_I$	Important region
$R_U$	Unimportant region



# Chapter 1

## Introduction

### 1.1 Machine Vision and Content-Based Image Retrieval

As digital images are becoming increasingly important across various fields, the need to recognize, classify and understand these images has increased. This need was the motivation behind the launch of the machine vision and image-understanding technologies.

To make the digital images useful in different applications, it is important to analyse these images and describe them automatically using image processing techniques and machine vision (MV). MV techniques use a variety of techniques to process, analyse and interpret images, which lead to a better machine understanding of the image contents. MV techniques convert images into a set of numerical values, known as features, and the process of converting the image into this set of values is known as feature extraction. For instance, the features might be texture, colour, edges, or boundaries. These features are used to describe the image and can be utilised in image matching and retrieval.

Retrieving images based on their labels, text description, or keywords is widely used in different situations, such as web applications. With such methods, the search depends on matching the query text with the description of the image, which is stored in the database. These techniques suffer from certain limitations, such as, the need for human intervention to extract the keywords for each image, which is a time consuming process. In addition, the keyword extraction process is subjective and depends on the human point of view

and how the human interprets the image and describes it [1]. Thus, alternative methods which can identify the visual contents of the image are strongly needed.

Content-based image retrieval systems (CBIR) are image retrieval systems that rely on the contents of the image in order to identify it. They are application-oriented techniques, and the definitions may differ from one application to another. For example, in crime prevention CBIR systems, if a fingerprint search is required, then the entire image would be considered and the systems would search for a specific image from a set of closely similar images. However, in the case of general search engines, when a person searches for a fingerprint, then how identical the query and the retrieved images are, is not relevant. This leads to the classification of similarity in the CBIR into three types as follows:

1. Conceptual similarity, in which the retrieved images have the same meaning of the query image. In this type, the meaning is more relevant than the similarity.
2. Perceptual similarity, in which the exact image is needed to be retrieved. For example, searching for a specific person in an image based on some other portrait.
3. Computational similarity, in which, regardless of the meaning or the contents of the image, some images might be computationally similar, based on the extracted features, even if they are not similar in meaning.

Most of the problems are due to the machine's semantic lack in image understanding, and because they utilize just the visual information of the image which is insufficient to give satisfactory results. Although some systems have introduced new features and better visual recognition capabilities, the results are still unsatisfactory.

Most well-known search engines, such as Google, Yahoo and Bing, present image search engines. Until recently, some search engines were using textual image search techniques that depend on the contents of the webpage which contains the queried image in order to retrieve the image. Others use tags and labels, where each image is described by a set of tags given by the user to improve the search process.

Recently, visual search engines have become available. These search engines accept an image from the user and search for images with similar visual features. However, the meaning of the image is not taken into consideration, causing the result to be semantically unacceptable.

## 1.2 Saliency Based Image Retrieval Systems (SBIR)

Not much effort has been done in the field of SBIR as compared to the tremendous amount of researches in the field of CBIR. In SBIR, the retrieval algorithms are applied only on the salient regions. This shall reduce the computations required in the retrieval process. Several image retrieval systems which utilise the saliency properties of the regions have been developed. The majority of the published algorithms have utilised Itti visual attention model [2] to identify the salient regions (or important regions) such as Hu et al. [3], Zdziarski and Dahyot [4], Ozyer and Vural [5], and de Carvalho Soares et al. [6]. After identifying the salient regions, various similarity measures were used to identify the contents of the salient regions and retrieve similar images.

Many algorithms were developed to be specialised and work on specific types of data such as Hu et al. approach [3] which was developed for flower image retrieval. Zdziarski and Dahyot [4] have presented another specialised approach; they applied their method on three classes of images, faces, houses and flowers. Natural images retrieval was the focus of Wang et al. [7] in which they presented a specialised algorithm to identify the nature of the images. Hua et al. [8] have used the principles of saliency to design a CBIR system to identify the images of lunar surface.

Colour and intensity were used by many authors to highlight the saliency of a region. Wan et al. [9] have used HSV colour space in highlighting the salient regions in the image. They adopted the change in intensity as a measure for saliency and used the colour features as a measure of similarity among images. The Colour Saliency Histogram (CSH) was presented by Lei et al. [10] as a measure of saliency. They have argued that the more concentrated pixels are corresponding to the colour the more visual stimulation is. After highlighting the salient regions, they reduced the number of colours by quantising the image and extract HSV histogram to be used as a similarity measure. Colour saliency extraction was used, also, by Chen and Wang [11]. Liang et al. have used saliency extracted from intensity, colour and orientation maps to cluster the images before the retrieval process [12]. They have used 1000 images in their test, of which 500 images were used to train the neural network to classify the images into two classes, while the other 500 images were used for testing purposes. An et al. [13] have also used HSV colour space in identifying the salient regions. In their approach, they have suggested the

use of colour contrast and the centre of the image to highlight the salient regions, then they used colour features and spatial binary map as similarity measures.

Liu et al. [14] have suggested the use of saliency maps to reduce the time consuming and the dimensions of the features vector of Scale Invariant Feature Transform (SIFT) algorithms. They have also used intensity and colour change as saliency extraction features and they used the features of the salient regions only for matching the images.

de Carvalho Soares et al. [6] have, also, used Itti et al.'s visual attention models to identify the foreground and background saliency to extend the bag-of-visual-words method by weighting the visual words according to their spatial locality which depends on the saliency extraction. Saliency and the bag-of-visual-words have, also, been used by some other authors such as Gao et al. [15], Giouvanakis and Kotropoulos [16].

In most of the algorithms discussed above, it was noted that the authors did not pay enough attention to the saliency extraction problems as they have applied their algorithms on simple images. Images with one object with smooth background and high foreground-background contrast have been used to verify the proposed algorithms. In addition, Itti visual attention approach, usually, highlights small spots of the image sequentially to identify the regions that grabs the human attention. Therefore, it is not suitable to extract the object of interest and instead it highlights only part of it. Using colour and intensity in saliency identification is another limitation of these algorithms as these features are insufficient to be used in saliency identification, as it will be discussed in this thesis.

### **1.3 Problems with MV and CBIR**

#### **1.3.1 The Semantic Problem**

The semantic gap between machine and human in image understanding is a big challenge in machine vision systems (MVS), and most of these systems suffer from this semantic issue. The majority of retrieval systems, which rely on visual features, suffer from retrieving irrelevant images or images with similar visual features but with a different meaning as shown in Figure 1-1. In this figure, it is apparent that the search was executed based mainly on the colour or intensity features, hence irrelevant images were retrieved. Another difficulty for these systems is the change in the results when the image is rotated. In Figure 1-2, when the non-rotated image in (a) is used as a query image, the returned

images were quite satisfactory. In contrast, as shown in (b), when the same image is rotated 90 degree, the retrieved images were very different and could not be considered as similar images to the one being queried. Similar problems have been observed in other visual search engines and the reasons for the erroneous results are being investigated. There are many reasons behind this semantic problem, such as a deficiency in the machine's knowledge, the uncertain nature of the image and the image's incompleteness.

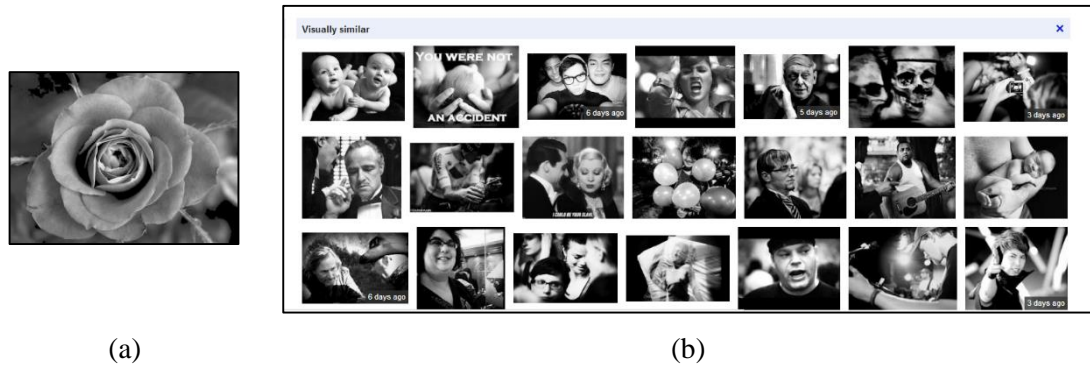


Figure 1-1: Results obtained from Google.com, (a) query image, (b) the retrieved images.

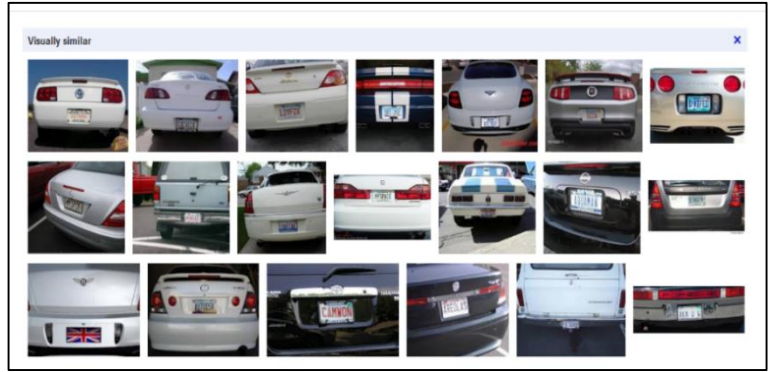
By studying human vision systems (HVS), better MVS can be developed by including some features from HVS such as attention and focusing. The main dissimilarity between HVS and MVS is the way in which the human sees the image. The human depends on the identification of concepts and objects to recognize the contents of an image, while the machine can only take in pixels and binary representations. Another significant dissimilarity is the knowledge, as humans use knowledge acquired through continuous learning to identify objects, whilst machines use only a limited set of features to identify objects.

The ability to learn is an important part of HVS; humans can learn new things very easily, whilst for the machine this is not possible even with the current availability of intelligent systems that present learning abilities. For example, even with the neural network, which has learning abilities, continuous learning is still a major issue, since it needs to be retrained each time from the beginning.

Finally, intentions, feelings, and emotions are extremely significant for HVS, while they are not available in traditional MVS. These HVS features are part of the recognition of the image since the human can recognize feelings like happiness, sadness and loneliness in the image and describe the image in this manner.



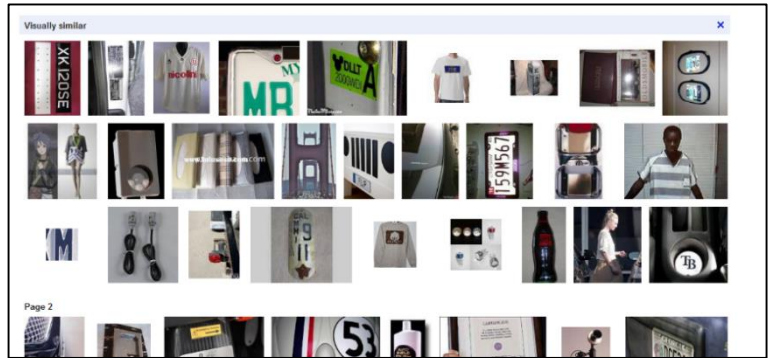
(a)



(b)



(c)



(d)

Figure 1-2: The effect of rotation in the query image on the results obtained from Google.com, (a) query image, (b) the retrieved images, (c) query Image rotated by 90 degree, (d) the retrieved images.

### 1.3.2 Whole versus Parts

The majority of retrieval systems recognize the image as a whole, especially if the image is large and presenting homogeneous details and objects. This causes erroneous results since it only takes into consideration the global features of the image, which might be similar for different images. Thus, recognition by parts or by objects is considered as a solution for these types of issues. One solution is to divide the image into smaller sub-images and extract the local features for each sub-image. The process of dividing the image into related regions is known as *image segmentation*. Various image segmentation techniques have been proposed based on colour, texture, adjacency, and other features. The drawback of these techniques is that they partition the image into small chunks or over-partition the image. Consequently, the recognition process will be computationally inefficient, as it needs to identify a large number of small parts. The significance of the

object in the image is important in solving such kinds of problems. Assigning a different significance level to each region is known as ‘region of interest extraction’.

## **1.4 Feasibility of Computational Intelligence**

The main two reasons for adopting computational intelligence in MV are the uncertainty in the image and the need to have adaptive systems that have learning abilities.

### **1.4.1 Uncertainty and Incompleteness of Images**

The uncertainty in images is another limitation for many applications. The overlap between the segments in segmentation process is an example of the uncertainty effect. This overlap disrupts the rule of segmentation which requires that the segments should be disjointed. Incompleteness and the shadow of the object can also be considered as part of the problem, as for example, some objects may present different shapes when rotated. One feasible solution for such issues is the use of uncertainty theory and fuzzy logic.

### **1.4.2 Adaptivity and Learning**

As images are dissimilar, one needs to consider systems that have the ability to learn new things and to adapt themselves to apply to different images. Furthermore, the need to classify image contents based on their similarity is another important motivation to consider computational intelligence as a viable approach to solving such problems. For example, Artificial Neural Network (ANN) may be utilized as a classifier given its limited yet ready learning ability. Additionally, Fuzzy Interfacing Systems (FIS) have the ability to learn new things by defining new rules.

## **1.5 The Motivation and the Goal**

Inspired by the human visual system (HVS), and given all the above-stated issues and limitations of existing techniques, the motivation for developing a new image retrieval system (IRS) is underlined. This IRS considers the way in which the human looks at the image and describes it. Not all details or objects in the image are important in HVS; only the most salient objects attract the human eye to look at them. Since this process needs some kind of intelligence, different artificial intelligence techniques, such as neural network and fuzzy logic will be considered in different stages of the proposed technique.

The core of the suggested system is that it utilizes content identification in a semantic way, meaning that it gives tags to the images based on their salient objects regardless of the visual features of the image. The following objectives shall be considered in the proposed system:

1. Saliency of objects.
2. Uncertainty nature of the image.
3. Adaptive learning and computational intelligence.
4. Automatic image description and tagging.

## **1.6 Contributions**

In this work, we are planning to achieve the following contributions and outcomes:

- 1- As we are planning to design an image retrieval system which utilises some of the important features of human visual system (HVS), we shall study the main features of HVS and specifically human attention principles. Consequently, human attention is represented in terms of point's saliency identification. We shall introduce a new bottom-up approach to extract the saliency of a region. The new saliency approach utilizes the principles of irregularity in some regions in an image to identify the saliency of the regions. The region is said to be salient if its contents are irregular as compared to other regions. Based on the aforementioned definition, an algorithm will be proposed to extract the salient region, and hence salient objects, from an image.
- 2- Converting salient points (or gaze points) into salient regions is another challenging issue which needs to be studied and analysed. An automatic clustering technique, which is appropriate for clustering salient points to form the regions of interest, will be developed. The proposed algorithm should be fully automatic; therefore, it needs an automatic stopping criterion, which will be discussed and developed as well.
- 3- Thresholding is one of the concerns in our investigation since it is crucial in many applications and algorithms. How to find a suitable thresholding value automatically will be investigated; additionally, based on the fact that the borders



between the regions are not well-defined, we will develop a fuzzy-based thresholding technique.

- 4- Image identification based on region saliency is discussed and an image identification algorithm developed. The image contents identification algorithm shall contain two parts; object identification and background identification, as given below:
  - a. In object identification, a Saliency Based Image Identification (SBII) technique is developed and discussed. The technique is based on identifying the salient objects only.
  - b. Image background identification is discussed and an algorithm for identifying the background of an image developed.
- 5- For evaluating the proposed algorithms and comparing them with the existing algorithms, we need to examine and analyse suitable evaluation techniques. Based on the available evaluation techniques and studying the pros and cons, we shall develop evaluation algorithms for saliency identification and image retrieval. These evaluation techniques are designed to be suitable to our applications.

## **1.7 Organization**

The thesis is organized as follows:

1. Chapter 1 introduces the reader to the main techniques and applications of image identification, to the necessity of salient-based system development, and indicates the directions (goals and objectives) in which the dissertations will explore. The primary limitations and problems of machine vision are described, with a view to motivating new research to overcome these challenges.
2. Chapter 2 contains brief information about the basic theoretical background necessary for deriving the developed techniques in this thesis. Topics such as HVS, MVS, clustering, computational intelligence and image retrieval techniques are presented in this Chapter.
3. Chapter 3 presents a brief discussion to the datasets that have been used in both saliency extraction algorithms and image retrieval. The chapter also includes

discussion to the methods that have been used to generate the ground truth data for saliency extraction in addition to the method used to extract the saccade points from the salient points.

4. In Chapter 4, the developed techniques of saliency identification and evaluation are presented, and the necessary theoretical background and derivations are presented as well.
5. Chapter 5 contains the proposed image identification method consisting of two parts: object identification and background identification. For both identification processes, newly developed techniques will be expounded upon and evaluated.
6. Chapter 6 will conclude with the recapitulation of the existing techniques and applications that were utilised in this research. Furthermore, it will debate the newly-developed system and provide an evaluation, offering recommendations for the improvement of salient-based image identification systems.

## **1.8 Publications**

The main publications concerned in this field are listed below:

1. M. Al-Azawi, Y. Yang and H. Istance, "Human Attention-Based Regions of Interest Extraction Using Computational Intelligence", in IEEE GCC 2015, Muscat 2015,
2. M. Al-Azawi, Y. Yang and H. Istance, "Irregularity-Based Image Regions Identification and Evaluation," Multimedia Applications and Tools, 2014.
3. M. Al-Azawi, Y. Yang and H. Istance, "Irregularity-Based Saliency Identification and Evaluation," in IEEE International Conference on Computational Intelligence and Computing Research, Madurai, 2013.
4. M. Al-Azawi, Y. Yang and H. Istance, "A New Gaze Points Agglomerative Clustering Algorithm and Its Application in Regions of Interest Extraction.," in IEEE IACC, 2014.
5. M. Al-Azawi, "Image Thresholding using Histogram Fuzzy Approximation," International Journal of Computer Applications, vol. 83, no. 9, pp. 36-40, 2013.

6. M. Al-Azawi, "Neural Network Based Automatic Traffic Signs Recognition: An Application on Intelligent Cars," in SETAC-2012, Emerging Trends in Advanced Computing, Ibri / Sultanate of Oman, 2012.
7. M. Al-Azawi and N. Ibrahim, "A New Edge Intersection-Based Salient points extraction and its application in computer vision," in NCBIT2014, Massana-Oman, 2014.
8. M. Al-Azawi and N. K. Ibrahim, "Bimodal Histogram Based Image Segmentation Using Fuzzy-Logic," in NCAIAE12, Massana, 2012.

Publications (7) and (8) were prepared jointly with Mrs. N. K. Ibrahim. The role of the co-author in publication (7) was executing the code available online for the ITTI and SUSAN algorithms using MATLAB and extract the results which have been used for benchmarking with the proposed algorithm.

In publication (8), the co-author has executed the existing thresholding algorithms code provided by the authors online using MATLAB and find the results of these methods and presented the paper in the conference.

## **1.9 Conclusions**

In this chapter, we have presented a brief introduction to the problem we are going to investigate and study. The main limitations and problems of the existing machine vision techniques, such as the semantic gap, erroneous identification due to the effect of the background, or identifying unimportant regions, have been addressed. Identifying these problems has provided us with motivations to carry out this research to overcome the primary challenges. The chapter also included a review of the feasibility of using computational intelligence techniques in image processing and how they can be utilised to improve the outcomes of image processing algorithms. The intended academic contributions are also presented in this chapter. The organization of the thesis along with a short description of each chapter is introduced to furnish the reader with the contents of the thesis. Finally, a list of relevant publications which came out while the work was being undertaken has been listed.

# Chapter 2

## Visual Perception

### 2.1 Introduction

This chapter main concern is image perception in human and what techniques are needed to be utilised for that purpose. Image perception in human is necessary for discussing the image perception in machine vision. Therefore, this chapter sets out to offer insight into the relevant theoretical concepts and background of the dissertation. State-of-the-art technologies and theories regarding Computational Intelligence (CI), CBIR, HVS, human attention and saliency, and Machine Vision (MV) are also covered in this chapter since they represent important parts in image perception. In addition, related image processing techniques, such as, features extraction, segmentation and thresholding are reviewed, studied and analysed since we are going to use these techniques frequently in different applications.

### 2.2 Visual Perception

Visual perception can be defined as the process of linking the image, or visual input, to previously existing models of the world. In a computer context, it is a system that accepts visual data (images) as input, performs the processing and recognition processes, and provides the image description. In order to achieve this, the machine requires different image representations. Ballard and Brown suggested four representations: (1) Generalized Image or iconic, (2) Segmented Image, (3) Geometric Representation, and (4) Relational Model [17].

To study the possibility of using features from human vision system and apply them in machine vision system, it is required to study the human vision system from different sides. Human attention is essential in our discussion; therefore, the main properties of human attention will be reviewed.

### **2.2.1 Human Vision System (HVS) and Machine Vision Systems (MVS)**

HVS is a very complex and integral system. For this system, the first part is the retina, which is responsible of sensing the light reflected from surfaces in the world and image formation [18]. In MVS, the eyes can be replaced with cameras, which work perfectly to capture the light reflected from surfaces and convert it into images. In HVS, goals and knowledge are high-level capabilities which guide the visual activities. In addition to high-level capabilities, low-level activities are also necessary [19] [17]. HVS uses features such as colour, texture and shape to recognize objects in a scene. It is not yet possible to model and replicate HVS in a machine, thus MVS does not model HVS but implements a system inspired by it.

#### **2.2.1.1 High-level vision and Low-level vision**

High-level capabilities like cognitive processes, geometric models, goals, and plans are very important in HVS and humans are heavily reliant upon them when it comes to understanding images. For example, when a human sees a person, the human depends upon knowledge, for example, the place and the time that person was seen, to identify the person and not upon the features of that person. However, high-level capabilities are dependent on low-level capabilities in order to understand an image.

The human also needs low-level features in order to understand an image. The human brain is trained to identify the texture like a brick wall, a cloudy sky, or a grass field [17]. Colours are identified by the human eye by sensing the reflection of light from an object, and it is also affected by the colours adjacent to the object in the image.

In contrast to MVS which is not specialized, the human brain can adapt itself according to the object, as it does in human face recognition, where the human can recognize facial features and identify the individual quite easily on account of this specialization.

How do humans recognize stimuli arriving at human retina as straight lines, squares, circles, or any other shape? The answer is the Law of Visual Reconstruction [20]. The

most substantial attempt to state this law is the Gestalt Laws. In Grouping Gestalt law, the points are grouped together to form a larger shape, as such in the case of computer vision where pixels are grouped together to form a larger shape. The formation of larger pixel shapes is a gestalt (group of points) according to their similarity in some spatial properties [20]. One may further differentiate between the two types of groups (gestalts); partial gestalts and global gestalts. For example, a gestalt forming a square is called a global gestalt, or rather; the square is a global gestalt that is constructed from recursive operations like the formation of the sides and corners. The sides and corners are then considered to be partial gestalts.

#### **2.2.1.2 Recognition by Components**

To overcome the obstacles encountered in identifying the image as a whole, recognition by components (RBC) was suggested, to simulate the human's recognition system. Several models have been proposed in this field, such as those of Biederman et al. [21] and Borges et al. [22]. In these models, the image is divided into components, mostly geometrical components like cylinders, cubes or arcs. Each object in the image is constructed from combining these components to form an object. As an example, if an arc is attached to a cylinder then it may form an object depending on the location of the arc. If it is attached to the side of the cylinder, it may form the shape of a cup, while if it is attached to the top of the cylinder it may form the shape of a bucket.

Dividing an object into atomic parts will not improve the recognition process since, and based on Gestalt law, it is not possible to identify the whole by identifying the parts but the parts by identifying the whole [23]. Therefore, using the approach of disassembling the image contents into small parts, identify these parts, and then identify the contents based on that is not an efficient way to use in image recognition. Nevertheless, it is efficient in describing regular shapes and objects which have regular shape.

#### **2.2.2 Human Colour Perception**

Human eyes contain two kinds of light sensitive cells in the retina; rods and cones. The rods are sensitive to motion and differentiate between light and dark. They also work when the light is low. The rods cannot detect colour; colour detection is the role of the cones, which function at higher level of light. Human eyes contain three types of cones

sensitive to three different spectra, resulting in trichromatic colour vision. These cones are labelled according to the wavelength they can sense, at Short (S), Medium (M), and Long (L) wavelengths.

Different colours can be sensed by combining the stimuli received in different cone types. The process of combining is not linear, so the eyes perform a complicated process in order to interpret colour. Two complementary theories of colour vision, the *trichromatic theory* and the *opponent process theory* were proposed to explain human colour perception. The first theory was proposed by Young and Helmholtz in the 19<sup>th</sup> century. It states that the retina's three types of cones are preferentially sensitive to blue, green, and red from which all other colours are obtained by combining them in different ratios. The second theory, which was proposed by Hering in 1872, states that the visual system interprets colour in an antagonistic way; red vs. green, blue vs. yellow, black vs. white.

## **2.3 Human Attention**

Human attention can be attracted by different stimuli using different senses such as hearing and visual senses. Stimuli such as loud sound which raised suddenly may attract the hearing sense and hence the human attention. In the same way, sudden change in light or colour luminance attracts the human visual attention. In general, it was argued in different literature that human attention is stimulated by different ways such as sudden change in the received information. In our research, we will focus our study on visual attention and its application in machine visual systems.

How can the humans identify the most important object in an image? Consider the image in Figure 2-1 (a) and think of an answer for the question: What is the important object in the scene? Based on a simple query that we have made, and as it was expected, the majority (90%) of people when asked the same questions have answered with 'car'. Then comes the next question, why is the car the important object in the image? Kadiyala et al. [24] [25] answered this question by referring this to the three main items that draw the attention of a human which are size, location and contrast. Accordingly, if any object is the largest object, positioned in the middle of the image and presents the highest contrast among other objects, then this specific object is more important than others.

According to human perception, the abovementioned theory is not always correct. Although it was asserted that in Figure 2-1 (a), the majority of observers would consider the car as the most important object, there still remains a minority which might not consider the car to be the most important object. Nonetheless, for the majority the car will represent the salient object. In the image the car is neither the largest object, nor in the middle of the image, nor the one with highest colour contrast. Hence, there is a need for alternative measures, based on what the human may use, to identify the salient objects.

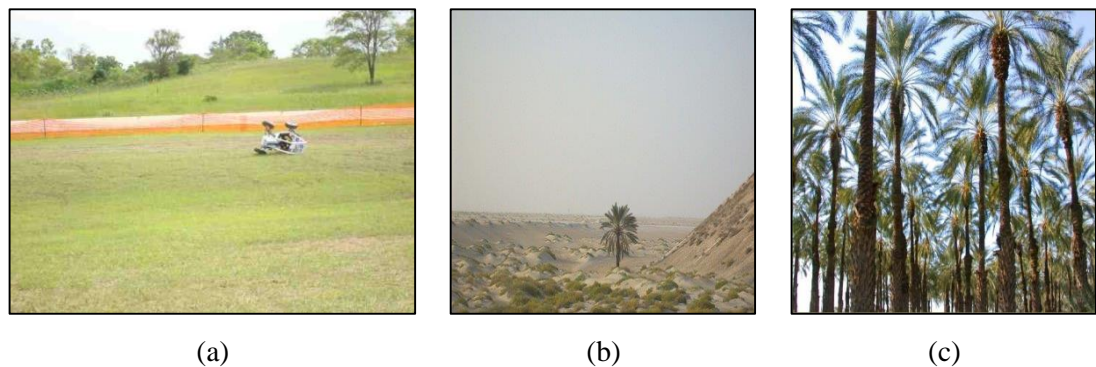


Figure 2-1: Different importance may be given for an object.

Moreover, object saliency detection remains a challenging problem due to the wide range of variability in the scale, location, orientation, and pose that real objects adopt [26].

In order to understand saliency in a better way we need first to understand human attention. In general, attention is very important to extract the regions of interest (ROI).

### 2.3.1 Attention

Human attention is the process of selecting a subset of the available information upon which to focus for enhanced processing and integration. Many authors studies the attention properties, both computationally and biologically. Most of the knowledge in this field can be obtained from the Computational Neuroscience field. More detailed information can be found in, for example, [18].

Human attention contains, mainly, three aspects and goes through these aspects in sequence as shown in Figure 2-2 [27].



### ***The Orienting Aspect***

When a human receives more than one stimulus input, the best way to select one stimulus is to orient the entire attention toward that stimulus only. Many types of orientation were proposed to explain the behaviour of the human or animal, such as, orienting reflex and goal-driven (endogenous). For the orienting reflex, abrupt actions capture the attention, such as, a large object with different colour in a smooth background. With goal-driven orientation, attention is oriented to a location in space or to an object voluntarily in a goal-driven manner, often based on a cue that indicates where to look. Goal orienting may need some symbolic cues, for example, labels, arrows and pointers.

When attention is captured by a certain object, oriented to a specific location, and then moved to another location it will not return to the previous location directly or it is inhibited to orient or return to the original location for a period of time; this is known as “Inhibition of return” [28] [2].

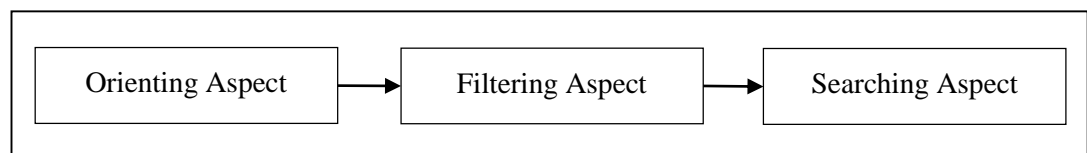


Figure 2-2: Attention phases (Aspects).

### ***The Filtering Aspect***

After orientation, the attention acts as a filter to remove information with less importance and focus on information with higher importance. For instance, in an image with an abrupt subject, like a bird in a sky, the human sees the bird while the image of the sky is blurred and not seen while looking at the bird. Hence, the human brain has filtered the image into two parts, one with higher importance (the bird) and one with less importance (the sky). The human has no ability to divide his attention among more than one stimulus, as the human cannot see the sky and the bird at the same time. This is known as the singularity of attention (SOA) principle.

### ***The Search Aspect***

When a human knows what to find but does not know where to find it in an image, attention will be involved in this search process. Search is classified as easy versus difficult search, and automatic versus controlled search.

Easy search (also known as pop-out or parallel search) is when the observer searches for an easily recognised item in a set of items. It will be easy for him to find the specific item regardless of the number of items [27].

Difficult search (also known as serial search) is when the observer needs to find a particular item based on a conjunction of features. The search will be slow and the required time to find the target will increase linearly with the number of items in the image, because one needs to focus on each item for a certain amount of time.

### ***Controlled versus Automatic Search***

Controlled search is usually utilised with the difficult search; in order to achieve more accurate search results, the same person may be required to perform each type of search. The process of repeating the search procedure will train the observer. Subsequently, the search process performed by the observer will become an automatic search rather than controlled. An automatic search is much faster than a controlled search as the training of the observer to perform some searching task will speed up and radically improve the results.

### **2.3.2 Pre-attentive Phase**

Pre-attentive processing is the unconscious accumulation of information from the environment. All information is considered as pre-attentive at the beginning, after that the brain processes and filters this information, assigning importance to each item based on the saliency of that item. Information with high saliency is more apt to undergo further attentive processing. Some data proceeds from pre-attentive to attentive, based on the saliency of the information and the goals and intentions of the observer. In an image, various properties may give an item the importance to pop-out from other items (destructors), such as, colour, shape, orientation, length, closure, size, curvature, density, contrast, and hue, as shown in Figure 2-3.

### **2.3.3 Attentive Phase**

After the pre-attentive stage, the attentive phase of image recognition is activated. This phase requires more time than the pre-attentive stage. Here, objects of interest or pop-out regions are recognized and the semantic relationships among the regions are extracted.

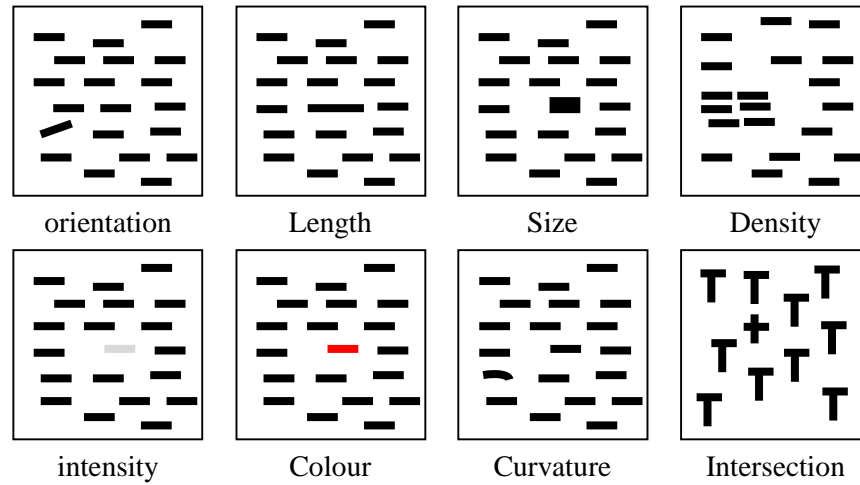


Figure 2-3: Pre-attentive visual pop-out features.

### 2.3.4 Top-Down and Bottom-Up Attention

Top-down attention and bottom-up are two important types of attention and was argued that these are independent as was discussed by Pinto et al. [29]. In bottom-up attention (also known as pure capture) or stimulus driven attention, the attention is captured voluntarily or due to some stimuli. In contrast, in top-down attention (also known as cognitive), the attention is derived by a goal or intention.

The bottom-up model uses the stimulus saliency and its contrast with the background. This model focuses on the visual saliency more than the conscious goal. In contrast, cognitive capture, emphasizes the idea that the observer knows what stimulus he searches for, implying that goals and intentions are present in the search process. The brain pays attention to stimuli which have features of the target item.

The goal of the observer is essential in the specification of the search method. For example, a walking man has different goals than a football player [30]. A man who is driving a car has the goal of watching surrounding cars, pedestrians and traffic signs, whilst not paying any attention to, for example, a bird in the sky. Thus, the goal of a CBIR user should be well defined and considered when designing the retrieval system. More detailed information can be found, for example, in [31] [29] [32].

### 2.3.5 Fixations and Saccades

Human vision behavioural studies consider fixations and saccades to be the two most significant functions of sight as they offer a proper understanding of human interest,

which in its turn offers insights into the most relevant points of an image. Fixation is where the human gazes at one point for a period of time, while saccade is the fast movement of the eyes from one point to another in an image. If the human fixates his eyes on certain regions then the most important points are found at that location. When the human moves his eyes very fast from one point to another then the saccade regions are generated, which are considered as unimportant regions or regions with less attention-capturing capacity.

Humans have a collection of passive mechanisms to reduce the amount of incoming visual information. For example, the signal stemming from the photoreceptors is compressed by a factor of about 130:1, before it is transmitted to the visual cortex [33]. The HVS has an active selection mechanism, involving eye movement. A saccade is a rapid eye movement allowing the vision to jump from one location to another. The purpose of this movement, which occurs up to three times per second, is to direct a small part of the HVS field into the fovea to achieve a closer inspection, which corresponds to a fixation.

### **2.3.6 Saliency**

Saliency can be seen as the computational representation of attention. Different literature have defined saliency in different ways. Some of them consider saliency from the photographer's point of view, as the photographer usually places the object of interest in the centre of the image. For others, the object of interest is the object with the higher level of contrast in colour and/or geometrical properties.

Usually, based on HVS which acts as a passive selector, by acknowledging some stimuli and rejecting others [33], the salient object is the one that captures the attention or attracts the HVS. HVS uses two stages to identify the objects, pre-attentive and attentive stages [34]. In pre-attentive stage, the local regions of image that present spatial discontinuity (pop-out features) are detected. In the attentive stage, relationships between these features are found and grouped together. Accordingly, both low-level and high-level features are required to identify the salient object. In addition, both local features of the object and global details of the image are required for the task. Local features may give visual importance to the object locally, which is corresponding to the pre-attentive recognition of the human, while global image details correspond to the attentive

recognition of the HVS. In the first stage (local features) the salient objects are identified, while, the importance is specified in relation with the environment (global image details). For instance, a tree is a salient object at the first sight, but its importance might be increased or decreased according to the background, if it is in a desert (Figure 2-1 (a)) or in a forest (Figure 2-1 (b)).

The automatic detection of salient image regions is important for applications such as adaptive content delivery, adaptive region-of-interest-based image compression, video summarization, progressive image, transmission, image segmentation, image and video compression [35], object recognition and content-aware image resizing [36].

## **2.4 Content – Based Image Retrieval**

As one of our goals in this research is to design an image identification system, it is important to discuss the content-based image retrieval systems (CBIR) since they represent the latest trends in image contents identification. This Section discusses the leading state-of-the-art techniques in image identification and contents understanding. Image identification is divided broadly into two categories: identification as a whole and identification by parts (contents). The identification phase presents two categories:

1. Same scene with different perspective (SSDP).
2. Different scenes with similar objects (DSSO).

The SSDP considers that different images for the same scene can be taken from different perspectives (such as view angle or camera location). These types of images contain the same object, but with different location, scale, lighting conditions or view angle. In contrast, in DSSO, the object might be in a different environment or background, making the identification more difficult than in SSDP.

Diverse features are used to identify the image contents (such as, texture, colour, statistical and geometrical features). The selection of suitable features in the identification process is crucial since they play an important role in image content identification. For example, colour features are not suitable in x-ray image identification since the image has only one band and no colours. Nonetheless, colour features are essential in the identification of natural scenes.

CBIR, also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR) [37], is any technology that aims at organizing the huge amount of image data based on the visual contents [38]. It utilises image processing and computer vision techniques to search for and retrieve images from a large database. Lehmann et al. [39] have defined CBIR as the process that aims to describe the complex object information of digital images by non-textual features, which are essential in efficient query processing. Hence, CBIR depends on the visual properties of the image to automatically identify contents.

Recently, in this field, numerous researches have been published. A simple search for the term ‘CBIR’ on IEEE and Google, in different time periods, retrieved the results shown in Figure 2-4. In this figure, the tremendous change in the number of researches is conspicuous.

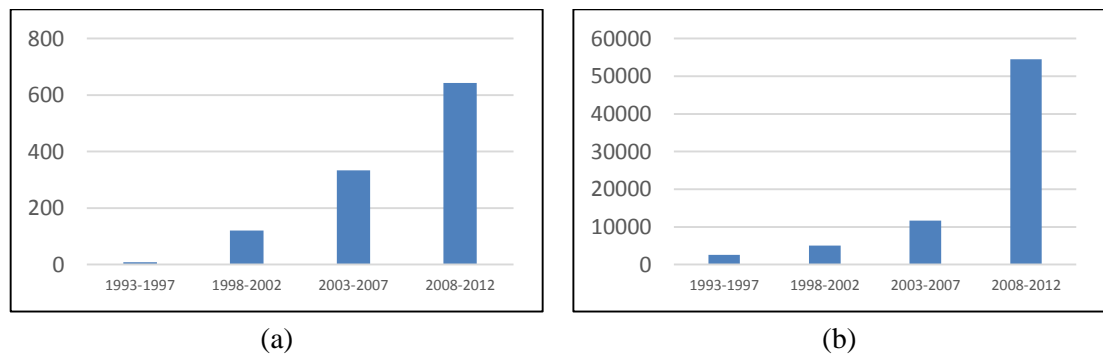


Figure 2-4: Number of CBIR researches vs. year in (a) IEEE, (b) Google.

The following is a survey of the most popular and state-of-the-art techniques in the field of CBIR. Several surveys were published on CBIR, such as ref [40], in which the author has presented a survey on the techniques that utilise the browsing model. He discusses how to classify the image in order to improve the retrieval process and offers the user a goal-oriented image browsing mode where the system searches for similar images.

Jaswal and Amit [41] address some of the challenges that CBIR experience. In addition, they discuss the types of features that can be used with CBIR. Refiee et al. [42] reviewed the semantic CBIR and the patch recognition issue. Furthermore, they discuss the use of supervised learning, unsupervised learning and relevance feedback. Singhai and Shandilya [38] review the three main features that are used in CBIR (colour, texture and shape), in addition to the Fuzzy-based CBIR systems. Jain and Singh [43] published a review which focuses on the clustering of the images in the database in order to improve

and simplify the retrieval process. Liu et al. [44] have published a review about the use of high-level semantic features in image retrieval. In other surveys, like [45], [46], [47], [48], [49] and [50], the authors have introduced almost identical definitions and discussed similar techniques such as, colour, texture, shape, and fuzzy based techniques.

Lehmanna, et al. [39] have applied CBIR to medical applications. Their model suggests the use of six layers of information modelling (raw data layer, registered data layer, feature layer, scheme layer, object layer, knowledge layer). The features used in the matching process are colour histogram and texture feature. Singhai and Shandilya (2010) used the edge histogram to capture the spatial distribution of edges in an image, in their efforts to build a universal CBIR system using low-level features. Measures like mean, median and standard deviation of RGB channels of colour histograms were used in comparing images, in addition to texture, features such as contrast, energy, correlation, and homogeneity are retrieved. Finally, they used the edge features including five categories: vertical, horizontal, 45° diagonal, 135° diagonal and isotropic [38].

Yang and Zhou [1] have divided the retrieval process into five steps: (1) Analysing the contents of the image to extract the feature of images; (2) Providing a fuzzy description of each image like samples or sketches; (3) Comparison step where the features of the query image are compared with the stored image in the image database; (4) Return the result to the user by similarity in a sequence from large to small. (5) Interactive with user, the user can select the best match and send his/her feedback to the system.

The histogram refinement technique is proposed by Pass and Zabih [51], and works by imposing additional constraints on histogram-based matching. They split the pixels in a given bucket into several classes, based on some local property. Within a given bucket, only pixels in the same class are compared. In addition, they proposed use of the colour coherence vector (CCV) which partitions each histogram bucket based on spatial coherence.

Liu, et al. [44] have used high-level semantics to improve the accuracy of the retrieval process. In their paper, they identify five major categories of techniques in narrowing down the ‘semantic gap’: (1) Using object ontology to define high-level concepts; (2) Using machine learning tools to associate low level features with query concepts; (3) Introducing relevance feedback (RF) into the retrieval loop for continuous learning of

user's intention; (4) Generating a semantic template (ST) to support high-level image retrieval, and (5) Making use of both the visual content of images and the textual information obtained from the Web for WWW (the Web) image retrieval.

Ref. [52] talks about the Regional-Based Image Retrieval (RBIR) which divides the image into regions and applies CBIR on each of these regions. Suhasini, et al. [53] proposed three types of histogram; the conventional colour histogram (CCH), the invariant colour histogram (ICH) and the fuzzy colour histogram (FCH). The conventional colour histogram (CCH) of an image is the frequency of occurrence of every colour in that image. The invariant colour histogram (ICH) depends on the colour gradients, and is used to overcome the effect of rotation/translation of an image. The fuzzy linking colour histogram (FCH) is used to address the problem of spatial relationship.

Thawari and Janwe (2011) compare CBIR on colour and texture features using the histograms. This comparison is based on the statistical parameters of the image. The parameters include mean, median and standard deviation of RGB channels of colour histograms. Then the texture measures, such as mean, second moment, third moment, forth moment, smoothness, and uniformity are retrieved [54].

Sharma et al. [55] have investigated two methods for describing the contents of images. The first one characterizes images by global descriptor attributes, while the second is based on the colour histogram approach. Cross-correlation value and image descriptor attributes are calculated prior to the histogram implementation, in order to make a more efficient CBIR system.

Nazari and Fatemizadeh [56] used Texture-Based CBIR in medical applications to discriminate between the normal and abnormal medical images based on features. The main indices are finding normal, abnormal and clustering the abnormal images to detect a certain two abnormalities: multiple sclerosis and tumour images to classify the database

## **2.5 Image Processing and Machine Vision**

Due to the advanced technology available nowadays and the improvement in transmission speed, image processing techniques such as filtering, enhancement, feature extraction and others are widely used in different applications. In order to solve problems



related to image processing, a chain of processing steps have been defined in various literature. This chain of processing is known as the image processing chain (IPC). The main blocks of the IPC are shown in Figure 2-5. First, the image is pre-processed using techniques such as filtering, noise removal and image reconstruction. The second stage is the feature extraction stage, in which different features are extracted from the image, like colour features, textures, edges and other low-level features. Based on the extracted features, the image can be segmented into well-defined, non-overlapped regions which can be categorized into objects and background. It is usually assumed that most of the segments represent objects. Hence, the next step is to identify these objects and ultimately, to understand the contents of the image.

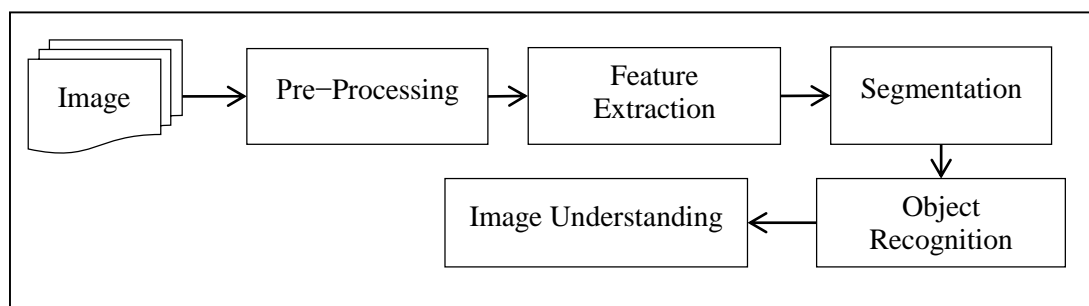


Figure 2-5: Image Processing Chain [57].

## 2.6 Image Feature Extraction

In order to understand and interpret the contents of an image, it is essential to extract some features that describe its contents. Features are measurements that can be extracted from the image like edges, statistical measurements, colour properties, texture or shapes. In machine vision applications, features should be meaningful and detectable. Features are classified into General features and Domain-specific features, where general features are application-independent, and domain-specific features are dependent on the applications (such as in face recognition and fingerprint recognition applications). Features can also be classified into low-level and high-level features.

### 2.6.1 Low-Level Features

Low-Level Features (LLF) are basic features that can be extracted directly from the image; like edges, texture, and colour. They are not easily utilised in image understanding since they give no information about the relationship between shapes in the image.

However, a set of high-level features that are useful in image understanding and contents recognition can be extracted from low level features.

### ***Edges***

Edges are important features and can be defined as the boundaries that separate two different regions, commonly assumed to be background and objects. Any sharp change in colours or intensity may produce an edge. Many edge detection algorithms rely on the luminance values of grey level images. First-order edge detection recognizes image intensity changes by looking for the minimum and the maximum in the first derivative of the image [58]. Many first order edge detection operators have been developed like Roberts' , Prewitt's and Sobel's edge detection operators. The second order derivative, such as Laplacian, is another method that highlights the edges, in which, if the change of the first derivative is high and zero if it is constant, then its value is greater.

Alternative methods were proposed to extract the edges, like Chebyshev polynomials method, which consists of nine  $3 \times 3$  kernels that can be used to extract edges with different directions [59]. Other statistical and computational approaches can be found in ref. [60], [61], [62], [63], [59] and [64].

### ***Corners***

The corner is a 2D structure in an image that can be represented mathematically. These are described through several junction types such as "Y", "X" and "T" corners. Corners do not change significantly with image changes, such as scaling, rotation and shifting, thus, they are important in applications like motion tracking and stereo matching.

Many methods for extracting corners exist and these can be categorized into three types: edge-based, raw-based and edge-curvature based.

### ***Lines***

A line is a straight edge, and is an important feature in extracting shapes and artificial objects in an image. The line can be extracted by finding the value of convolving the image with some kernels.

### ***Spots***

A spot is a small circle in a square template of size  $3 \times 3$ . In other words, it represents a closed contour of edges. The spot can be extracted by convolving the image with a circular impulse response array.

### ***Colour***

Colour features are widely used to describe the contents and the nature of an image. The most well-known representation of colour features is the histogram. The image histogram is the distribution of colours over image pixels, or simply the frequency of occurrence of colours in the image. The main advantage of using the colour histogram is that it is not affected by the rotation, scaling, and shifting of the image. The prime weakness of using the histogram is that it loses the space information of the colour [1]. The colour histogram  $\vec{h}$  can be defined as a vector by:

$$\vec{h} = \langle h_1, h_2, h_3, \dots, h_n \rangle \quad 2-1$$

where  $h_i, i = 1, 2, \dots, n$  is number of pixels of colour level  $i$  in the image and  $n$  is the number of intensity levels in an image. The number of intensity levels is equal to  $2^b$  where  $b$  is the number of bits used to represent the colour in an image.

Colour moments, such as, the mean (first order), variance (second order), and skewness (third order) are used to describe the histogram:

$$\mu_i = \frac{1}{N} \sum_{j=1}^N p_{ij} \quad 2-2$$

$$\sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^N (p_{ij} - \mu_i)^2} \quad 2-3$$

$$s_i = \sqrt[3]{\frac{1}{N} \sum_{j=1}^N (p_{ij} - \mu_i)^3} \quad 2-4$$

where  $p_{ij}$  is the value of  $i^{th}$  colour component of the image pixel  $j$ , and  $N$  is the number of pixels in the image.

### ***Texture Features***

Another issue of histogram or colour matching is that different images may have similar histograms (colour distribution). Moreover, different objects may have similar colours scales, such as, the sky and the sea. In order to overcome this issue, texture is required. When it refers to the description of what the image's texture is based on, we usually adopt the texture's statistic feature, structure feature, and spectral features [38]. Although texture is not well defined like colour feature, it gives a competent description to the contents of the object in the image like cloud, trees, bricks, and fabric).

Texture features can be obtained using Gabor filter, wavelet transform or local statistics measures. Among the six Tamura features of coarseness, directionality, regularity, contrast, line-likeness, contrast and roughness, the first three are more significant. The latter features are related to the previous three and do not add much to the effectiveness of texture description [44].

According to Haralick [65], the main ways of analysing the texture can be categorised into statistical, structural, and spectral approaches. In statistical approach, Moments of Intensity and Grey Level Cooccurrence Matrices (GLCM) are the most common methods to describe the texture. Moments of Intensity measures method is one of the statistical approaches that can be used to describe the nature of the image's texture. The following moments can be used in describing the structure of the image contents:

1. The first moment or the mean intensity,
2. The second moment or the variance, which describes how similar the intensities are within the region,
3. The third moment or the skew which describes how symmetric the intensity distribution is about the mean, and
4. The fourth central moment or kurtosis, describing how flat the distribution is.

$$\mu = \frac{1}{N \times M} \sum_{j=1}^N \sum_{i=1}^M p_{ij} \quad 2-5$$

$$\sigma_i = \sqrt{\frac{1}{N \times M} \sum_{j=1}^N \sum_{i=1}^M (p_{ij} - \mu)^2} \quad 2-6$$

$$s_i = \sqrt[3]{\frac{1}{N \times M} \sum_{j=1}^N \sum_{i=1}^M (p_{ij} - \mu)^3} \quad 2-7$$

$$k_i = \sqrt[4]{\frac{1}{N \times M} \sum_{j=1}^N \sum_{i=1}^M (p_{ij} - \mu)^4} \quad 2-8$$

The second important method to describe the texture is the GLCMs, which describe the relation of each grey intensity level with its neighbours [66]. GLCMs are extracted from the grey-level intensity image based on the joint probability distributions of pairs of pixels. The GLCM can be specified in matrices of relative frequencies  $f(i, j)$  with which two neighbouring pixels separated by a distance  $\delta$  and angle  $\theta$  occur on the image, one with grey level  $i$  and the other with grey level  $j$  [67]. The relative frequencies of grey level pairs of pixels separated by a distance  $\delta$  in the direction  $\theta$  combined to form a **relative displacement vector**  $\mathbf{D} = (\delta, \theta)$ , which is computed and stored in the GLCM  $G$ . Figure 2-6 shows an example of the extraction of GLCM from the grey levels given in (a), the GLCM in (b) was extracted using a displacement of one and angle of  $0^\circ$ . (c) was extracted using a displacement of one and an angle of  $180^\circ$ . From the figure it is noted that given the number of grey levels in (a) is 6 then the size of the co-occurrence matrix is that of  $6 \times 6$ . Another relevant matter to consider is that the matrices with opposite angles and similar displacements are transpose to each other for example,  $G(x, y, \delta, \theta) = G^T(x, y, \delta, \theta + 180^\circ)$ .

1	2	1	2	3	4	5
2	3	4	5	4	3	2
3	4	5	4	4	3	2
4	5	3	4	2	1	2
2	3	2	4	1	2	3
2	1	1	1	2	2	3
3	3	2	2	1	1	1
(a)						
	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	4	4	0	0	0
2	0	4	2	5	0	0
3	0	0	1	1	4	0
4	0	1	1	2	1	4
5	0	0	0	1	2	0
(b)						
	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	4	4	0	1	0
2	0	4	2	5	0	0
3	0	0	5	1	2	1
4	0	0	0	4	1	2
5	0	0	0	0	4	0
(c)						

Figure 2-6: GLCM extraction, (a) intensity values, (b)  $GLCM(x, y, 1, 0^\circ)$ , (c)  $GLCM(x, y, 1, 180^\circ)$ .

Other statistical measures can be used to describe the texture in images such as Grey Level Run-Length Matrix where the number of the successive pixels with similar grey

levels is measured to give description to the texture. Some measures are extracted with Short Runs Emphasis, Long Runs Emphasis or Grey Level Non-uniformity. The main drawback of this technique is its sensitivity to noise.

### **2.6.2 High Level Features**

High-Level Features (HLF) or semantic features can be extracted or constructed from low-level features like edges, corners, and lines. Several high-level features have been suggested as will be shown in the following discussion.

Points pattern techniques are example of converting low level features into high level features. After the point detection process, the resultant low-level points feature can be used to form high-level features such as objects. Intrinsic structure finding is one of the methods that can convert LLF into HLF where the points' structure is described in accordance with the structure.

Border tracking is the second important technique for getting high-level features from low-level features. In border tracking, the border of the set of points is traced to give the external shape of the object.

Geometric primitive fitting and geometric primitive extraction are essential techniques to extract high-level features. A geometric primitive is a curve or surface that can be described by an equation with a number of free parameters. Geometric primitive fitting establishes if a set of points is described perfectly by the given primitive or not. Geometric primitive extraction should intelligently find the appropriate primitive for fitting the set of points. Furthermore, in geometric primitive extraction, not all points given are trustworthy given the consideration of outliers.

Shape descriptors are used to identify the contents of an image; these features are invariant with the image shift, resizing, or rotation. Hence, these might be used in the image matching process. Some shape features are not constant with the changes in image such as the area, which might be changed through scaling. In other words, relative measures require to be utilized. In ref. [68], Mehtre et al. introduced a good comparative study of the shape-based CBIR. 3D shapes and depth have been used recently in the matching process of similar shapes in two different images [69].

## 2.7 Image Segmentation

Image segmentation is an important component in computer vision applications such as robot vision, computer pattern recognition, document analysis and understanding. It is the process of dividing the image into non-overlapping, homogenous and connected regions [70]. The dividing process depends on attributes such as luminance amplitude for grey images, and colour components for colour images [59]. The segmentation process starts with a pre-processing stage which includes noise removal, image smoothing, background correction and sharpening [71]. Currently, there exist many types of segmentation in use and these approaches use different image processing techniques. Ref. [72], [73], [74] are examples of the existing image segmentation techniques. Some segmentation algorithms depend on luminance in grey images where the colour image is converted to a grey image and then these algorithms are applied. Other techniques, which consider the colour image in the segmentation process give better results as these are analogous to human perception.

Segmentation algorithms can be classified as either local or global. Local segmentation only accounts for the pixels' values and those of their neighbouring pixels, while global segmentation accounts for the whole of the image [75].

The segmentation process divides the image region  $R$  into a set of homogenous, non-overlapped, connected sub-regions  $R_i$ . The sub-regions need to satisfy the following criteria:

1. The union of all sub-regions forms the original image
2. The regions  $R_i$  should be connected for all  $i = 1, 2 \dots N$ .
3. The regions  $R_i$  should be homogeneous for all  $i = 1, 2 \dots N$ .
4. Different adjacent regions  $R_i$  and  $R_j$  should be disjoint  $R_i \cap R_j = \emptyset$ .

### 2.7.1 Image Thresholding

Thresholding techniques are the earliest techniques to be widely used in various applications. Thresholding is a computationally efficient and fast process for converting a grey image into a binary image. If  $\mathcal{I}(x, y)$  is an image and  $\mathcal{B}(x, y)$  is the binary image after thresholding then:

$$B(x, y) = \begin{cases} 1, & \text{if } I(x, y) \geq T \\ 0, & \text{if } I(x, y) < T \end{cases} \quad 2-9$$

where  $T$  is a threshold value, one represents a pixel on the object and zero represents the background. The selection of the value of  $T$  is a crucial issue in such techniques. Therefore, many such techniques have been proposed, such as local, global, and adaptive thresholding techniques [76]. Al-Rawi and Stephan [77] utilized Genetic Algorithms (GA) in finding the optimum threshold value, with the use of conventional histogram techniques in extracting the threshold value. In this instance, the role of GA is for optimization purposes, which can be achieved by conventional optimization techniques. Savakis [78] utilized foreground and background clustering for thresholding documents. Furthermore, the author used adaptive algorithms in thresholding documents. The use of Artificial Neural Networks (ANN) in thresholding is also present in the works of other authors, such as [79] [80] [81]. Papamarkos et al. [79] suggested a multi-thresholding technique using ANN in image thresholding. Hamza et al. [80] used ANN in binarizing documents. Lastly, some survey researches were published to discuss the diverse array of thresholding techniques such as Ref. [82].

### ***Bimodal histogram***

The bimodal histogram is a method through which the value of  $T$  can be found. A bimodal histogram is a histogram with two peaks and a valley. One of the peaks is the background, while the second peak is the object. To separate the object from the background the value of  $T$  is extracted from the histogram by taking the valley value. Figure 2-7 shows the result of applying the above technique.

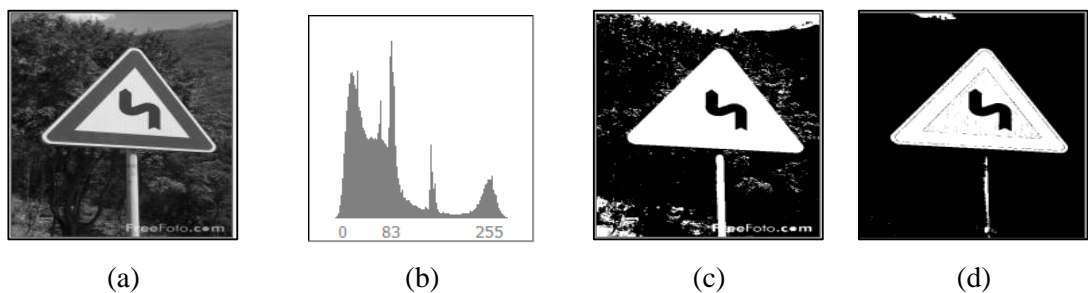


Figure 2-7: Thresholding technique using histogram analysis. (a) Original grey image; (b) histogram of the grey image; (c) selecting arbitrary threshold (100), (d) selecting the threshold value equal to the valley (200).



As may be observed, Figure 2-7 (a) represents the original grey image, (b) provides that the valley value is 200 (this value becomes the threshold for the image segmentation) and (d), shows the separation between background and object. However, in (c) where an arbitrary valley value of 100 is set, the results are not as precise and the separation is partial. The main disadvantage of this technique is its applicability is limited to images that contain only a sole object, or separated objects with similar attributes, since it is a global thresholding technique.

One of the applications of this technique is in thresholding typewritten images where the nature of the images is bimodal by default. These contain black colours against a light background, for which the threshold is calculated with the use of the bimodal histogram [59].

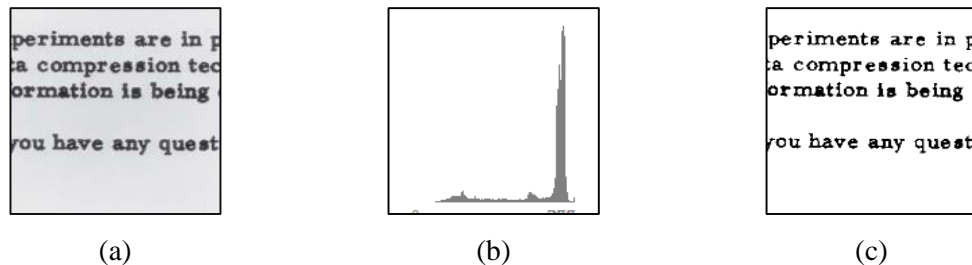


Figure 2-8: Typewritten text thresholding with bimodal technique, (a) original grey image, (b) histogram of the grey image, (c) selecting the threshold value equal to the valley (200).

### ***Multilevel luminance thresholding***

If the bimodal histogram thresholding technique is used, and repeated until a non-bimodal histogram is reached, a better segmentation is obtained. This technique was proposed by Tomita et al. [83] and is known as multilevel luminance thresholding technique. It is applicable to images with multi-peaks, meaning when the histogram presents more than two peaks. First, the lowest valley is taken as the threshold and the image is separated into two images: dark image and light image. Following this, the process is repeated for each and every new image generated through the lowest valley threshold. The repetition is carried out until the images present a unimodal histogram.

## **2.7.2 Colour Image Segmentation**

Colour image segmentation is the process of extracting two or more homogenous and disjoint regions from a colour image. This extraction depends on some of the features

extracted from the image such as colour, texture and other geometric or visual features. The pixels in each region should be connected and satisfy uniformity criteria.

As in grey level image processing, pixel-based techniques are mainly divided into two types: histogram-based and cluster analysis-based. In histogram-based techniques, the segmentation process uses the surrounding intervals for pixels starting from one or more maxima in the histogram [70].

The histogram is used in finding the threshold value that separates the object from its background. If there is more than one object then the multi-level bimodal histogram thresholding is applied. In colour images, similar techniques are utilized with the distinction that in colour images each pixel is described by three components rather than one component. These three components depend on the assigned colour representation system for example, RGB or HSx. The one-dimensional histogram is widely employed in colour image segmentation where the image is converted through various methods into a one band image. The simplest method is by converting the image into grey image by finding the average of the three colours R, G, and B as shown in Figure 2-9.

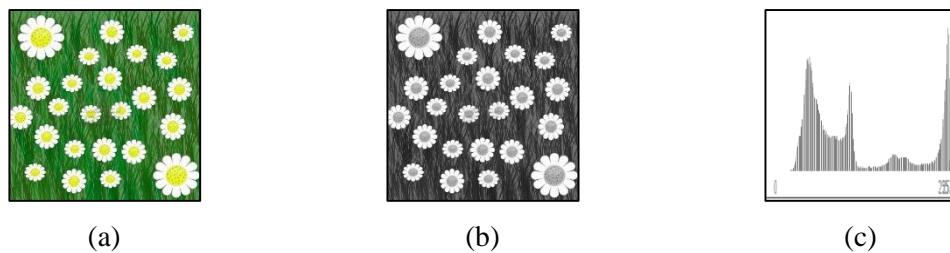


Figure 2-9: Converting colour image into one band image: (a) original image; (b) RGB averaging conversion; (c) histogram of (b).

## 2.8 Clustering

The process of organizing objects into groups based on elements' similarity in certain features is essential for different applications such as data mining, pattern recognition, and image analysis. The main two techniques of such purposes are clustering and classification techniques. However, although both techniques are based on the same process, they are not identical. In clustering, the groups have no labels and the clustering process is only to divide the set of objects into smaller subsets (clusters). The elements of each subset share common features. The process decides the number of clusters and the number of elements in each cluster. Clustering is an unsupervised process since it

does not need user intervention in any form. Classification also performs a clustering process but in a supervised manner, as the classification process labels certain elements. In clustering, the clusters must be different in order to have distinct clusters, implying different measures must be used for the differentiation process, such as, location.

Estivill-Castro states that the cluster cannot be exactly defined, that is why there are many different clustering algorithms. Most of the algorithms are application-oriented, i.e. they differ from each other based on the application at hand. The algorithm that is suitable for use in a particular problem may not work efficiently in other problem, so the selection of appropriate algorithm for a particular problem is a crucial issue

Many researches were published on this topic such as: [84], [85], [86] [87], [88], [89], [90] and [91]. Ref. [85] discusses a multitude of clustering algorithms including the algorithms that utilize computational intelligence techniques, such as, ANN, uncertainty, and evolutionary techniques. ANN clustering is widely discussed in terms of unsupervised learning algorithms as a means for clustering (such as, the self-organizing map) [92] [93] [94]. Density is applied in object clustering algorithms; Amini et al. [95] have proposed the density-grid clustering technique, where the space is divided into grids and then the regions classified based on their density; they have applied this algorithm in data streaming clustering. Torn [96] has used density clustering to solve global optimization problems. Fuzzy clustering is another attractive field of research: the most popular of these algorithms are Fuzzy C-Mean FCM, and Probabilistic C-Mean PCM [97] [98].

### **2.8.1 Clustering Types**

Clustering techniques can be classified generally into five types: Exclusive, Overlapping, Hierarchical, Centroid-based and Probabilistic clustering as shown in Figure 2-10. Exclusive clustering means that an object belongs to only one cluster. In contrast, in the overlapping clustering, each object may belong to more than one cluster as in Fuzzy clustering. Hierarchical clustering is an iterative process divided into agglomerative and divisive clustering. In agglomerative, all objects are considered as clusters, then based on some similarity criteria similar clusters are united together to form a larger cluster; this is also known as bottom-up clustering. The divisive clustering is the opposite, as all objects

are considered to be a single cluster and based on some dissimilarity measure the cluster is divided into smaller clusters. This is also known as top-down clustering [99].

For the centroid-based techniques, first the number of clusters ( $k$ ) is set. Following, the distances among the objects and the centroid of the  $k$ -clusters are calculated. The object will be a member of the nearest cluster, the new centroid is calculated, and the process is repeated until all points become members of one of the clusters. For the probabilistic techniques, clusters are defined as objects belonging most likely to the same distribution. Specifying the number of clusters is also a limitation of such algorithms since they require user intervention to specify this value.

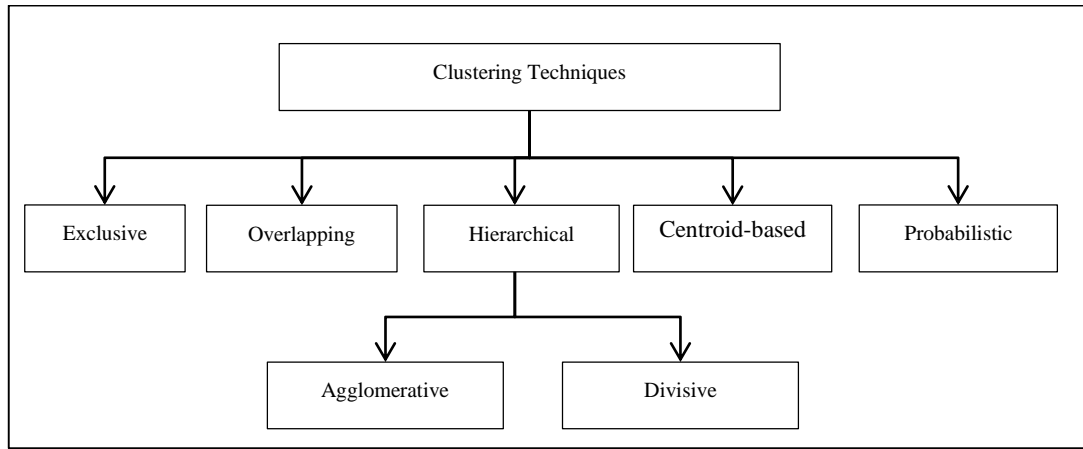


Figure 2-10: Clustering techniques.

### 2.8.2 Basic Models of Clustering

We shall start with deriving the necessary equations and modelling for the clustering techniques. First in hard clustering, a specific object should belong to only one cluster based on the clustering criterion. Let us assume that the set  $P$  contains  $N$  objects ( $p_i$ ) which are needed to be clustered such that:

$$P = \{p_i | 0 < i \leq N\} \quad 2-10$$

Moreover, let us assume further that the objects may belong to one of the  $M$  clusters  $c_j$  such that:

$$C = \{c_j | 0 < j \leq M\} \quad 2-11$$

Then we may define the mapping  $Y$  as follows:

$$Y: P \rightarrow C$$

$$Y = \{\rho_j \mid 0 < j \leq M\}$$
2-12

The object  $p_i$  is said to be a member of cluster  $c_j$  if and only if  $\rho_j(D(p_i, p_j)) = 1$  with  $p_j \in c_j$ . The function  $D(p_i, p_j)$  is a distance function, and  $\rho_j$  is a step function as given below:

$$\rho_j(x) = \begin{cases} 1 & \text{if } D(p_i, p_j) < T \\ 0 & \text{if } D(p_i, p_j) \geq T \end{cases}$$
2-13

where  $T$  is a predefined threshold.

As an example of the hard clustering technique, consider the objects shown in Figure: 2-11. In this figure, there are three clusters and six objects. The shape is the feature that is used in the clustering process.

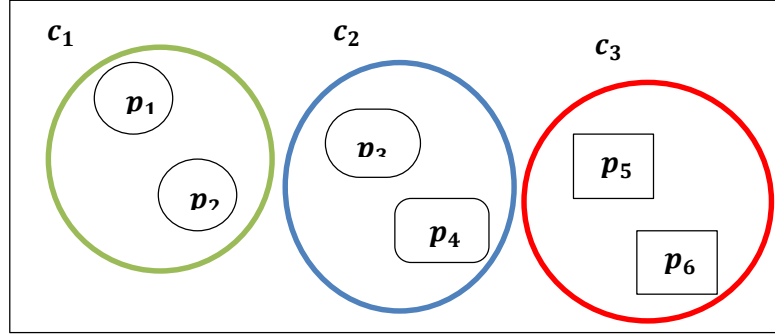


Figure: 2-11: Hard-clustering example.

$$P = \{p_1, p_2, p_3, p_4, p_5, p_6\}$$

$$C = \{c_1, c_2, c_3\}$$

$$Y = \{\rho_1, \rho_2, \rho_3\}$$

$$\rho_1(D(p_1, p_1)) = 1,$$

$$\rho_1(D(p_2, p_1)) = 1,$$

$$\rho_1(D(p_3, p_1)) = 0,$$

$$\rho_1(D(p_4, p_1)) = 0,$$

$$\rho_1(D(p_5, p_1)) = 0,$$

$$\rho_1(D(p_6, p_1)) = 0$$

$$\rho_1(P) = \{1, 1, 0, 0, 0, 0\}$$

In the same way the values of  $\rho_j$  are calculated and they will be:

$$\rho_2(P) = \{0, 0, 1, 1, 0, 0\},$$

$$\rho_3(P) = \{0, 0, 0, 0, 1, 1\}.$$

The obtained distance matrix ( $D$ ) between the objects is given in Figure 2-12 (a). It is clear from the figure that the matrix is symmetric; this is because  $D(p_i, p_j) = D(p_j, p_i)$

thus, the correspondent elements above and below the diagonal are equal. By applying the function  $\rho: D \rightarrow S$  the similarity matrix ( $S$ ) given in Figure 2-12 (b) is obtained.

	$p_1$	$p_2$	$p_3$	$p_4$	$p_5$	$p_6$
$p_1$	1	0.9	0.3	0.2	0.1	0.1
$p_2$	0.9	1	0.3	0.2	0.1	0.1
$p_3$	0.3	0.3	1	0.9	0.3	0.3
$p_4$	0.2	0.2	0.9	1	0.4	0.4
$p_5$	0.1	0.1	0.3	0.4	1	0.9
$p_6$	0.1	0.1	0.3	0.4	0.9	1

(a)

	$p_1$	$p_2$	$p_3$	$p_4$	$p_5$	$p_6$
$p_1$	1	1	0	0	0	0
$p_2$	1	1	0	0	0	0
$p_3$	0	0	1	1	0	0
$p_4$	0	0	1	1	0	0
$p_5$	0	0	0	0	1	1
$p_6$	0	0	0	0	1	1

(b)

Figure 2-12: Distance Matrices, (a)  $D$  before applying thresholding, (b)  $S$  after thresholding

In general, the above discussion can be expressed using vectors representation as follows:

$$\hat{P} = [p_1 \quad p_2 \quad \dots \quad p_N] \quad 2-14$$

$$\hat{C} = [c_1 \quad c_2 \quad \dots \quad c_M] \quad 2-15$$

$$\hat{Y} = [\hat{\rho}_1 \quad \hat{\rho}_2 \quad \dots \quad \hat{\rho}_M]^T \quad 2-16$$

$$\hat{\rho}_k = [\rho_{k,1} \quad \rho_{k,2} \quad \dots \quad \rho_{k,N}]$$

$$\rho_{k,l} \in \{0,1\}, l = 1, 2, \dots, N$$

$$\hat{C} = \hat{Y} \otimes \hat{P} \quad 2-17$$

$$\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_M \end{bmatrix} = \begin{bmatrix} \rho_{1,1} & \rho_{1,2} & \dots & \rho_{1,N} \\ \rho_{2,1} & & & \\ \vdots & & \ddots & \\ \rho_{M,1} & & & \rho_{M,N} \end{bmatrix} \otimes \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_N \end{bmatrix} \quad 2-18$$

The vector multiplication  $\otimes$  in Equation 2-18 differs from traditional multiplication in that the addition is replaced with a relative operation such that the expression  $c_1 = p_1 \oplus p_5 \oplus p_7$  for example means that the objects  $p_1, p_5, \text{ and } p_7$  are members in the class  $c_1$ .

To form larger clusters the process described above is repeated several times until a reasonable number of clusters has been obtained. In order to apply the above algorithm on clusters we need to convert the cluster into an object. One possible conversion is made by taking the average of the objects in the cluster. We shall define here the cluster centre of gravity CCG; this CCG (Centroid can be used also) can be extracted by calculating the average of the objects in the cluster. CCG will replace the two objects in that cluster, i.e. the cluster itself will be treated as an object.

In the above discussion, it was assumed that the variable  $\rho_{kl}$  can take either zero or one only, since it was assumed that the object may take one of two cases; it either belongs to

the cluster or does not belong to it. In the soft clustering techniques, the situation is different, it is assumed that an object may belong to more than one cluster with different membership value, thus, the values of  $\rho_{kl}$  may take any value between 0 and 1.

Getting back to the example discussed above, we shall analyse the problem from a different point of view. Consider the objects shown in Figure 2-13. In this figure there are three classes and six objects. The shape is the feature that has been used in the clustering process.

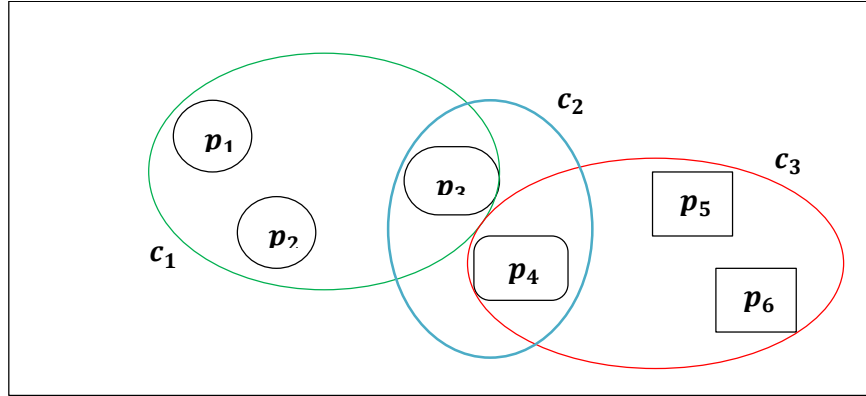


Figure 2-13: Soft-clustering example.

$$P = \{p_1, p_2, p_3, p_4, p_5, p_6\}$$

$$C = \{c_1, c_2, c_3\}$$

$$Y = \{\rho_1, \rho_2, \rho_3\}$$

$$F = \{f_1, f_2, f_3\}$$

$$\rho_1(D(p_1, f_1)) = 0.8,$$

$$\rho_1(D(p_2, f_1)) = 0.8,$$

$$\rho_1(D(p_3, f_1)) = 0.2,$$

$$\rho_1(D(p_4, f_1)) = 0.1,$$

$$\rho_1(D(p_5, f_1)) = 0,$$

$$\rho_1(D(p_6, f_1)) = 0$$

$$\rho_1(P) = \{0.8, 0.8, 0.2, 0.1, 0, 0\}$$

with  $F = \{f_1, f_2, f_3\}$  as the cluster feature set which can be used as a measure of similarity of the object with the contents of the cluster.

In the same way the values of the mappings  $\rho$  are calculated, as given in the following table:

	$p_1$	$p_2$	$p_3$	$p_4$	$p_5$	$p_6$
$c_1$	0.8	0.8	0.2	0.1	0	0
$c_2$	0.2	0.2	0.7	0.7	0.2	0.2
$c_3$	0	0	0.1	0.2	0.8	0.8

Figure 2-14 shows the graphical representation for the mapping function that maps the objects to their corresponding clusters.

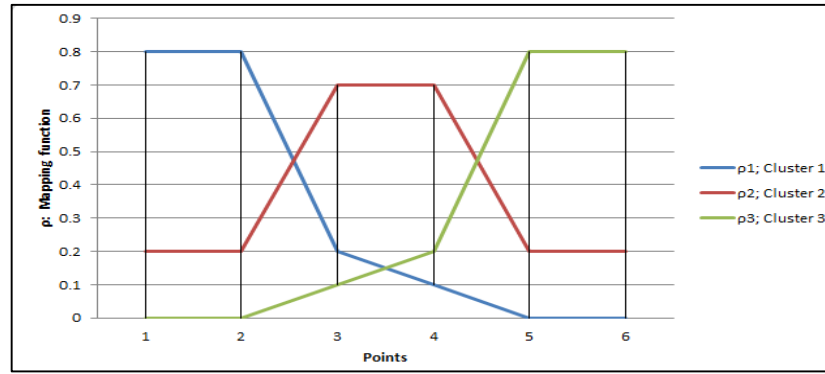


Figure 2-14: Mapping of objects to clusters using soft (Fuzzy) clustering.

It is clear from the above discussion that there is a need to specify the minimum possessive value that specifies whether the object belongs or not to a specific cluster. In the above example we have assumed that the minimum membership value is 0.2 thus cluster  $c_1$  for example will contain only the objects ( $p_1, p_2, \text{and } p_3$ ).

### 2.8.3 Sequential Clustering Technique (SCT)

In this clustering technique, each object  $p$  is compared to other objects, and the object with minimum distance will be linked to the object  $p$ . Both objects are marked as checked and excluded from the next comparison. As an example on this method, consider the points shown in Figure 2-15. In this figure, (a) shows the points that need to be clustered and the results of applying the sequential clustering technique is shown in (b).

The main drawbacks of this technique are first, it will produce a total of  $\frac{N}{2}$  clusters, and second, the object that is tested first will decide the nearest point. We shall refer to this as the *first point selection problem* (FPSP). From Figure 2-15, it is clear that the distance  $\overline{p_1 p_2}$  is the minimum distance; then there is a connection between these points. For the point  $p_2$ , the point  $p_3$  should be the nearest point, but according to the algorithm  $p_1$  is the closest point since we skip it from the comparison after it gets engaged with another point, thus we got  $p_4$  as the closest point to  $p_3$  and not  $p_2$ . Based on the above discussion we may get inappropriate results. As shown in Figure 2-15, we got three clusters instead of two.



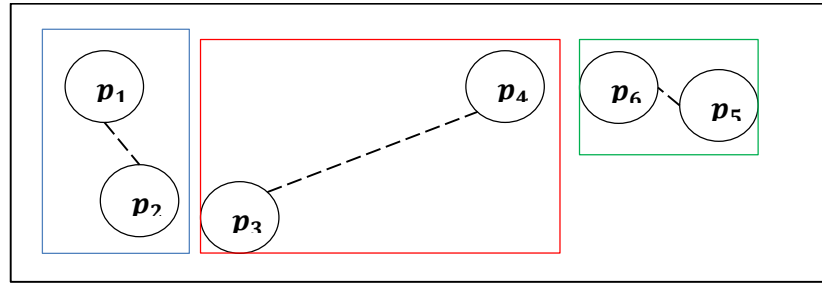


Figure 2-15: SCT clustering results.

An example of real data is shown in Figure 2-16. In this figure, the process was started with point 1, which got point 2 as the closest neighbour, and both points 1 and 2 shall not be considered in the next comparison. Another example of the erroneous results is in the point number 12. From the figure it is clear that points 10, 11, 13, and 14 are closer to 12, but it was selected by 7 since point 7 will search for its closest neighbour before other points; thus 12 will share the same cluster with 7.

In order to overcome the two limitations mentioned above, a refined sequential clustering technique (RSCT) will be discussed here. The problems manifested themselves due to the exclusion of the connected objects from further comparison. In this technique, we shall define the cluster centre of gravity (CCG) which is the average of the features of the two connected objects. Instead of excluding the connected objects, the connected objects will be replaced by one object with their average features; the CCG.

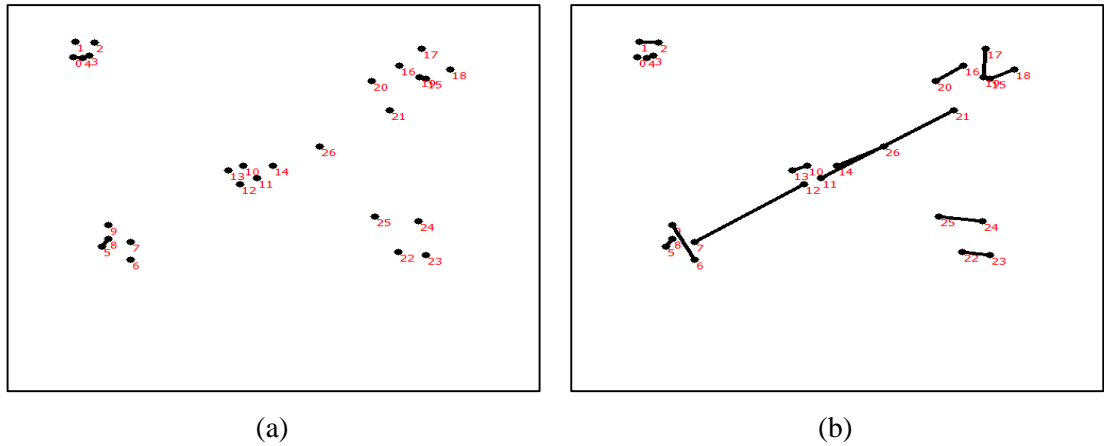


Figure 2-16: Clustering process, (a) points to be clustered, (b) results of applying sequential clustering.

From Figure 2-17, it is clear that the distance between the points  $p_1$  and  $p_2$  is minimum, thus they should belong to the same cluster, and in the same way the rest of clusters will be constructed. The points  $p_1$  and  $p_2$  are replaced by the point  $p_{12}$  and points  $p_3$  and  $p_4$

are replaced by  $p_{34}$  and so on. The above operations are repeated again on the new set of points to construct larger clusters.

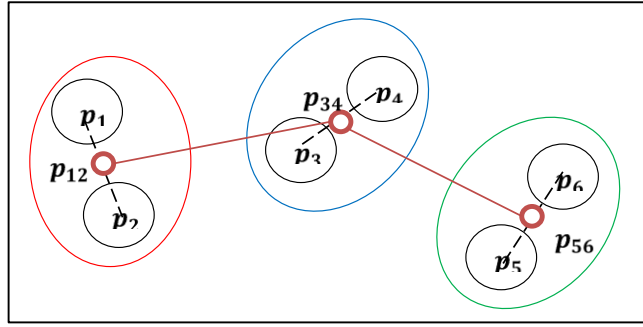


Figure 2-17: RSCT process.

Figure 2-18 shows the results of applying the refined sequential clustering technique, it is clear from the figure that the number of clusters here is not equal to  $\frac{N}{2}$ , but still there is a limitation which is that without a predefined threshold value, all points will be connected to form a single cluster. Threshold can be used to cluster the set of objects into more than one cluster, so let us define a threshold value  $T$ . Any distance larger than this threshold should be ignored and the points with that distance do not belong to the same cluster. Figure 2-18 shows the effect of the selection of the value of  $T$  on the number of clusters and the size of each cluster. The results obtained from this technique are very much better than those obtained from the SCT, but still the effect of the first point selection (FPSP) affects the results.

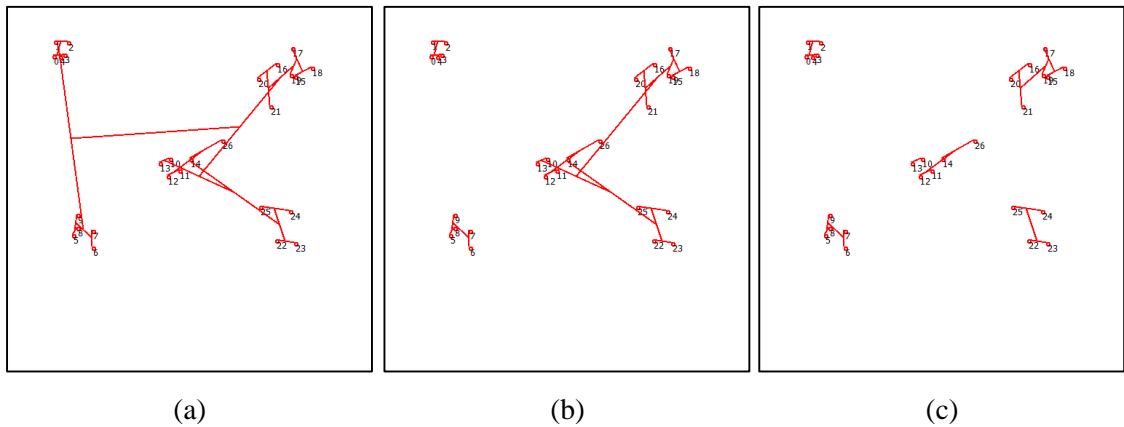


Figure 2-18: Clustering process, results of applying RSCT, (a) no threshold is used, (b)  $T=200$ , (c)  $T=100$ .

### 2.8.4 Parallel Clustering Technique (PCT)

In order to overcome the FPSP which is due to the sequential nature of the above two techniques, a modified technique can be used. This technique is named as ‘parallel’ since it compares all the points simultaneously and at the same time i.e. in parallel. In this technique, the distance matrix  $D$  is calculated first before the comparison process, as was the case in the previous techniques. After the distance matrix has been calculated, the comparison among the distances in the matrix shall take place. Any two points with minimum distance shall be considered as points in the same cluster. Consider the following distance matrix:

$$D = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1N} \\ d_{21} & d_{22} & \dots & d_{2N} \\ \vdots & & \ddots & \vdots \\ d_{N1} & & \dots & d_{NN} \end{bmatrix} \quad 2-19$$

As discussed above, the matrix is symmetrical, so it is clear that there is no need to consider the elements below the diagonal. Another important property of the matrix  $D$  is that there is no need to consider the diagonal points since they represent the distance between the object and itself and usually equal to zero. Thus, only the elements above the diagonal will be considered in the calculations. The distance matrix will be as follows:

$$D = \begin{bmatrix} 0 & d_{12} & \dots & d_{1N} \\ 0 & 0 & \dots & d_{2N} \\ \vdots & & \ddots & \vdots \\ 0 & & \dots & 0 \end{bmatrix} \quad 2-20$$

For each object  $p$ , the minimum distance shall be extracted and the object corresponding to that minimum distance will share the object  $p$  the same cluster.

#### **Thresholding**

In this technique, there is no need to state a pre-specified thresholding value; instead, we will use the number of clusters. When the number of clusters has been specified then the threshold shall be specified automatically.

Let  $n_T$  be the number of clusters, then the value of the threshold  $T$  can be extracted from the distance matrix, which is equal to the  $(N - (n_T + 1))th$  element in the ascending sorted distances list. This can be achieved by sorting a list containing all the distance

values in ascending order and selecting the  $(N - (n_T + 1))th$  distance as the threshold. This threshold is used to separate the clusters from each other.

Figure 2-19 shows the results of applying the PCT with different numbers of clusters. Specifying the number of clusters rather than the threshold value was also used in the k-mean clustering technique.

### ***Reduction in Calculations***

Because of the symmetrical properties of the distance matrix, and the fact that the object does need to be compared with itself, the number of calculations could be considerably reduced. Since the number of objects is  $N$  and the comparison is performed between two objects every time, then the total number of comparisons needed without the reduction of the matrix is equal to  $N^2$  since each object needs to be compared with all other objects including itself. This number of comparisons can be reduced if comparing the object with itself is excluded. In that case each object will be compared with  $(N - 1)$  objects and hence the total number of comparisons becomes  $(N \times (N - 1))$ . With the consideration of symmetry of the distance matrix and ignoring the comparison of the object with itself then the number of comparisons will be equal to  $(N \times (N - 1))/2$ .

### **2.8.5 Iterations Stopping Criteria**

As clustering is an unsupervised iterative process, a criterion at which the iterations should be terminated needs to be determined, so that the algorithm stops. There are three possible stopping criteria: thresholding, specifying the number of clusters and automatic. For the first two criteria, the user specifies a reasonable thresholding value or a reasonable number of clusters, while in the third criterion the algorithm stops automatically without any interference from the user.

The selection of the threshold value is a crucial issue which drastically affects the results, therefore it should be selected carefully. The main issue with thresholding is that there is no theoretical justification for the selection of the threshold, while manual selection requires the user to provide a suitable threshold value, which limits its applicability. Selecting a small threshold value will increase the number of clusters and will over-cluster the objects which means forming a large number of clusters with a small number

of objects, while setting it to a large value results in a small number of clusters with a large number of objects in each cluster.

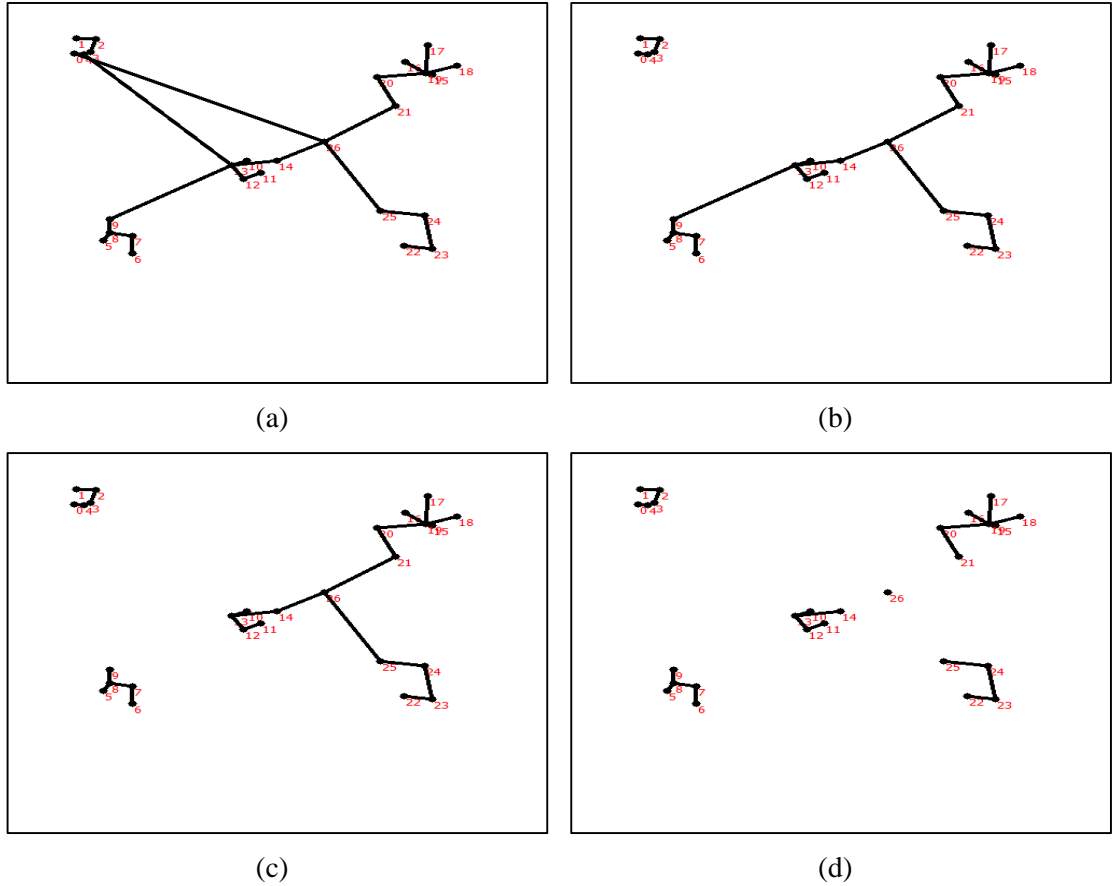


Figure 2-19: Results of applying PCT with different number of clusters  $n_T$ , (a)  $n_T = 1$ , (b)  $n_T = 2$ , (c)  $n_T = 3$  (d)  $n_T = 6$ .

The selection of the number of clusters is used in k-mean methods seed points selection. Such techniques require the user to select a suitable number of clusters manually. For both thresholding selection and number of clusters selection, human interference is necessary which transform the clustering into a supervised process.

Some efforts have been made to find an automatic stopping criterion at which the clustering process stops [100], such as, using density of points distribution [96], where the space is divided into finite grids and then the density at each grid is measured to specify the cluster size. The process needs to specify the size of the grids and the accepted density in each grid to be considered or ignored. Mojena [91] has suggested extracting the clusters for all values of  $k$  from one to  $N$ , and then finding the mean of all values of  $k$  to find the optimum value [91], but it is clear that the process is computationally

inefficient. Hartigan [101] has calculated the stopping criterion function for two successive values of  $k$  and compares the results. If there is no change, then the process should be terminated. Pedersen and Kulkarni [102] have suggested a method that combines the above two methods with a third statistical method to find the optimum stopping criterion. Nonetheless, figuring out a general stopping criterion is not feasible since there is no possible representation for the points distribution in the space unless a thorough study of the nature of the problem is presented first.

### 2.8.6 Application on Points Clustering

As an application for the above technique is points clustering, each point will be compared to other points; in this case, since we are dealing with points, then the feature that will be compared is the coordinates of the point  $(x, y)$ . The points are said to be related to the same cluster if the distance between them is minimum. Figure 2-20 shows the result of applying the above technique on real data.

Figure 2-20 (a) shows the points (salient points), (b) shows the points after removing the background for clarity reasons, (c) shows the result of applying SCT, (d), (e), and (f) show the results of applying RSCT with different values for the threshold value  $T$ . The images in (g) and (h) show the results of applying PCT for different values of clusters number  $n_T$ .

## 2.9 Computational Intelligence

According to [103], computational Intelligence (CI) is the “simulation of human intelligence on a machine, so as to make the machine efficient enough to identify and use the right piece of knowledge at any given step of solving a problem”. The main goal of CI is to add the ability of learning and thinking to computer applications. CI presents many applications in different fields like expert systems, speech and natural language understanding, intelligent control, image understanding and computer vision applications [104]. Production systems, swarm intelligence, neural nets and genetic algorithms are examples of widely used techniques in CI applications. Machine learning is the most important part of CI, where the machine can adapt its knowledge according to the input it gets from the environment. One of the central applications of CI algorithms is in the

field of image processing applications, especially in the field of image recognition and computer vision.

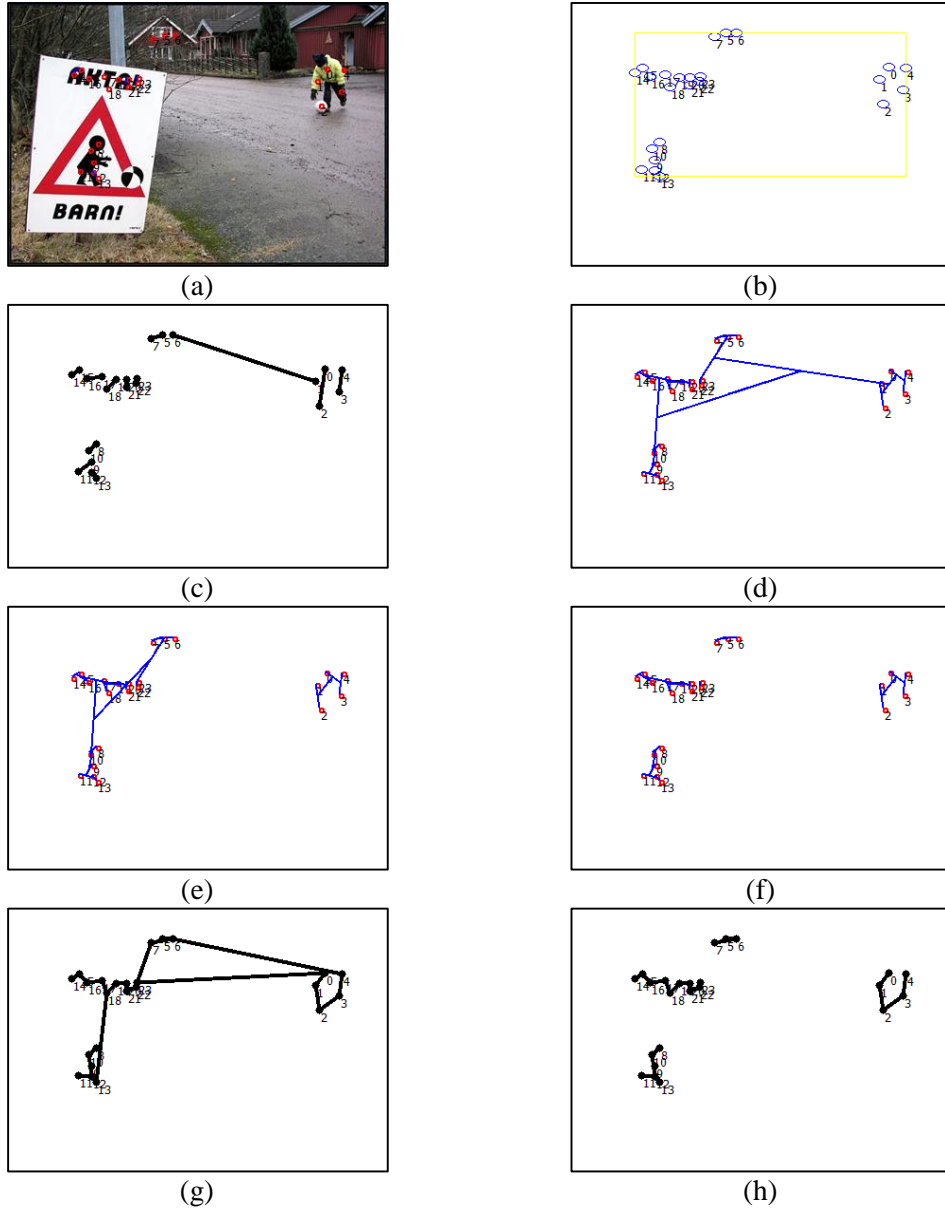


Figure 2-20: Points clustering example, (a) gaze points, (b) gaze points after removing the background, (c) SCT, (d) RSCT  $T = \infty$ , (e) RSCT  $T = 100$ , (f) RSCT  $T = 50$ , (g) PCT  $n_T = 1$ , (h) PCT  $n_T = 4$ .

### 2.9.1 Machine Learning Techniques

Machine learning utilizes the available information to improve the machine's understanding, through which its performance is improved. There are three main categories of machine learning techniques: supervised learning technique (SLT), unsupervised learning technique (ULT) and reinforcement learning technique (RLT).

In SLT, also known as Associative Learning, the machine is provided with a set of inputs and the corresponding desired output. The training process is improved by finding the error ratio of the actual output and the desired output. The machine gathers some knowledge from the training process in such a way that it can give correct responses to similar inputs. Examples of SLT are inductive learning and analogical learning.

ULT is used if the desired output is not available. For this learning category, the machine is provided with input only and it should update itself to generate classes for the similar objects or the objects with similar features.

RLT is an intermediate form of the supervised and unsupervised learning. The learning machine does some action on the environment and gets a feedback response from the environment. The learning system grades its action good or bad based on the environmental response and accordingly adjusts its parameters [103].

## **2.9.2 Artificial Neural Networks**

One of the most important techniques in CI is the Artificial Neural Networks (ANN) or Neural Networks (NN) that are a simulation to the neurons in biological brains. Biological brains are highly complex and nonlinear computing devices. Neurons are the primary elements of the biological brains and consist of four main structural components: the dendrites, the cell body, the axon, and the synapse. The dendrites act as receptors that receive signals from several neighbourhood neurons and they pass these signals on to a little thick fibre called dendron. The received signals are processed in the cell body and the resulting signal is transferred through a long fibre named an axon. At the other end of each axon exists an inhibiting unit called a synapse. This unit controls the flow of neuronal current from the originating neuron to the receiving dendrites of neighbourhood neurons [103] [105].

Neural networks are very powerful computational devices due to their parallelism, and fault and noise tolerance capabilities. Most of the computational operations are performed in the training phase, which is needed just once, followed by generalization which is a computational efficiency process.

ANN can be classified based on the number of its layers as Single Layer (SLANN) and Multi-Layers (MLANN). Single layer perceptron is an example on the SLANN, which



contains only one layer that links the input with the output. SLANN applications are limited to some simple linear classifications and it fails in nonlinear classification. Multilayer perceptron is an example of MLANN, which is a feed forward network consisting of more than two layers. For example, it consists of one or more hidden layers between the input and output layers [106] [107]. With multilayer perceptron, more complex problems can be solved which are not solvable with single layer. In contrast to single layer, multilayer can separate the space into more complex decision regions given that every layer is trained in the same training algorithm of single layer perceptron. Lastly, the number of nodes assigned to each layer varies from one layer to another.

Learning or stimulation is defined as the process of changing the free parameters of ANN by adapting the connections weight between the neurons or synapse. Many techniques have been proposed such as: Error Correction Learning (ECL), Memory-Based Learning (MBL), Hebbian Learning (HL), Competitive Learning (CL), Boltzmann Learning (BL) and Associative Memory Learning (AML).

In the last few decades many ANN models were developed, out of which the most well-known ANN is Back Propagation Artificial Neural Network (BPANN) which is a feed-forward ANN that uses supervised training. For a given input pattern, the output vector is estimated by forward pass. Following, the error value is calculated and propagated back through the network to update the weights within the network. Hopfield Artificial Neural Network (HANN) is another well-known model which consisting of a number of fully interconnected binary nodes which at any given time represent a certain state. The HANN maps binary or bivalent input sets on binary output sets. It is suitable for use with images that have binary values, like edge maps and binary images, in which only two grey levels are used. HANN utilizes supervised learning techniques in which different patterns are used to train the network for each input with which there should be an associated output.

Self-Organizing Map ANN (SOMANN) is a type of vector quantization method. SOMs are trained in an unsupervised manner with the goal of projecting similar d-dimensional input vectors to neighbouring positions (nodes) on an m-dimensional discrete lattice. The update is gained by only one node, which is the winner node. After training, the input space is subdivided into  $q$  regions, corresponding to the  $q$  nodes in the map. The learning

process of SOM is known as competitive learning, in which the output neurons compete amongst themselves to be activated, with the result that only one is activated at any one time. This activated neuron is called a winner-takes-all-neuron or the winning neuron.

### ***Application of ANN in Image Processing***

ANNs are widely used in the field of image processing in various applications starting from pre-processing to image recognition. ANN provides image processing techniques with nonlinearity which is important in image processing as most of the image features are nonlinear and require nonlinear processing. Since its incipient stages, Neural Computing has been used for pattern recognition and image classification purposes. Images can be used as input to the neural network in different ways; one of these is by considering the complete image as an input vector to the ANN. Another way is by using some of the features extracted from the images as input to the ANNs, such as GLCM and edges.

Pre-processing is an application of ANN in image processing. An image can consist of image reconstruction and/or image restoration. To perform pre-processing, ANNs have been applied in the form of optimization of traditional pre-processing functions, approximation of mathematical models, and classification of a pixel based on its neighbours.

Optimization can use Hamming ANN (HANN); however, mapping the actual problem to the energy function of the HANN might be difficult. In order to overcome this, the original problem needs modifications. Although having managed to map the problem appropriately, the HANN is a useful tool in image pre-processing, although convergence to a good result is not guaranteed.

Filtering is another pre-processing technique that utilizes ANN. Both supervised and unsupervised training can be used to perform such operations. In the supervised learning technique, the ANN is trained using some templates and the ideal output of the ANN. For example, assume that an algorithm to perform low-pass filtering is used. This removes the noise from the image; noise could be a light pixel in a dark region or vice versa. After training is completed, the classification process starts, in which each sub-image is used as an input to the ANN and the value of the centre pixel is the output.

### 2.9.3 Uncertainty and Fuzzy Intelligence

Fuzzy logic is a representation of uncertainty in which each logic operation like OR, AND, and NOT is represented in a way that it differs from the True/False binary space [108]. Let  $X = \{x\}$  be a set of points and  $\mu_A(x)$  be the characteristics function or membership function, then  $\mu_A(x)$  associate each  $x$  a real value in the interval  $[0, 1]$ . The function  $\mu_A(x)$  measures the membership of  $x$  in the class  $A$ . The closer the value is of  $\mu_A(x)$  to one, the higher the probably of its being a member of  $A$ . All the binary logical operations are defined according to the above definition like OR operation is replaced by MAX, AND is replaced by MIN, and NOT is replaced by complement, the detailed theory of which is found in ref. [108].

Fuzzy sets are an extension to the idea of crisp sets, in which an element is a member in a class if the value of the membership function is one; and not a member in the class if it is zero. According to this definition, an element is either a member of the class or not a member at all. Fuzzy sets propose a ‘partial’ membership where the strength of membership is measured. Partial membership allows for an element to be a member in more than one class with different membership values.

#### *Linguistic Variables and Terms*

Fuzzy Linguistic variables (FLVs) are the input or output variables of the system. FLVs values are words or sentences from a natural language, instead of numerical values. For example temperature values can be described by linguistic terms as too cold, cold, warm, hot and too hot, these terms describe a specific feature of the temperature expressed as:  $T(t) = \{ \text{too cold , cold, warm, hot, too hot} \}$ .

The If–Then Rules specify a relationship between the input and the output in fuzzy logic. Fuzzy relations present a degree of presence or absence of association between the elements of two or more sets. Assume  $U$  and  $V$  are two universal sets; a fuzzy relation is a set in the product space and is characterized by the membership function. A fuzzy relationship  $R(U, V)$  is a subset of the cross product  $U \times V$ .  $R$  is characterized by the membership function  $\mu_R(x, y)$  where  $x \in U, y \in V$  and  $\mu_R(x, y) \in [0, 1]$ . The If-Then Rule assumes that “If  $x$  is  $A$  Then  $y$  is  $B$ ” for all  $x \in U$  and  $y \in V$  and has a membership function  $\mu_{A \rightarrow B}(x, y) \in [0, 1]$ . The **If** part of the rule “ $x$  is  $A$ ” is the *antecedent* or

*premise*, while the *Then* part of the rule, “*y is B*” is the *consequent* or *conclusion*. The If–Then Rule involves two steps: the first step is to evaluate the antecedent, which involves fuzzifying the input and applying any necessary fuzzy operators, and the second is implication or applying the result of the antecedent to the consequent, which essentially evaluates the membership function.

Fuzzy Inference System (FIS) is a system that nonlinearly maps the input data into a scalar output using fuzzy rules that involved the membership function, if–then rules, fuzzification and other fuzzy specific operations. FIS consists of four blocks:

1. The Fuzzifier maps the input into corresponding fuzzy memberships.
2. Inference engine is responsible on the mapping of input fuzzy sets to output fuzzy sets.
3. The Defuzzifier maps output fuzzy sets into a crisp number.
4. The rulebase contains linguistic rules that are provided by experts.

Regions and features in an image are not crisply defined due to the nature of the image and pixels’ values, thus there are some levels of fuzziness in extracting features. Any decision taken at any level will affect the higher level. Hence, uncertainty should be considered when moving from one level to another. In other words, the processes performed at higher levels retain as much information as possible from the input image with least uncertainty. To make this idea clearer, consider the process of identifying an object from its background. If the first stage, which is – in this case – an edge detection process, involves some uncertainty due to ill-definition of the edges due to noise, then overlapping with other objects, or even shadow, is passed on to the following stage. This uncertainty may affect the stages of processing and recognition. The incertitude in an image may be defined in terms of greyness ambiguity or spatial ambiguity or both. Greyness ambiguity is defined as the indefiniteness in deciding whether the pixel is white, grey, or black, while spatial ambiguity is a geometrical ambiguity and may occur in the shape or geometry, as in finding the centroid or shape edges of a region.

The image  $I$  of size of  $N \times M$  can be represented as an array of fuzzy singletons. Each member in this array has a value of membership denoting its degree of possessing some property, such as brightens, which can be represented in fuzzy notation as:

$$I = \{\mu_x(p_{mn}) : m = 1, 2, \dots, M; \text{ and } n = 1, 2, \dots, N\} \quad 2-21$$

where  $\mu_x(x_{mn})$  is the degree of possessing such property  $x$  by the  $(m, n)th$  pixel, and can be defined using global, local or spatial information [109].

## 2.10 Conclusions

In this chapter, the theoretical background of the research is introduced and discussed to furnish the necessary theoretical background for the rest of the theoretical derivation of the thesis. Topics such as CBIR and CI are highlighted and reviewed as they are required in the proposed algorithms. Furthermore, connective elements, such as human attention and saliency were presented, given that the purpose of this research is to present a system which utilizes these features in image contents identification.

To support the technologies explained in the specialized literature, wherever possible, experimental results obtained from implementing various algorithms are presented.

# Chapter 3

## Data Collection and Analysis

### 3.1 Introduction

In this chapter, different standard datasets shall be discussed and analysed. Standard datasets are used to simplify the benchmarking between the proposed algorithms and the existing algorithms. Mainly two kinds of datasets are discussed and used; saliency dataset and image retrieval datasets. The saliency datasets are used to test the saliency extraction algorithms, while the retrieval datasets are used to test the image retrieval algorithms.

In addition, the developed datasets shall be discussed as well, and the code provided by the authors of the existing algorithms that are going to be used for benchmarking is tested against the developed datasets.

### 3.2 Datasets

#### 3.2.1 Saliency Datasets

The developed algorithms are needed to be tested using different datasets. In order to benchmark the proposed algorithms (such as the saliency extraction algorithm) with other existing algorithms, the following standard datasets are used:

1. MSRA Salient Object Dataset [110]

2. MIT Saliency Dataset, [111]
3. Image and Visual Representation Group IVRG Dataset [112]
4. LI, Jian Dataset [113] [114]
5. Web Object Saliency Database [115];
6. Photos were collected by the author.

This research dataset is constructed with the assistance of the above mentioned datasets. The selected images in the dataset are classified in four classes as follows:

- 1- The first class (C1) contains images with a single large uniform object placed in the centre.
- 2- The second class (C2) contains images with a single isolated object but not uniform nor in the centre.
- 3- The third class (C3) contains images with multiple isolated salient objects and with fewer details in the background.
- 4- The fourth class (C4) contains images with multiple objects with overlapping and with coarse background.

It is clear that the difficulty of saliency detection increases with the classes, i.e. C4 contains the most difficult images. The defined classes are constructed based on the saliency difficulty, where C1 presents the least complex saliency to detect while C4 presents the highest level of difficulty.

The selected dataset contains the datasets used in most of the saliency identification algorithms such as [116], [113], [117], [118], [119], [120], [121], [122], [123] and [124].

The ground truth data is constructed using two experiments. The first one is the manual ground truth data, which is obtained from the human's labelling of the images. The second one is measured with eye-tracking data, in which the ground truth data is collected with eye-trackers.

### **3.2.2 Image Retrieval Dataset**

The retrieval process was applied on the WANG dataset [125]. This dataset was used by most of the image retrieval algorithms. WANG dataset is a subset of 1,000 images of the

Corel stock photo database, which are manually selected and classified into 10 classes with 100 images each. The images in each class are relevant to each other, but at the same time, present some similarity with images in other classes. This database is available online at: <http://wang.ist.psu.edu/docs/related/>. However, WANG dataset suffers from a certain drawback, which is that it was organized to give a high retrieval rate as it contains many similar images. It is not feasible to test the proposed algorithm using this dataset. Figure 3-1 shows examples on the high retrieval rate of the WANG dataset using the histogram feature as a similarity measure. From this figure, it is clear that the retrieved images are very similar to the query image. This is because the dataset was designed to have 100 images from each class in 10 different classes. To make our experimental test more realistic a new dataset has been constructed containing different images with different contents yet with similar features. The traditional retrieval algorithm was tested against the new dataset and the results compared with the proposed algorithms

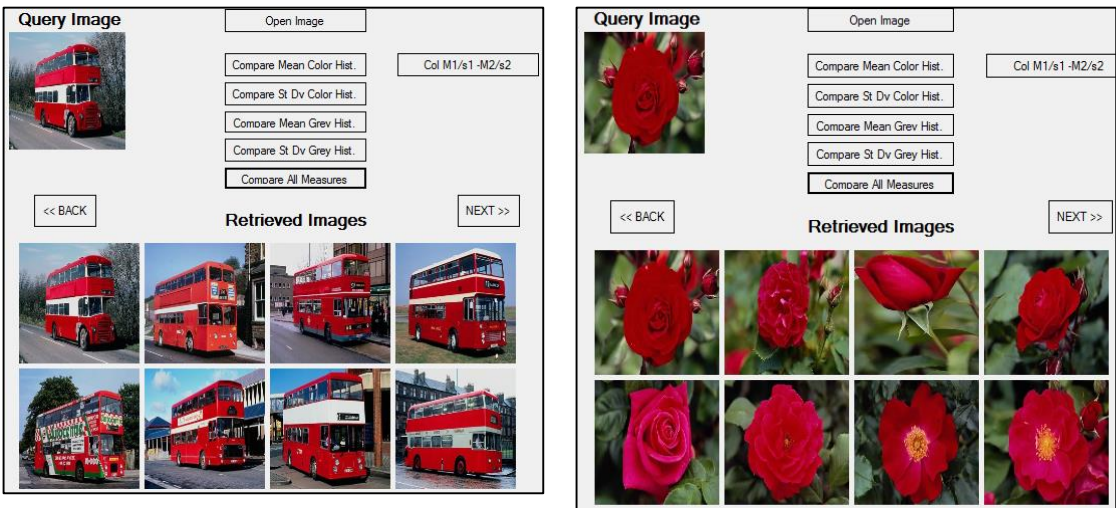








Figure 3-1: High retrieval rate of WANG dataset.

The new dataset considers the variation of the image contents and contains similar object in different background or different objects in similar background. Table 3-1 shows examples of the contents of the constructed dataset. From the table it is clear that the contents are different but the colour features contents are similar for each case.



Table 3-1: Different image contents with similar image features.

Case	Image 1	Image 2	Image 3
1			
2			

### 3.3 Saliency Ground Truth Data

This Section presents a discussion about the ground truth data which is used in saliency algorithms evaluation and benchmarking. The ground truth data includes two categories; manually labelled salient regions and eye tracking data. The first category contains images that were labelled manually by people to highlight the salient regions, while the second category contains images that were labelled by eye tracking device. Thus, we shall refer to the three types of data as follows:

- 1- Manual Ground Truth Data (MGTD), which contains interest points that have been extracted manually by humans using a friendly used software designed as part of this work. The experiment description is given in the following section.
- 2- Eye Tracking Ground Truth Data (ETGTD), which contains interest points that have been extracted using an eye tracker device.
- 3- Computationally Extracted Data (CED), which contains interesting points that have been extracted using computational techniques such as the proposed algorithm results, as will be explained in the next chapter.

#### 3.3.1 Manual Ground Truth Data (MGTD)

This type of data is useful as ground truth data to evaluate the computationally extracted salient regions. The data is collected manually by using the experiment that was designed for this purpose. The following is the description and the results of the experiment that we have been carried out to collect the manual data.

### ***Manual Data Collection Experiment:***

In this experiment, software with friendly interfacing was designed and the participants were requested to intentionally mark the salient points in the image.

Figure 3-2 (a) shows the interface of the software that was used in manual salient regions extraction. The participant uses the mouse to click on the regions he/she finds interesting; (b) shows the clicks that were done by a participant, and (c) shows the points after clustering.

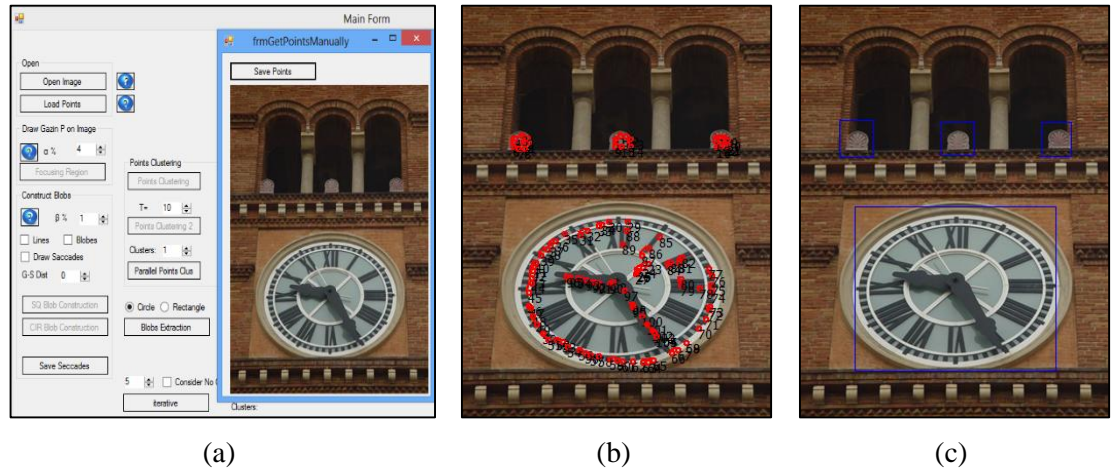


Figure 3-2: Manual points extraction, (a) software interface, (b) points clicked by the user, (c) clustered points.

In this experiment, images from different classes and with different sizes, levels of saliency, difficulty and nature were used. These images were displayed to the participant on a 14-inch HD screen. 34 participants participated in this experiment. The diversity of the participants was taken into account so as to get a realistic feedback. The charts in Figure 3-3 show the distribution of the participants based on their gender, age, nationality, and education level. This diversity is required to collect feedback that reflects different points of view.

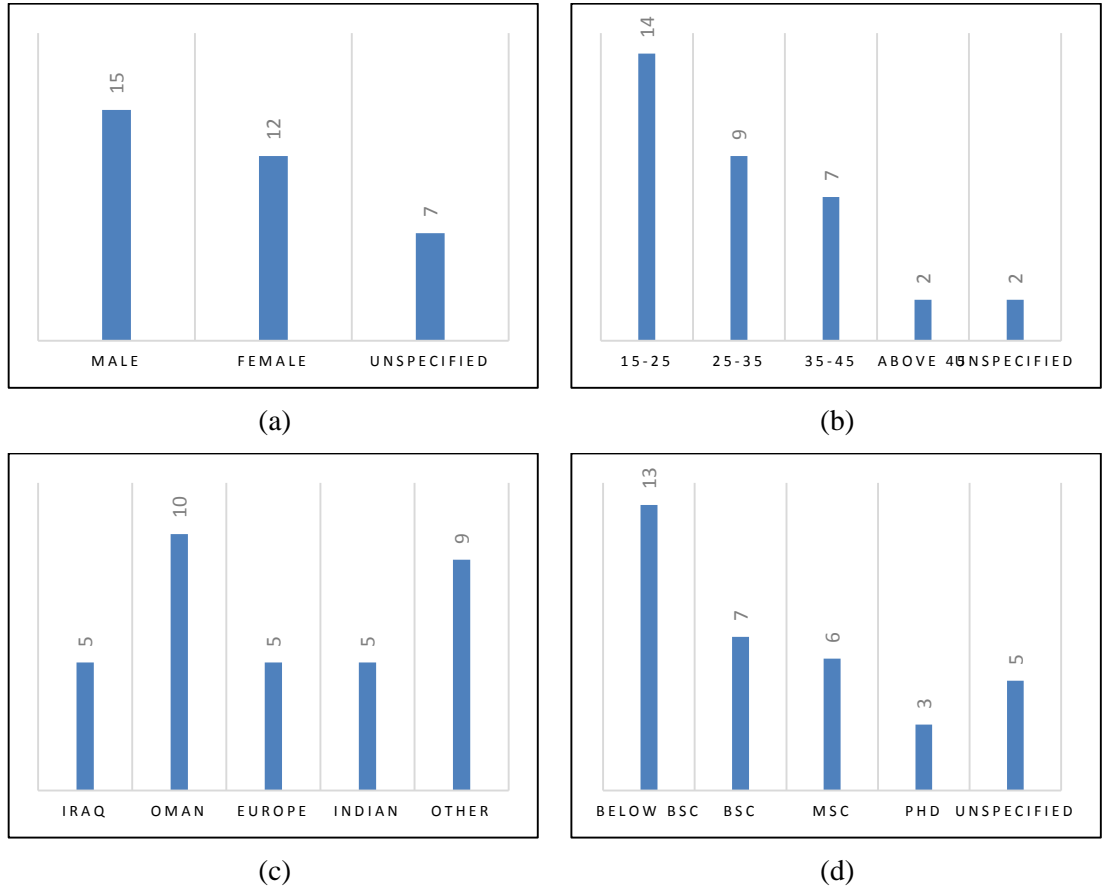


Figure 3-3: The distribution of the participants based on (a) gender, (b) age, (c) nationality, and (d) education level.

Since it is not possible to find an intersection between the sets of points provided by the participants because it is not possible for two participants to click on the same point  $(x, y)$ , we may consider the intersection of the surrounding region instead. Usually a region of size of  $7 \times 7$  is a good choice to represent each point. This is because humans do not look at a single point and when they click on a point, they usually look at the surrounding small region. If we represent the surrounding region with a circle centred at the point specified by a user, then the average point can be extracted by averaging the centres of the intersected circles.

Consider the centres of the intersected regions are given by the set  $P_I$ , which contains the points  $P_i(x_i, y_i)$  where, i.e.  $P_I = \{P_i \mid i = 1, 2, \dots, k\}$  then the result of the intersection, will be a point  $P(x, y)$  such that:

$$x = \frac{1}{k} \sum_{i=1}^k x_i \quad , \quad y = \frac{1}{k} \sum_{i=1}^k y_i \quad 3-1$$

### 3.3.2 Eye Tracking Ground Truth Data (ETGTD)

Eye trackers are widely used nowadays in different applications; these devices are used to collect the gaze points from the user. These gaze points can give an excellent impression of the interest of the user as a human can pay attention only to important (based on his point of view) parts of the scene. In our research, the gazing data has been collected using an eye tracking device and OGAMA Version: 4.2.4470 software (available online at <http://www.ogama.net>) to convert the gaze points into spatial coordinates. The collected data, using OGAMA, contains much information, such as the user name, the fixation points, the gaze points, the fixation time, etc. as shown in Figure 3-4 (d). This data can be used to determine where the interest of the user lies by tracking his eyes' gazing points, and also can be used to identify the important (salient) regions in an image by taking the average of the gaze data for several users.

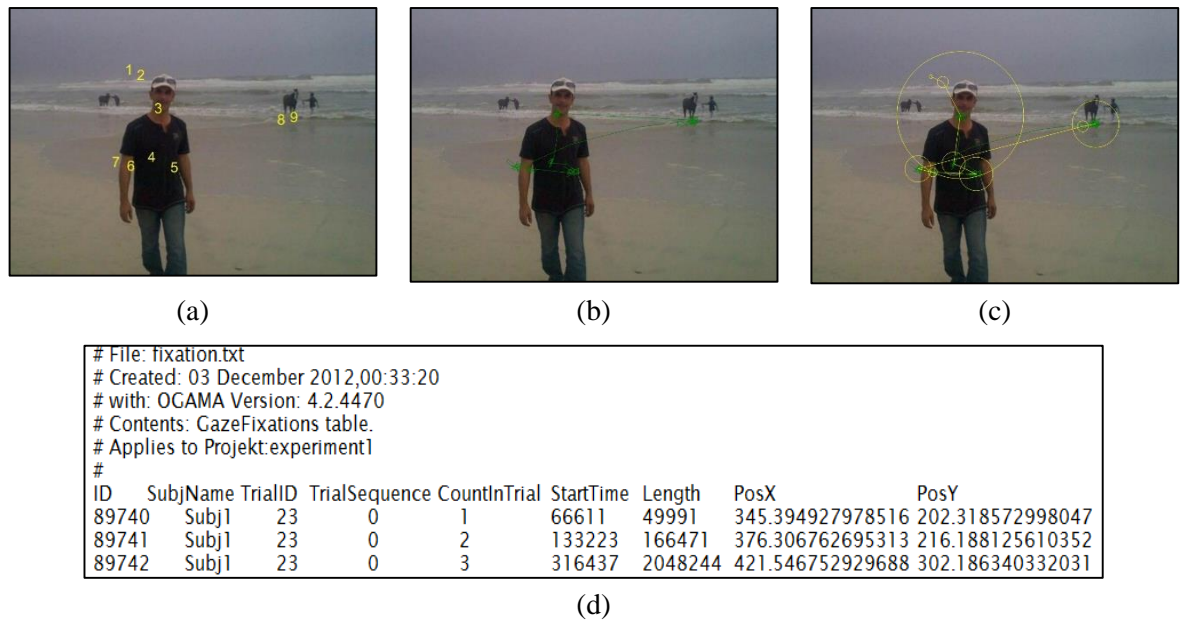


Figure 3-4: The data obtained from OGAMA, (a) the gazing points, (b) the fixation points, (c) the regions of interest, (d) the gazing raw data.

The adjacent points can be connected together to form the region of interest as shown in Figure 3-4 (c). In this figure, the yellow circles represent the regions of interest based on the fixation points. The diameter of the circle depends on the number of fixation points and the time consumed at each point, the larger the number of fixation is, the larger the diameter of the circle is.

In this research, we shall use the gaze points obtained from the eye tracker to identify the user-oriented regions of interest. The data that will be used is the gaze points' coordinates  $(x, y)$  and the time consumed at that point  $t$ .

For each image, there will be a set of interesting points that can be represented by the vector  $\hat{P}$ . Each interesting point consists of two components; the coordinates and the corresponding time. This vector can be represented as follows:

$$\hat{P} = [< p_0, t_0 > \quad \cdots \quad < p_i, t_i > \quad \cdots \quad < p_n, t_n >]^T \quad 3-2$$

where the ordered pair  $< p_i, t_i >$  is the interesting point,  $p_i = (x_i, y_i)$  is the coordinate of the gaze point, and  $t_i$  is the time consumed at that point. We shall refer to  $p_i$  as the gaze point and to the ordered pair  $< p_i, t_i >$  as the interesting point. The obtained points shall be clustered and used as ground truth data.

Participants were selected to achieve a good diversity in age, gender, education level, and nationality. This diversity is required to have a realistic and general data. Figure 3-5 shows the distribution of the participants based on their gender, age, nationality, and education level. The images were displayed on a 14-inch, HD LED screens with resolution of  $1280 \times 768$ .

The images that were used in the test have been selected carefully to achieve diversity in nature and contents, although some images contained the same object in a different background or environment.

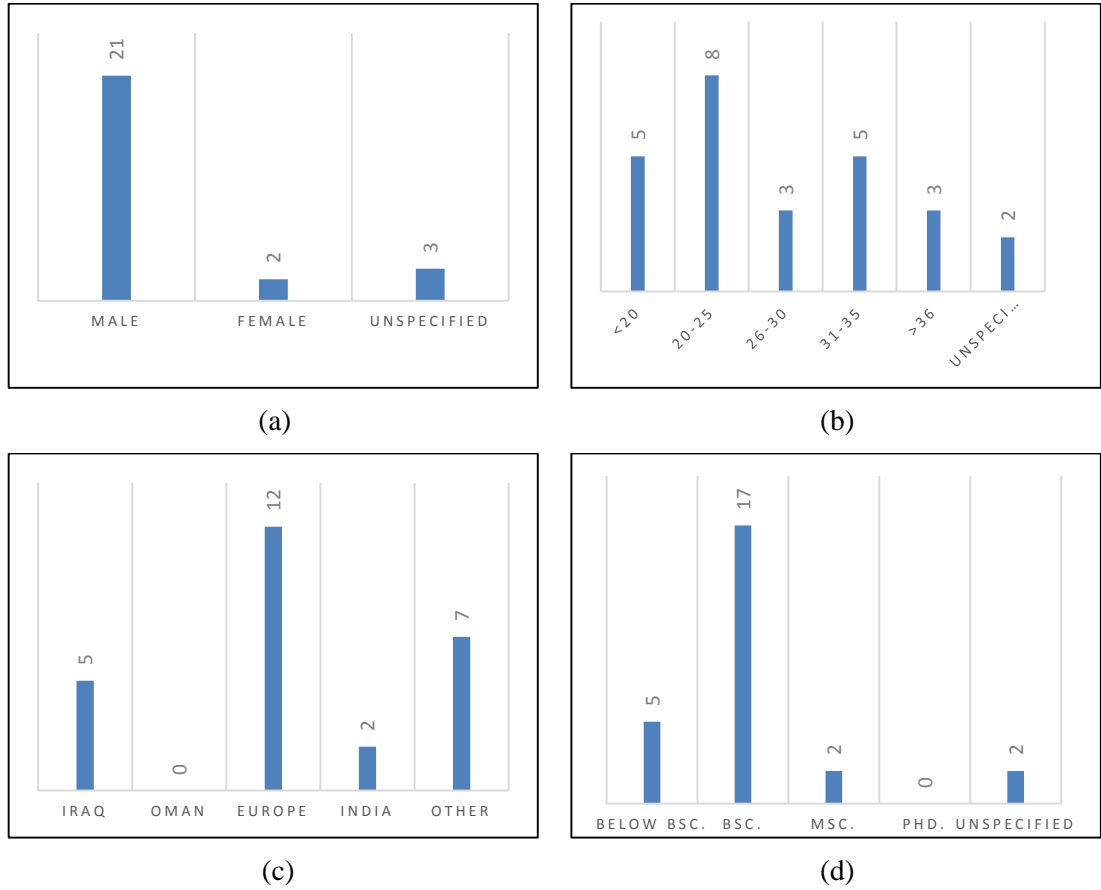


Figure 3-5: The distribution of the participants based on (a) gender, (b) age, (c) nationality, (d) education level.

### 3.4 Benchmarking Data

The proposed algorithms were benchmarked with the existing state-of-the-art algorithms to verify the importance of the developed algorithms. The benchmarking process has been performed in two ways; either by using the code provided by the authors or by using the results published by them. The main algorithms that we used in benchmarking are given below:

- 1- Itti et al., (IT) [2], the code is available online at [126].
- 2- Ma and Zhang, (MZ) [117], the results are available online at [127].
- 3- Harel et al., (GB) [118], the results are available online at [127].
- 4- Hou and Zhang, (SR) [119], the results are available online at [127].
- 5- Achanta et al., (AC) [120], the results are available online at [127].
- 6- Achanta et al., (FT) [116], the code is available online at [128].

- 7- Li et al., (HFT) [113], the code is available online at [129].
- 8- Achanta and Susstrunk, (MSSS) [121], the code is available online at [130].
- 9- Fang et al., (CO) [124], the code is available online at [131].

### 3.5 Saccade Points Extraction

The second important set of points, which need to be extracted are the saccade points. Saccades, as defined earlier, are points along the eyes' moving path upon which the user does not focus. In other words, any point that is not a gaze point is nominated to be a saccade point. The importance of extracting the saccades is to identify the regions that are not important to the user.

Since any point on the path between two gaze points can be considered as a saccade point, it is possible to select the midpoint between any two salient points as a saccade point. This technique suffers from a certain limitation, which is the possibility of having a saccade point very close to a salient point or inside a region of interest (ROI). Therefore, this way of extracting the saccade points is not efficient; instead, one can extract the saccade points after clustering the salient points and forming the ROIs. The saccade points can then be extracted as the midpoints between two ROIs. Another refinement that can be used to improve the results is by thresholding the distance between the gaze points. If the distance between the salient points is small then there is a strong possibility of getting a saccade point inside a region of interest, so by filtering the saccade points based on the distance between each two salient points, one can avoid such problems.

With these modifications, saccade points will be specified only inside the focusing region. In order to include regions outside the focusing region, we shall use the corners of the image as virtual salient points for this purpose only. Therefore, there should be two thresholding values to threshold the distance between the points as given below:

$$\begin{aligned}
 T_{dw} &= \gamma W \\
 T_{dh} &= \gamma H \\
 0 &< \gamma < 1
 \end{aligned}
 \tag{3-3}$$

where  $T_{dw}$  is the threshold of the minimum horizontal distance between a saccade point and its nearest salient point, and  $T_{dh}$  is the threshold of the minimum vertical distance

between a saccade point and its nearest salient point.  $W$  and  $H$  are the width and height of the image, and  $\gamma \in \mathbb{R}$  is a fraction tuning factor.

Figure 3-6 shows the result of extracting the saccade points using the suggested technique. From the figure, it is clear that most of the saccade points fall in the unimportant regions except for two point which fall on the salient object because of the eye tracking data.

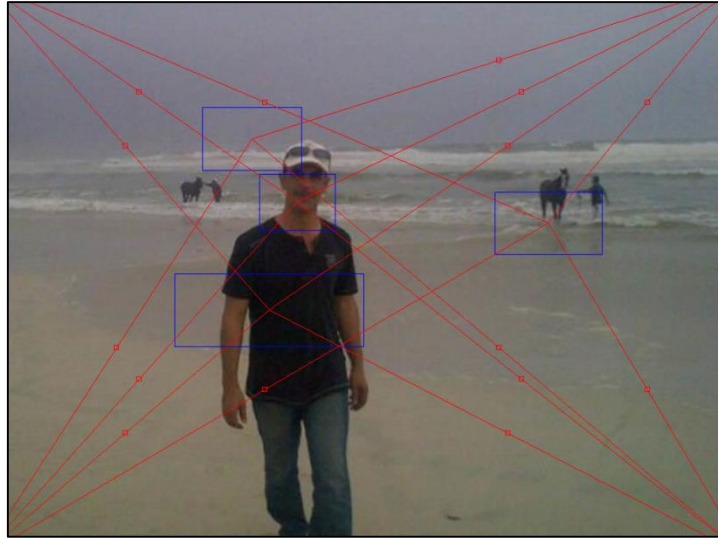


Figure 3-6: Extracting the saccade points  $\gamma = 0.5$ .

### 3.6 Conclusions

In this chapter we have discussed different issues relating to data collection, such as, standard datasets used both in saliency extraction and in image identification algorithms. A discussion about the datasets has been presented and main pros and cons of the available datasets have been identified. In addition, we have discussed the ground truth data and two methods for extracting ground truth salient points and regions have been designed and the results have been discussed.

The main limitation of WANG standard dataset, which is used in image content identification, has been discussed and a dataset, which can give more reasonable results, has been constructed.

Finally, as saccade points are important in regions of interest identification, a method to extract the saccade points from the salient points has been introduced and the necessary software has been designed and implemented.



# Chapter 4

## Saliency Identification and Evaluation

### 4.1 Introduction

The majority of traditional image retrieval systems recognize the image as a whole based on features like colour distribution, texture, and shape. Using these features may give inaccurate results since some pictures with different contents can have similar colour distribution or similar texture measures, as illustrated in Figure 4-1. This figure shows two different images with similar colour distribution, and statistical measures. The histograms and statistical measures are not identical but they are close enough to mark the two images as similar. The same thing may happen with texture or shape recognition. Therefore, considering the image as a whole will certainly generate some irrelevant results.

Recognition by component was used to overcome the aforementioned problem, which dramatically improved the results. Image segmentation which divides the image into a set of different, related regions (segments) is required with such techniques. Various image segmentation techniques have been proposed based on colour, borders, or texture. After the segmentation process, each segment will go through the recognition process, and finally, the descriptions of the segments are gathered to form the overall description of the image.

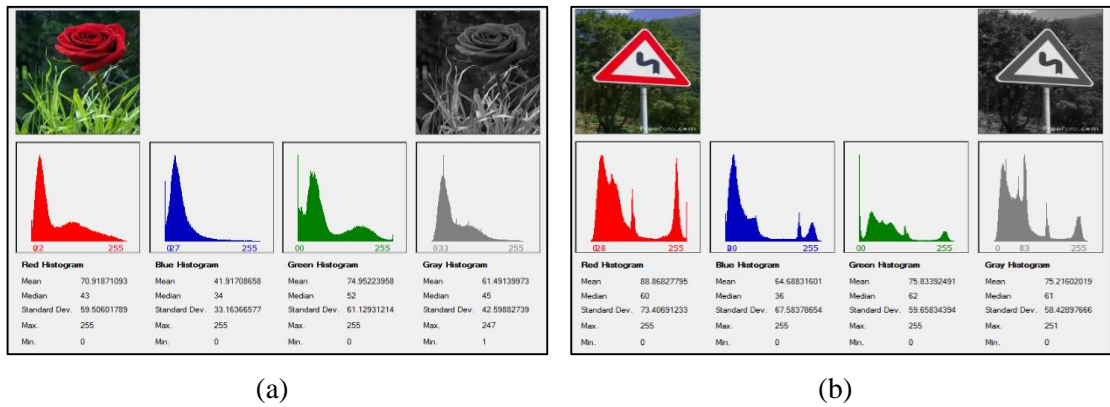


Figure 4-1: Two different images with similar colour distribution and statistical measures.

Image segmentation techniques may divide the image based on the visual contents. As a result, the object itself may be divided into smaller chunks, causing identification to become more complex. In addition to the above problem, unimportant segments, like bush background for example, will also be considered. Figure 4-2 shows the result obtained from segmenting an image based on the colour features.

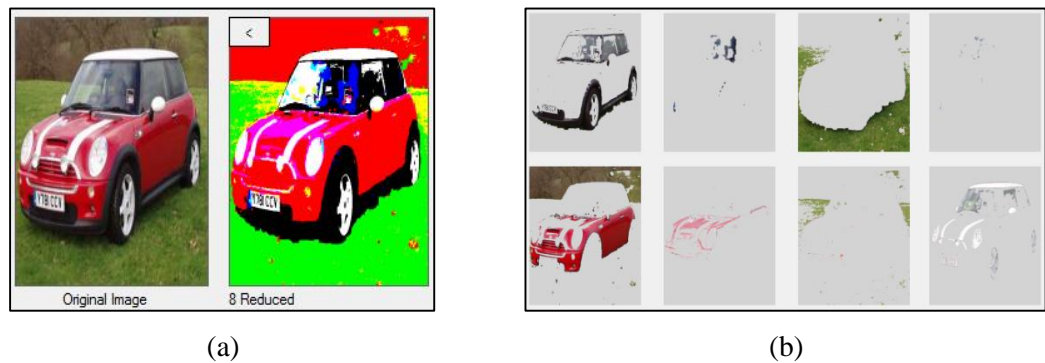


Figure 4-2: Colour-based image segmentation (a) original image, and its reduced colour image, (b) segments obtained by applying colour based segmentation.

There are many problems with the segmentation given above such as, it has partitioned the image to small parts, including the object itself. Also, it has produced some segments that are not important enough to be recognized, and finally there is some overlap between the object and its background. In order to overcome the abovementioned limitations, one should consider techniques that only highlight the important objects and keep the parts of each object connected together. Saliency-based segmentation can be used for this purpose, in which only salient regions or important regions are identified and described based on the saliency extraction technique.

## **4.2 Salient Points and Regions Extraction**

Saliency identification and extraction is the process of extracting the most important (salient) regions from an image based on how abrupt they are as compared to other parts of the image. Many literatures divide salient point extraction techniques into various categories. The major categories are [26]:

1. Bottom-Up Feature-Based Approaches (BUFB).
2. Top-Down Knowledge-Based Methods (TDKB).
3. Template Matching Methods (TM).
4. Appearance-Based Methods (AB).
5. Integration of Parts Detectors (IOP)

In the first approach, BUFB, the technique aims at finding structural features that exist even when the pose, viewpoint, or lighting conditions vary and then use them in the detection procedure. TDKB aims to construct the knowledge first, then, search for interesting points according to the knowledge database. TM uses patterns of the objects and performs the correlation between the patterns and the input images to search for similar objects. AB uses templates to train the algorithm to identify objects of special importance. The IOP approach depends on dividing the object into smaller objects, detecting these parts and integrating them together to form the object of interest.

Another broad classification of the salient point extraction techniques was proposed by Toet [36]. He classified the techniques as biologically-based, purely computational, or a combination of both.

### **4.2.1 Wavelet-Based Techniques**

Wavelet is a multiresolution representation that expresses image variations, giving information about variations in the signal at different scales. It was used by several authors, who applied the principles of wavelet transform to extract the salient points such as [132], [133], [134], [135] and [136]. Loupiaz et al. suggested the use of the orthogonal Haar wavelet in extracting the salient points in an image [132]. They reasoned that a high wavelet coefficient at a coarse resolution corresponds to a region with high global variations. In their method, they find the variation at a lower level and track this variation to the original image size; this tracking is performed from one level to the upper one.

Song et al. proposed the use of wavelet to identify the salient points in a colour image; they converted the colour image into its three bands and calculated the wavelet for each band, from which the salient points can then be extracted [134].

The use of wavelet may give good results in a non-homogeneous image, as in the cameraman example that was used in [132], while it is not efficient with images with homogeneous details or images with high texture nature.

#### **4.2.2 Location-Based Saliency**

The location of the object was a method adopted by many authors to extract the salient objects from an image. Kim et al. [25] and Lou et al. [122] are examples of authors who use this approach. In both the aforementioned papers, the authors have used the central position and colour contrast as the features that give importance to the object. They also assumed that photographer always puts the important object in the middle of the image. This is not always correct, since salient object is not necessarily in the centre of the image.

#### **4.2.3 Background Suppression**

Davis & Sharma have suggested the use of borders in detecting humans in thermal images [137]. They assumed that different objects give off different radiation, and humans will give off higher thermal radiation than most objects. They reduced the background by extracting the regions of interest, then they located the objects' boundaries, and finally they used a threshold value to divide the image into object and background.

Zhang et al. present an object/background segmentation method that uses object templates to segment a new image and extract the points of interest in that image [138]. They propose the use of what they call 'characteristics points', which are the salient points in a template, and found a correspondence between these points and the salient points. This approach requires a pre-knowledge of the objects in the image.

In the above approaches, it is clear that one would need a good knowledge of the nature of the image, and that these approaches are very much domain-oriented, best used in a search for a specific object or a person. In addition to the above challenges, one may note

that the algorithms use pre-defined objects to identify them. This is a significant limitation, since with time the object database will become very large indeed.

#### **4.2.4 Corner-Based Techniques**

Geometric features, such as corners, were also used to identify the points' saliency. Corner detectors were originally designed for robotics and shape recognition because they are an excellent feature in that they remain constant with any scaling, shifting and rotation. Corners were first adopted as a measure of saliency by Schmid and Mohr as part of their effort to identify interest points locally [26] [139]. Several corner detectors can be used, such as Harris, Moravec and SUSAN. These detectors are the most well-known approaches for extracting the corners from an image, more details are available in a wide range of resources such as [140] [141] [142]. There are some limitations associated with corner-based techniques; one is that they have difficulties with non-geometrical objects and another is their difficulty with textured images such as grass.

Loupias et al. [132] have criticized the use of corner-based techniques and highlighted the limitations of using this method. The limitations they presented are first, important visual features are not necessarily going to be corners, and second, many corners may gather in a small region, like in the area of texture in an image.

#### **4.2.5 Feature Maps-Based Saliency**

Feature maps are the maps of features extracted from an image. The saliency of a point or a region is measured based on the combination of these maps. Low-level features such as colour, intensity, texture, and orientation are commonly used for this purpose.

Early work in this field was done by Koch and Ullman [143] and Itti et al. [2]. Itti *et al.* developed a model to extract the regions of attention, which is considered one of the most popular models in its field. In their model, they used image hierarchy, i.e. different resolution levels of the image using a Gaussian pyramid. They used intensity, colour, and orientation features to distinguish the salient regions. The features were calculated by a set of centre surround operations. Forty-two feature maps were generated from these features; 6 for intensity, 12 for colour, and 24 for orientation. The saliency map is extracted by combining the above maps. Inhibition of return was used to prohibit the algorithm from considering the same salient object more than once.

The main drawback of this technique is that it uses so many feature maps (42). In addition to the use of the image pyramid, these two drawbacks may well affect the speed of the algorithm. The third significant drawback is that it extracts the interest regions sequentially, resulting in one region at each iteration using the winners-takes-all paradigm, which increases the computation time. In addition to the above drawbacks, there is another limitation related to our application, which is that the algorithm usually just focuses on one part of the object.

#### **4.2.6 Frequency Spectra-Based Saliency**

Frequency domain was used to extract the saliency of the objects in many literatures such as [144], [145], [113], [116], [146], and [147]. In such approaches, the authors considered that salient points are usually of high change in the frequency domain both in magnitude and in orientation. Bruce et al. [144] suggest the use of magnitude to extract the salient regions in an image. They divide the image into sub-images and use Fourier Transform to convert the sub-image to the frequency domain. After this, they calculate the magnitude of the image from the real and imaginary parts of the spectral image and consider the regions with high frequency as salient regions. In ref. [145], these authors also considered the magnitude of the spectral image to extract the saliency map. They considered that the image contains two parts, prior knowledge and innovation. The prior knowledge is the redundant information in the image and the innovation is the salient information. The saliency is extracted by taking the Fourier Transform for the image, extracting the log spectrum, then subtracting the redundant part, and finally converting it back to spatial domain to get the saliency map. The main limitation of such techniques is that they consider regions with high frequency as salient regions. Regions of high frequency are regions with high local changes such as edges, and edges are not necessarily salient regions.

#### **4.2.7 Colour-Based Saliency**

Colour-based saliency identification has received much attention, such as by [122], who considered colour and spatial information as measures of saliency. In their method, they gave the regions with less colour repetition in the image a higher saliency. In addition, they gave the region close to the centre a higher saliency value. In [123], the authors

considered the colour and pattern as a measure of saliency. In [124], the authors used colour in addition to texture and intensity information to identify the saliency.

It was noticed that most of the publications consider rarity of colour as a measure of saliency; additionally, in most of the proposed algorithms, the author uses other features such as location, texture, and intensity along with the colour information.

### **4.3 Evaluation of Saliency**

How good is the saliency extraction algorithm? Moreover, how suitable is it for the application at hand? For any developed algorithm, there should be a measure to evaluate how good it is, and how suitable for the application at hand, and one should answer these two questions carefully. First, we need to set up some rules to evaluate the saliency extraction algorithms. These rules should satisfy the idea of saliency and attention; also, they should consider the requirements of the application at hand. Not much effort has been made in this field, although there are many surveys which have tried to investigate the most state-of-the-art evaluation methods, such as [148], [36], [149], [150], [151] and [152]. Some of the literature evaluates the results qualitatively or visually, based on the observers' points of view. This kind of evaluation is the easiest method and depends totally on the feedback collected from observers; there is no fixed measure in this method [153] [154]. Area Under the Curve (AUC) was used for this purpose, in which the saliency map is converted to a binary image and then the AUC is calculated and compared to the AUC extracted from the ground truth data [155] [149] [152] [156] [157]. Tatler supported the idea of empirical evaluation of the saliency since they were more interested in mimicking the natural behaviour of human eyes [30]. Precision and Recall is another method was used in many publications such as [158], [159], [116], [160] and [147] to evaluate the saliency algorithms. Correlation-based measures, in which the correlation between the saliency map generated by the saliency extraction algorithms and those generated by using fixation data, [152] [149] [161] [162] [163] [164] were also widely used. Some other techniques were proposed such as Least Square Index [152], String-Edit Distance [152], Earth Mover's Distance [152] [155] [151] and Receiver Operating Characteristics [152] [156] [165].

### 4.3.1 Previous Work

Many methods were proposed to perform the evaluation process; in the following sections we shall discuss the state-of-the-art methods.

#### 1. Correlation-Based Measures

Based on the idea that the linear correlation between two variables gives an impression about how similar the variables are, or how strong the relationship is between them, correlation measures were used to evaluate the saliency extraction algorithms. Correlation gives a single scalar value to describe the similarity between the variables. The Correlation Coefficient (CC) has a value falling in the range of  $[-1,1]$ . The higher the value of the absolute value of the coefficient, the better the linear relationship between the variable is [33] [149]. Masciocchi et al. have suggested the following equation to extract the value of CC [163].

$$CC(P, H) = \frac{\sum_x \sum_y P(x, y) H(x, y)}{\sqrt{\sum_x \sum_y P(x, y)} \sqrt{\sum_x \sum_y H(x, y)}} \quad 4-1$$

where  $P(x, y)$  is the extracted points set and  $H(x, y)$  is the ground truth points set.

#### 2. Receiver Operating Characteristics (ROC)

ROC was used by many publications, such as in [152] [156] [165]. Both maps, the ground truth and the extracted maps, need to be thresholded and binarised using different thresholding values. Two quantities are extracted from the above maps, which are, the true positive rate (TPR) and the false positive rate (FPR).

Let  $H \in \{0,1\}$  and  $P \in \{0,1\}$  represent the ground-truth and the extracted binary maps after thresholding respectively, then TPR and FPR can be calculated as follows [149]:

$$\begin{aligned} TPR &= \frac{N(H \cap P)}{N(H)} \\ FPR &= \frac{N(P) - N(H \cap P)}{T(H) - N(H)} \end{aligned} \quad 4-2$$

where the operator  $N(.)$  represents the number of ones and  $T(.)$  represents the total number of elements. TPR and FPR are calculated for different values of threshold, then the relation between them is plotted and the area under the curve is calculated.



The inherent limitation of ROC, however, is that it only depends on the ordering of the fixations (ordinality) and does not capture the metric amplitude differences [152]. Another limitation according to Judd et al. [151] and Zhao et al. [152] is that, as long as the hit rates are high, the area under the ROC curve is always high regardless of the false alarm rate. While an ROC analysis is useful, it is insufficient to describe the spatial deviation of predicted saliency map from the actual ground truth map. If a predicted salient location is misplaced, but misplaced either close to or far away from the actual salient location, the performance should be different for each.

### 3. The Kullback-Leibler Divergence (KLD)

In contrast to the above techniques, which were for finding the similarity between the extracted and the ground truth maps, KLD computes the dissimilarity between probability density functions of the ground truth  $H$  and extracted maps  $P$ . The KLD is given by the following equation [33]:

$$KL(p|h) = \sum_x p(x) \log \left( \frac{p(x)}{h(x)} \right) \quad 4-3$$

The value of KLD is very small whenever the probability density is close for both maps, and reaches zero if they are identical.

### 4. Precision and Recall (PAR)

Precision (positive predictive value) and recall (sensitivity) was used by many publications, such as [158], [159], [116], [160], and [147], to measure how close the extracted results from the ground-truth were.

In the field of information retrieval, precision is the fraction of retrieved documents that are relevant to the search, while recall is the fraction of the documents relevant to the query that are successfully retrieved. These can be found using the following formula [166]:

$$\begin{aligned} Precision &= \frac{|\{Relevant Retrieved\} \cap \{Total Retrieved\}|}{|\{Total Retrieved\}|} \\ Recall &= \frac{|\{Relevant Retrieved\} \cap \{Total Retrieved\}|}{|\{Relevant Retrieved\}|} \end{aligned} \quad 4-4$$

To make the precision and recall given in Eqn. 4-4 suitable for use in evaluating the saliency extraction algorithm, we shall use the modification as follows [122] [167]:

$$\begin{aligned} Precision &= \frac{\sum_{(x,y)} P(x,y) \cdot H(x,y)}{\sum_{(x,y)} P(x,y)} \\ Recall &= \frac{\sum_{(x,y)} P(x,y) \cdot H(x,y)}{\sum_{(x,y)} H(x,y)} \end{aligned} \quad 4-5$$

where  $P(x,y)$  and  $H(x,y)$  are the extracted and ground truth maps respectively.

F-Measure is used to evaluate the overall performance, and is defined as the weighted harmonic mean of precision and recall, and is given by [159] :

$$F - Measure = \frac{(1 + \beta) \times Precision \times Recall}{\beta \times Precision + Recall} \quad 4-6$$

Where  $\beta$  is a tuning factor, in ref. [159] it was set to 0.5, and 0.3 in ref. [167] to weight the precision with respect to recall.

#### 5. Earth Mover's Distance (EMD)

EMD was used in some publications, such as [152] [155] [151] [168] [169] [170]. This is based on the minimal cost that must be paid to transform one distribution into another, or in other words, it is the minimal cost that must be paid to transform one histogram into another [168].

#### 6. Normalized Scanpath

This technique was used by Zhao et al. [152] [171], which they used to evaluate saliency values at fixated locations. Scanpath is the eye movement from one fixation to another. These paths can be compared in different ways, such as **vector-based metric**. The vector-based metric measure was adopted because it considers the shape of the scanpath, the length of the scanpath, the direction of the scanpath saccades; the position of fixations and the duration of fixations.

#### 7. String-Edit Distance (SED)

SED is another well-known measure to evaluate the saliency. In this technique, the image is divided into grids and each grid will be given a letter; the sequence of the fixations at each region will be denoted by a letter. After this process, we shall get a string; adjacent

fixation points will give two successive similar letters. These similar letters are removed and then the string obtained from the computational extraction method is compared with the string obtained from empirical experiments. The number of differences between the two strings will represent the measure of how good the computational method is. The smaller this number is, the better the saliency method. The advantage of this method is that it retains the order of the fixations and is easy to calculate, while the main drawback is that it does not consider the fixation duration [152], [172] and the size of the grid should be specified reasonably. In addition, it was designed to be used only with scanpath extraction methods and it does not work with extracting the salient object.

## 8. Info Contents

Info Contents is another method which was introduced by Schmid et al. [173] to evaluate point detectors. This measure depends on the extracted features. It measures the distinctiveness of local features computed at extracted points. It is assumed that the higher the information content of extracted points, the better they are for indexing. The information content is computed as the entropy of local features' distribution for a set of extracted points.

### 4.3.2 Proposed Evaluation Measures

In order to develop an evaluation technique that is suitable for the application at hand, one needs to first identify the main features, constraints, and requirements. Based on our application, the computational saliency extraction process is used to extract the salient objects in a scene, which means extracting the important regions from the image, from which the important objects are extracted. This process can be regarded as a segmentation process. Thus, the order of the salient points is not of great importance in our application at this stage.

To evaluate the salient regions extraction process, we shall consider the extracted region rather than the salient points; this is because we are interested more in the objects located in a specific region. These objects will be identified in the identification phase. Measures such as the centroid of the region, density of points, area of the region, and importance of the region can be used to evaluate the salient regions extraction process.

To evaluate the obtained results, we shall compare them with the ground truth data obtained both from manual labelling and from eye-tracking data. The evaluation measure can be viewed as similarity or dissimilarity measures between the extracted data and the ground truth data.

#### 4.3.2.1 Distance measures properties:

Consider the Relation  $D(X, Y)$  which finds the distance between two sets of data  $X$  and  $Y$  which belong to the power set of the universal set  $U$ , then  $D(X, Y)$  can be defined as a relation from  $U^2$  to  $R$  such that:

- 1-  $\forall (X, Y) \in U^2 \wedge X \neq Y, D(X, Y) > 0$
- 2-  $\forall X \in U, D(X, X) = 0$
- 3-  $\forall (X, Y) \in U^2 \wedge X \neq Y, D(X, Y) = D(Y, X)$
- 4-  $\forall (X, Y) \in U^2, \text{if } D(X, Y) = 0 \Leftrightarrow X = Y$

#### 4.3.2.2 Similarity measures properties:

Consider the Relation  $S(X, Y)$  which finds the similarity between two sets of data  $X$  and  $Y$  which belong to the power set of the universal set  $U$ , then  $S(X, Y)$  can be defined as a relation from  $U^2$  to  $R$  such that:

- 1-  $\forall (X, Y) \in U^2 \wedge X \neq Y, S(X, Y) < M$  with  $M$  is an arbitrary large number and it is the maximum possible similarity measure.
- 2-  $\forall X \in U, S(X, X) = M$
- 3-  $\forall (X, Y) \in U^2 \wedge X \neq Y, S(X, Y) = S(Y, X)$
- 4-  $\forall (X, Y) \in U^2, \text{if } S(X, Y) = M \Leftrightarrow X = Y$

The process of comparing the similarity between the two sets can be achieved either by minimizing the distance measure or maximizing the similarity measures.

$$S(x, y) = 1 - \frac{D(x, y)}{D_{max}} \quad 4-7$$

where  $D_{max}$  is the maximum possible distance between the vectors.

In general, let  $P$  be the extracted points set and  $H$  as the ground truth data such that:

$$\begin{aligned}
P &= \{p_i \mid p_i = (x_i, y_i) \wedge i = 1, 2, \dots, N\} \\
H &= \{h_i \mid h_i = (x_i, y_i) \wedge i = 1, 2, \dots, M\}
\end{aligned}
\tag{4-8}$$

It was noticed that it is not possible to compare the points individually, since it is not possible for two points to fall on the same x, y coordinates; accordingly, we may use the centroid of the data to compare the regions.

#### 4.3.2.3 Single Centroid Distance (SCD)

In single centroid distance (SCD), the centroids of the two saliency maps are calculated then compared. The distance between these two centroids represent the dissimilarity between the two maps. The centroid of the map can be extracted as the mathematical average of the salient points in the map and can be calculated as follows:

$$\begin{aligned}
C_{Px} &= \frac{1}{|P|} \sum_{i=1}^{|P|} x_i \\
C_{Py} &= \frac{1}{|P|} \sum_{i=1}^{|P|} y_i \\
C_P &= (C_{Px}, C_{Py}) \\
C_{Hx} &= \frac{1}{|H|} \sum_{i=1}^{|H|} x_i \\
C_{Hy} &= \frac{1}{|H|} \sum_{i=1}^{|H|} y_i \\
C_H &= (C_{Hx}, C_{Hy}) \\
D(P, H) &= \sqrt{(C_{Px} - C_{Hx})^2 + (C_{Py} - C_{Hy})^2}
\end{aligned}
\tag{4-9}$$

where  $|P|$  and  $|H|$  are the cardinalities of the extracted and the ground truth salient points sets respectively, and  $C_P$  and  $C_H$  are the centroids of the extracted and ground truth data respectively and  $D(P, H)$  is the distance between the two centroids.

Figure 4-3 shows an example of how to calculate the centroid of a set of points. In this Figure, (a) shows the computationally extracted salient points and (b) shows the same points after removing the background for clarity only. (c) shows the calculated centroid

from the set of points (the yellow filled circle), and (d) same as (c) after removing the background to illustrate the process in a clearer way.

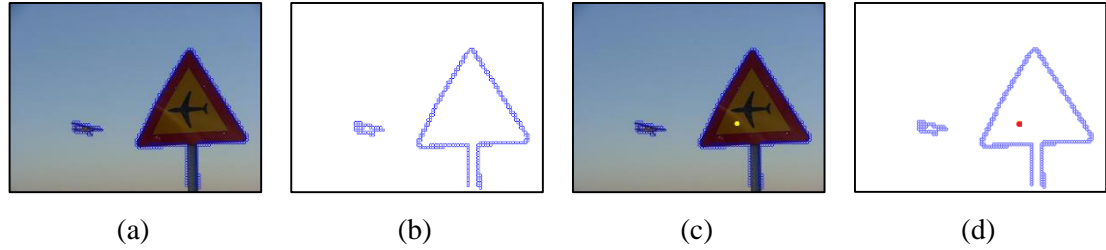


Figure 4-3: The extraction of the centroid from a set of points, (a) the points on the original image, (b) the same points on a white background, (c) the centroid, (the yellow circle), (d) the centroid and points on a white background.

From this Figure 4-3, it is clear that there are more than one salient region, but with only one centroid. This centroid was calculated for all points, not for regions, which is one of the limitations of this method, hence, we shall improve it in the next Section.

The comparison between two sets of points, such as salient points or gaze points, can be achieved using Equation 4-9. Figure 4-4 shows the result of comparing two sets of salient points. The first set was extracted using the local irregularity saliency extraction method, which will be discussed later, and the second set was obtained using manual labelling. In this figure, the yellow points are the automatically extracted salient points and the blue points are the salient points that were obtained using manual labelling.

In this example, the distance between the two sets of points, automatically extracted and ground truth salient points, is very small (8.122) which means that the sets of points are very similar, in spite of the differences in the point distribution of the two sets.

Another possible comparison to evaluate the salient point extraction technique is by comparing the extracted data with the one obtained from the eye-tracking device. Figure 4-5 shows the distance between the extracted data and the eye-tracking data. The distance between the two centroids was (30.1), which is also an acceptable value given that the maximum possible distance between the points is  $\sqrt{W^2 + H^2}$  where W and H are the width and the height of the image respectively. In our example the maximum distance is  $\sqrt{300^2 + 400^2} = 500$ .

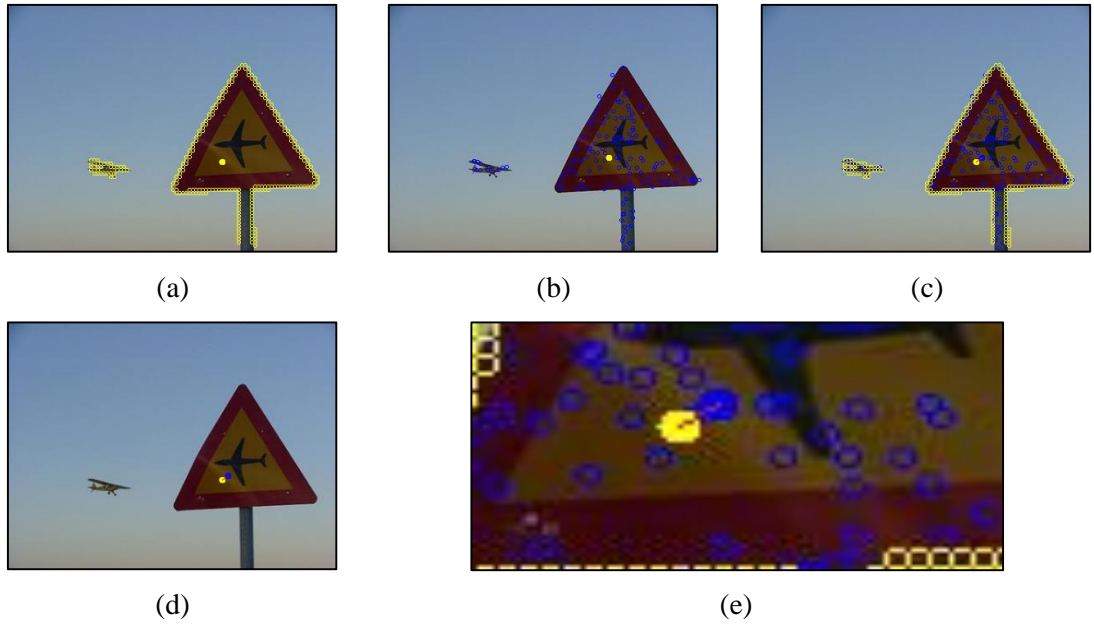


Figure 4-4: Comparing two points sets, (a) automatically extracted points, (b) manually labelled points, (c) the distance between centroids, (d) the same distance in (c), (e) the distance between the centroid magnified.

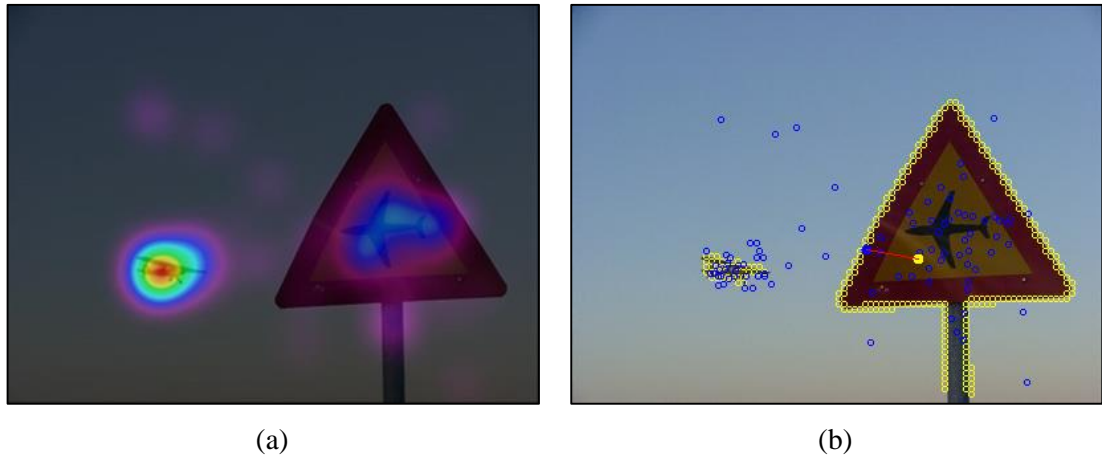


Figure 4-5: The distance between the centroid of the computationally extracted data and the eye-tracking data, (a) fixations obtained from the eye tracking data, (b) the distance between the centroids.

#### 4.3.2.4 Region-Based Centroid Distance (RBCD)

From the above discussion, it was obvious that the SCD compares all the points at the same time, which can be affected by the falsely detected salient points (FDSP). Here, we shall consider the distance of the centroid after clustering, i.e. after forming the regions of interest from the points of interest.

Let  $P_i \subseteq P$  be the set of adjacent points that can be grouped to form a region  $R_{P_i} \subseteq R_P$ , where  $i = 0, 1, \dots, N_P$ , and  $N_P$  is the number of the regions that were computationally extracted from the image. In the same way, we may define  $R_{H_j} \subseteq R_H$ , where  $j = 0, 1, \dots, N_H$ , and  $N_H$  is the number of the regions that were manually extracted from the image or those that correspond to the ground truth data.

By considering the above assumptions, we shall obtain sets of centroids instead of a single centroid.  $C_P$  and  $C_H$ .  $C_P$  is the set of centroids corresponding to the regions computationally extracted with cardinality of  $N_P = |C_P|$ .  $C_H$  represents the set of centroids corresponding to the ground truth regions with cardinality of  $N_H = |C_H|$ .

The distance between the two sets of regions can then be found using the sum of the distances between the corresponding regions, i.e.

$$D(P, H) = \sum_{i=1}^{N_P} \sqrt{(C_{Px_i} - C_{Hx_i})^2 + (C_{Py_i} - C_{Hy_i})^2} \quad 4-10$$

Equation 4-10 is correct if and only if the number of regions in both ground truth and extracted sets is the same, which is not always the case. Instead, again, we may find the average centroid for each set and compute the distance between these two centroids.

$$\begin{aligned} C_{Px} &= \frac{1}{N_P} \left( \sum_{i=1}^{N_P} C_{Px_i} \right), C_{Py} = \frac{1}{N_P} \left( \sum_{i=1}^{N_P} C_{Py_i} \right) \\ C_{Hx} &= \frac{1}{N_H} \left( \sum_{i=1}^{N_H} C_{Hx_i} \right), C_{Hy} = \frac{1}{N_H} \left( \sum_{i=1}^{N_H} C_{Hy_i} \right) \\ D(P, H) &= \sqrt{(C_{Px} - C_{Hx})^2 + (C_{Py} - C_{Hy})^2} \end{aligned} \quad 4-11$$

This method suffers from a certain limitation, which is that all regions have the same weight and their effect on the distance is the same, i.e. there is no difference between a small and a large region. Even regions with single points, which might be falsely detected, will have the same effect as the important regions.



#### 4.3.2.5 Weighted Region-Based Centroid Distance (WRBCD)

To overcome the problem uncovered in the latter method, we shall refine the RBCD by considering the weighted average i.e. giving different weights to different regions. This weight is based on the importance of the region as given in the following equation

$$C_{Px} = \frac{1}{N_P \sum_{i=1}^{N_P} w_{P_i}} \left( \sum_{i=1}^{N_P} w_{P_i} C_{Px_i} \right)$$

$$C_{Py} = \frac{1}{N_P \sum_{i=1}^{N_P} w_{P_i}} \left( \sum_{i=1}^{N_P} w_{P_i} C_{Py_i} \right)$$

4-12

$$C_{Hx} = \frac{1}{N_H \sum_{i=1}^{N_H} w_{H_i}} \left( \sum_{i=1}^{N_H} w_{H_i} C_{Hx_i} \right)$$

$$C_{Hy} = \frac{1}{N_H \sum_{i=1}^{N_H} w_{H_i}} \left( \sum_{i=1}^{N_H} w_{H_i} C_{Hy_i} \right)$$

where  $w_{P_i}$  and  $w_{H_i}$  are the weights corresponding to the regions  $R_{P_i}$  and  $R_{H_i}$  respectively.

One possible measure for the importance of the regions is the density of points, in which the centroids are weighted based on the number of points that have participated in the calculation of that centroid. This suggestion may not be precise enough since in cases where the points on the boundaries specify the important region, in such cases, the density will not give any impression about the importance of the region. Thus, the area of the region can be used instead of the density, in which the area of the region is used to weight the average of the centroids as shown in Equation 4-13.

Equation 4-13 suffers from the effect of a certain geometrical problem, that is, the effect of the weighting process will cause regions with small areas to be closer to the origin (0,0), (top-left corner in the case of image) which will drag the centroid to the centre as shown in Figure 4-6. In this figure, it is clear that in (a) the overall centroid is placed in the centre distance between the two regions in spite of the difference in the importance of the two regions based on the area of the region. (b) shows the effect of the

domination of the larger region in such a way that the other regions' dimensions were considered close to the origin, thus the overall centroid was dragged closer to the origin.

$$\begin{aligned}
a_i &= \mathcal{A}(R_i) \\
C_{Px} &= \frac{1}{N_P \sum_{i=1}^{N_P} a_{P_i}} \left( \sum_{i=1}^{N_P} a_{P_i} C_{Px_i} \right) \\
C_{Py} &= \frac{1}{N_P \sum_{i=1}^{N_P} a_{P_i}} \left( \sum_{i=1}^{N_P} a_{P_i} C_{Py_i} \right) \\
C_{Hx} &= \frac{1}{N_H \sum_{i=1}^{N_H} a_{H_i}} \left( \sum_{i=1}^{N_H} a_{H_i} C_{Hx_i} \right) \\
C_{Hy} &= \frac{1}{N_H \sum_{i=1}^{N_H} a_{H_i}} \left( \sum_{i=1}^{N_H} a_{H_i} C_{Hy_i} \right)
\end{aligned} \tag{4-13}$$

Where  $a_i$  is the area corresponding to region  $R_i$  and  $\mathcal{A}(\cdot)$  is a function to extract the area of the region.

In order to overcome this problem, an equation can be derived geometrically to correct the effect mentioned above. In the derived equation, the distance between the two points (not the average) is calculated, then the midpoint between the two points will be weighted, i.e. first we find the weighted midpoint then we find the centroid. The derived equations are given below:

$$\begin{aligned}
a_i &= \mathcal{A}(R_i) \\
C_{Px}(n+1) &= C_{Px_1}(n) - \frac{C_{Px_1}(n) - C_{Px_2}(n)}{2} \left( 1 - \frac{a_{P1}(n) - a_{P2}(n)}{a_{P1}(n) + a_{P2}(n)} \right) \\
C_{Py}(n+1) &= C_{Py_1}(n) - \frac{C_{Py_1}(n) - C_{Py_2}(n)}{2} \left( 1 - \frac{a_{P1}(n) - a_{P2}(n)}{a_{P1}(n) + a_{P2}(n)} \right) \\
C_{Hx}(n+1) &= C_{Hx_1}(n) - \frac{C_{Hx_1}(n) - C_{Hx_2}(n)}{2} \left( 1 - \frac{a_{H1}(n) - a_{H2}(n)}{a_{H1}(n) + a_{H2}(n)} \right) \\
C_{Hy}(n+1) &= C_{Hy_1}(n) - \frac{C_{Hy_1}(n) - C_{Hy_2}(n)}{2} \left( 1 - \frac{a_{H1}(n) - a_{H2}(n)}{a_{H1}(n) + a_{H2}(n)} \right) \\
a_P(n+1) &= a_{P1}(n) + a_{P2}(n) \\
a_H(n+1) &= a_{H1}(n) + a_{H2}(n)
\end{aligned} \tag{4-14}$$

where  $C_{Px_1}(n)$ ,  $C_{Py_1}(n)$  and  $a_{p1}(n)$  are the  $x$  and  $y$  coordinates of the centroid and the corresponding area of the first computationally extracted region at iteration ( $n$ ). Similarly,  $C_{Px_2}(n)$ ,  $C_{Py_2}(n)$  and  $a_{p2}(n)$  are the  $x$  and  $y$  coordinates of the centroid and the corresponding area of the second computationally extracted region at iteration ( $n$ ). The same definitions are applicable on the parameters of first and second ground truth regions  $C_{Hx_1}(n)$ ,  $C_{Hy_1}(n)$ ,  $a_{H1}(n)$ ,  $C_{Hx_2}(n)$ ,  $C_{Hy_2}(n)$  and  $a_{H2}(n)$ .

As the process is iterative,  $n$  was used to indicate the iteration and takes values between 0 and the number of distances among the regions ( $|R| - 1$ ). Since we compare two regions in each iteration, then the first region shall take the index (1) and (2) is given to the second region.

In the above equations, at  $n = 0$ ,  $C_{Px_1}(0)$  and  $a_{p1}(0)$  are the  $x$ -coordinate of the centroid and the area of the first region respectively and  $C_{Px_2}(0)$  and  $a_{p2}(0)$  are the  $x$ -coordinate of the centroid and the area of the second region respectively.  $C_{Px}(n + 1)$  is the  $x$ -coordinate of the centroid of the resultant region which is obtained from the two regions centroids'  $x$ -coordinates  $C_{Px_1}(n)$  and  $C_{Px_2}(n)$  with the corresponding areas  $a_{p1}(n)$  and  $a_{p2}(n)$  respectively. The area corresponding to the resultant region ( $a_p(n + 1)$ ) is calculated by finding the sum of the tow areas  $a_{p1}(n)$  and  $a_{p2}(n)$ . At the next iteration,  $C_{Px_1}(n)$  takes the value of  $C_{Px}(n + 1)$  and  $a_{p1}(n)$  takes the value of  $a_p(n + 1)$ . The values of  $C_{Py}(n + 1)$ ,  $C_{Hx}(n + 1)$ ,  $C_{Hy}(n + 1)$  and  $a_{H1}(n)$  are calculated In the same way. The iterations are repeated until all regions are considered and only one centroid is obtained.

Figure 4-6 (c) shows the result of applying equation (4-14), which shows that the centroid is very much closer to the large region centre than it is to the small region.

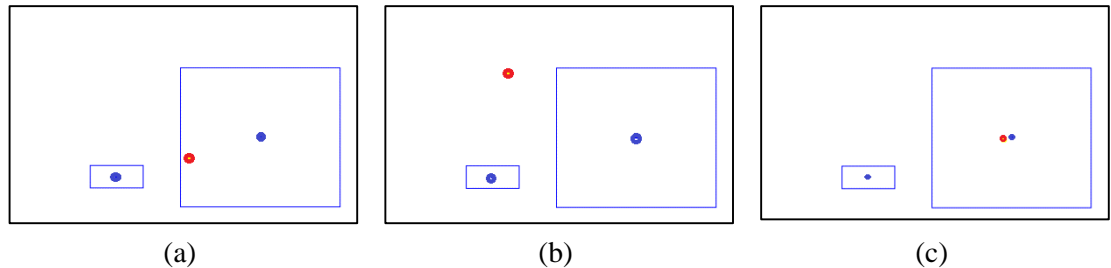


Figure 4-6: The effect of the weighting of the centroids on the overall centroid (a) without weighting, (b) with weighting, (c) weighted using equation 4-14

#### 4.3.2.6 Exclusive OR – Based Distance (XORD)

Another possible measure is by taking the exclusive OR (XOR) between the two saliency maps. In this technique, the pixels in the salient regions are set to one, while others are set to zero. Pixel-wise XOR is applied between the two maps. Identical pixels will produce zero and different pixels will produce one. Two measures can be extracted from the XOR process; the similarity XOR measure (XORS) and the distance XOR measure (XORD):

$$\begin{aligned} XORS &= \frac{1}{HW} \sum_{y=0}^H \sum_{x=0}^W (1 - [p(x,y) \oplus h(x,y)]) \\ XORD &= \frac{1}{HW} \sum_{y=0}^H \sum_{x=0}^W [p(x,y) \oplus h(x,y)] \end{aligned} \quad 4-15$$

where  $h(x,y)$  and  $p(x,y)$  are the value of the ground truth map and the extracted map respectively and  $\oplus$  is the exclusive OR operation.

To prove that the XORD can be used as a distance measure we shall prove that the four distance measures properties given in Section 4.3.2.1 are satisfied as follows:

$$1- XORD(X,Y) > 0$$

From the definition of the XOR operation, the result of applying this operation is either one or zero, therefore, the distance measure cannot be less than 0

$$2- XORD(X,X) = 0$$

Again, from XOR definition, similar bits shall give zero and different bits shall produce one, and since all bits in the sets are equal then the bitwise XOR shall produce all zeroes.

$$3- XORD(X,Y) = XORD(Y,X)$$

Since the XOR operation is commutative, i.e.  $p(x,y) \oplus h(x,y) = h(x,y) \oplus p(x,y)$  then XORD is inherently commutative.

$$4- \text{if } XORD(X,Y) = 0 \Leftrightarrow X = Y$$

As all the bits shall produce zeroes, the bits in both sets are similar, which yields to having two similar sets i.e. equal.

Figure 4-7 shows an example on the application of the XORD. In this figure, (a) and (b) show the maps we want to compare, (c) and (d) the maps after binarisation, and (e) is the result of applying the XOR on the maps in (c) and (d).

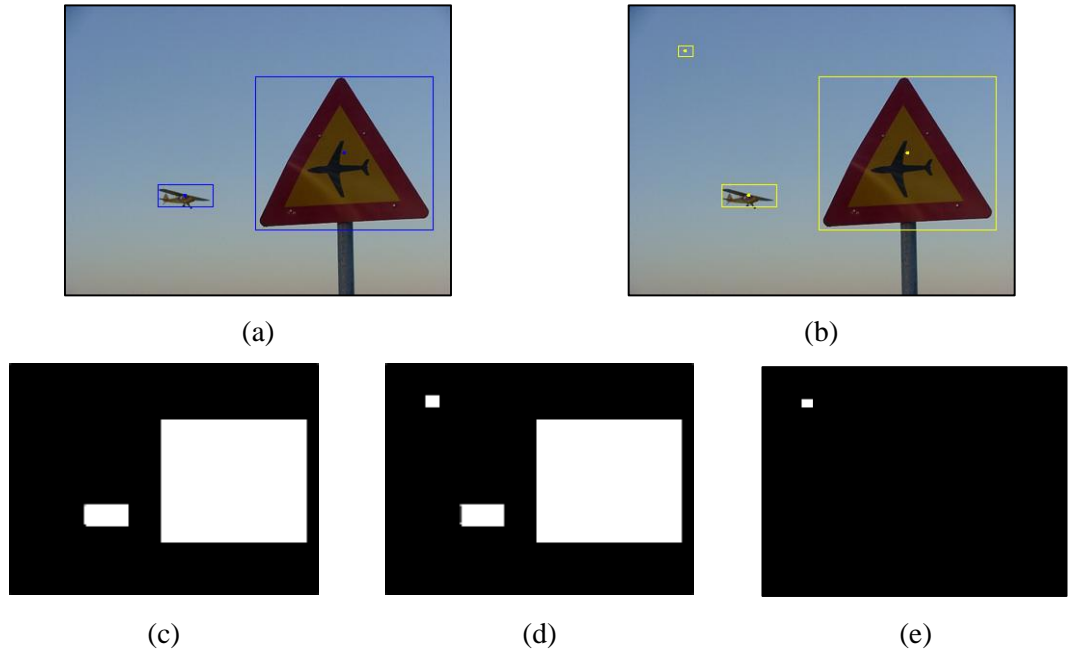


Figure 4-7: Comparing two different regions using XORD, (a) first saliency map, (b) second saliency map, (c) binary saliency map of (a), (d) binary saliency map of (b), (e) the result of applying the XOR between the two maps in (c) and (d).

The effect of falsely detected salient points (FDSP) on the four measures given above, SCD, RBCD, WRBCD, and XORD has been studied and compared with the Precision (Prec), Recall (Rec) and F-Measure (FM) on different images and the following experiments were designed to study this effect. The first experiment was designed to study the effect of falsely detected regions with different sizes, while the second one was designed to study the effect of the distribution of the FDSP on the accuracy of the distance measure. The FDSP were added purposely and in well-studied positions.

#### ***Experiment 1: Studying the effect of the quantity of the FDSP***

In this experiment, the effect of the size or density of FDSP has been studied to measure the accuracy of the distance measures. The points were added gradually to the original set of points in adjacent positions. The distances between the new set of points and the original one were measured using the four methods described above, in addition to the Prec, Rec and FM measures.

Figure 4-8 shows a sample of the FDSP in which different numbers of points were used to study their effect. In this figure, (a) shows the added points, while (b) shows the points after clustering and forming the salient regions. The images at the top left in both (a) and (b) are the ground truth image without any added FDSP.

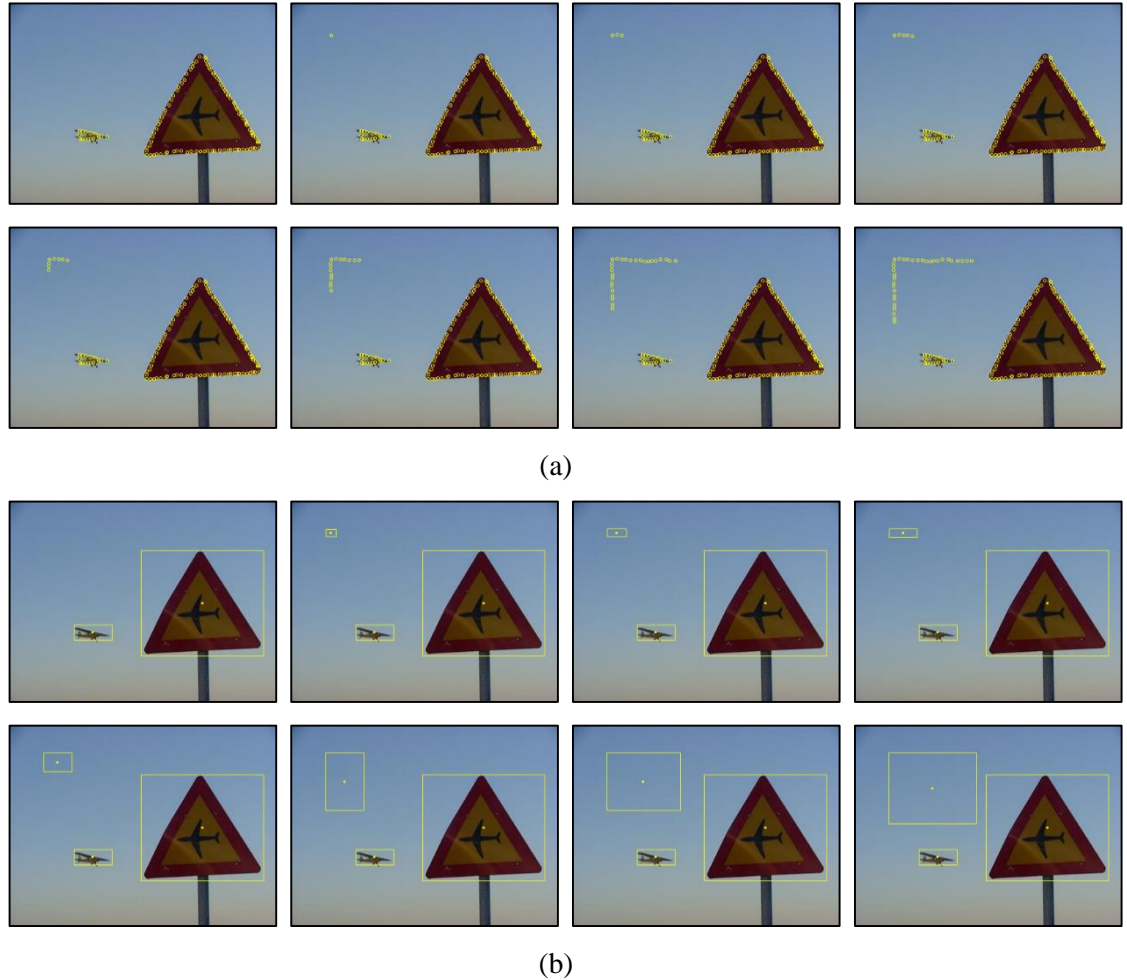


Figure 4-8: Sample of the falsely detected salient points that were used in the experiment, (a) points, (b) regions.

The distances between the ground truth data, which is the one in the top left, and the sets of data that contain FDSP were calculated using the methods explained above.

Table 4-1 gives the values of the distance with the increase in the number of FDSP. The number of FDSP is increased with the cases. Case 0 shows the result of comparing the image with itself without adding any points.

To make the curves comparable, we shall find the normalised values for the distance by dividing all the distances in one method by the maximum value and then scaling it. Since Prec, Rec and FM, measures decrease with the increase of points as they are similarity

measures, then we have converted them into distance measures by subtracting them from 100 after normalization. This is why we have used 100-Prec, 100-Rec and 100-FM. Table 4-2 gives the results of the distances given in Table 4-1 after normalisation.

Table 4-1: The values of the distance measures with the increase of the number of FDSP.

Case	SCD	RBCD	WRBCD	XORD	Prec	Rec	FM
0	0	0	0	0	1	1	1
1	2.5	65.3	65.3	1.185345	0.868734	0.9791	0.891949
2	6.3	63.6	63.6	2.5	0.777782	1	0.819823
3	9.1	61.5	61.5	3.922414	0.724227	0.9882	0.771811
4	15.2	60	60	8.649425	0.711142	1	0.761932
5	26.3	55.13	55.13	24.58333	0.617367	1	0.67716
6	32.125	52.2	52.2	35.83333	0.554322	1	0.617869
7	36.6	49.2	49.2	50.0431	0.487357	1	0.552748
8	40.3	45	45	68.57759	0.416025	1	0.480822
9	45.8	40.5	40.5	100.5172	0.332159	1	0.392677

Table 4-2: The scaled normalised values of the distance measures with the increase of the number of FDSP.

Case	SCD	RBCD	WRBCD	XORD	100-Prec	100-Rec	100-FM
0	0	0	0	0	0	0	0
1	4.347826	98.93939	2.363636	1.179245	19.59196	99.9	17.71326
2	13.04348	96.36364	4.945455	2.487136	33.16686	0	29.53714
3	19.56522	92.42424	7.636364	3.90223	41.1601	55.96	37.408
4	33.04348	90.90909	16.18182	8.604917	43.11309	0	39.02749
5	57.17391	83.33333	40.25455	24.45683	57.10945	0	52.9246
6	69.56522	78.78788	54.54545	35.64894	66.5191	0	62.64441
7	79.56522	74.24242	67.27273	49.78559	76.51389	0	73.31995
8	87.6087	68.18182	81.81818	68.2247	87.16043	0	85.11109
9	99.56522	61.36364	100	100	99.67778	0	99.56112

Figure 4-9 shows the curves of the distance with the increase of the number of falsely detected salient points FDSP. From this figure, one can notice the following:

- 1- In SCD the curve is changing slightly with respect to the change of points. In general, the measure is acceptable but more sensitive to FDSP.
- 2- The RBCD curve is increased rapidly in the first few FDSP's and then it starts decreasing with the increase of the FDSP's, which gives a wrong impression about the distance between the ground truth data and the data that contains FDSPs.

- 3- The WRBCD curve increases with the increase in the number of the FDSP's, and the increase is stable. This result is more stable than SCD.
- 4- The XORD is less sensitive to FDSPs, so it is more accurate for use in evaluating saliency extraction techniques.
- 5- The (100-Prec) measure increases with the increase of FDSP's but from the graph it is clear that it is very sensitive to the change of FDSP's. The same thing is applicable to (100-FM) measure as FM is derived from the Prec and Rec measures.
- 6- The Rec measure did not sense the change in FDSP's in a stable way in this experiment. This is because based on equation 4-5 the numerator of the equation is equal to the denominator as the results of performing the logical AND between the two images eliminates the effect of FDSPs and produces the ground truth image.

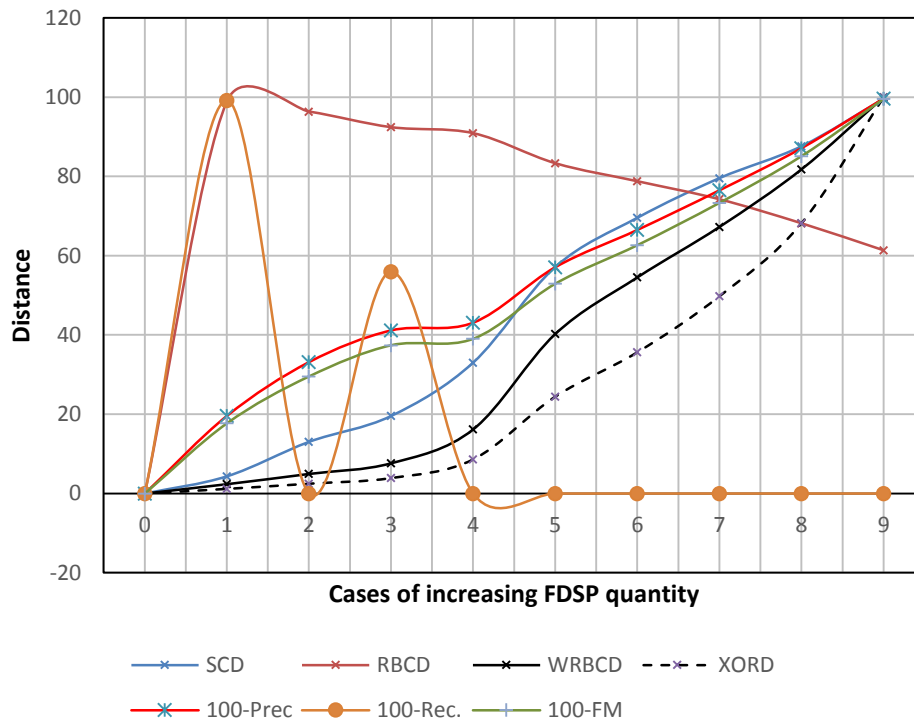


Figure 4-9: The effect of the number (quantity) of the falsely detected salient points FDSP.

### ***Experiment 2: Studying the effect of the distribution of the FDSP***

In this experiment, the effect of the distribution of the FDSP's has been studied to measure the accuracy of the distance measures. The points were added gradually in



different positions to the original set of points. The distances between the new set of points and the original one were measured using the methods described above.

Figure 4-10 shows a sample of the FDSP ; different numbers of points were used to study their effect. In this figure, (a) shows the points distribution, while (b) shows the points after clustering and forming the salient regions.

The distances between the ground truth data, which is the one in the top left, and the sets of data that contain FDSPs were calculated using the four methods explained above in addition to the Prec, Rec and FM. Table 4-3 gives the values of the distances with the increase of the number of FDSPs and the change in their location.

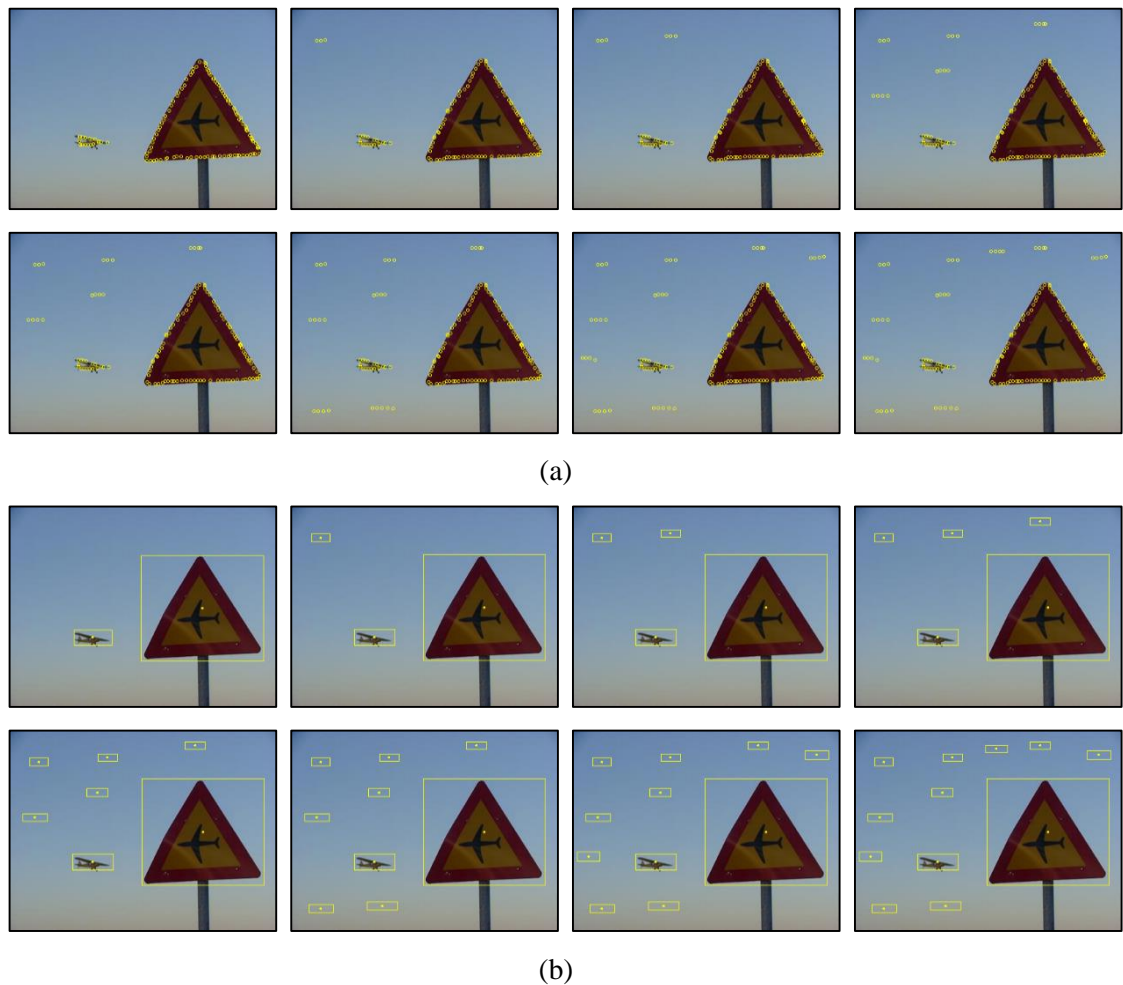


Figure 4-10: Sample of the falsely detected salient points that were used in the experiment 2, (a) points, (b) regions.

Table 4-3: The values of the distance measures with different points distribution of FDSPs.

Case	SCD	RBCD	WRBCD	XORD	Prec	Rec	FM
0	0	0	0	0	1	1	1
1	12.1	69.2	3	724	0.947886	1	0.959425
2	16.2	86.1	4.1	1676	0.829534	1	0.863503
3	17.8	88.7	5.7	2006	0.811282	1	0.848222
4	24.6	93.3	8.7	2413	0.789735	1	0.830009
5	29.7	96.4	9.98	2797	0.670319	1	0.725516
6	33.15	89.4	12.1	3241	0.605925	1	0.666541
7	36.3	80.3	13.8	3792	0.642749	0.988247	0.699156
8	42.16	88.7	17.2	4269	0.719149	1	0.768989
9	39.1	74.3	16.7	4732	0.548474	1	0.612271

Again, the data in the above table should be normalised to make them comparable with each other. The normalised data is in Table 4-4.

Table 4-4: The normalised values of the distance measures with different points distribution of FDSPs.

Case	SCD	RBCD	WRBCD	XORD	100-Prec	100-Rec	100-FM
0	0	0	0	0	0	0	0
1	28.80952	72.08333	17.44186	14.25758	11.54233	0	10.46566
2	38.57143	89.6875	23.83721	33.00512	37.75551	0	35.20696
3	42.38095	92.39583	33.13953	39.50374	41.79808	0	39.14829
4	58.57143	97.1875	50.5814	47.51871	46.57043	0	43.84605
5	70.71429	100.4167	58.02326	55.08074	73.01909	0	70.79797
6	78.92857	93.125	70.34884	63.82434	87.2812	0	86.0096
7	86.42857	83.64583	80.23256	74.67507	79.12532	100	77.59711
8	100.381	92.39583	100	84.06853	62.20396	0	59.58509
9	93.09524	77.39583	98.23529	93.18629	100.00	0	100.00

Figure 4-11 shows the graphs of the distance with the increase of the number of falsely detected salient points FDSPs and the distribution.

From Table 4-4, and the curves in Figure 4-11, one can notice that RBDC, (100- Rec), (100-Prec) and (100-FM) are not stable. The distance should change with the change of the FDSPs but from the figure, in some cases, such as 7 and 9 in RBCD, 9 in SCD, and 8 in (100-Prec) and (100-FM) the value of the distance was decreasing when it should be increasing. This is because all the regions have the same weight so when some regions are in the opposite position there is some kind of balancing so their effect is reduced. The WRBCD and XORD curves are better than the others.

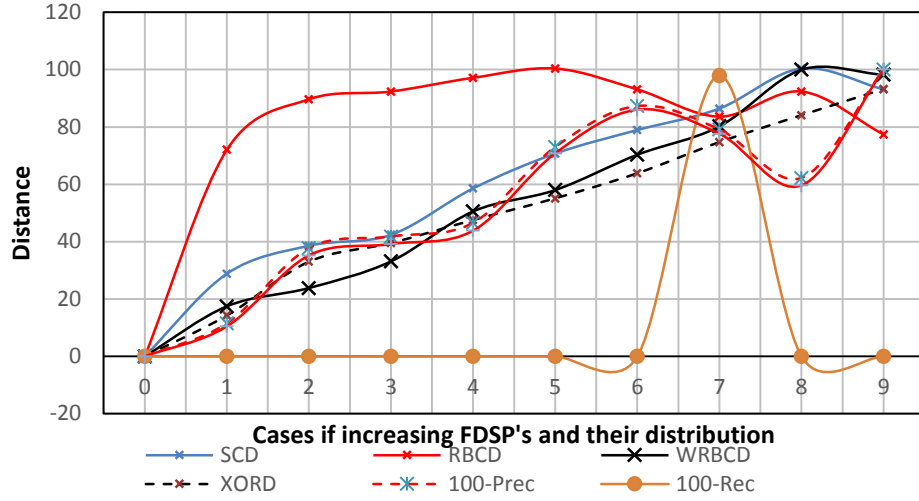


Figure 4-11: The effect of the distribution of the falsely detected salient points FDSP.

#### 4.4 Salient and Gaze Points Clustering

This Section's main concern is the salient and gaze clustering techniques, for which different clustering techniques have been developed and discussed. Clustering is an important part of most of the saliency extraction algorithms, especially in the algorithms that aim to extract salient regions from the image. In applications like identifying salient regions in an image, salient points need to be clustered to form the salient regions.

#### 4.5 Blobs – Based Clustering (BBC)

Part of the work in this Section was published in IEEE IACC 2014 [174]. The technique is based on merging the points to form a region, thus it is suitable for salient points and gaze points clustering. Based on the theory of human vision and image formation in human eyes, it is concluded that the human does not gaze at a point with  $x, y$  coordinates but gazes on the surrounding region instead [175] [176] [177], we shall develop a clustering technique to form a salient region from a set of salient points, or gaze points. In this technique, we shall define the cloud of points, which can be defined as the set of the adjacent points with a reasonable distance between them.

The distance between the points, in both vertical and horizontal directions, should be considered in identifying the point neighbourhood. Therefore, there should be some threshold value  $T$ , which is used in this identification.

$$\mathbb{C} = \{p_i | D(p_i, p_j) \leq T, \quad \forall j \neq i\} \quad 4-16$$

where  $\mathbb{C}$  is the cloud that contains all points with distance  $D(p_i, p_j)$  less than or equal to a given threshold  $T$ .

Selecting the value of  $T$  is a crucial issue, since it will affect the cloud size and hence the region size. The selection of this value could be relative to the size of the image. One possible way to find its value is by taking it as a ratio from the image dimensions (width  $W$  and height  $H$ ). The Euclidian distance can be used as a measure for the distance between the two points, thus the distance can be calculated as follows:

$$D(p_i, p_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

$$p_i = (x_i, y_i) \quad , \quad p_j = (x_j, y_j)$$
4-17

The threshold is calculated as follows:

$$T = \alpha \sqrt{H^2 + W^2} \quad , \quad 0 < \alpha < 1$$
4-18

where the tuning factor  $\alpha$  is a fraction and represents the ratio that finds the values of  $T$ .

The filtering process needs to compare each point with all other points, which is a computationally costly process since it follows the combination rule as given below:

$$nC_r = \frac{n!}{(n-r)! r!}$$
4-19

where  $n$  is the number of the points and  $r = 2$ , since each comparison shall be carried out between two points. If the number of gazing points is equal to 100, then the total number of comparisons will be (4950).

Another challenge one may face in this method of comparison is that, if there is more than one cloud, the process will be even more complex since the points need to be grouped in different clouds and there should be an iterative process. To avoid such complexity we suggest the use of blobs extraction as given in the following discussion.

The suggested algorithm utilizes the technique of blobs construction and extraction. In this technique, the points are replaced with a blob (a rectangle or a circle) centred at the given point. These blobs are grouped together to form the cloud which in its turn will be used to construct the region of interest. The size of the region of interest is varied, based on the selected distance between the points in the cloud. The blob size is specified as a ratio of the image dimensions.

$$hb = \beta H$$

$$wb = \beta W$$

$$0 < \beta < 1$$

4-20

where  $hb, wb$  are the height and width of the blob respectively, and  $\beta$  is the tuning factor.

Figure 4-12 shows the result of applying the proposed technique with different values of  $\beta$ . In (b) it is clear that the points are isolated and there is a large number of interesting regions. In (c) the number of regions is reasonable and the overlapped blobs are with similar features. Finally, in (d) the size of the blob is very large and hence most of the points are overlapped which will result in a very small number of interesting regions. The case in (d) is useful in defining the focusing region, which can be defined as the region that contains all the regions of interest, or all the regions that the user gazes inside.

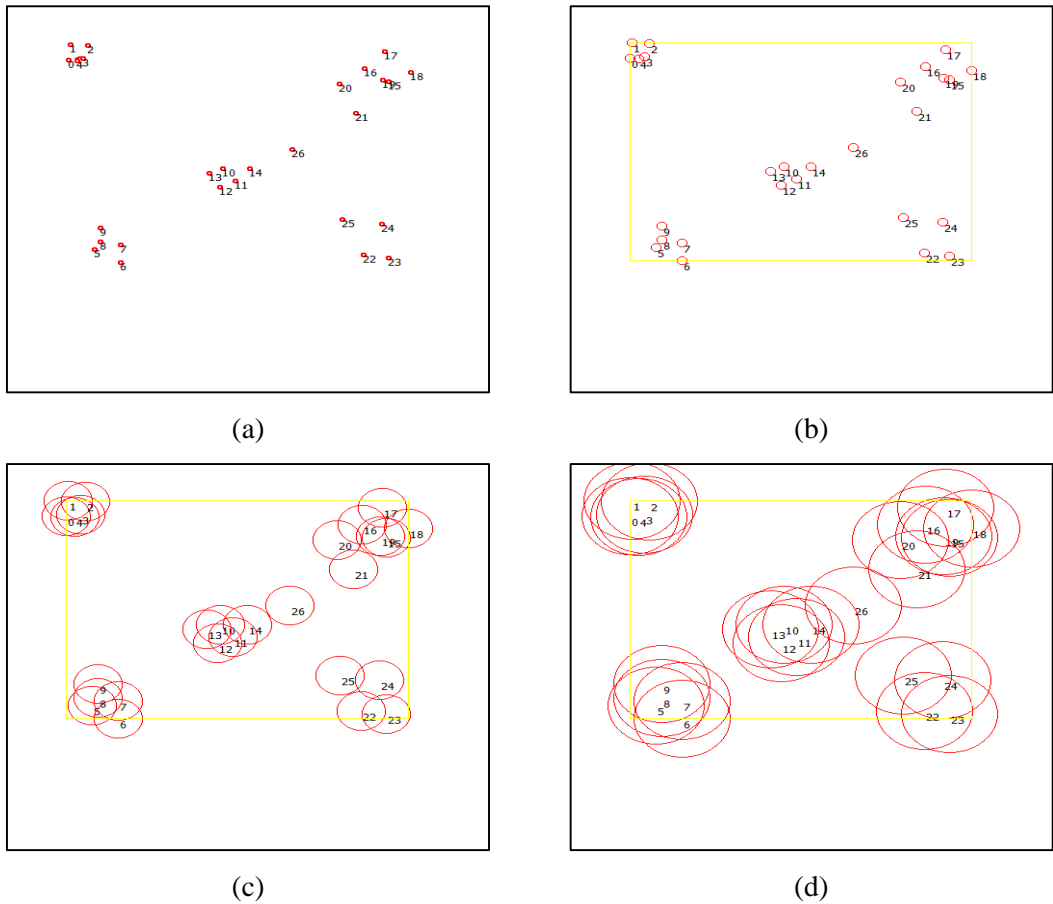


Figure 4-12: The selection of the size of the blob, (a) points, (b)  $\beta = 0.02$ , (c)  $\beta = 0.10$ , (d)  $\beta = 0.20$ .

Clouds are constructed by grouping the overlapped blobs. This grouping process uses boundary tracing to form the union of the blobs to form the cloud. The clouds are then

converted into interesting regions by drawing a regular shape (rectangle, polygon, or circle) which contains the cloud. The minimum and maximum values for each coordinate are extracted, and then these values are used to specify the surrounding shape.

Figure 4-13 illustrates the process described above. As shown in this figure, the points are replaced with circles or ellipses as shown in (a). The overlapped circles are joined together to form the cloud as shown in (b). By tracing the boundaries of the cloud the extreme values ( $X_{min}$ ,  $X_{max}$ ,  $Y_{min}$ , and  $Y_{max}$ ) can be extracted as shown in (c). These points are used to specify the location and the dimensions of the surrounding shape which represents the interesting region.

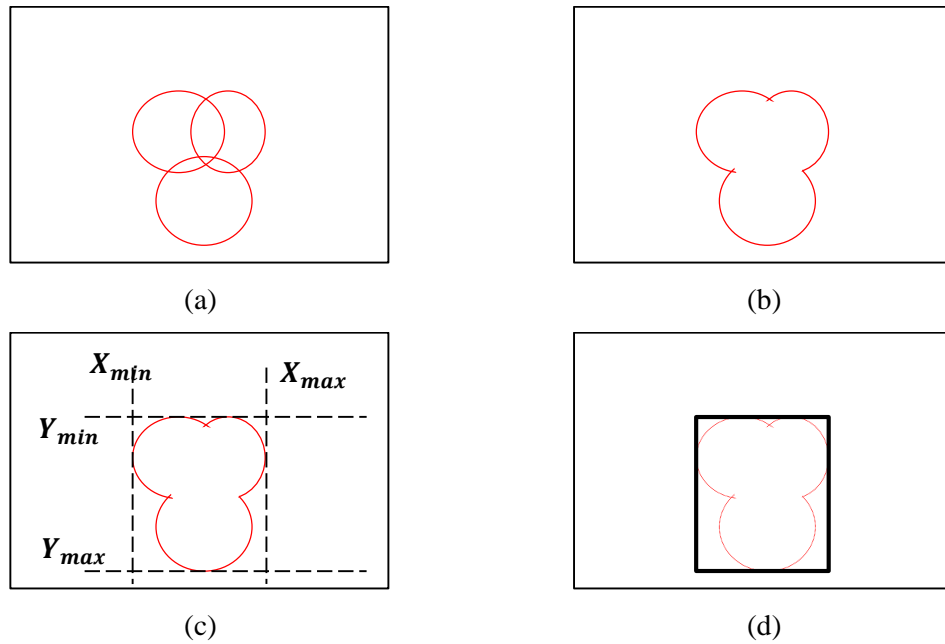


Figure 4-13: Union of overlapped blobs to form an interesting region, (a) overlapped blobs, (b) merging the blobs to form a region, (c) extracting the dimensions of the region, (d) the interesting region.

The results of applying the above technique are given in Figure 4-14, which shows the results with different values of  $\beta$ , and the effect it has on the number and size of the regions of interest. Both rectangles and circles have been used as blobs. From the figure, it is obvious that the value of  $\beta$  plays an important role in cloud size and regions of interest. In (a),  $\beta$  value was set to 0.02 which means that the dimensions of the blobs are 2% of the image dimensions, i.e.  $(0.02 \times W) \times (0.02 \times H)$ . This value is very small and no overlaps among the points were detected. In (b) the value of  $\beta$  was set to 0.1,

which has given better results, and larger clouds were constructed. In (c) the value of  $\beta$  has been selected to be relatively high resulting in clouds that are relatively large.

Figure 4-15 shows an example of the application of the above technique upon real data collected from the eye tracker. In this figure, it is clear that the size of the regions of interest is very close to the size of the cloud. Subsequent selection of the value of  $\beta$  will affect the size of the interesting regions as well as the cloud sizes.

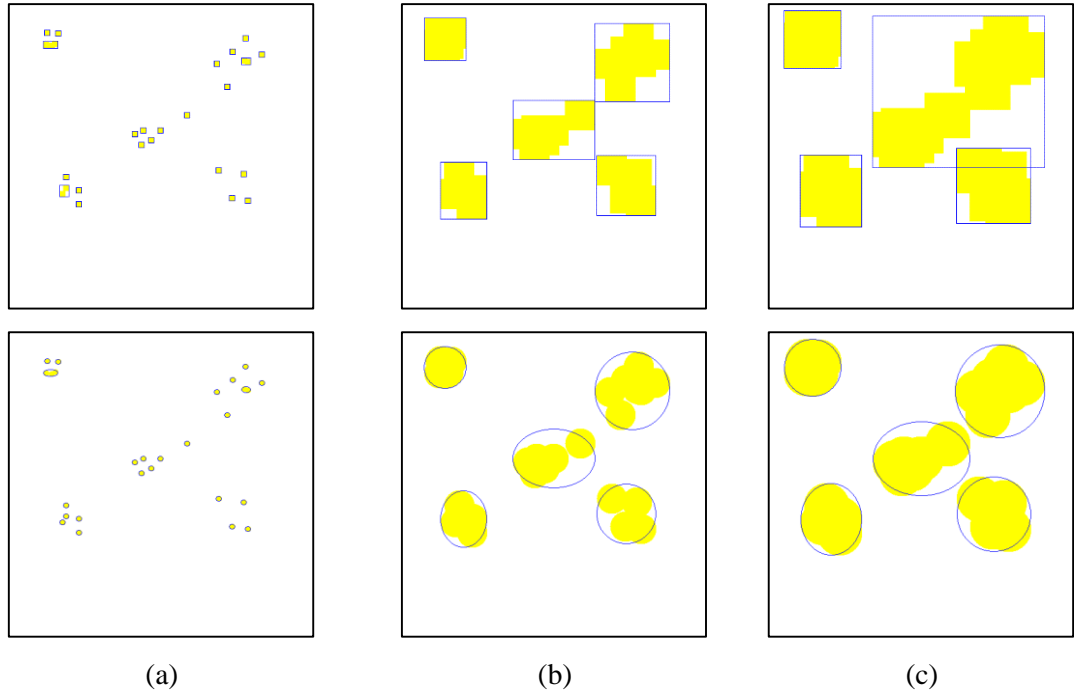


Figure 4-14: Converting gazing points to blobs and constructing clouds from the overlapped blobs, (a)  $\beta = 0.02$ , (b)  $\beta = 0.10$ , (c)  $\beta = 0.15$ .

#### 4.6 Iterative Blobs-Based Clustering (IBBC)

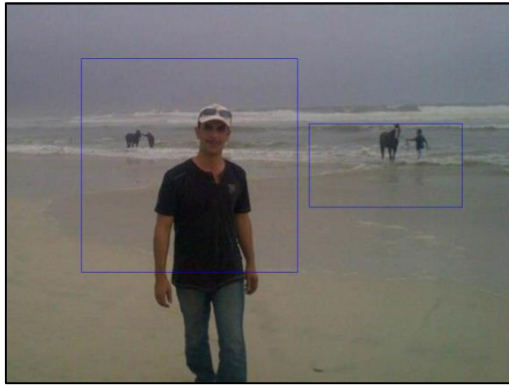
In this Section, a modified blob-based clustering technique is developed, in which the concepts of bottom-up, hierarchical, agglomerative techniques are considered. In this discussion, we shall use a circle to represent each point, although a rectangle or an ellipse can be used as well. In Section 4.5, the size of the blobs was constant and it was found from a ratio of the image height and width, which will also limit the process from being unsupervised since the user needs to identify the value of  $\beta$ .



(a)



(b)



(c)

Figure 4-15: Converting gaze points to blobs, (a)  $\beta = 0.02$ , (b)  $\beta = 0.10$ , (c)  $\beta = 0.20$ .

The blobs' dimensions are changed with iterations, i.e. for a circle the radius at iteration  $n$  is increased by an increment  $\delta_r$ . The increment could be any integer value, based on the size of the image, however, for safety reasons we will select the increment as one. The radius and the area of the blob changes with iterations can be expressed as follows:



$$\begin{aligned}
r(n+1) &= r(n) + \delta_r \\
r(n) &= r(0) + n\delta_r \\
r(0) &= 1 \\
a(n) &= 2\pi r(n) = 2\pi n\delta_r
\end{aligned}
\tag{4-21}$$

where,  $r(n)$  is the radius at iteration  $n$ ,  $r(0)$  is the radius at  $n = 0$ , and  $a(n)$  is the area of the blob at  $n$ .

The number of clusters will start with  $N$ , i.e. at iteration  $n = 0$ ; the number of clusters is equal to the number of points. This number shall change with iteration, it will be reduced by half of the number of intersected regions since each two overlapping and intersected regions will produce a new single region.

$$\begin{aligned}
N_c(n) &= N - N_\cap(n)/2 \\
N_c(0) &= N
\end{aligned}
\tag{4-22}$$

where  $N_c(n)$  is the number of clusters at iteration  $n$  and  $N_\cap(n)$  is the number of overlapped regions at iteration  $n$ .

In addition, we shall define the cluster mean distance as the average of the distances inside clusters, i.e. the distances between the points in a particular cluster, as follows:

$$\mu_i(n) = \frac{1}{\left(N_i^c(n)\right)^2} \sum_{j=1}^{N_i^c(n)} \sum_{k=1}^{N_i^c(n)} D(p_j, p_k) ; i = 1, 2, \dots, N_c(n)
\tag{4-23}$$

where  $\mu_i(n)$  is the mean of the distance inside the  $i^{th}$  cluster at iteration  $n$  and  $N_i^c(n)$  is the number of points in the  $i^{th}$  cluster at iteration  $n$ .

The standard deviation inside the cluster shall be calculated as follows:

$$\sigma_i(n) = \sqrt{\frac{1}{\left(N_i^c(n)\right)^2} \sum_{j=1}^{N_i^c(n)} \sum_{k=1}^{N_i^c(n)} \left(D(p_j, p_k) - \mu_i(n)\right)^2} ; i = 1, 2, \dots, N_c(n)
\tag{4-24}$$

We shall define  $\mu^c(n)$  and  $\sigma^c(n)$  as the internal distance and internal standard deviation respectively, which are the average of the mean distances and standard deviation for all clusters at iteration  $n$ , and  $\omega^c(n)$  is the normalized internal distances average as follows:

$$\mu^c(n) = \frac{1}{N_c(n)} \sum_{k=1}^{N_c(n)} \mu_k(n) \quad 4-25$$

$$\sigma^c(n) = \frac{1}{N_c(n)} \sum_{k=1}^{N_c(n)} \sigma_k(n) \quad 4-26$$

$$\omega^c(n) = \frac{\mu^c(n)}{\sigma^c(n)} \quad 4-27$$

External distance and standard deviation can be defined as the statistical measures between the centroids of the clusters and can be calculated as follows:

$$\begin{aligned} x_i(n) &= \frac{1}{N_i^c(n)} \sum_{j=1}^{N_i^c(n)} x_j(n) \quad ; i = 1, 2, \dots, N_c(n) \\ y_i(n) &= \frac{1}{N_i^c(n)} \sum_{j=1}^{N_i^c(n)} y_j(n) \quad ; i = 1, 2, \dots, N_c(n) \\ p_i(n) &= (x_i(n), y_i(n)) \\ \mu^x(n) &= \frac{1}{(N_c(n))^2} \sum_{j=1}^{N_c(n)} \sum_{k=1}^{N_c(n)} D(p_j(n), p_k(n)) \\ \sigma^x(n) &= \sqrt{\frac{1}{(N_c(n))^2} \sum_{j=1}^{N_c(n)} \sum_{k=1}^{N_c(n)} \left( D(p_j(n), p_k(n)) - \mu^x(n) \right)^2} \\ \omega^x(n) &= \frac{\mu^x(n)}{\sigma^x(n)} \end{aligned} \quad 4-28$$

where  $x_i(n)$  and  $y_i(n)$  are the average of the  $x$  and  $y$  coordinates of the cluster  $i$  that form the point  $p_i(n)$ .  $\mu^x(n)$  and  $\sigma^x(n)$  are the external mean and standard deviation between the centroids of the clusters.

It is clear that there is a direct relationship between the change in internal and external distances and standard deviations, and for every change in the internal measures, there should be a change in the external measures. This is because the process of merging points together will reduce the number of clusters and increase the number of points found in any one of the clusters. This change in the number of clusters and number of points in clusters will affect both internal and external statistical measures. It was noticed

that the value of  $\omega^c(n)$  increases with iterations while the value of  $\omega^x(n)$  decreases with iteration since the number and arrangement of the clusters' centroids are changed; thus we shall define the difference between the two measures as follows:

$$\varepsilon(n) = |\omega^c(n) - \omega^x(n)| \quad 4-29$$

In the case of no change in the clustering process, the change of  $\varepsilon(n)$  with respect to the iterations should be zero, i.e.  $\Delta\varepsilon(n)/\Delta n = 0$ , This case will be considered as the stopping criterion. The results of applying the above technique are given in Figure 4-16. In this figure, different measures have been tested with respect to iteration and studied to find the best stopping criterion. In the same figure, (a) shows the points that need to be clustered, (b) shows the result after applying the proposed algorithm. (c) and (d) show the internal and external distance changes with iterations respectively. It is clear from these figures that the external distance is very much higher than the internal distance, so in order to keep internal and external distances within the same range, we have divided them on the internal and external standard deviation given in (e) and (f).

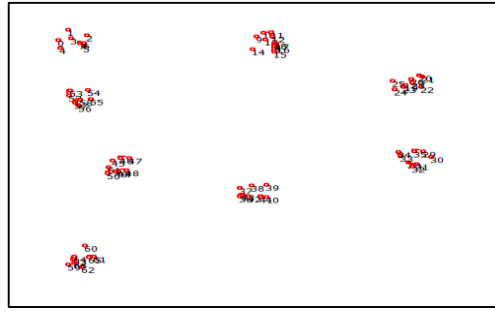
The normalized internal distance  $\omega^c(n)$  increases with iterations since the internal distance increases and the standard deviation change is smaller than the change of the distance since the variation of the points in clusters is not large. In contrast, the external distance varies inversely with the iteration since the change in standard deviation is large as compared to the change of the external distance, as shown in (g). In the same figure, the difference  $\varepsilon(n)$  starts decreasing from its large value until it reaches the minimum point and then it starts increasing again. When the change of  $\varepsilon(n)$  i.e.  $\Delta\varepsilon(n)/\Delta n$ , as shown in (h), becomes zero, that means that no further clustering can be achieved and we shall consider this case as the optimum case.

The proposed technique has improved the extraction of the regions of interest based on interest points' data in different ways; the obtained results are quite satisfactory and the extracted regions do indeed represent the most interesting regions.

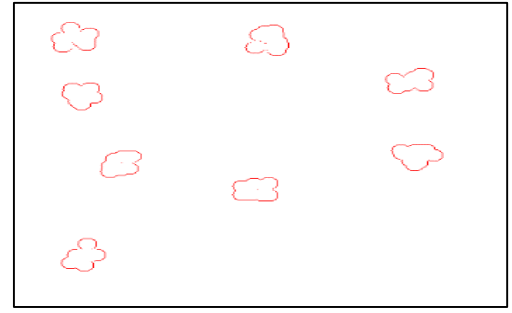
The main advantages of the proposed method are; it does not need to compare each point with all other points, since all points (blobs) will grow at the same time and they will be merged with the neighbouring points in a process similar to water drops merging. Thus, it needs fewer computation processes since during the growing process more than one point may get connected to clusters at the same time. The second important achievement

is that it does not need any human intervention, since the stopping criterion is automatically specified.

The algorithm was applied on about 500 different points combinations. Some of the combinations were artificial, to simulate the cases that were not obtained by using the eye-tracking device. Table 4-5 shows some of the points combinations and the results of applying the proposed algorithm. Table 4-6 shows the results of applying the proposed algorithm on real data. From the results listed in the two tables it is clear that there are some clusters with a small number of points and sometimes with only one point, which can easily be filtered since they might come from noise or task-free gazing.



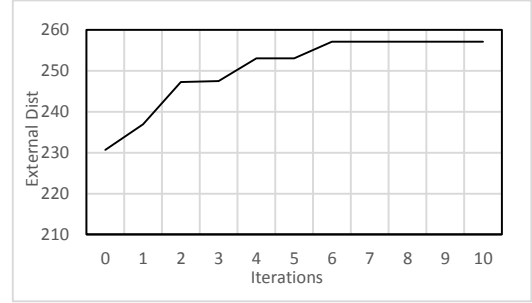
(a)



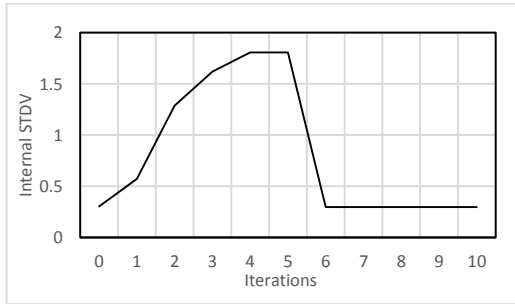
(b)



(c)



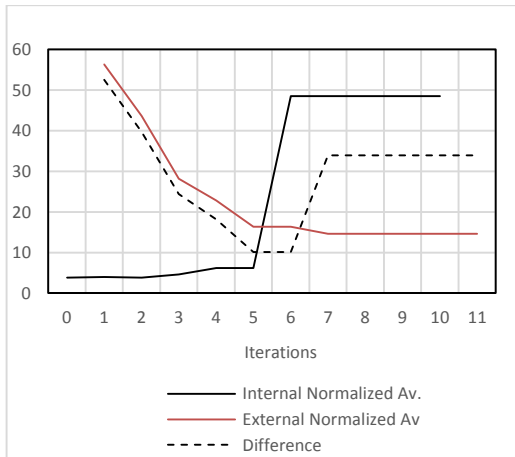
(d)



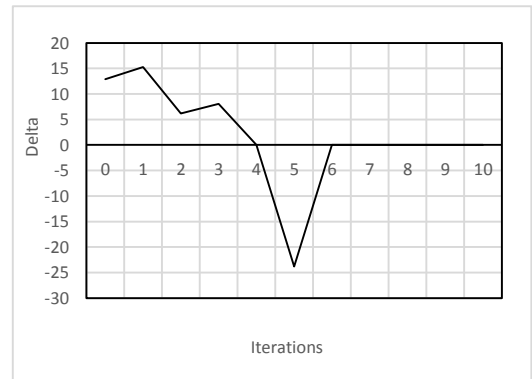
(e)



(f)



(g)



(h)

Figure 4-16: Changes of variables with iterations, (a) points to be clustered, (b) clustered points, (c) internal distance, (d) external Distance, (e) internal standard deviation, (f) external standard deviation, (g)  $\omega^x(n)$  vs.  $\omega^c(n)$  and  $\varepsilon(n)$ , (h)  $\Delta\varepsilon(n)/\Delta n$ .

Table 4-5: Results of applying the proposed algorithm after removing the background for visual clarification.

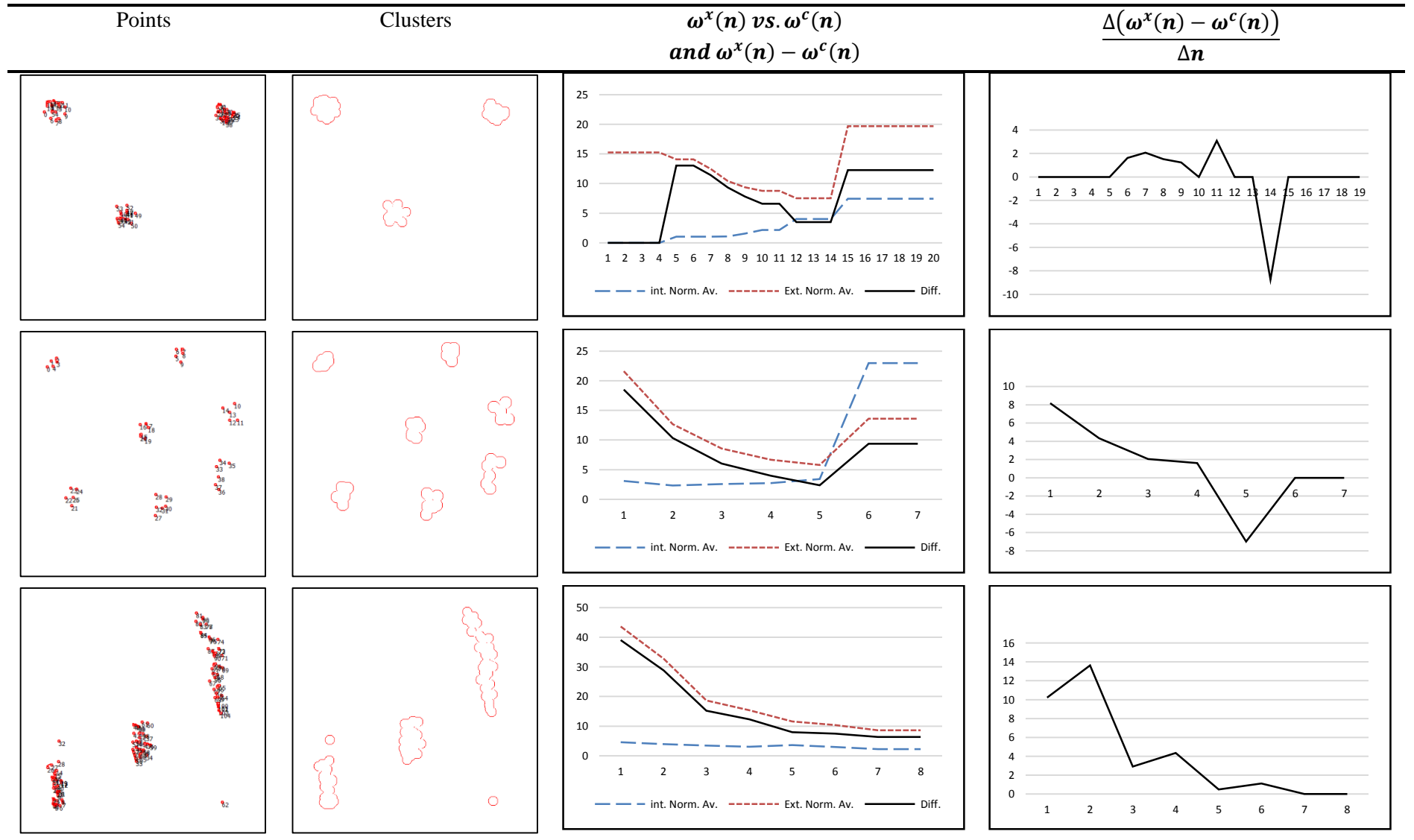


Table 4-6: Results of Applying the proposed algorithm.

Gaze Points	Extracted regions using the proposed technique	Region approximation to ellipse	information
			Image Size: 400 x 400 Gaze Points: 27 Increment $\delta_r = 2$ Iterations: 5 Blob size: 18 pixel
			
			
			
			Image Size: 400 x 400 Gaze Points: 92 Increment $\delta_r = 2$ Iterations: 4 Blob size: 16 pixel
			Image Size: 400 x 400 Gaze Points: 46 Increment $\delta_r = 2$ Iterations: 8 Blob size: 24 pixel
			Image Size: 400 x 400 Gaze Points: 41 Increment $\delta_r = 2$ Iterations: 8 Blob size: 24 pixel

The above clustering algorithm is necessary to cluster gaze points or salient points to form interest regions. The gaze points can be obtained from eye trackers, while salient points can be obtained from saliency extraction algorithms. In the following Section, we shall discuss the proposed saliency extraction algorithm which utilizes the above clustering technique.

## 4.7 Irregularity-Based Saliency Identification (IBSI)

Parts of this section were published in Multimedia Applications and Tools Journal, 2014 [178] and in IEEE ICCIC 2013 [179].

In the previous sections, we have furnished the necessary theory and proposed the algorithms necessary for developing the new saliency extraction algorithm that will be discussed in this Section.

Based on the discussion in Section 4.2 above, one may notice some important problems need to be considered when thinking about developing a new saliency extraction technique. Some problems are due to the nature of the image, such as the effect of texture when extracting the salient regions using corners, and the high frequency regions in regions of less importance in the case of frequency. Another important problem, which is the pre-knowledge of the nature of the image and its contents before extracting the salient regions, should also be treated as well. One suggested solution for these limitations, which is part of our study, is to use the information that can be extracted from the image itself to recognize the importance of the objects in it. Properties, such as contrast, location and rarity have been used to highlight regions as salient regions as was discussed earlier. In rarity, we have argued that rarity of colour, texture, and other features does not necessarily mean that the regions with those rare features are the salient regions. The same thing is applicable to contrast: contrast alone cannot give an impression about the saliency of the object and it needs more features considered together with it to determine its saliency. In location-based techniques, we have discussed that salient objects do not need to be in the middle of the image. The same thing is true of other features.

It was noticed that the salient region is one which is different from the surrounding environment, in structure, colour, and intensity. We shall refer to that as irregularity of the region. Irregular regions are the regions that are different from the surrounding background, such as a fine region in a coarse background or a region with a different structure as compared to the background. Hence, we shall consider the irregularity as the main measure of saliency.

Irregularity can be defined as any irregular part or region in the image both locally and globally, in contrast to the frequency or texture. In the case of frequency, regions with



high variation, such as edges, are considered as high frequency regions, which in their turn are considered as salient regions. In our research, we shall argue that saliency is neither a local property nor a global property alone; one should consider the irregularity both locally and globally, therefore we shall define the salient region as follows:

**Definition:** A region is said to be salient if its contents are irregular as compared to other regions in the image.

The information in the image shall be divided into two parts, important and unimportant, in other words, salient regions and non-salient regions.

$$H(I) = H(R) + H(S) \quad 4-30$$

Where  $I$  is the image,  $H(.)$  is an information contents measure,  $R$  is the regular or unimportant regions in the image, and  $S$  is the salient or the important regions. Many possible measures can be used to measure the information contents of a dataset, such as entropy and statistical measures.

By considering that the salient points occur randomly in an image, and based on the nature of the image, the pixels' values can be treated as random variables. Therefore, we shall consider the statistical measures, such as expected value and variation measure, to measure the formality and saliency of a region in an image. If the expectation value is very close to the pixels' values in the case of regular regions, and the variation measure value is small, then the measure of variation (irregularity) can be derived from these two measures.

#### **4.7.1 Statistical Measures as Descriptors**

Statistical measures are widely used in different applications and image analysis is one of these applications. The main advantages of using statistical measures in image processing techniques lie in the simplicity of the calculations and the fact that they do not get affected much by the variations of the image. Many statistical measures have been proposed and used, such as, mean, variance, standard deviation, skewness, etc.

##### ***Expected Value***

The difference between the pixels' values and the expected value ( $\mu$ ) in a specific region ( $r$ ) can be considered as a measure of variation, since it gives a low value in uniform

regions and a high value in non-uniform regions. Therefore, the first order variation measure ( $v_f$ ) shall be considered as the first measure of irregularity and is given by:

$$v_f = \sum_{x,y \in r} |\mathcal{I}(x,y) - \mu|$$

$$\mu = \frac{1}{h \times w} \sum_{y=1}^h \sum_{x=1}^w \mathcal{I}(x,y)$$
4-31

where  $\mathcal{I}(x,y)$  is the image intensity value at location  $(x,y)$  and  $h$  and  $w$  are the height and width of the region  $r$  respectively.

In the following discussion we shall prove that the expected value is changing with the irregularity of the image. We shall derive the relations and equations in the one-dimensional space, and then the equations can be extended to be applicable for two dimensions.

Let us assume that the pixel in the window is represented by:

$$p_i = p_m + \delta_i, \quad i = 1, 2, \dots, N$$
4-32

Where  $p_i$  is the pixel value,  $p_m$  is a reference value, in our case it is the minimum pixel value in the window, and  $\delta_i$  is the difference between the pixel  $p_i$  and the reference pixel value  $p_m$ . The mean can be calculated as:

$$\mu = \frac{1}{N} \sum_{i=1}^N p_i$$
4-33

By substituting the value of  $p_i$ , we shall get:

$$\mu = \frac{1}{N} \sum_{i=1}^N (p_m + \delta_i)$$
4-34

$$\mu = \frac{1}{N} \sum_{i=1}^N p_m + \frac{1}{N} \sum_{i=1}^N \delta_i$$
4-35

Since  $p_m$  is a constant then  $\sum_{i=1}^N p_m$  is equal to  $Np_m$

$$\mu = \frac{1}{N} \left[ Np_m + \sum_{i=1}^N \delta_i \right]$$
4-36

$$\mu = p_m + \frac{1}{N} \sum_{i=1}^N \delta_i \quad 4-37$$

The variation measure in Equation 4-31 then becomes:

$$\begin{aligned} v_f &= \frac{1}{N} \sum_{i=1}^N |p_i - \mu| \\ v_f &= \frac{1}{N} \sum_{i=1}^N \left| p_m + \delta_i - \left( p_m + \frac{1}{N} \sum_{i=1}^N \delta_i \right) \right| \\ v_f &= \frac{1}{N} \sum_{i=1}^N \left| \textcolor{red}{p}_m + \delta_i - \textcolor{red}{p}_m - \frac{1}{N} \sum_{i=1}^N \delta_i \right| \\ v_f &= \frac{1}{N} \sum_{i=1}^N \left| \delta_i - \frac{1}{N} \sum_{i=1}^N \delta_i \right| \end{aligned} \quad 4-38$$

The value of the first order variation  $v_f$  is very small in uniform regions as the difference between the pixels is small and it tends to be zero if the difference between these values is constant and equal.

From Equation 4-38, it is obvious that if the difference between the pixels' values is small, the value of the variation measure tends to be small; and if the difference is constant then the value of the variation measure shall reach zero, as given below:

$$\begin{aligned} v_f &= \frac{1}{N} \sum_{i=1}^N \left| \delta_i - \frac{1}{N} \sum_{i=1}^N \delta_i \right| \\ \delta_1 &= \delta_2 = \dots = \delta_N = \delta \\ v_f &= \frac{1}{N} \sum_{i=1}^N \left| \delta - \frac{1}{N} \sum_{i=1}^N \delta \right| \\ v_f &= \frac{1}{N} \sum_{i=1}^N \left| \delta - \frac{1}{N} N\delta \right| \\ v_f &= \frac{1}{N} \sum_{i=1}^N |\delta - \delta| = 0 \end{aligned} \quad 4-39$$

In contrast, if the value of  $\delta_i$  is large then the variation measure value shall be very high.

The maximum possible value for the variation measure is equal to  $\frac{M}{2}$ , where M is the maximum intensity value. This is because the maximum value of  $v_f$  shall occur when the term  $|p_i - \mu|$  is large for all i, the maximum possible difference that can occur is between the maximum and minimum possible values. If we assume the minimum possible value is zero, (since we are talking about grey levels and no negative values are allowed) and the maximum value is M, then the maximum the difference could be is M. With this assumption, half of the values should be zeroes and half of them should be M; this leads to an average value of  $(M - 0)/2$  as the average of the gray levels in the image.

From the above discussion, one can observe that the maximum possible value for  $v_f$  is equal to  $M/2$  as given below:

$$\begin{aligned}
 v_f^* &= \frac{1}{N/2} \sum_{i=1}^{N/2} |0 - \mu| + \frac{1}{N/2} \sum_{i=\frac{N}{2}+1}^N |M - \mu| \\
 v_f^* &= \frac{1}{N/2} \sum_{i=1}^{N/2} \left| 0 - \frac{M}{2} \right| + \frac{1}{N/2} \sum_{i=\frac{N}{2}+1}^N \left| M - \frac{M}{2} \right| \\
 v_f^* &= \frac{1}{N/2} \sum_{i=1}^{N/2} \left| -\frac{M}{2} \right| + \frac{1}{N/2} \sum_{i=\frac{N}{2}+1}^N \left| \frac{M}{2} \right| \tag{4-40} \\
 v_f^* &= \frac{1}{N} \sum_{i=1}^N \left| \frac{M}{2} \right| \\
 v_f^* &= \frac{1}{N} \left( N \cdot \frac{M}{2} \right) = \frac{M}{2}
 \end{aligned}$$

### ***Deviation Measure***

Another possible measure, which can be extracted from the image to describe the variation in a region, is the standard deviation. The two dimensional discrete standard deviation can be calculated using the following equation:

$$\sigma = \sqrt{\frac{1}{h \times w} \sum_{y=1}^h \sum_{x=1}^w (\mathcal{I}(x, y) - \mu)^2} \tag{4-41}$$

The value of the standard deviation in the regular region is very small and tends to be zero if the difference between the pixels' values is constant and equal. Again, we shall show that this measure is very small when the region is smooth and with no big variation as follows:

$$\begin{aligned}
 \sigma &= \sqrt{\frac{1}{N} \sum_{i=1}^N \left[ (p_m + \delta_i) - \left( p_m + \frac{1}{N} \sum_{i=1}^N \delta_i \right) \right]^2} \\
 \sigma &= \sqrt{\frac{1}{N} \sum_{i=1}^N \left[ \delta_i - \frac{1}{N} \sum_{i=1}^N \delta_i \right]^2} \\
 \sigma &= \sqrt{\frac{1}{N} \sum_{i=1}^N \left[ \delta_i^2 - 2\delta_i \frac{1}{N} \sum_{i=1}^N \delta_i + \left( \frac{1}{N} \sum_{i=1}^N \delta_i \right)^2 \right]}
 \end{aligned} \tag{4-42}$$

Again, for smooth regions the differences between pixels' values is very small and close or similar in values. To demonstrate, we shall assume the values are similar and equal to  $\delta$ , i.e.  $\delta_1 = \delta_2 = \dots = \delta_N = \delta$ , then the standard deviation value shall be as follows:

$$\begin{aligned}
 \sigma &= \sqrt{\frac{1}{N} \sum_{i=1}^N \left[ \delta^2 - 2\delta \frac{1}{N} \sum_{i=1}^N \delta + \left( \frac{1}{N} \sum_{i=1}^N \delta \right)^2 \right]} \\
 \sigma &= \sqrt{\frac{1}{N} \sum_{i=1}^N \left[ \delta^2 - 2\delta \frac{1}{N} N\delta + \frac{1}{N^2} (N\delta)^2 \right]} \\
 \sigma &= \sqrt{\frac{1}{N} \sum_{i=1}^N [\delta^2 - 2\delta^2 + \delta^2]} = 0
 \end{aligned} \tag{4-43}$$

The maximum possible value for the standard deviation is equal to  $\frac{M}{2}$ , where M is the maximum intensity value. As discussed earlier, the maximum value shall occur when half of the pixels are zeroes and half of them are maximum possible intensity value (M).

$$\begin{aligned}
\sigma^* &= \sqrt{\frac{1}{N/2} \sum_{i=1}^{N/2} (0 - \mu)^2 + \frac{1}{N/2} \sum_{i=\frac{N}{2}+1}^N (M - \mu)^2} \\
\sigma^* &= \sqrt{\frac{1}{N/2} \sum_{i=1}^{N/2} \left(\frac{M}{2}\right)^2 + \frac{1}{N/2} \sum_{i=\frac{N}{2}+1}^N \left(M - \frac{M}{2}\right)^2} \\
\sigma^* &= \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{M}{2}\right)^2} = \sqrt{\frac{1}{N} \cdot N \cdot \left(\frac{M}{2}\right)^2} = \frac{M}{2}
\end{aligned} \tag{4-44}$$

The variation measure can then be extracted as a function of the mean, standard deviation, and the pixel value, i.e.

$$v = f(\mathcal{I}(x, y), \mu, \sigma) \tag{4-45}$$

One possible equation is by taking the product of the factors given above such that:

$$\begin{aligned}
v &= |\mathcal{I}(x, y) - \mu| \cdot \sigma \\
v^* &= \left(\frac{M}{2}\right)^2 \\
v^{norm} &= \frac{(|\mathcal{I}(x, y) - \mu| \cdot \sigma)}{\left(\frac{M}{2}\right)^2}
\end{aligned} \tag{4-46}$$

where  $v^*$  is the maximum possible value for the irregularity measure and  $v^{norm}$  is the normalized variation measure.

#### 4.7.2 Local Saliency Identification (LSI)

This section will discuss the irregularity in the image locally, for which the image shall be divided into sub-images (windows). The irregularity measure shall be applied to each window and then windows with a high irregularity measure shall be highlighted as salient regions while other windows will be marked as background.

Let  $\mathcal{I}(x, y)$  be the image intensity at position  $(x, y)$ , where  $x = 1, 2, \dots, W, y = 1, 2, \dots, H$ . The image can be represented as a set of pixels  $\mathbb{I}$ , which may be defined as:

$$\begin{aligned}\mathbb{I} &= \{p_{xy} \mid x, y \in \mathbb{N} \wedge x = 1, 2, \dots, W \wedge y = 1, 2, \dots, H\} \\ p_{xy} &= \mathcal{I}(x, y)\end{aligned}\tag{4-47}$$

where  $p_{xy}$  is the pixel's value at location  $(x, y)$  and  $\mathbb{N}$  is the set of natural numbers.

The set  $\mathbb{P}_{\mathbb{I}}$  can be defined as power set of  $\mathbb{I}$  which contains all the subsets  $\mathbb{s}$  that can be extracted from the set  $\mathbb{I}$ . The definition of  $\mathbb{P}_{\mathbb{I}}$  is given by the following formula:

$$\mathbb{P}_{\mathbb{I}} = \{\mathbb{s} \mid \mathbb{s} \subseteq \mathbb{I}\}\tag{4-48}$$

According to our application, the main constraint is that: all the elements in a given subset should belong to a connected region i.e. they should be in a given neighbouring area.

The sub-image (window) is a subset of the image with elements that belong to the same region. The size of this sub-image is  $w \times h$  and shall be denoted as region  $\mathbb{r}_{ij}$ .

Considering that the regions are disjointed, they can be defined as follows:

$$\begin{aligned}\mathbb{r}_{ij} &= \{p_{kl} \mid p_{kl} \in \mathbb{I} \wedge j \times w < l \leq (j+1)w \wedge i \times h < k \leq (i+1)h\} \\ \mathbb{R} &= \left\{ \mathbb{r}_{ij} \mid 1 \leq j \leq \frac{W}{w} \wedge 1 \leq i \leq \frac{H}{h} \right\} \\ \mathbb{R} &\subseteq \mathbb{P}_{\mathbb{I}}\end{aligned}\tag{4-49}$$

Assuming that there is no overlap between adjacent regions, the total number of regions is  $\frac{W}{w} \times \frac{H}{h}$ . In most cases there will be an overlap between the regions, and if we denote the overlapping value by  $\delta$ , then the total number of regions is  $\frac{W}{w-\delta} \times \frac{H}{h-\delta}$ .

Furthermore, we shall define a new space, which is the description space, in which; each region  $\mathbb{r}_{ij}$  will go through features extractors to convert it into a set of measures. We shall define mapping  $\psi$  from the regions space  $\mathbb{R}$  to the description space  $\mathbb{D}$  as given below:

$$\psi: \mathbb{R} \rightarrow \mathbb{D}\tag{4-50}$$

Where  $\mathbb{D}$  is the set of descriptions of the regions in  $\mathbb{R}$  and is given by:

$$\begin{aligned}\mathbb{d}_{ij} &= \psi(\mathbb{r}_{ij}) \\ \mathbb{D} &= \left\{ \mathbb{d}_{ij} \mid 1 \leq j \leq \frac{W}{w} \wedge 1 \leq i \leq \frac{H}{h} \right\}\end{aligned}\tag{4-51}$$

$\mathbb{d}_{ij}$  might be a single value or a set of values.

If we consider  $\psi(\cdot)$  as the variation or irregularity measure then it can be expressed as follows:

$$\begin{aligned}\mathbb{d}_{ij} &= \psi(\mathbb{r}_{ij}) = f(p_{xy}, \mu_{ij}, \sigma_{ij}) \\ v_{ij} &= \sum_{y=1}^h \sum_{x=1}^w |p_{xy} - \mu_{ij}| \sigma_{ij}; \forall p_{xy} \in \mathbb{r}_{ij} \\ \mathbb{d}_{ij} &= \{v_{ij}\}\end{aligned}\tag{4-52}$$

Where  $v_{ij}$  is the measure of variation at region  $\mathbb{r}_{ij}$ , in this case  $\mathbb{d}_{ij}$  shall be represented by one value which is  $v_{ij}$ . In order to make the values acceptable and comparable, we shall normalize the value of  $\mathbb{d}_{ij}$  by dividing them by the maximum as follows:

$$\begin{aligned}\mathbb{d}^* &= \text{Max}(\mathbb{D}) \\ \mathbb{d}_{ij}^* &= \frac{\mathbb{d}_{ij}}{\mathbb{d}^*}\end{aligned}\tag{4-53}$$

$\mu_{ij}$  and  $\sigma_{ij}$  are the mean and standard deviation of the pixels in region  $\mathbb{r}_{ij}$  and are calculated as follows:

$$\begin{aligned}\mu_{ij} &= \sum_{y=-\frac{h}{2}}^{\frac{h}{2}} \sum_{x=-\frac{w}{2}}^{\frac{w}{2}} \frac{1}{h \times w} \mathcal{I}(x+i, y+j) \\ \sigma_{ij} &= \sqrt{\frac{1}{h \times w} \sum_{y=-\frac{h}{2}}^{\frac{h}{2}} \sum_{x=-\frac{w}{2}}^{\frac{w}{2}} (\mathcal{I}(x+i, y+j) - \mu_{ij})^2}\end{aligned}\tag{4-54}$$

It is clear that the value of  $|\mathcal{I}(x, y) - \mu_{ij}|$  will be higher in irregular regions than that for the uniform region since in the case of uniform regions, the values of the intensity in that region will be close to the value of the expected value (mean). In contrast, for a region with higher variation there is higher difference between the mean and the pixels' values.

### ***Numerical Example***

In order to clarify the idea empirically we shall consider the following numerical example, showing two different regions, formal and irregular. In Figure 4-17, (a) and (b) show the pixels' values for a regular region (R) and irregular or salient region (S) respectively. The mean was calculated for the two regions and the values were  $\mu_R = 121$



for a regular region in (a) and  $\mu_S = 121$  for an irregular region in (b). Although the values are the same, the differences from the pixels' values are different as shown in (c) and (d).

	1	2	3	4	5
1	120	125	122	120	125
2	125	120	121	119	118
3	122	121	120	119	118
4	121	122	121	120	119
5	123	124	125	124	123

(a)

	1	2	3	4	5
1	120	200	210	150	180
2	110	200	120	90	120
3	120	80	120	100	20
4	10	200	10	200	10
5	120	150	100	140	150

(b)

	1	2	3	4	5
1	1	4	1	1	4
2	4	1	0	2	3
3	1	0	1	2	3
4	0	1	0	1	2
5	2	3	4	3	2

(c)

	1	2	3	4	5
1	1	79	89	29	59
2	11	79	1	31	1
3	1	41	1	21	101
4	111	79	111	79	111
5	1	29	21	19	29

(d)

	1	2	3	4	5
1	2.2	8.8	2.2	2.2	8.8
2	8.8	2.2	0	4.4	6.6
3	2.2	0	2.2	4.4	6.6
4	0	2.2	0	2.2	4.4
5	4.4	6.6	8.8	6.6	4.4

(e)

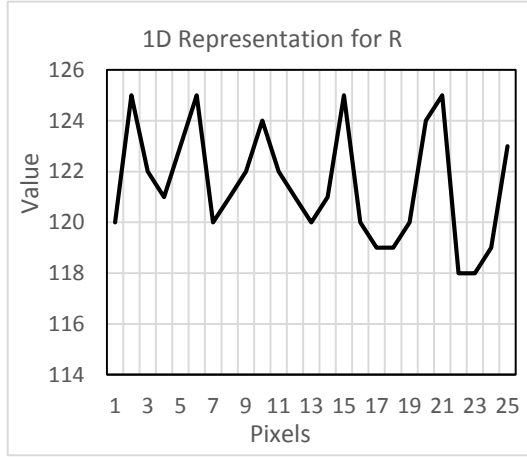
	1	2	3	4	5
1	60	4740	5340	1740	3540
2	660	4740	60	1860	60
3	60	2460	60	1260	6060
4	6660	4740	6660	4740	6660
5	60	1740	1260	1140	1740

(f)

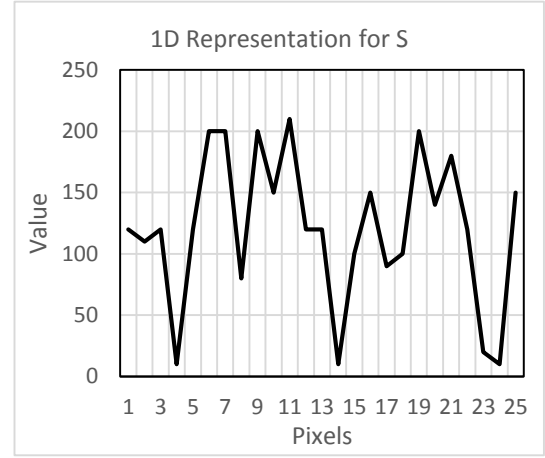
Figure 4-17: Numerical example on formal and irregular regions, (a) formal region, (b) irregular region, (c) difference from mean for (a), (d) difference from mean for (b), (e)  $|\mathcal{J}(x, y) - \mu_R| \times \sigma_R$  of (a), (f)  $|\mathcal{J}(x, y) - \mu_S| \times \sigma_S$  of (b).

The standard deviation was calculated for both regions R and S and the values were:  $\sigma_R = 2.2$  and  $\sigma_S = 60$ , from which it is clear that the value is high for irregular regions and small for regular regions. The variation for both regions was calculated for (a) and (b) and the values were:  $v_R = 101.2$  and  $v_S = 69235$ .

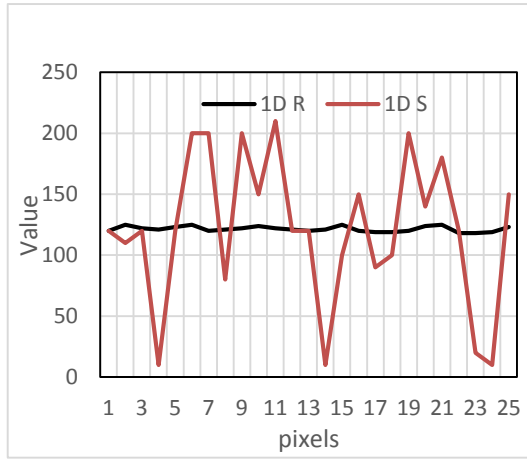
From the example, it is clear that the variation was very high for the irregular region and very small relatively for the formal region. The above example can be represented graphically by converting the region into a one-dimensional array and drawing the values against the index as shown in Figure 4-18.



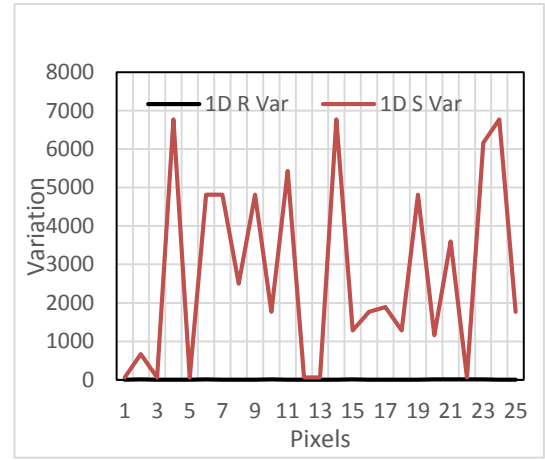
(a)



(b)



(c)



(d)

Figure 4-18 one dimensional representation of (a) R, (b) S, (c) R and S, (d) 1D  $|J(x,y) - \mu_R| \times \sigma_R$  of (a) and  $|J(x,y) - \mu_S| \times \sigma_S$  of (b).

Figure 4-19 shows the value of the original and intensity (grey) images, the mean image and the difference between the intensity image and the mean image. The Cartesian space representation shows that the values corresponding to the salient objects are higher than other values.



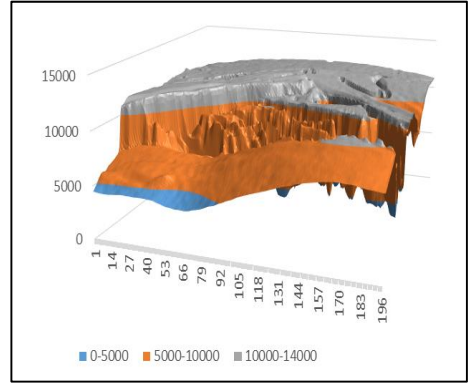
(a)



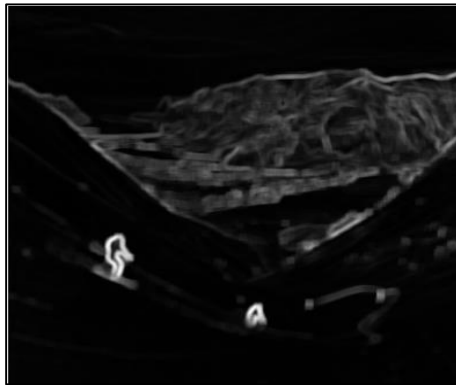
(b)



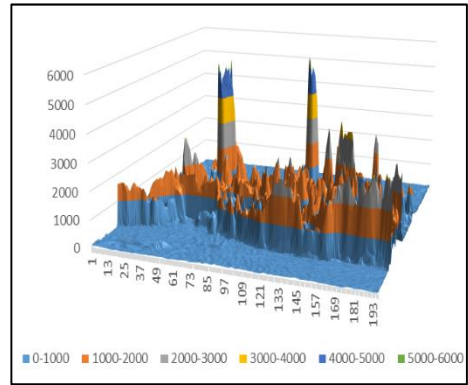
(c)



(d)



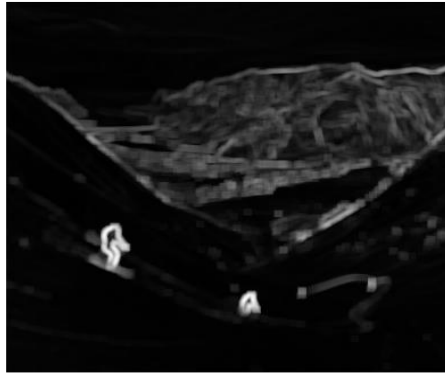
(e)



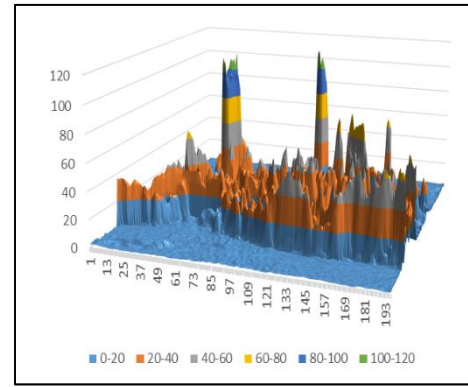
(f)

Figure 4-19: Image and space representation, (a) original image, (b) intensity image, (c) mean image, (d) Cartesian representation of the mean image, (e) difference between intensity and mean images, (f) Cartesian representation of the difference image.

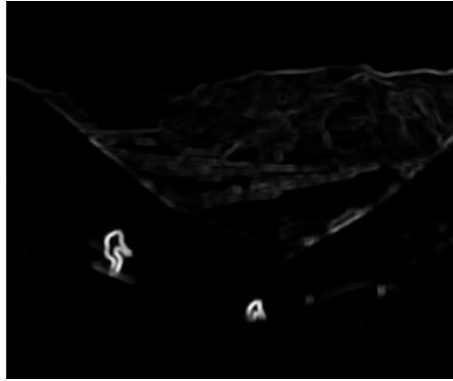
Figure 4-20 shows the results of applying the above-described method on an image. It is clear from the figure that the values corresponding to the salient objects have become even higher and more distinguishable.



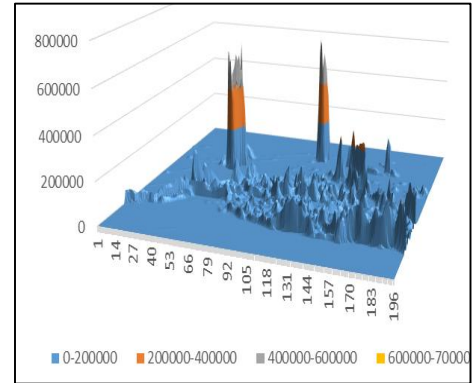
(a)



(b)



(c)



(d)

Figure 4-20: The results of applying the irregularity measures on an image, (a) standard deviation image, (b) Cartesian representation of the standard deviation image, (c) variation values, (d) Cartesian representation for the variation.

Figure 4-21 shows examples of applying the irregularity measures on images from the dataset with different sizes. The window size was selected  $0.02W \times 0.02H$  with overlapping of  $0.01W$  horizontally and  $0.1H$  vertically between the windows.

#### 4.7.2.1 Window Size Selection

The selection of the window (sub-image) size is an important issue since it dramatically affects the results. The size is very much an application-oriented issue, since different applications may need different window sizes. Most researchers select the window size manually based on the application at hand such as in [34].

Kim et al. have suggested using a large window in the centre of the image [25]. They argue that the window should be large, located near the centre, and contain significant colour and texture characteristics. This selection was suitable for their application in which they suggest a method to extract the salient object from the centre of the image.

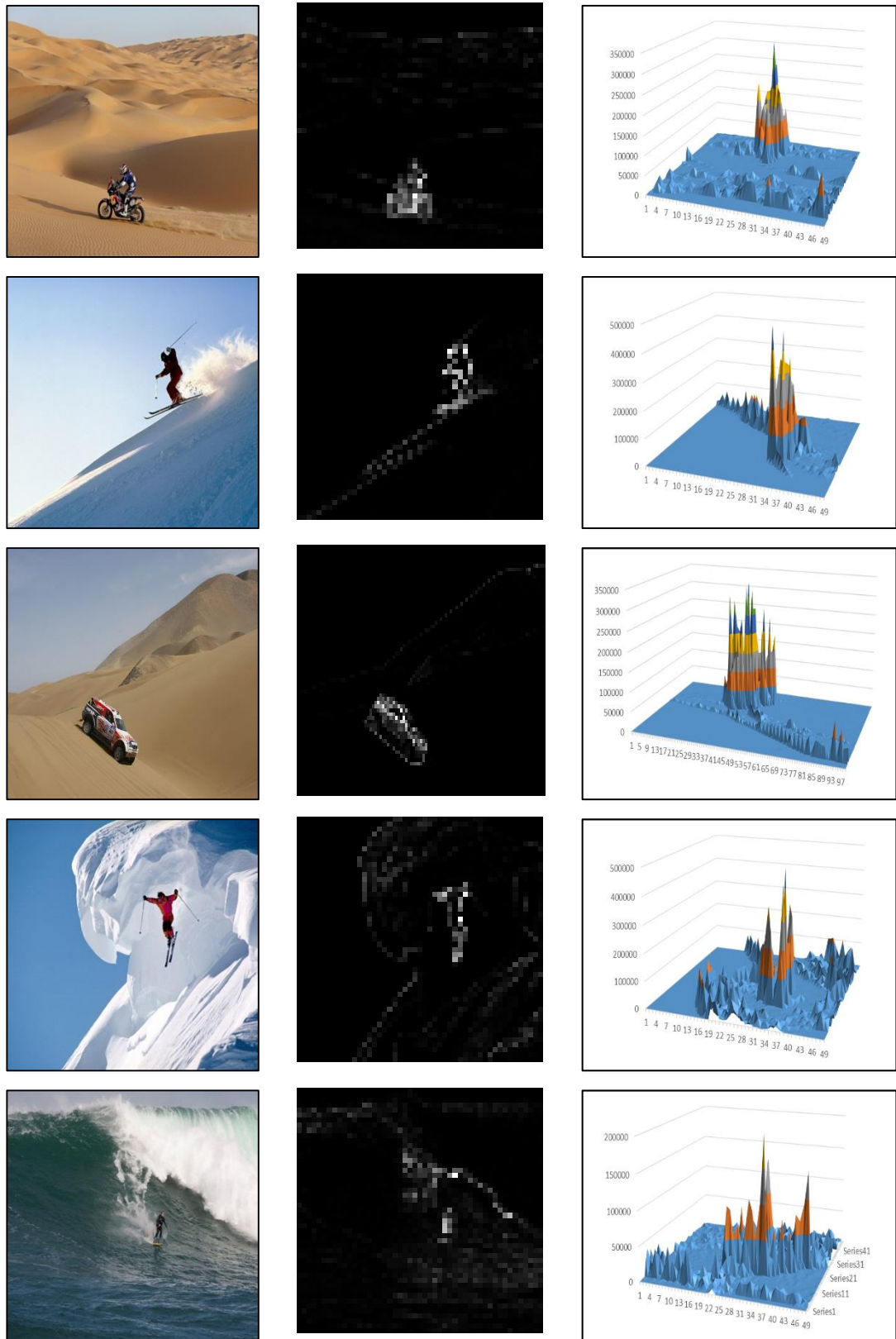


Figure 4-21: Examples of applying the irregularity on images to extract the salient object.

In their efforts to develop a template-based segmentation algorithm using salient points, Zhang et al. have used a window size of  $16 \times 16$ . The image size they used was  $1024 \times 768$ . In finding saliency in noisy images, Kim et al. used  $7 \times 7$  local windows [157]. Banerjee and Kundu use  $5 \times 5$  and  $3 \times 3$  in their application of edge-based feature in CBIR [180]; also in CBIR, an  $8 \times 8$  windows was used in localizing CBIR by Rahmani et al. [181]. Many other researchers have used different window sizes based on the application for which they need the windows

In general, one can conclude that there is no certain limitation to the selection of the window size other than the application at hand. Small windows, such as  $5 \times 5$  and  $3 \times 3$ , shall highlight fine details and they are useful in applications such as edge detection, noise reduction, and image smoothing, while large size windows may be suitable for highlighting large details such as objects. Without any pre-knowledge of the nature of the objects, one cannot specify which size of window is the best to use. Thus, and based on the above discussion, one can select optimum window size based on the application, and the image size. In our case, selecting a small size window will highlight textured regions as salient regions, which is not the desired result, since we are trying to avoid regular textured regions.

Although in most publications the reason behind using a specific sub-image size remains largely unclear, we have tried to find a suitable value for most of the image types that we are going to use. The images were selected for their good diversity in the nature of the objects in that image. Some images have large single objects, and some other images have multiple objects, both small and large. Texture was another feature we considered in our dataset collection for specifying the size of the sub-image empirically. Images with high, medium, and low texture have also been selected for testing.

Different sub-image (window) sizes have been tested, with constant overlapping between the sub-images for each image. For every size the salient regions are extracted and compared with the manually extracted ground truth regions using the XOR method described above. The test was performed on images of different sizes and with different saliency difficulty levels.

The error for each sub-image size is calculated for each image and then the average of these error values is calculated and plotted against the window size to find the size corresponding to the minimum error.

Figure 4-22 (a) shows the relation between the sub-image size and the distance measured using XOR, with overlapping between the windows with a value of one. From the figure, it is clear that the minimum distance was at 2% ( $0.02W \times 0.02H$ ), 2.5% ( $0.025W \times 0.025H$ ) and 3% ( $0.03W \times 0.03H$ ). Thus, the best selection for the window size should be at one of these values. Since the computation increases with the size of the window, then the best selection for the sub-image size is ( $0.02W \times 0.02H$ ).

### ***The effect of overlapping between sub-images***

The overlapping between adjacent sub-images (windows) is another important issue which needs to be taken care of, since it affects the accuracy and the implementation time. Thus, another test was performed to measure this effect. In this test, the size of the window was set to a fixed value and the overlapping between the windows was set to different values, i.e.  $\delta_w \in \{1, 2, \dots, w\}$  and  $\delta_h \in \{1, 2, \dots, h\}$ , where  $w$  and  $h$  are the width and the height of the window respectively, and  $\delta_w$  and  $\delta_h$  are the overlapping between the adjacent windows horizontally and vertically respectively. The same set of images that was used in the previous test is used here.

The XOR was calculated to all the images in the dataset with fixed window size of ( $0.02W \times 0.02H$ ) and different overlapping values. Figure 4-22 shows the relationship between the overlapping value and the XOR. From this figure the best value of the overlapping is at  $0.01H$ ,  $0.01W$ . When the overlapping value is  $0.05H$ , more details will be extracted or identified as interest points. If the overlapping between windows is large, then some details may be lost and some are not identified. Thus, and based on the curve given in Figure 4-22 (b), the best value for windows overlapping is at 50% of the window size itself, i.e. half of the window size is common between each adjacent window.

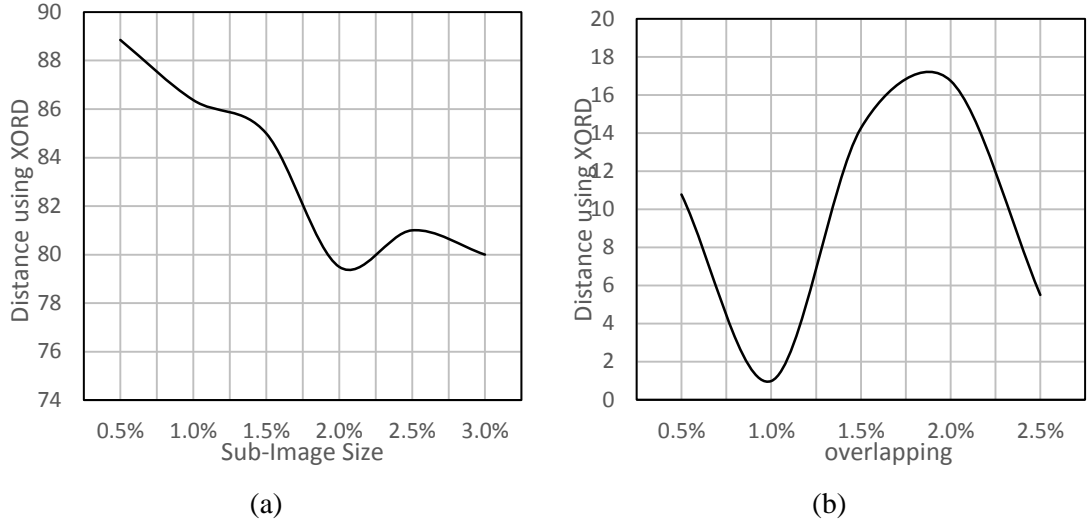


Figure 4-22: Average distance measured using XOR distance, (a) vs. sub-image size, (b) vs. overlapping.

#### 4.7.2.2 Multiple Scales

Some algorithms use hierarchical representation, which means extracting the saliency in different resolution levels, or different image scales, such as in [2], [132] [133] [134] [135] [136], [182], [120], [183] and others, while some other algorithms operate on one level, or single scale, such as in [153], [184], [185], [186], and others.

In our research, we shall adopt single scale of the image to extract the salient points to avoid unnecessary calculations, by reducing the size of the image and then tracking the salient points in upper levels. In order to get more information on the sub-image we increase the window size and keep the image size fixed.

#### 4.7.2.3 Unimportant Details Suppression and Thresholding

The main proposition of pre-attention is that attention may be captured by salient pop-out objects or regions; in this case, extracting the salient object means removing details of less importance. Based on the local descriptor there, regions are divided into different classes as given in the following equation:

$$\mathbb{R} = \mathbb{R}_I \cup \mathbb{R}_U \quad 4-55$$

where  $\mathbb{R}_I$  is the set of all important regions and  $\mathbb{R}_U$  is the set of all unimportant regions.

To separate these two regions, thresholding can be used, in which one should select a suitable threshold value to be used to isolate the important from the unimportant regions.



In thresholding, a suitable value should be extracted from the available data to isolate the important regions from the unimportant ones. This thresholding value should be selected carefully since it plays an important role in discriminating the important regions from the unimportant. Small thresholding values may lead to consideration of some unimportant regions as important and vice versa. Regions with description values larger than, or equal to, a predefined threshold value  $T$ , are considered as important regions and other regions will be considered as unimportant, as shown in the following equation:

$$\begin{aligned}\mathbb{R}_I &= \{r_{ij} \mid d_{ij}^* \geq T\} \\ \mathbb{R}_U &= \{r_{ij} \mid d_{ij}^* < T\}\end{aligned}\tag{4-56}$$

The value of  $T$  can be derived from the available data in the description set. Many techniques can be used to extract the value of  $T$ . One possible technique is by the use of histogram thresholding as histogram gives good information about the distribution of the grey levels over pixels.

Since most of the algorithms assume the input image is a dark object in a light background, or vice versa, thus the histogram is expected to have two peaks and one valley. The grey level corresponding to the valley, which represents the minimum point between the two peaks, is considered as the threshold value. This thresholding technique is known as bimodal histogram thresholding (BMHT).

Table 4-7 gives the normalized variation values distribution of the image given in Figure 4-20 (c). In this table, one can see that most of the values fall in the dark region, which represents the unimportant part of the image, while there are very fewer values in the lighter region or the important (salient) part of the image. This is an expected result as we predict the salient points or the salient regions will be less than the uniform regions. From the table one may notice that the valley of the histogram falls into the value of 70, which means this value can be used as the threshold.

The main limitation of this technique is that the borders between the objects and the background have been assumed to be well-defined, which due to the varied nature of images is not always correct. The definition of the borders or the edges is not always crisply defined due to the uncertainty in the image which might be because of shadow or fading in the image. Therefore, Fuzzy principles should be applied here as shown below.

Table 4-7: The distribution of the normalized variation measures for the image given in Figure 4-20 (c)

Bin	0	10	20	30	40	50	60	70	80	90	100
Frequency	0	2334	40	10	5	3	2	1	3	0	3

#### ***Fuzzy-Based thresholding techniques (FBT)***

Fuzzy-Based thresholding techniques have been used in various literatures. Cheng et al. [187] proposed using the c-partition entropy fuzzy to select the threshold value. Maximum entropy principle and fuzzy c-partition method have been used to select the threshold value(s) associated with the maximum entropy of the fuzzy c-partition. They assumed a fixed and known number of clusters and classified the pixels into these clusters. Entropic thresholding was also used by Wang et al. [188] also Tao, Tian, and Liu [189] and others. Yong, Chongxun, & Pan [190], and Junwei et al. [191], used c-mean with fuzzy logic as a clustering algorithm in their papers published in 2004 and 2007 respectively. Tizhoosh suggested the use of ultrafuzzy (fuzzy II) algorithms to find the threshold value, with which he improves the resultant clustered image [192]. Prasad et al. suggested a fast unsupervised thresholding algorithm based on  $\pi$  membership function [193]. In their method, Prasad et al. considered a fuzzy  $\pi$ -function to transform the fuzzy intensities into normally-distributed intensities. The intensities of an image are transformed to an interval  $[0, 1]$  by  $\pi$ -function in terms of a standard S-function. The values of the function represent the degrees of the closeness in terms of intensities. The function is, therefore, used to locate the intensities of object and background. The selection of a crossover point which is the arithmetic mean of the image could be viewed as an object-background classification problem.

#### ***Fuzzy-Based Bimodal thresholding techniques (FBMT)***

Al-Azawi and Ibrahim have applied the principle of membership function of the fuzzy logic in the bimodal histogram thresholding technique [194]. The method was a modification to the bimodal thresholding technique. Figure 4-23 (a) shows the traditional bimodal thresholding technique, in which the value of T was selected as the grey level value corresponding to the valley of the histogram. The main problem with the method given in (a) is that it assumes that the borders between the object and its background are well-defined and crisp, which is not the case in real images. (b) in the same figure shows the use of the Fuzzy membership function in extracting the value of T.

In the proposed thresholding technique, the histogram is represented by the set  $\mathcal{H}$ , which is given by:

$$\mathcal{H} = \{x_i \mid i = 0, 1, \dots, n - 1\} \quad 4-57$$

where  $x_i$  is the histogram value corresponding to the grey level  $i$ , and  $n$  is the number of grey levels in the image.

As we want to extract the object from the background then the object can be defined as a subset from  $\mathcal{H}$  as follows:

$$\begin{aligned} \mathcal{O} &= \{ \langle x, \mu_o(x) \rangle, x \in \mathcal{H} \} \\ \mathcal{O}^T &= \{x_i \mid \mu_o(x) \geq T\} \\ \mathcal{O} &\subseteq \mathcal{H} \end{aligned} \quad 4-58$$

where  $\mathcal{O}$  is the fuzzy set of grey levels for the object,  $\mu_o(x)$  is the membership function which indicates how the grey level belongs to the object and  $\mathcal{O}^T$  is a crisp set that contains pixels belonging to the object.

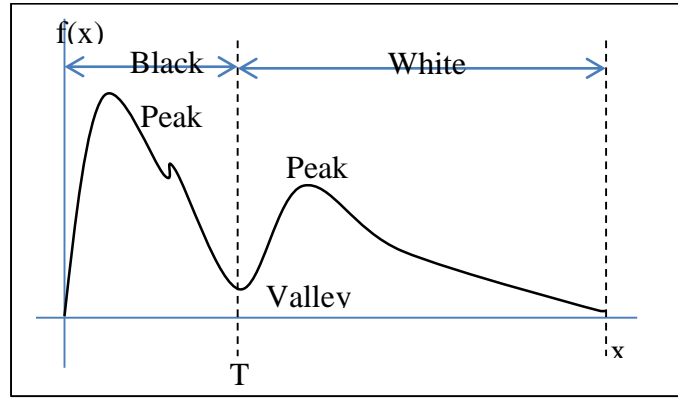
In the same way, we shall define the background fuzzy set as follows:

$$\begin{aligned} \mathcal{B} &= \{ \langle x, \mu_B(x) \rangle, x \in \mathcal{H} \} \\ \mathcal{B}^T &= \{x_i \mid \mu_B(x) < T\} \\ \mathcal{B} &\subseteq \mathcal{H} \end{aligned} \quad 4-59$$

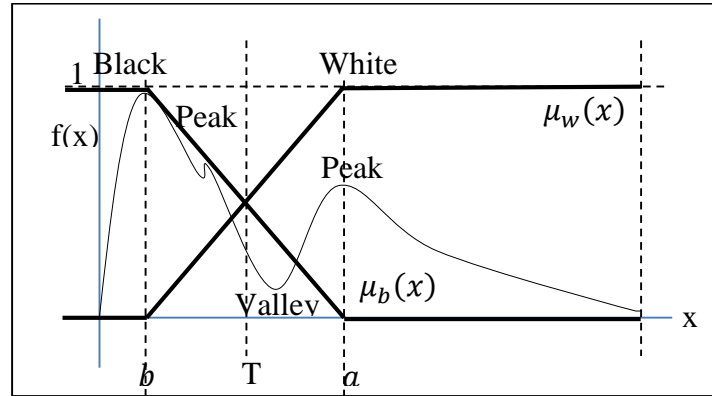
where  $\mathcal{B}$  is the fuzzy set of grey levels for the background,  $\mu_B(x)$  is the membership function which indicates how the grey level belongs to the background and  $\mathcal{B}^T$  is a crisp set that contains pixels belonging to the background.

$$\mu_B(x) = 1 - \mu_o(x) \quad 4-60$$

Both fuzzy sets are normal since the height of each of them is unity, i.e.  $h(\mu_o) = h(\mu_B) = 1$ , where  $h(\cdot)$  is the height of a fuzzy set, which is the largest membership value attained by any point. As the object might be a white region in a black background or vice versa, we shall use the white and black membership functions instead of object and background membership functions.



(a)



(b)

Figure 4-23: Bimodal thresholding technique, (a) normal technique, (b) the application of fuzzy membership function in this technique.

Based on the discussion above, the two membership functions;  $\mu_b(x)$  which measures how much the pixel is black, and  $\mu_w(x)$  which measures the white membership value shall be used. Thus, the luminance space will be divided into two subsets B and W. According to this figure, the histogram is divided into three regions rather than two regions, as in crisp thresholding. One of the regions is the black region and the second one is the white region. The third region, which is the overlapping between the black and white regions, has been generated due to the uncertainty in image luminance values and it may represent fading or shadow. The value of  $T$  can be found in different ways, such as, the intersection of the two lines of the membership functions, known as the crossover point, by calculating the entropy of the regions, and using CI techniques such as Genetic Algorithms and Neural Networks. The membership functions can be derived from the figure as follows:

$$\mu_w(x) = \begin{cases} 1 & x \geq a \\ \frac{x-b}{a-b} & b < x < a \\ 0 & x \leq b \end{cases} \quad 4-61$$

$$\mu_b(x) = \begin{cases} 1 & x \leq b \\ \frac{a-x}{a-b} & a < x < b \\ 0 & x \geq a \end{cases} \quad 4-62$$

The crossover of the two membership functions is given by  $x^* = \frac{a+b}{2}$ . This value can be used as the threshold value  $T$ .

By using the  $\alpha$ -cut definitions, the following definition can be presented:

**Definition:** The white region set  $R_w$  is the set that contains all regions with grey level  $x \in W$  and  $\mu_w(x) \geq \alpha_w$  with  $\alpha_w = \mu_w(T)$  and the set of the black regions  $R_B$  is the set that contains all regions with grey level  $x \in B$  and  $\mu_b(x) > \alpha_b$  with  $\alpha_b = \mu_b(T)$ .

This theory can be modified to be used with saliency thresholding by considering two membership functions. Here we shall define the linguistics set  $L = \{NS, LS, S, VS\}$  that can be used to describe the level of saliency of the region based on the Fuzzy thresholding process. The ranges and the meaning of each linguistic variable are given in Table 4-8.

Table 4-8: Fuzzy linguistic variables meanings and ranges.

FLV	Meaning	Range
NS	Non Salient Regions	$0 \leq x < b$
LS	Less Salient Regions	$b \leq x < T$
S	Salient Region	$T \leq x < a$
VS	Very Salient Regions	$a \leq x \leq 100$

### **Empirical Example**

The following is an example of applying the above-described method on a sample image. The image is of size  $400 \times 400$ , thus windows size of  $8 \times 8$  with step size of 4 has been selected. Figure 4-24 shows the results of applying the saliency enhancement technique on the image given in (a). The obtained image in (b) still needs to be thresholded to isolate the important regions from the unimportant ones.



Figure 4-24: The application of saliency enhancement, (a) original image, (b) the saliency enhanced image ISM.

The values of the histogram of the ISM are given in the Table 4-9 which shows the bins, the frequency of occurrence (F), the log scaled frequency (LF), the normalized log scaled frequency (NLF) and the smoothed NLF (SNLF).

Table 4-9: The distribution of the normalized variation measures for the image given in Figure 4-24

Bin	F	LF	NLF	SNLF
0	100	2	0.479501	0.545931
5	678	2.83123	0.678789	0.570505
10	203	2.307496	0.553224	0.551599
15	58	1.763428	0.422783	0.447727
20	34	1.531479	0.367173	0.39135
25	40	1.60206	0.384095	0.35074
30	18	1.255273	0.300952	0.304609
35	9	0.954243	0.22878	0.224692
40	4	0.60206	0.144344	0.162505
45	3	0.477121	0.11439	0.124375
50	3	0.477121	0.11439	0.11439
55	3	0.477121	0.11439	0.148432
60	8	0.90309	0.216516	0.186562
65	9	0.954243	0.22878	0.196547
70	4	0.60206	0.144344	0.17249
75	4	0.60206	0.144344	0.13436
80	3	0.477121	0.11439	0.13436
85	4	0.60206	0.144344	0.110302
90	2	0.30103	0.072172	0.072172
95	1	0	0	0.024057
100	1	0	0	0

Figure 4-25 shows the graphical representation for the histogram data given in the above table. The range of the data is very large and since we are interested more in which frequency is higher, we shall reduce the difference in range by drawing the histogram in Log scale.

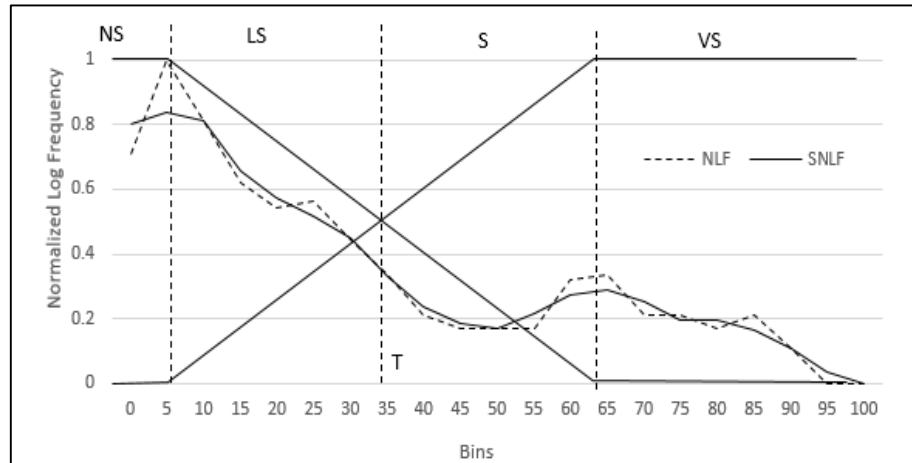


Figure 4-25: Fuzzy membership function representation of the histogram.

Table 4-10 shows the threshold ranges. For salient regions extraction we shall consider the S and VS regions and suppress the NS and LS regions.

Table 4-10: The ranges of the salient regions.

FLV	Meaning	Range
NS	Not Salient Regions	$0 \leq x < 5$
LS	Less Salient Regions	$5 \leq x < 35$
S	Salient Regions	$35 \leq x < 65$
VS	Very Salient Regions	$65 \leq x \leq 100$

By using the threshold values in Table 4-10, which were extracted from the histogram given in Figure 4-25, the results in Figure 4-26 were obtained. From the figure, (a) shows the non-salient regions (white) which are in the range of (0 to 5). It is clear that these regions contain most of the redundant information such as the sky and the snow. (b) shows the less salient regions, (c) shows the salient regions and (d) shows the very salient regions.

Since saliency is not crisp and fuzzy, as explained earlier, then the user can select the level of saliency he/she wants to use. Selecting very salient regions may result in highlighting only those parts of the objects that are very salient, or only areas where the values of irregularity measures are very high. On the other hand, if we select the non-

salient regions, the unimportant details may be included in the resultant saliency map. Therefore, the best threshold choice is made by considering both salient and very salient regions.

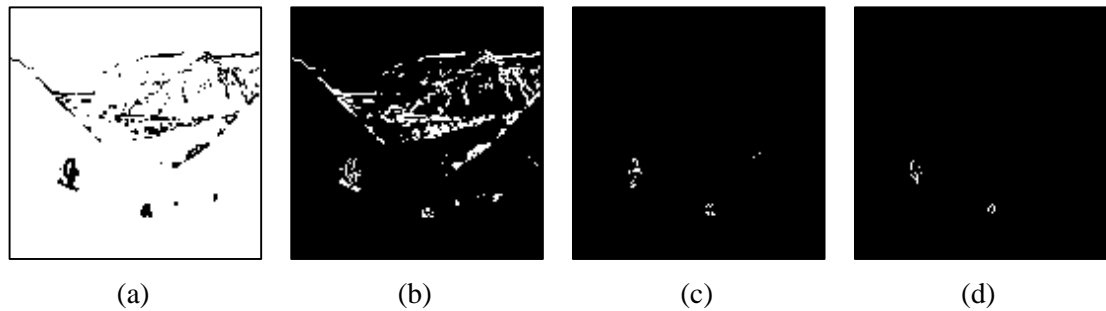


Figure 4-26: Applying Fuzzy bimodal histogram thresholding, (a) NS region, (b) LS regions, (c) S regions, (d) VS regions.

With the threshold obtained above, the results of applying the algorithm are as shown in Figure 4-27.

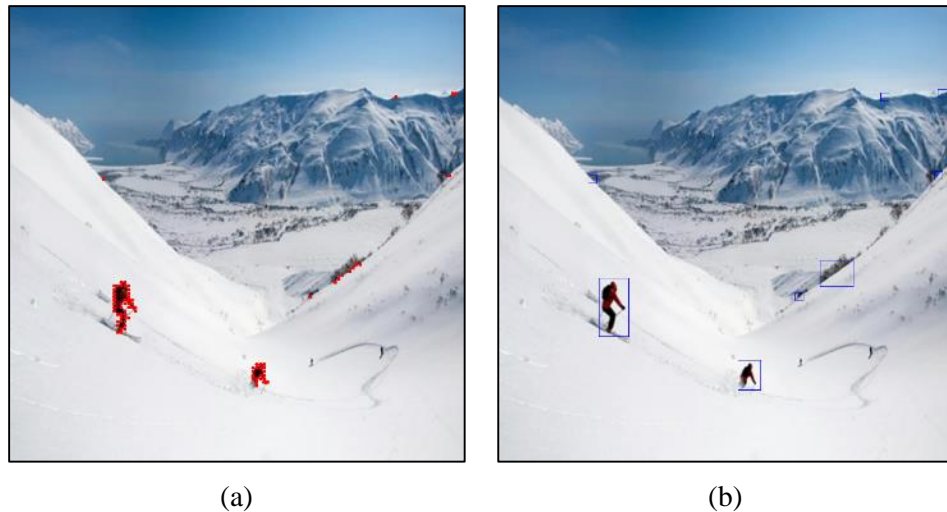


Figure 4-27: Applying the thresholding technique on the saliency enhanced image, (a) important regions, (b) important object.

In some images, the object itself may contain regular regions. This may cause the algorithm to highlight the borders of the object as shown in Figure 4-28.

Figure 4-28 shows an example of an object containing a regular region inside. The obtained borders will form the salient region in the points-clustering and regions-forming phases. Figure 4-28 (b) and (c) show the process of points clustering and merging using the iterative blobs-based clustering technique (IBBC), which will be discussed later, and (d) shows the extracted objects.



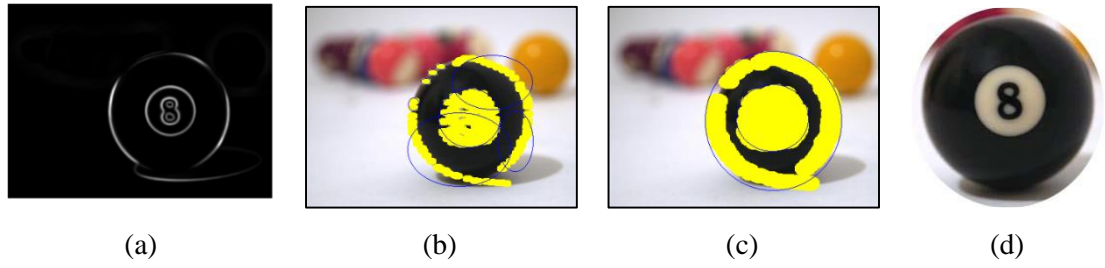


Figure 4-28: Illustration of a case where there is a regular region inside the object, (a) irregularity map, (b) points clustering and merging using IBBC after 4 iterations, (c) same as (b) but when IBBC stops after 8 iterations, and (d) the extracted objects.

In cases like the one described above, the algorithm may look like an edge detector while it is not, because edge detectors try to find thin edges and to cover all the objects in the image, while in irregularity maps we extract the salient and irregular regions only. Figure 4-29 shows the difference between the edge map obtained from popular edge detectors such as Canny and Sobel, and the map obtained from the local irregularity saliency extraction algorithm. Some other edge detectors were tested and it was found that they too were different from the local irregularity map for the same reasons as above.

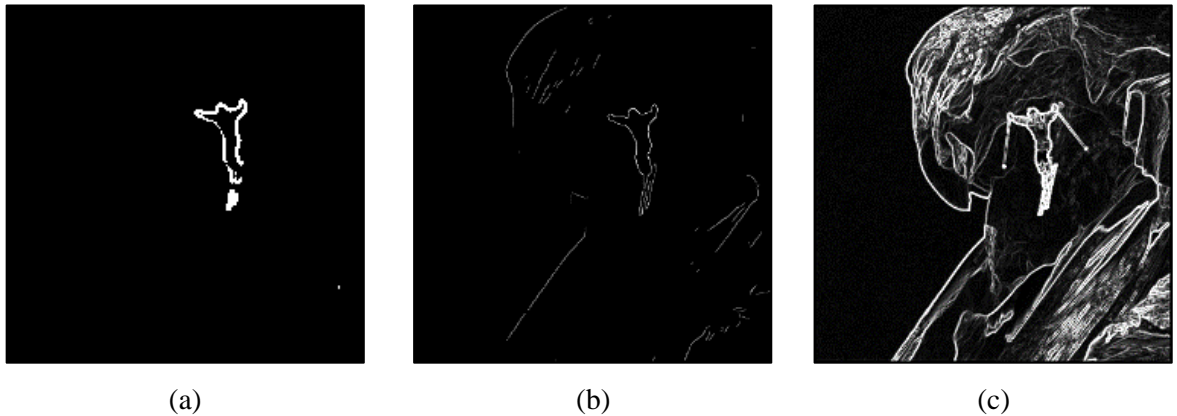


Figure 4-29: Comparison with edge detector, (a) irregularity map, (b) edge map obtained by Canny edge detector, (c) edge map obtained from Sobel edge detector.

#### 4.7.2.4 Results

This section contains the experimental results from applying the LSI algorithm. Table 4-11 shows the results obtained from applying the LSI algorithm. This figure shows the main steps of the LSI. The class column gives the class of the image as it was classified in data collection. The saliency map shows the saliency maps obtained from the image after thresholding. The last two columns show the salient region obtained from clustering the points using the IBBC technique, and the salient object which is extracted by taking the borders of the blobs.

#### 4.7.2.5 Benchmarking

The LSI algorithm was used to extract the salient regions, and the MSRA dataset has been used at this stage since it was adopted by most of the saliency identification algorithms. A comparison with different state-of-the-art saliency extraction techniques was carried out. The evaluation was performed by comparing the obtained saliency map with the ground truth (GT) maps. The similarity with the ground truth data was calculated for each image in the dataset and then the average was extracted for each algorithm and compared to other methods.

Table 4-11: Experimental results obtained from LSI.


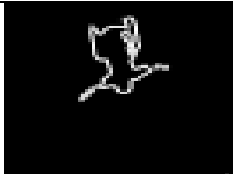















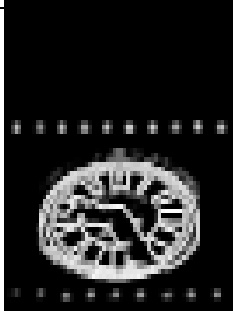


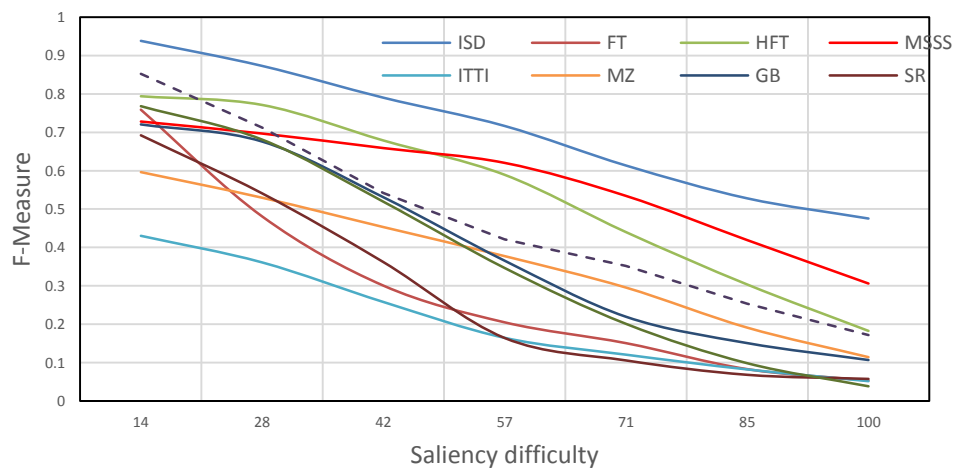
#	Class	Image	Saliency Map	Salient Region	Salient Object
1	C2				
2	C2				
3	C2				
4	C2				
5	C3				

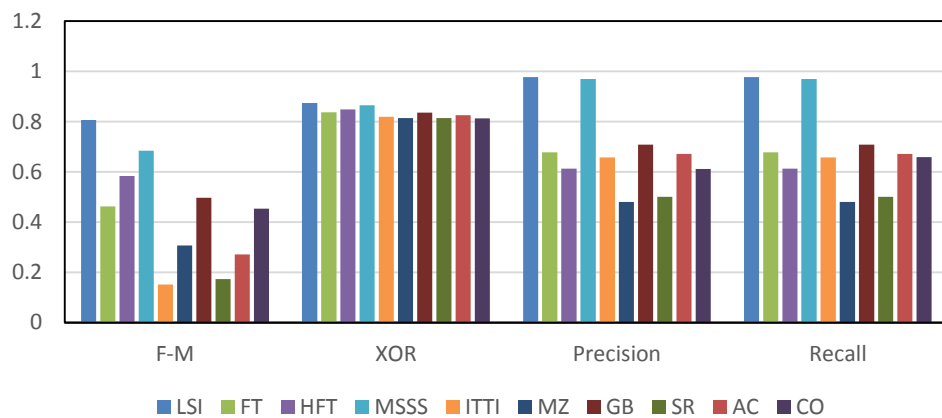
Figure 4-30 (a) shows the F-Measures curves obtained from comparing the proposed method with IT [2], MZ [117], GB [118], SR [119], AC [120], FT [116], HFT [113], MSSS [121], and CO [124] [147]. The F-measure was calculated for each method using Equations (4-5) and (4-6).

From the F-measure curves shown in Figure 4-35 (a), it is clear that the F-measure curve corresponding LSI is higher than other methods, which means that the results that we have obtained are quite satisfactory. The minimum curve is the one corresponding to Itti method; this is because Itti is more concerned with the sequence of extraction.

Figure 4-30 (b) shows the benchmarking of LSI with other algorithms using the adopted evaluation measures.



(a)



(b)

Figure 4-30: LSI benchmarking with state-of-the-art algorithms, (a) F-Measure, (b) different measures comparison.

The same comparison was performed on images from different classes and the obtained results are as follows:

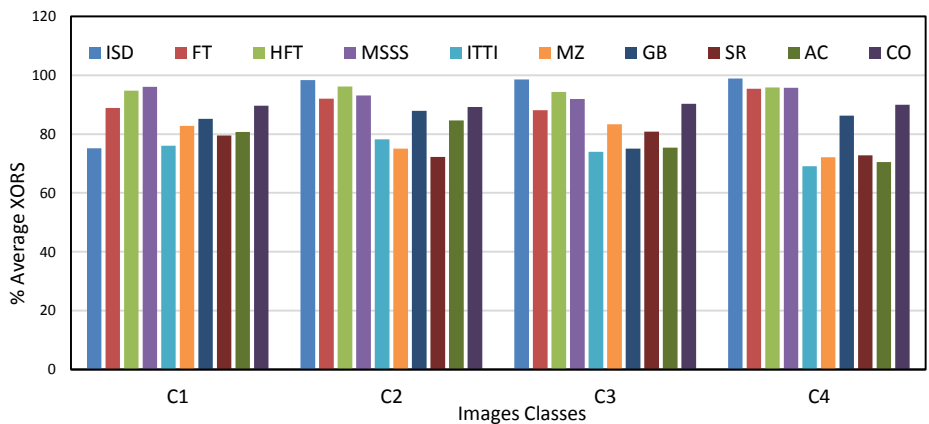

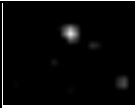


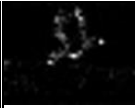








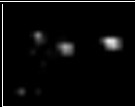
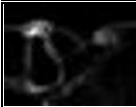










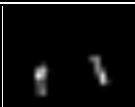





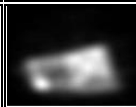

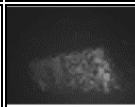






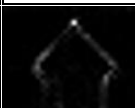


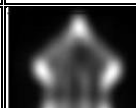







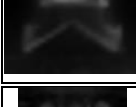



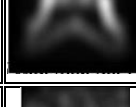







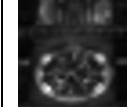





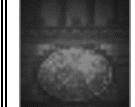

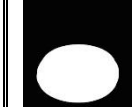



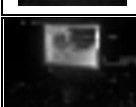
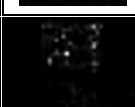


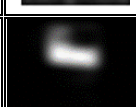





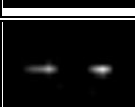

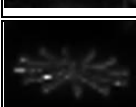










Figure 4-31: XORS measure for different saliency extraction methods and for different images classes.


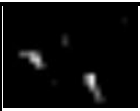

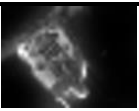
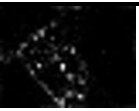




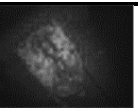




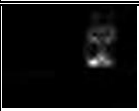

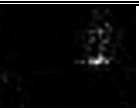


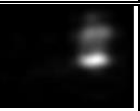





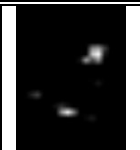








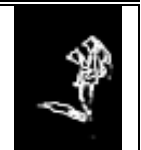


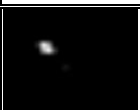
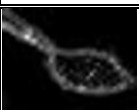
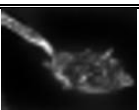



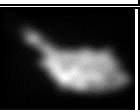







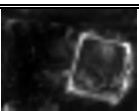
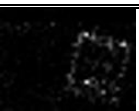




















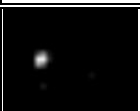

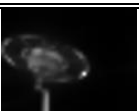











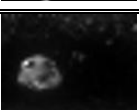
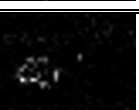
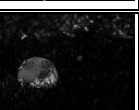

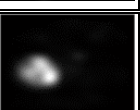

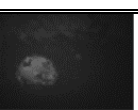


Figure 4-31 shows the XOR similarity XORS rather than distance XORD to show the efficiency of the proposed algorithm as compared to other algorithms.

From Figure 4-31, it is clear that ISD is not the best algorithm when applied on C1, whilst it has given excellent results when applied on the other three classes, as discussed earlier. In addition, and for further illustration, a qualitative comparison has been performed with the aforementioned methods and is given in Table 4-12.

Table 4-12: Saliency maps obtained from different algorithms.

#	Class	Image	ITTI	MZ	GB	SR	AC	FT	HFT	MSSS	CO	ISD	GT
1	C2												
2	C2												
3	C2												
4	C2												
5	C3												
6	C2												
7	C2												
8	C4												



9	C2												
10	C2												
11	C2												
12	C2												
13	C3												
14	C4												
15	C2												
16	C3												

### 4.7.3 Global Saliency Identification (GSI)

By applying the irregularity principle on the entire image, it was possible to improve the saliency detection. The resultant saliency enhanced image  $G(x, y)$  can be obtained from subtracting the mean of the image from every pixel and multiplying it by the standard deviation as shown below:

$$G(x, y) = \frac{|I(x, y) - \mu| \times \sigma}{\max_{x,y}(G(x, y))} \times 255 \quad 4-63$$

In equation 4-63, every pixel value in  $G(x, y)$  is divided by the maximum value in order to normalize the values since different images have different ranges of values.



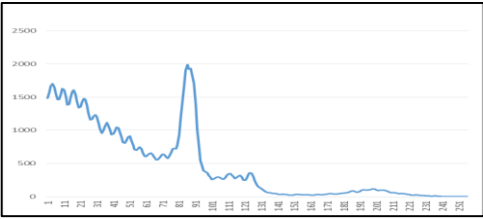

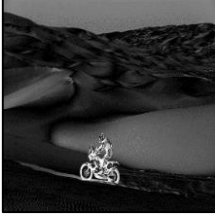
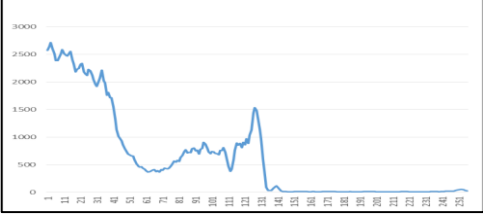


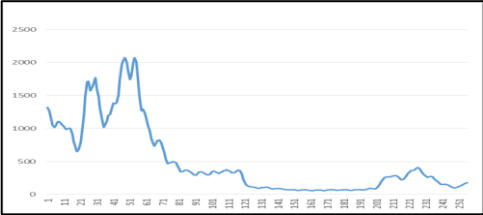


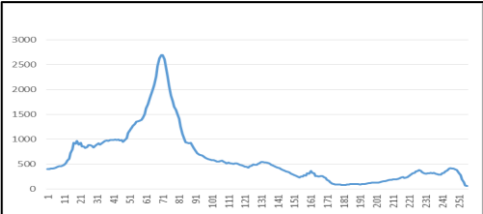

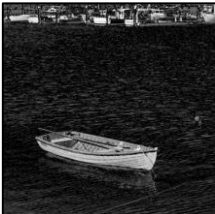
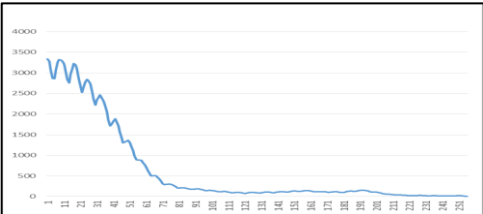





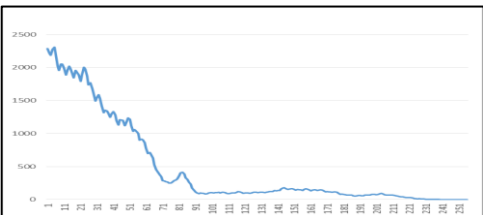
The results of the global saliency enhancement are shown in Table 4-13 which shows the saliency enhanced images and their corresponding histogram. The reason behind extracting the histogram is to show the intensity distribution of the images.

#### ***Thresholding***

From the histogram, it is clear that its dark values are very high as compared to the light values, and usually there is a peak within the dark region. In addition to that, there is a small peak in the light region, which represents the salient region. The histograms of the images given in the above table were smoothed, to overcome the useless ripples without affecting the general shape of the histogram.

From Figure 4-32, it is clear that even after smoothing, the main shape of the histogram remains the same. The white region size of the histogram gives a good impression of the size of the salient region, hence, one can use this fact to find the value of the threshold that can be used to separate the important regions from the unimportant ones. In the histogram given in Figure 4-32, the value of the threshold was found using FBMT described above and it is equal to 135. The result of applying this threshold value on the corresponding image is given in Figure 4-33.

Table 4-13: Global Saliency Enhancement.

Original Image	Global Saliency	Histogram
		
		
		
		
		
		
		



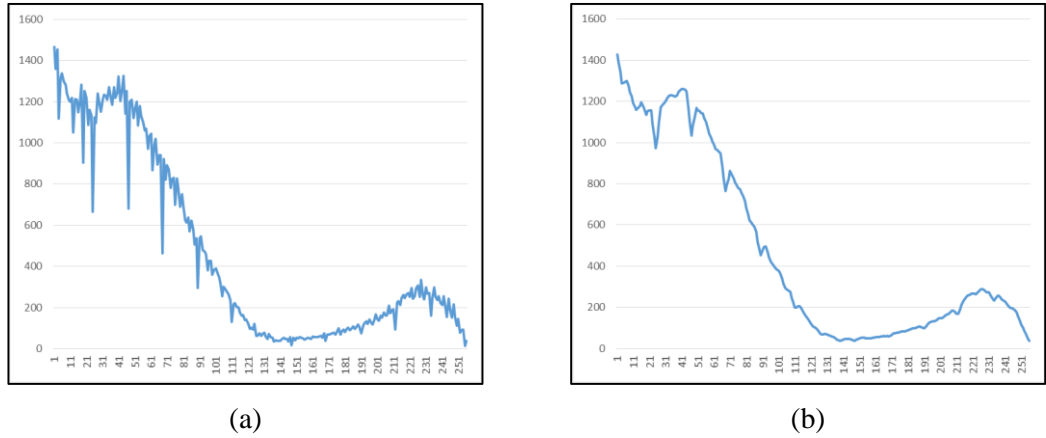


Figure 4-32: Histogram Smoothing, (a) Original Histogram, (b) Smoothed histogram.

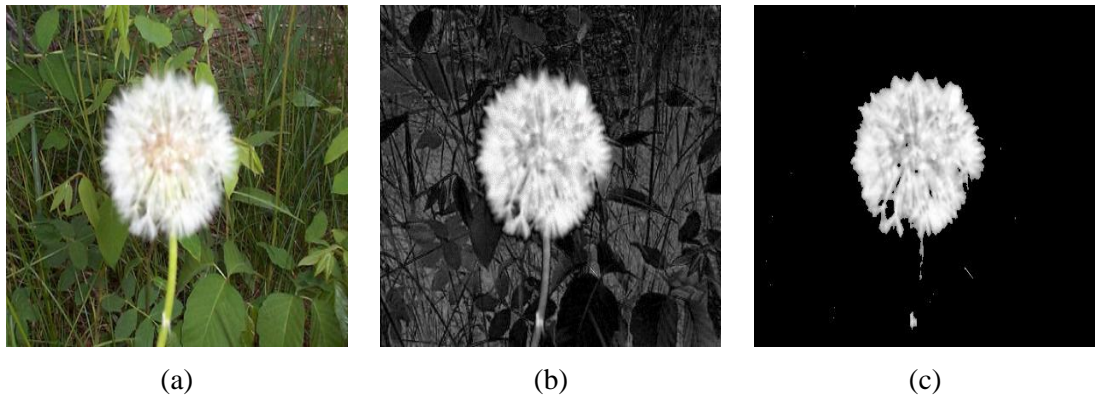


Figure 4-33: Global Salient Region Extraction, (a) original image, (b) saliency enhanced image, (c) the salient object after thresholding.

Table 4-14 shows the result of the images in Table 4-13 after thresholding.




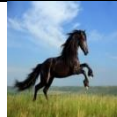


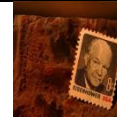

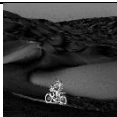







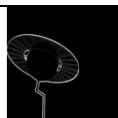
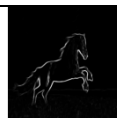
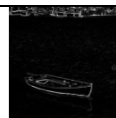
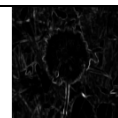

Table 4-14: Global saliency enhancement.

Original Image							
Global Saliency Extraction							
Thresholded Image							

Table 4-15 shows a comparison between the results obtained from applying the saliency enhancement globally and locally. From the table it is obvious that the global

enhancement highlights the salient regions as a whole, while in local it is mostly the borders that are highlighted. In addition, the local enhancement highlights the salient regions in a better way, and the contrast between the salient regions and the background is discriminated in a better way.

Table 4-15: A comparison between global and Local Saliency.

Original Image							
Global Saliency Extraction							
Local Saliency Extraction							

#### 4.7.4 Two Steps Saliency Identification (TSSI)

In this Section, an integrated system which uses both saliency extraction techniques, LSI and GSI, is proposed. The general structure is given in Figure 4-34.

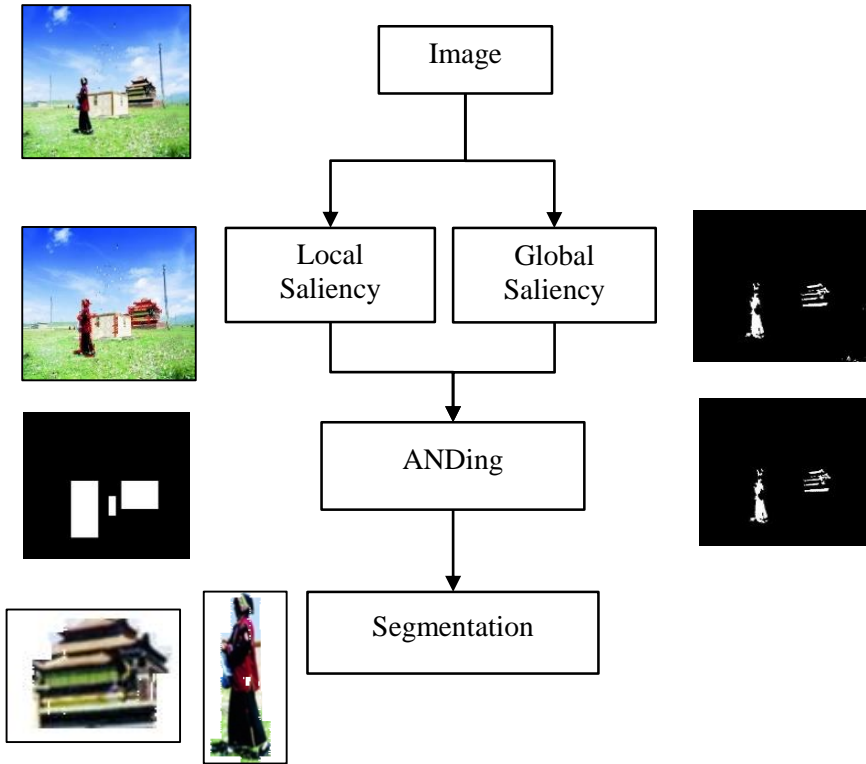


Figure 4-34: TSSI application on an image.

In TSSI, the image undergoes two saliency identification processes; LSI and GSI. In LSI, the saliency in the structure of the object is identified since it uses small sub-images or windows in calculating the variations. On the other hand, the GSI highlights the saliency of the objects in which the contrast in intensity is pronounced. The region is salient if it satisfies both local and global saliency criteria, which means, it should be salient locally and globally, thus we use ANDing operation between them.

#### 4.7.5 Improving the TSSI

To improve the results obtained from TSSI, several refinements can be used; the following subsections include the main enhancements.

##### 4.7.5.1 Falsely detected salient points filtering

Although many thresholding techniques have been tested, and the adopted thresholding techniques have given the best results, still some falsely detected salient points (FDSP) may occur due to the texture or noise. These points may affect the results and cause incorrect regions extraction. Consequently, isolated points filtering is necessary to overcome this problem. A low pass filter can be used for this purpose but it will not work efficiently since we work with binary images. However, thresholding after low pass filtering will only augment the thresholding process, which we want to reduce to a minimum.

The binary majority filter (BMF) can work well to solve the above problem since it will improve the resultant binary saliency image in two ways: firstly, it reduces the effect of FDSP by removing the single or minor points from the region under consideration, and secondly, it improves the object in the region by filling the holes in the region. In BMF, each pixel is replaced by zero or one based on the majority of the surrounding pixels' values. Figure 4-35 shows a numerical example of the BMF. In this figure, the number of white pixels is 3, while the number of black pixels is 6, thus the centre pixel shall be replaced with black no matter if its original colour was white or black.

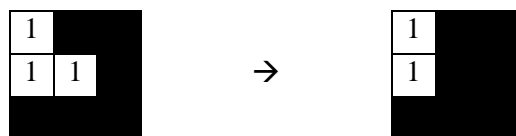


Figure 4-35: Numerical example of applying BMF.

Figure 4-36 shows an example of applying the BMF on an image after applying the thresholding process. From this figureFigure 4-36, it is clear that the noise and the FDSP have been filtered and removed from the image.

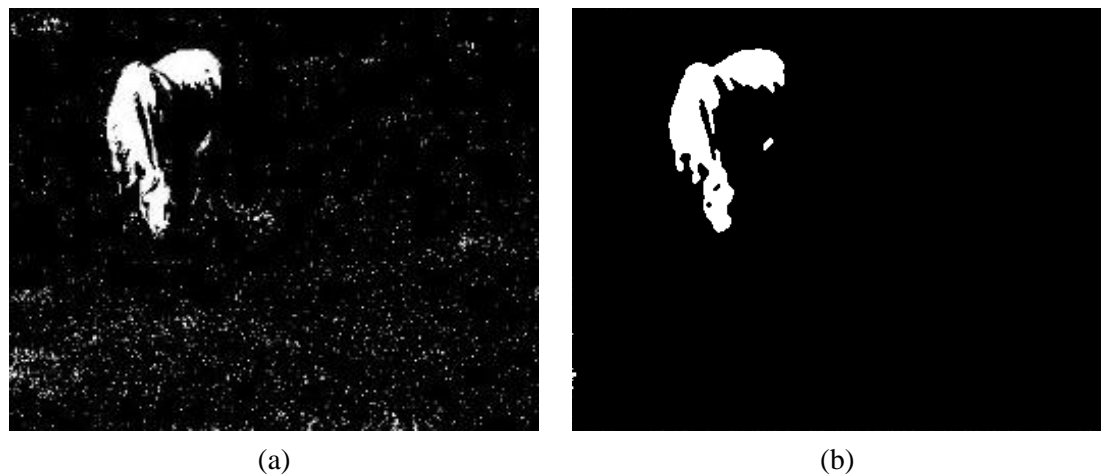


Figure 4-36: Example of applying BMF on real image, (a) binary image with FDSP, (b) the resultant image after applying BMF.










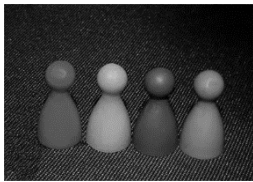
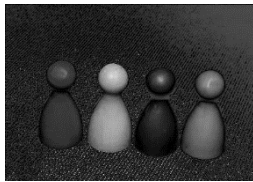
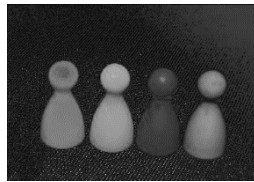




#### 4.7.5.2 Similar intensity problem

One common problem with using the grey image (intensity) rather than colour image is that although it reduces the computation burden, there is inevitably some loss of information. The information is lost due to some different colours having the same intensity, since the intensity is calculated as the average of the three colour bands (red, green, and blue). For example, the colours  $c1 = (100, 50, 0)$  and  $c2 = (0, 0, 75)$  shall have the same intensity since the average of both is the same which is 75. Therefore, the colour contrast may lack representation when using intensity.

In order to overcome this problem, the irregularity enhancement process shall be applied on each colour band separately. This will highlight the irregularity in the colour so that the problem of having the same intensity value for different colours shall be overcome.

Table 4-16 shows examples of such cases. In case 1, the intensity of the flower is with less irregularity than the shadow has, thus the shadow has been highlighted as the salient region. In case 2, the object and the background have close intensity values, therefore, the object could not be recognized using intensity, while it was possible to recognize it using colour. The same phenomenon is visible with the rest of the cases; so using colour instead of intensity can drastically improve the results.

Table 4-16: The effect of colour contrast on improving the results.

	Original image	Intensity Image	Saliency enhanced image using intensity	Saliency enhanced image using colour
1				
2				
3				
4				

#### 4.7.5.3 Object isolation improvement

Another problem raised with TSSI is the problem of extracting or isolating the object from the background. This problem occurred because with GSI, not all the details of the objects are highlighted, hence some details might be considered as not salient. In addition, LSI produces regions with regular shapes such as rectangles or ellipses, while most of the objects have irregular shapes.

Figure 4-37 shows an example of the above-mentioned limitation. The image in (a) in the figure was extracted using LSI. In this method, it is noticed that all the details of the object are maintained but the background is still part of the object. The remaining background affects the features measured in the recognition phase. In (b) in the same figure, the object was extracted using TSSI. From the figure, it is clear that most of the details have been removed but the general structure of the object was retained.

From the above discussion, it was noticed that both LSI and TSSI suffer from either an increase in unimportant details as in LSI, or from a loss of important information from the object as in TSSI.

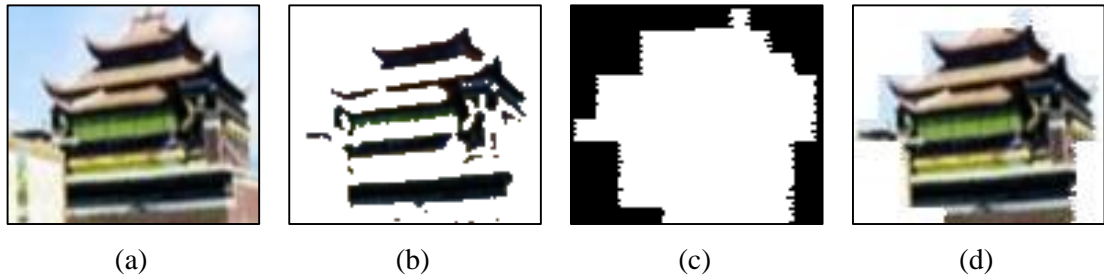


Figure 4-37: Extracting the object from the image, (a) using the region extracted from LSI, (b) using TSSI, (c) mask, (d) using RTSSI.

In order to overcome these limitations or at least reduce their effect, we shall improve the object extraction process by applying holes-filling and blobs-extraction algorithms. In the hole-filling process, if a small black region is surrounded by a white region, then it will be converted to white. The BMF has done part of this task by replacing some pixels with whatever is the majority value of the surrounding pixels.

The second important improvement, by using the blobs extraction, is to extract the object's boundaries and then fill the area using region-growing techniques. The mask shown in Figure 4-37 (c) was obtained using the blobs refinement. This mask is used to identify the object by ANDing the region with the mask. The obtained object after the ANDing process is given in (d).

Figure 4-38 shows the results of applying the three methods discussed above. In this figure, ground truth (GT) image is given on the left, then TSSI, LSI, and the refined two-step saliency identification RTSSI. The histogram mean and standard deviation measures were taken as examples to show the difference of the features when compared to GT.

The distances from the GT image were calculated and it was noticed that the minimum distance was obtained with the RTSSI, which means the suggested refinements have improved the results.







	GT	TSSI	LSI	RTSSI
Image				
mean	152	195	141	156
Stdv.	91	98	81	92
Distance	0	43.5	14.9	4.1

Figure 4-38: Comparison of the three methods of salient objects extraction.

The following graph shows a comparison among the three aforementioned methods of salient objects extraction for both the individual image and for a set of images. From the graphs it is clear that the minimum distance, and hence the best result, shall be obtained with RTSSI.

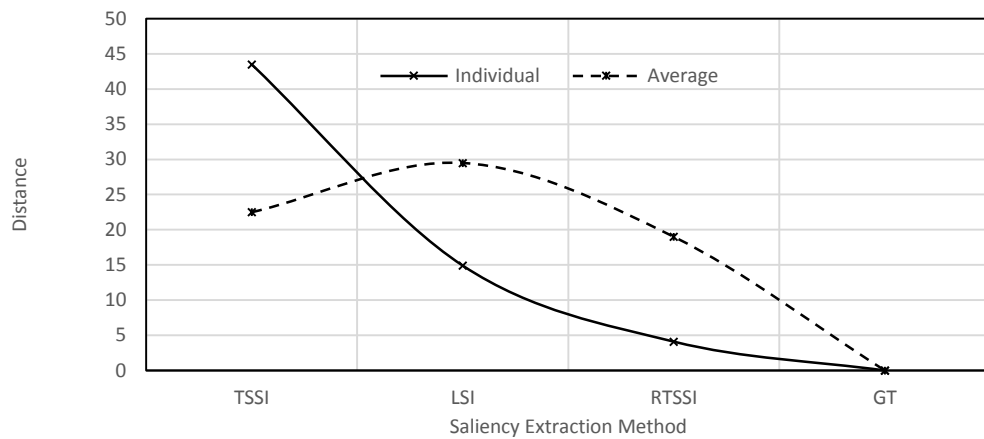


Figure 4-39: The distance between different saliency identification methods and the ground truth data.

#### 4.7.6 Results

In order to study the feasibility and the performance of the proposed algorithm, both qualitative and quantitative benchmarking has been performed, The Figure 4-40 shows the entire RTSSI algorithm processes block diagram step by step. In this figure, the image will go through two phases; local saliency extraction and global saliency extraction. In the local saliency extraction phase, the salient regions that contain the salient object are identified and cropped to remove all unimportant details. The results of this phase are a set of regions that contain a salient object. In the second phase, the global salient objects

are extracted, after which the falsely detected points are removed, and then the object mask is identified. The obtained masks are ANDed together to form the final mask which is used to extract the salient objects from the image.

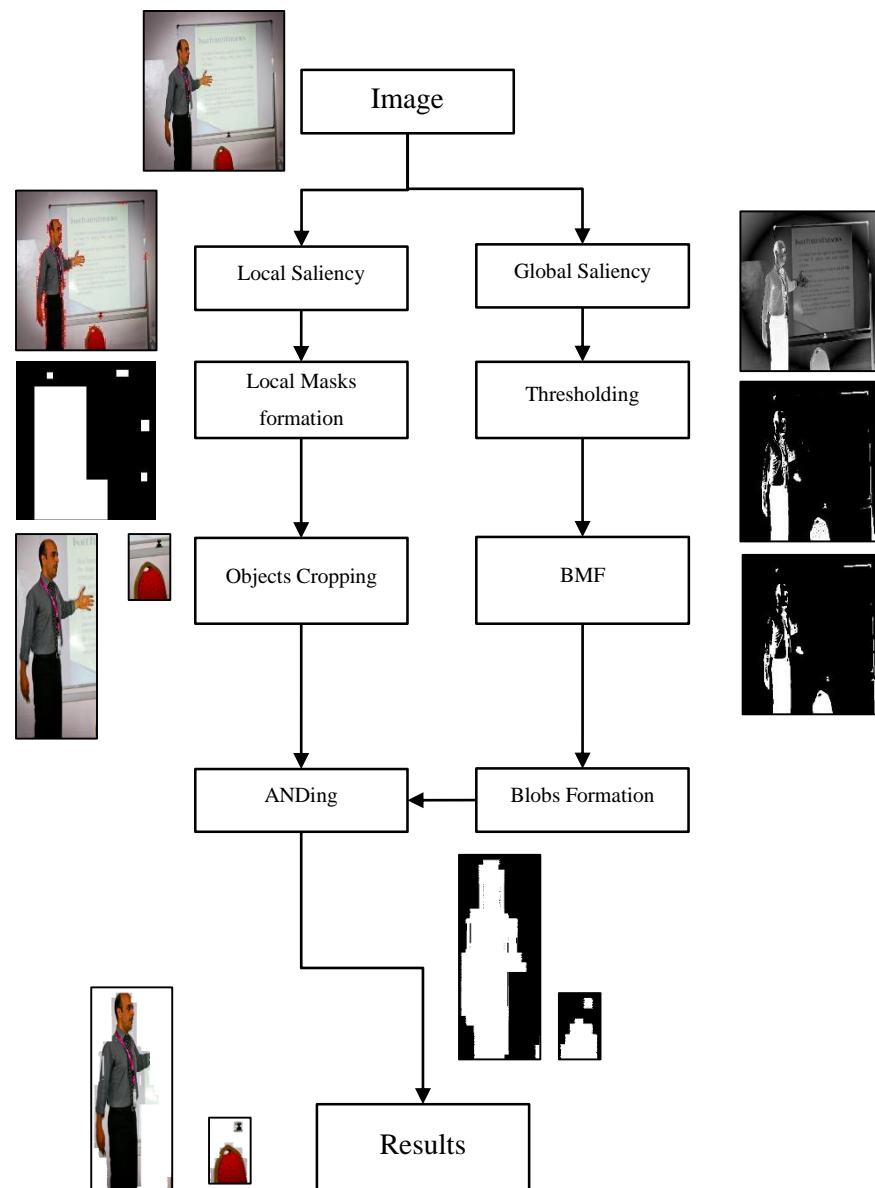


Figure 4-40: Stages and processes of the refined two stage saliency identification (RTSSI) block diagram.

Figure 4-41 shows a detailed example for further illustration.

It was noticed that the region containing the salient object in some cases is larger than the object itself. This is because of the irregular shape of the object. This may cause part of the background to be included with the object, which may affect the features of the



object. In order to overcome this kind of problem, global saliency identification was integrated with LSI to form RTSSI.

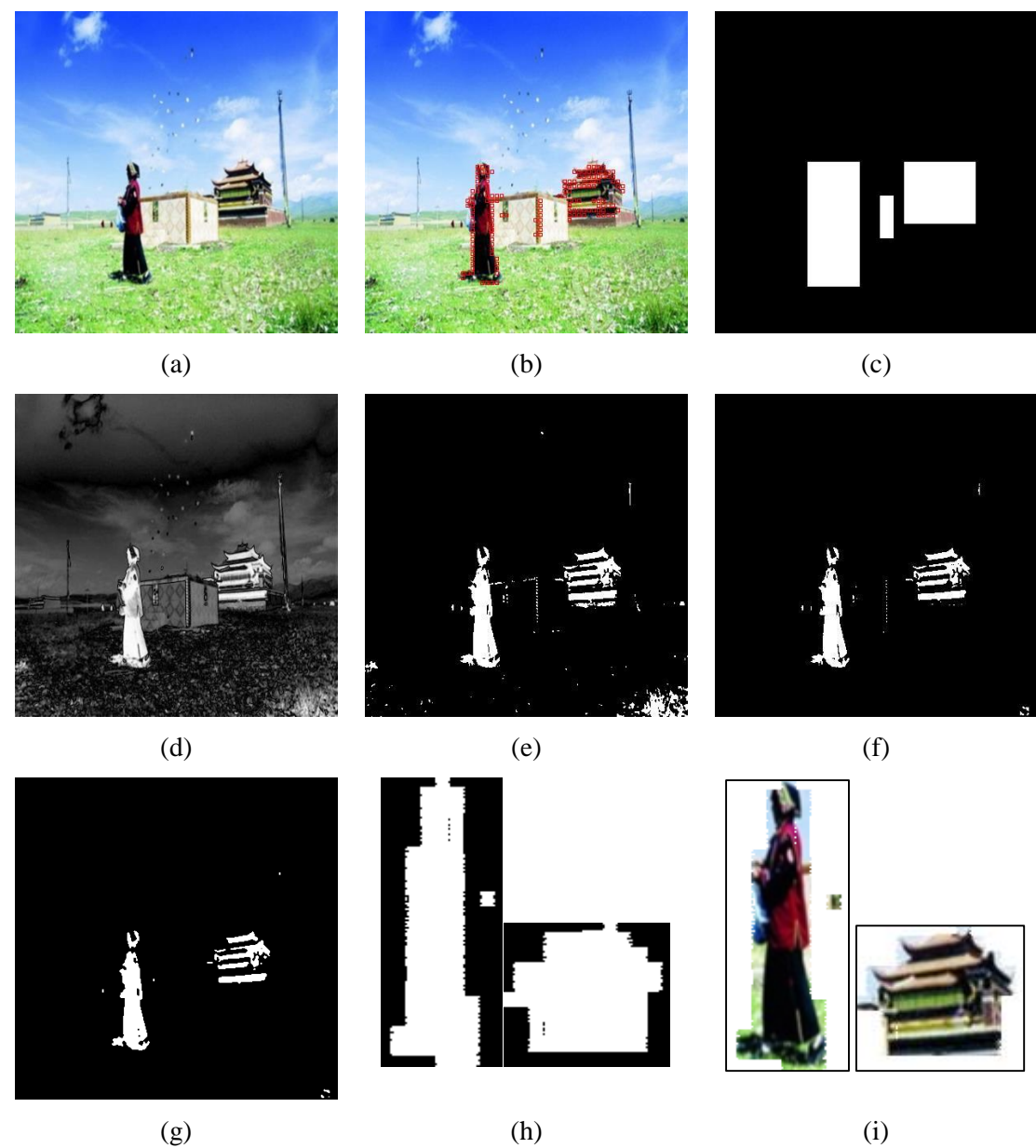


Figure 4-41: Example of applying the RTSSI algorithm, (a) original image, (b) local saliency identification, (c) local masks, (d) global saliency enhancement, (e) binary saliency map after thresholding, (f) ANDing the masks in (c) and the binary image in (e), (g) applying BMF, (h) extracting the masks, (i) the obtained objects.

Table 4-17 shows the steps of object extraction using RTSSI. The table shows the original image, the Local Saliency Mask (LSM), the Global Saliency Mask (GSM), the Refined Saliency Mask (RSM) that was obtained from applying the RTSSI, and the extracted objects.

Table 4-17: Steps of applying RTSSI.





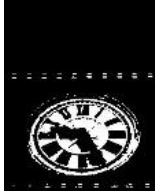


image	LSM	GSM	RSM	Object
				
				
				
				
				
				
				
				
				
				
				

Table 4-18 shows a qualitative comparison between FSM and RSM against the ground truth data GT.

Table 4-18: Qualitative comparison with the ground truth masks.







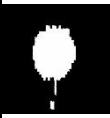
































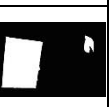
























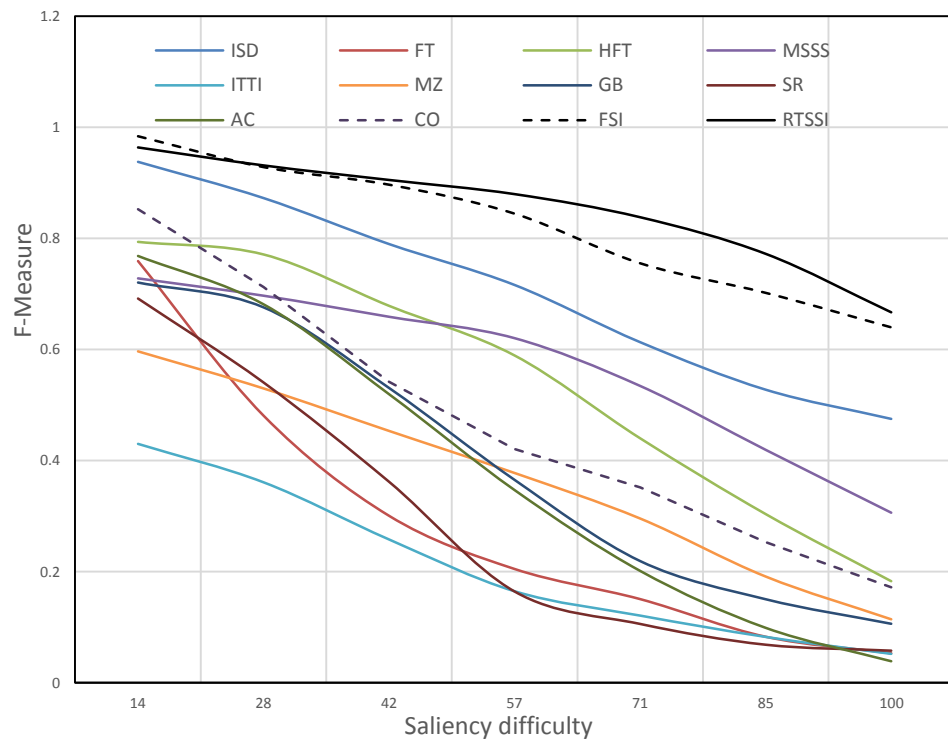
image	FSM	RSM	GT
			
			
			
			
			
			
			
			

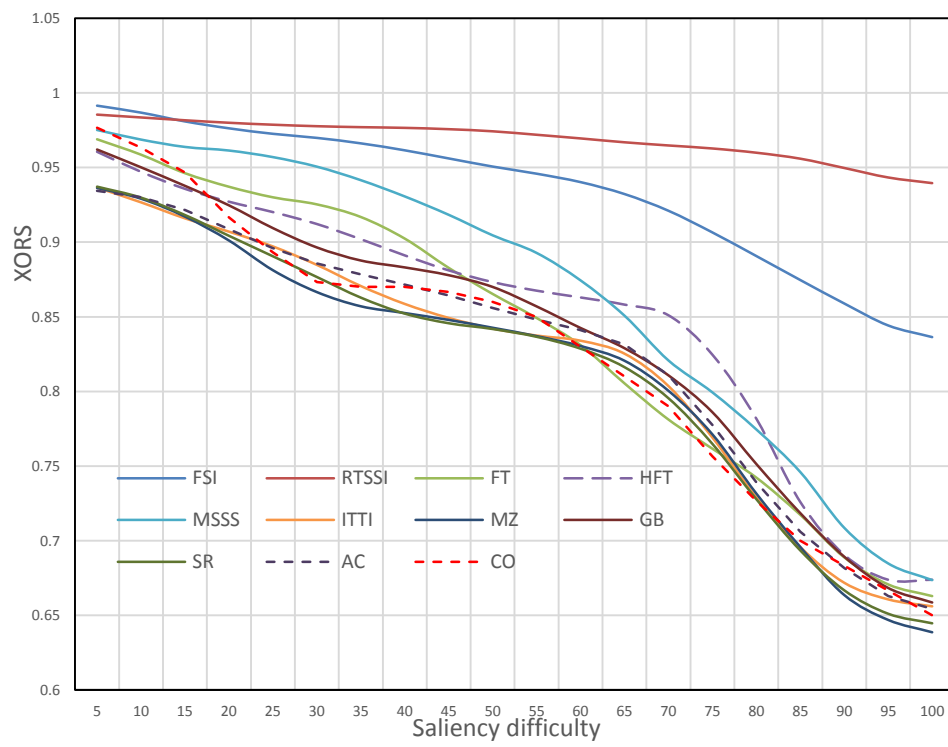
image	FSM	RSM	GT
			
			
			
			
			
			
			
			

#### 4.7.7 Benchmarking

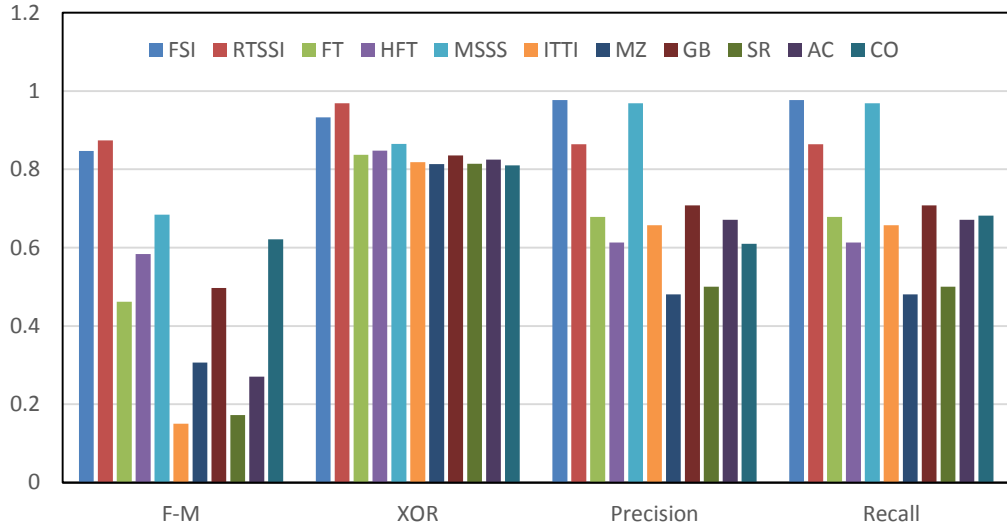
The efficiency of the proposed algorithm has been tested and measured against state-of-the-art methods which are IT [2], MZ [117], GB [118], SR [119], AC [120], FT [116], HFT [113], MSSS [121], and CO [124] [147]. Figure 4-42 graphically shows the qualitative comparison of the proposed method and the aforementioned methods. From the graph it is clear that the proposed algorithms produced better results.



(a)



(b)



(c)

Figure 4-42: Quantitative comparison with different saliency extraction techniques, (a) F-Measure curves, (b) Exclusive OR curves, (c) average measures comparison.

## 4.8 Neural Network Based Salient Points Clustering

Artificial Neural Networks (ANNs) can be used for identifying the regions of interest and the unimportant regions using the data obtained either from the eye trackers or from the salient points extraction algorithms. The neural networks need to be trained first using a set of images, and then they can be used for identifications. The image is divided into sub-images and every sub-image will be identified using the trained neural network.

Since both input and target output sets can be provided to the ANN, supervised learning techniques, such as back-propagation learning technique can be used to classify the salient and non-salient regions. Multilayer perceptron can be used for this purpose as it can separate the space into more complex decision regions than single layer; every layer is trained in the same training algorithm as single layer perceptron. The number of nodes in each layer is different from one layer to another. The input to the neural network will be the set of descriptions of the sub-images surrounding the salient and saccade points. Let us assume that the set of salient regions in the image is  $\mathbf{S}$ , the set of non-salient regions is  $\mathbf{U}$ , and the neutral regions be  $\mathbf{N}$  then the set of all regions  $\mathbf{P}$  is the union of these three sets i.e.

$$\begin{aligned}
\mathbf{P} &= \mathbf{S} \cup \mathbf{U} \cup \mathbf{N} \\
n_p &= |\mathbf{P}|, \quad n_s = |\mathbf{S}|, \quad n_u = |\mathbf{U}|, \quad n_n = |\mathbf{N}| \\
n_p &= n_s + n_u + n_n \\
\mathbf{P} &\subset \mathbf{I}
\end{aligned}
\tag{4-64}$$

where  $n_p, n_s, n_n$ , and  $n_u$  are the cardinalities of the sets  $\mathbf{P}, \mathbf{S}, \mathbf{N}$  and  $\mathbf{U}$  respectively. Furthermore, we shall define a mapping  $\psi$  from the set  $\mathbf{P}$  to the set of real numbers. This mapping extracts the measures which shall be used in identifying the regions, i.e.

$$\begin{aligned}
\psi: \mathbf{P} &\rightarrow \mathbf{R}^n \\
\hat{f} &= \psi(p), \quad \forall p \in \mathbf{P}
\end{aligned}
\tag{4-65}$$

where  $n$  is the number of features in the features vector and  $\hat{f}$  is a vector of features that are extracted from the sub-image surrounding the points  $p$ . This vector of features will be used in training the neural network to identify the rest of the regions.

The training data for the neural network shall contain two different types of sample input;

1. The vectors of features that are extracted from the region surrounding the interest points  $\hat{f}^s$ .
2. The vectors of features that are extracted from the region surrounding the unimportant points  $\hat{f}^u$ .

Since the training of such kinds of neural nets is supervised training, which means that we need to provide the network with both the input and desired output vectors, then the neural network will have only one output to specify the saliency level of the rest of the sub-images in  $\mathbf{N}$ . The output vector  $\hat{y}$  will have a real value between zero and one. During the training phase, it will be given one for the important regions and zero for the unimportant regions.

For the purpose described above, we shall use the back propagation ANN with the following specifications:

- The number of input nodes is equal to the number of features that are used in describing or identifying the sub-image.

- The number of nodes in the hidden layer is equal to one-half of the number of nodes in the input layer. This is a reasonable number to produce the required nonlinearity of the network.
- The network will have one output to identify the saliency level.
- The bipolar sigmoid function shall be used as an activation function.

$$g(x) = \frac{1 - e^{-\alpha x}}{1 + e^{-\alpha x}} \quad 4-66$$

where  $\alpha$  is a tuning factor that may affect the slope and the convergence of the function. Assigning a large value to  $\alpha$  will result in a curve similar to that of the threshold function, while small values reduce the slope and make the curve slowly change from minimum to maximum. The optimal value should be in the range of unity; value such as 2 may increase the convergence speed and retains the nonlinearity property of the function. Bipolar sigmoid was selected instead of the sigmoid function to increase the range of the output values to between  $-1$  and  $1$ , while it is only between  $0$  and  $1$  in the regular sigmoid function.

The features that have been used in this test are the average of the three colour bands, more about the features will be discussed in the identification section.

Figure 4-43 shows the result of applying the neural network in identifying the important regions using the important points. As shown in the figure, the red circles represent the regions surrounding the important points and the blue circles represent the regions surrounding the extracted saccade points. The sub-images are sized  $7 \times 7$ . The neural network was trained using the set of regions given in Figure 4-43 (b) and then the trained neural network was used to identify all other sub-images with an overlapping of six pixels between them, as shown in (c) which also shows the map of the neural network output, and (d) which shows the identified important regions.

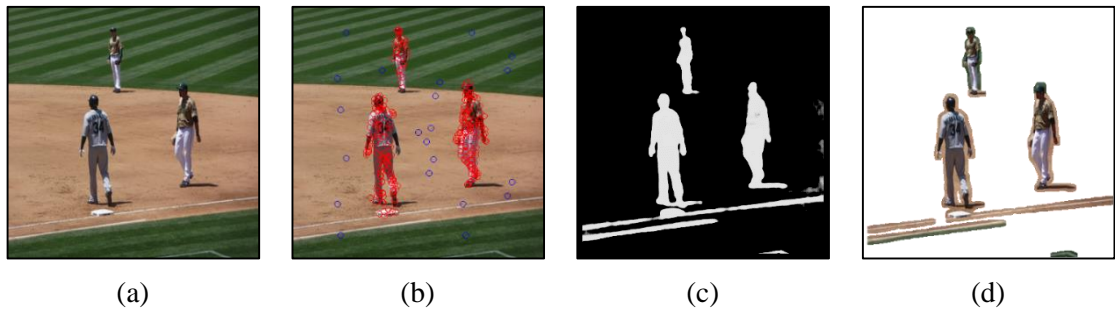


Figure 4-43: The application of neural network on salient regions extraction, (a) original image, (b) important points (red) and saccade points (blue), (c) saliency map, (d) the extracted objects.

The same trained neural network can be used to extract the salient region in similar images as shown in Figure 4-44.


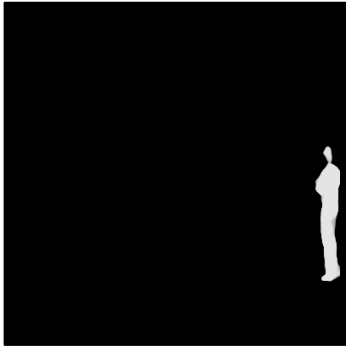
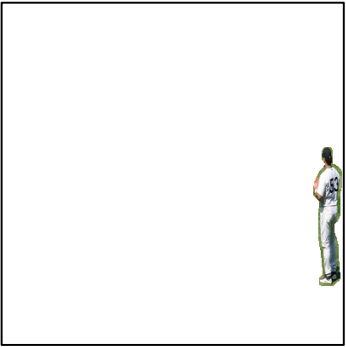

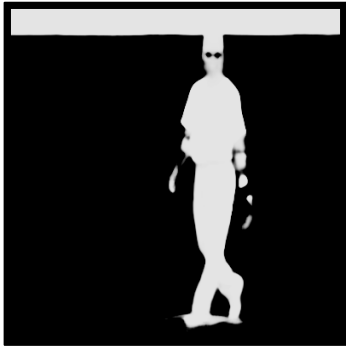

Original Image	Saliency Map	Salient objects
		
		

Figure 4-44: The application of the trained neural network on salient regions extraction on images similar to the image in Figure 4-43.

This approach suffers from the disadvantage of the long computation time required in ANN training, because for each image the ANN has to be trained based on the fixations and saccades corresponding to that image. The main application for this method is in overcoming the problem that we have faced in the aforementioned irregularity approaches, which is, when the algorithms were applied on the images from class C1, the results were not quite satisfactory. This is because the images in C1 contain regular



large objects in the middle of the image. In such cases, the irregular part of the image is the small background and not the object. To overcome this problem we shall consider the points close to the centre of the image as the salient points and train the neural network to identify the rest of the object.

## **4.9 Conclusions**

In this chapter, the main existing approaches to extracting saliency have been reviewed and analysed. The merits and demerits of each approach have been studied in order to develop a new saliency identification algorithm. A novel saliency extraction algorithm has been developed which uses the irregularity in the regions as a measure of saliency. Both local and global saliency identification approaches have been developed, from which a two-stage saliency extraction algorithm has been built. All the saliency approaches have been tested against standard datasets. In the local irregularity approach, a clustering algorithm was needed to cluster the salient points to form the salient regions; therefore, a clustering technique was proposed which is suitable for clustering salient and gaze points. In addition, ANNs have been tested and used to cluster the salient points and their efficiency and limitations were reviewed. Automatic fuzzy thresholding was another contribution presented in this chapter. In this approach, Fuzzy logic is used to isolate salient points from non-salient points.

Lastly, saliency evaluation approaches were developed in this chapter. Most existing saliency evaluation approaches have been reviewed and studied and the new saliency approaches which have been developed are suitable for use with our application.

# Chapter 5

## Saliency Based Image Contents Identification (SBICI)

### 5.1 Introduction

In this Chapter, we shall discuss the identification phase of the image contents based on the salient region. A Saliency-Based Image Contents Identification (SBICI) technique shall be developed and presented in this Chapter. This technique will be compared with the existing CBIR techniques. Several techniques will be discussed, and compared with the proposed algorithm. In addition, a background identification algorithm which utilizes the texture features and computational intelligence techniques has been developed.

The algorithms were implemented using Microsoft Visual Studio C# as the programming language and Microsoft SQL server as the database server. The WANG image database has been used first in the tests since it was adopted by most of the CBIR techniques, and it is easier to benchmark the results. Thereafter, another image dataset was constructed containing images with more variation and diversity in their contents.

In order to evaluate the efficiency of the retrieval algorithm we shall discuss most of the retrieving evaluation methods such as the precision and recall curves. In addition, an evaluation technique will be suggested, which will be considering not only how relevant the retrieved images are, but also their order of retrieval.

## 5.1 Image Contents Identification

Almost all traditional image contents identification techniques start with feature extraction. Features such as colour, shape and texture, can be used for this purpose. The features should be invariant with respect to the changes in the image. As discussed above, some images need to be identified as a whole, while others are identified based on the contents of their regions. The basic image retrieval system is given in the diagram shown in Figure 5-1.

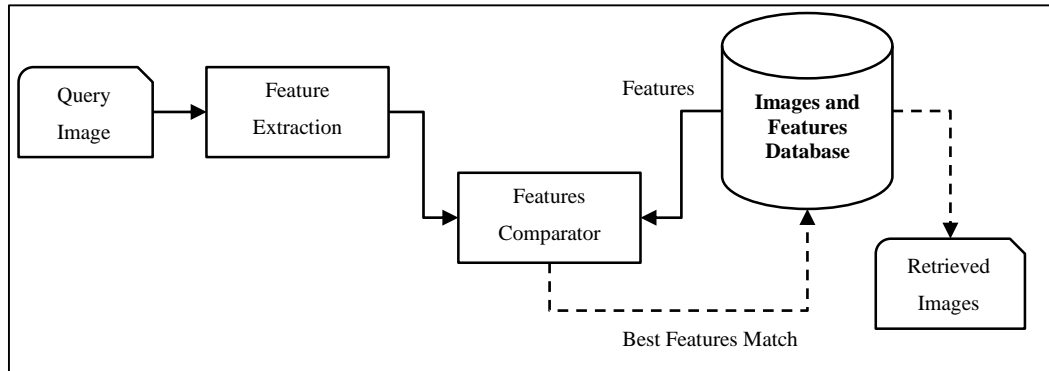


Figure 5-1: CBIR basic diagram.

In the diagram shown in Figure 5-1, the image undergoes a feature extraction process in which some measures are extracted from the image. These measures are compared with the features stored in the database and images corresponding to higher similarity measures will be selected as the relevant images. In either case, whether we are searching for an image as a whole or by parts, the image or the region will be converted to a set of measures. First, we shall define the region  $R$ , which may cover only a part of the image, a few parts, or the entire image. The mapping  $\psi$  is defined as a feature extractor that maps the image contents in  $R$  into an array of measures or descriptors  $D$ ,

$$D = \psi(R) \quad 5-1$$

$D$  is calculated for the query image and for the images in the database, then in order to find the matched images for the query image, these measures are compared, either by finding the maximum similarity or the minimum distance.

## 5.2 Colour – Based CBIR

Colour distribution is one of the most important features of the image and is widely used in different image applications. One of these applications is in image identifications. Some measures can be extracted from the pixel colour values and used for identifying the contents of the image. Statistical operations, such as mean, median, standard deviation, etc. can be used for this purpose. The information obtained from this method does not give sufficient information about the contents of the image; instead, a histogram can be used.

Numerous algorithms have used the colour histogram feature in investigating the contents of an image. The histogram features can be utilized to extract measures that can be used to compare images with each other. Histograms offer a good description of the contents of the image and are not affected much by the rotation, scaling, or transition of the image. However, they do suffer from a certain constraint in that they give no information about the spatial relationship between the pixels.

There are several possible ways to compare the histograms of two images, i.e. find the similarity or dissimilarity between them, and, in general, these can be divided into two types: Bin-by-Bin Distance BBD and Cross-Bins Distance CBD.

### 5.2.1 Bin-by-Bin Distance (BBD)

Minkowski distances was used widely in many publications such as [195], [196] and others. It can be used to find the distance between two histograms bin by bin as given in the following equations:

$$\begin{aligned} H_Q &= \langle h_{q0}, h_{q1}, \dots, h_{qn-1} \rangle \\ H_P &= \langle h_{p0}, h_{p1}, \dots, h_{pn-1} \rangle \\ D(H_Q, H_P) &= \sqrt[r]{(h_{q0} - h_{p0})^r + (h_{q1} - h_{p1})^r + \dots (h_{qn-1} - h_{pn-1})^r} \end{aligned} \quad 5-2$$

Where  $H_Q$  and  $H_P$  are the histograms corresponding to the images Q and P respectively and  $D(.,.)$  is the function used to find the distance between two vectors.  $r$  is any integer value greater than 0. The Euclidean distance is a special case of the Minkowski distance where  $r = 2$ .

Another well-known similarity measure is the intersection between the histograms (IBH), in which, if there is an intersection between them then having similarity between the images is possible. The simplest form of the intersection measure between two histograms is given below:

$$U(H_Q, H_P) = \sum_{i=0}^{n-1} \text{Min}(h_{qi}, h_{pi}) \quad 5-3$$

This measure is high if the intersection between the histograms is high, i.e. if the corresponding components in the two histograms are close to each other. In contrast to the BBD, which measure the dissimilarity between the histograms, IBH measures the similarity between them. IBH is very easy to calculate, but it is not accurate since it depends on the number of elements in a certain bin. Siggelkow suggested the normalized intersection in which he modified equation 5-3 to become as follows [197]:

$$U_{norm}(H_Q, H_P) = \frac{1}{N_Q} \sum_{i=0}^{n-1} \text{Min}(N_Q \cdot h_{qi}, N_P \cdot h_{pi}) \quad 5-4$$

where  $N_Q$  and  $N_P$  are the number of pixels in images Q and P respectively. He suggested this normalization to overcome the problem that may occur due to the difference in the size between the images.

Another measure, which uses the BBD principles, is the Kullback-Leibler divergence, which is calculated as follows:

$$KLD(H_Q, H_P) = \sum_{i=0}^{n-1} h_{qi} \log \left( \frac{h_{qi}}{h_{pi}} \right) \quad 5-5$$

As per [197], this measure is not symmetric and is numerically unstable.

Figure 5-2 shows the results obtained using the Minkowski distances between the query image and the images stored in the database. (a) This figure shows the query image for which similar images are being sought in the database, (b) shows the results obtained from comparing the red histogram. Similarly, (c), (d) and (e) show the results obtained from comparing green, blue, and grey histograms respectively. It is clear from the figure that the retrieval process is poor for many reasons, one of which is that it is very sensitive to the shift in the histogram, perhaps due to the change in lighting conditions.



(a)



(b)



(c)



(d)



(e)

Figure 5-2: Experimental results of CBIR based on histogram comparison using Minkowski distances between (b) red histogram, (c) green histogram, (d) blue histogram, and (e) grey histogram.

The average accuracy in the previous test was 48%, which is very low if we consider that the WANG database contains many images with similar colours and lighting conditions. This low accuracy is due to the low probability of having similar values in corresponding bins, even if the images are similar but not the same. Thus, a better and more robust measure is needed, one which can find the relationship between the histograms regardless of their shift or scale. Cross-bin is one of the better approaches for achieving this.

### 5.2.2 Cross-Bin Distance (CBD)

In BBD, the main drawback is when there is some shift in the histogram due to changing the lighting conditions or luminance. In such cases, this shift will give high distance or dissimilarity even when the images are similar. In order to overcome these kinds of

problems, CBD measure can be used. In contrast to BBD, CBD compares different bins rather than the corresponding bins only [197].

Histogram statistical measures can be considered as one of the CBD measures since it finds some measures that can be extracted from the histogram as a whole. The main statistical measures are the mean, standard deviation, and skewness, which were discussed in section 2.6.1 and given equations 2-2, 2-3 and 2-4.

Since the results obtained from applying CBD are very close to each other – almost the same – we shall only discuss the results obtained from one of them. We shall consider the statistical measures in the test and compare the results with the results obtained from BBD. Mean and standard deviation for each band in addition to the intensity were used in the comparison. The features vector  $\mathcal{F}$  is given by:

$$\mathcal{F} = \langle \mu_R, \mu_B, \mu_G, \mu_I, \sigma_R, \sigma_B, \sigma_G, \sigma_I \rangle \quad 5-6$$

where:  $\mu_R$ ,  $\mu_B$ ,  $\mu_G$ , and  $\mu_I$  are the mean of the red, blue, green, and intensity histograms respectively and  $\sigma_R$ ,  $\sigma_B$ ,  $\sigma_G$ ,  $\sigma_I$ , are the standard deviation of the red, blue, green, and intensity histograms respectively.

The average precision of applying the CBD on the WANG database is better than that of BBD and in the range of 51% but yet it is still not good enough.

### 5.2.3 Global vs. Local Histogram

Both types of colour histogram, Local Colour Histogram (LCH) and Global Colour Histogram (GCH), can be used in the identification process. In GCH, the colour distribution of the entire image is calculated and then compared to other images. In such case it is quite possible to have different images with similar GCH, thus, LCH can be used to give better results. In LCH, the image is divided into regions and the histogram is calculated for each region, and then the set of obtained histograms will be calculated with other images' local histograms. The main issue with LCH is the specification of the size and location of the regions. Heidemann suggests the specification of such local regions be based on the Harris saliency map. In addition, he proposes the use of a global window with dimensions of 90% of the width and the height surrounding the centre of the image, assuming that most of the information is covered by this global window [198].

### 5.3 Texture – Based CBIR

Texture is used in CBIR to give information about the structure of the contents of an image. As different image contents may have a similar colour distribution, texture can be used to discriminate between these images. Many measures can be used to identify the texture of an object but the Moments of intensity and grey level co-occurrence matrices (GLCM) are mostly used in CBIR applications.

GLCM are used to describe texture since they measure the relationship between the adjacent pixels displaced by a certain distance. For instance, let us define  $f_{ij}$  as the probability of occurrence of the intensity value  $i$  at position  $(x, y)$  and the adjacent pixel with intensity value  $j$  at position  $(x + \delta_x, y + \delta_y)$ . The value of the displacement  $\delta$  is an integer number larger than, or equal to, one. Usually GLCM are described in terms of displacement and angle, i.e.  $GLCM = f(x, y, \delta, \theta)$ , where  $\delta$  is the displacement and  $\theta$  is the angle between the pixels. Usually  $\theta$  takes standard values such as, 0, 90, 180, 270, and 360 degree. Instead of using the angle  $\theta$  we can use two different displacements  $\delta_x$  and  $\delta_y$ . The GLCM (G) is then defined as follows:

$$G = \begin{bmatrix} f_{00} & f_{01} & \cdots & f_{0N} \\ f_{10} & & & \\ \vdots & & & \\ f_{1N} & & & f_{NN} \end{bmatrix}$$

$$f_{ab} = \frac{1}{(H - \delta_y)(W - \delta_x)} \sum_{i=1}^{H-\delta_y} \sum_{j=1}^{W-\delta_x} g(a, b) \quad 5-7$$

$$g(a, b) = \begin{cases} 1 & \text{if } I(x, y) = a \wedge I(x + \delta_x, y + \delta_y) = b \\ 0 & \text{otherwise} \end{cases}$$

where G is a GLCM with size of  $N \times N$ , and N is the number of grey levels in the image.

Haralick et al. [199] suggested 14 features describing the two-dimensional probability density function  $f$ , of these measures. Five measures are widely used in literatures, these measures are Angular Second Moment (ASM) ( $\mathcal{A}$ ), Contrast ( $\mathcal{C}$ ), Correlation ( $\mathcal{C}$ ), Inverse Difference Moment (IDM) ( $\mathcal{M}$ ) and Entropy ( $\mathcal{E}$ ) as shown below [200]:



$$\mathcal{A} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f_{ij}^2 \quad 5-8$$

$$\mathcal{C} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-j)^2 f_{ij} \quad 5-9$$

$$\mathfrak{C} = - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{(i-\mu_x)(j-\mu_y)}{\sqrt{\sigma_x \sigma_y}} f_{ij} \quad 5-10$$

$$\mathcal{M} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{f_{ij}}{1 + (i-j)^2} \quad 5-11$$

$$\mathcal{E} = - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f_{ij} \log f_{ij} \quad 5-12$$

Where  $\mu_x$  and  $\mu_y$  are the mean values; and  $\sigma_x$  and  $\sigma_y$  are the standard deviation values of the matrix and can be calculated as follows [201]:

$$\mu_x = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} i f_{ij} \quad 5-13$$

$$\mu_y = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} j f_{ij} \quad 5-14$$

$$\sigma_x = \sqrt{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i - \mu_x)^2 f_{ij}} \quad 5-15$$

$$\sigma_y = \sqrt{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (j - \mu_y)^2 f_{ij}} \quad 5-16$$









Tsaneva *et al.* [200] analysed these measures as shown in Table 5-1.

Table 5-1: interpretation of GLCM [200].

Texture feature	Interpretation
ASM	Increases with regularity of the texture
Contrast	Related to contrast of the texture. Also known as “Sum of squares variance”
Correlation	A statistic of texture
IDM	Related to contrast of the texture. Also known as “Homogeneity”
Entropy	Increases with irregularity of the texture

The results in Table 5-2 show the GLCM extracted from each image and the corresponding statistical measures that have been extracted from the GLCM. Because of the nature of images, the shape of the GLCM is expected to be concentrated in the adjacent colours.


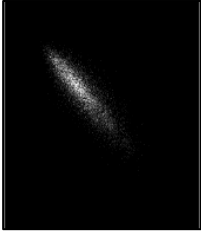
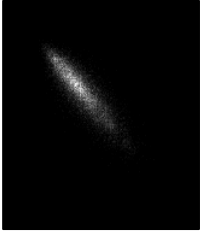
Table 5-2: Experimental Results of GLCM and the corresponding measures of images.

image				
GLCM				
$\mu_x$	100.1734	120.759	171.529	170.174
$\mu_y$	99.87	120.85	171.45	170.20
$\sigma_x$	23.6	27.54	20.23	17.571
$\sigma_y$	23.5	27.54	20.15	17.47
$\mathcal{A}$	0.0004218	0.00045	0.0012	0.0010
$\mathcal{C}$	151.76	167.08	115.17	40.56
$\mathfrak{C}$	-20.405	-24.511	-17.34	-16.363
$\mathcal{M}$	0.1105	0.1587	0.222	0.233
$\mathcal{E}$	3.4754	3.498	3.243	3.128

### *The effect of symmetry*

Most of the algorithms suggest the use of symmetric GLCM which means that elements above the diagonal of the matrix are equal to the elements below the diagonal or in other words  $\text{GLCM}(i, j) = \text{GLCM}(j, i)$ . After converting the GLCM to a symmetric matrix the result in Table 5-3 is obtained. From the results listed in this table, it is clear that there is not a big difference between the measures obtained for symmetric and asymmetric GLCMs.

Table 5-3: The result of GLCM and statistical measures of the symmetric GLCM.

image	CCM	$\mu_x$	$\mu_y$	$\sigma_x$	$\sigma_y$	$\mathcal{A}$	$\mathcal{C}$	$\mathfrak{C}$	$\mathcal{M}$	$\mathcal{E}$
		100.1734	99.87	23.6	23.5	0.0004218	151.76	-20.405	0.1105	3.4754
	Asymmetric									
		100.0220	100.02	23.617	23.617	0.00037	151.762	-20.404	0.1105	3.5477
	Symmetric									

#### Using the GLCM in CBIR

The GLSM can be used in identifying the contents of an image since it can recognize the regions with texture inside, like sky, cloud, grass, fabric, skin, etc. The algorithm was applied in identifying the texture images in a database. The Euclidean distance was used to find the distance between the texture features extracted from GLCM.

$$D(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad 5-17$$

Figure 5-3 shows the results obtained by applying the above algorithm. From the figure, it is noted that the results are not very accurate since the differences between the GLCM measures are very large, as shown in Table 5-2 and Table 5-3. For example, the ASM is in the range of  $10^{-3}$ , the contrast range is a few hundred up to thousands, the correlation is in the range of tens, IDM in  $10^{-1}$  range, and the entropy is in the range of unity. This makes one or more of the features dominant and may eliminate the effect of other measures, thus, the results of the comparison may not be accurate.

In order to overcome this problem, some kind of normalization is required. Some normalization processes may be achieved by dividing the features by the maximum value of all the 5 features. That is not applicable here since this may make small values even

smaller. The second common normalization is made by dividing the features by the sum of the measures, which is also not applicable for the same reason. In order to normalize the features one needs to find a common factor between each feature for one image with the same feature for all other images. To make the distance normalized we shall suggest the distance given in the following equation:

$$D(q, d_i) = \frac{1}{\sqrt{5}} \left[ \left( 1 - \frac{\text{Min}(\mathcal{A}_{di}, \mathcal{A}_q)}{\text{Max}(\mathcal{A}_{di}, \mathcal{A}_q)} \right)^2 + \left( 1 - \frac{\text{Min}(\mathcal{C}_{di}, \mathcal{C}_q)}{\text{Max}(\mathcal{C}_{di}, \mathcal{C}_q)} \right)^2 + \left( 1 - \frac{\text{Min}(\mathcal{G}_{di}, \mathcal{G}_q)}{\text{Max}(\mathcal{G}_{di}, \mathcal{G}_q)} \right)^2 + \left( 1 - \frac{\text{Min}(\mathcal{M}_{di}, \mathcal{M}_q)}{\text{Max}(\mathcal{M}_{di}, \mathcal{M}_q)} \right)^2 + \left( 1 - \frac{\text{Min}(\mathcal{E}_{di}, \mathcal{E}_q)}{\text{Max}(\mathcal{E}_{di}, \mathcal{E}_q)} \right)^2 \right]^{1/2} \quad 5-18$$

Where  $q$  is the query image and  $d_i$  is an image with index  $i$  from the database images.

In the above equation, the maximum value for the term  $\frac{\text{Min}(\cdot)}{\text{Max}(\cdot)}$  is one, and the minimum value for the same term is 0 when the difference is extremely high, thus the distance will be  $\sqrt{5}$ . Thus to make it unity we shall divide the above equation by  $\sqrt{5}$ . The general format for the above equation is as follows:

$$D(p, q) = \sqrt{\frac{1}{N} \sum_{i=1}^N \left( 1 - \frac{\text{Min}(p_i, q_i)}{\text{Max}(p_i, q_i)} \right)^2} \quad 5-19$$

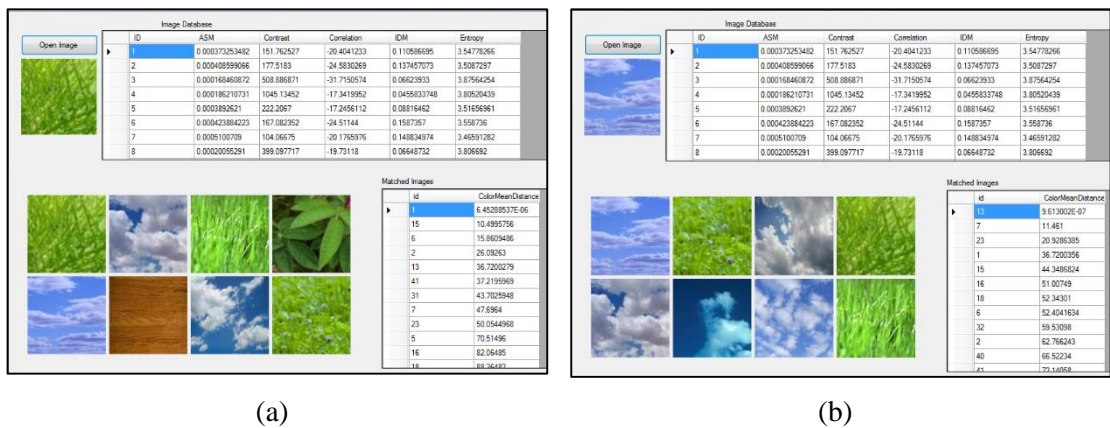


Figure 5-3: Finding the match for the query image using the formula given in Equation 5-17, (a) grass texture query image, (b) cloud texture query image.

Figure 5-4 shows the result by the improved distance measure given in Equation 5-19.

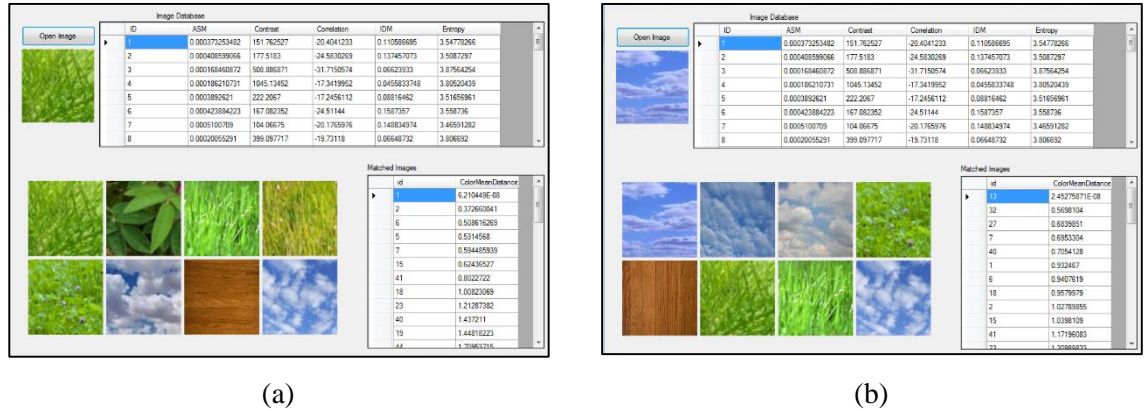


Figure 5-4: finding the matched for the query image using the formula given in Equation 5-19 , (a) grass texture query image, (b) cloud texture query image

The above qualitative comparison showed that the results in the second case are better than those that were obtained in the first case. In order to make the comparison clearer, a quantitative comparison shall be performed in Table 5-4 using the weighted efficiency evaluation measure (WEEM) which is discussed in Section (5.6).

Table 5-4: Quantitative comparison between the result obtained using equations 5-19 and 5-21.

Case	Figure	Equation	Average WEEM
1	5-4	5-17	0.59
2	5-5	5-19	0.7

## 5.4 Saliency Based Image Retrieval (SBIR)

Figure 5-5 shows the basic diagram of the SBIR system. In this diagram, the image first goes through the saliency regions extraction process to produce a set of salient regions. For each salient region, the set of features is extracted using an appropriate feature extractor. Every region in the query image will be compared with all regions in the images in the database. Images with the best match of features will be retrieved as the relevant images.

In SBIR system, the image shall be divided into two main categories; objects and backgrounds. The objects are extracted from the salient regions, and the backgrounds are extracted from the non-salient regions. Both parts are fed into the second stage, which is the identification process. Two different identification algorithms are used for this

purpose; the first one for identifying the objects and the second one for identifying the background.

In the aforementioned system, we shall introduce a retrieval technique that utilizes the principle of saliency in image retrieval. In the proposed technique, only salient regions in the images are compared, which means that we are not going to match the entire image but only a few regions of it. This will give better results since in some cases, the unimportant regions might be dominant and the effect of the salient region will be very small. For example, consider the case in which one needs to search for a ball in a field using a colour histogram as the feature. In this case, the surrounding environment's effect on the histogram is very much higher than that of the ball, thus the retrieved images will be more relevant to the green grass than to the ball.

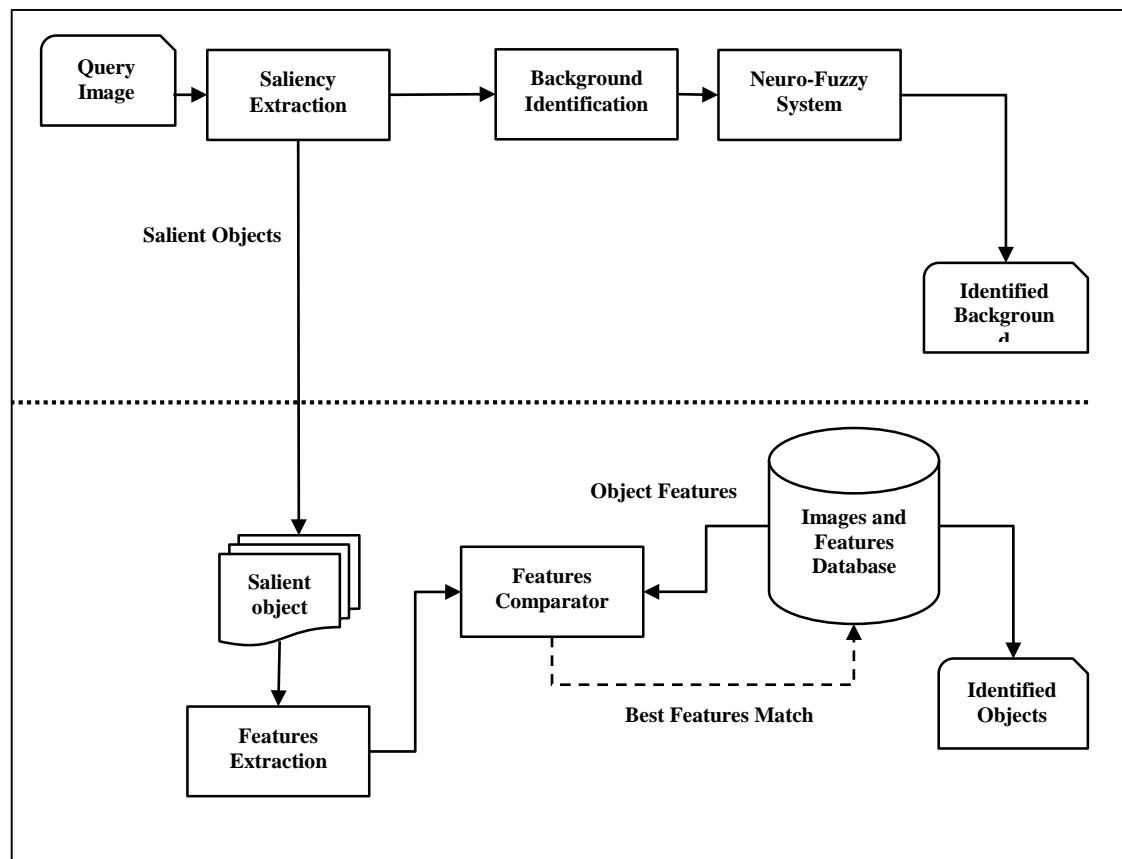


Figure 5-5: SBIR System.

The definitions of important and unimportant regions need to be identified first. Important regions are the regions that contain objects that need to be identified, while unimportant regions are the regions that contain unimportant details such as a

background. For example, if we consider saliency as the measure of importance, then regions with more salient points are more important than others with less salient points.

Let us define  $\mathbb{R}$  as the set of regions that can be extracted from the image  $\mathbb{I}$ , then  $\mathbb{S}$  will be the set of salient regions in that image and  $\mathbb{U}$  is the set of all non-salient regions. The Set  $\mathbb{S}$  is a subset from  $\mathbb{R}$  and the set  $\mathbb{R}$  contains both regions  $\mathbb{S}$  and  $\mathbb{U}$  i.e.  $\mathbb{R} = \mathbb{S} \cup \mathbb{U}$ .

Let us further define  $s_i$  as the salient region then  $\mathbb{S} = \{s_i | i = 0, 1, \dots, N\}$ , where  $N$  is the number of salient regions in the image. For each image, the set of salient regions is extracted and matched with other regions in the images stored in the database.

The main advantage of using Saliency-Based Image Contents Identification (SBICI) is that it focuses on the features from the objects rather than the background. As discussed before, we have considered that the information in a scene ( $\mathbf{H}$ ) is divided into two parts; important ( $\mathbf{H}_I$ ) and unimportant ( $\mathbf{H}_U$ ). Important information is the information contained in the object (in the region of interest) while most of the unimportant information is in the background, thus  $\mathbf{H} = \mathbf{H}_I + \mathbf{H}_U$ . The information contents can be extracted using the following function:

$$\begin{aligned}\mathbf{H} &= \psi(\mathbf{R}), \\ \mathbf{H}_I &= \psi(\mathbf{R}_I), \\ \mathbf{H}_U &= \psi(\mathbf{R}_U)\end{aligned}\tag{5-20}$$

where  $\psi(.)$  is the feature extractor, and  $\mathbf{R}$  is the region we want to extract the features from. The main problem here is that the feature extractor usually merges the features of the important and unimportant regions together to find the features of  $\mathbf{R}$ , in other words without suitable segmentation  $\mathbf{H}$  is not equal to the sum of the important and unimportant regions information, i.e.

$$\begin{aligned}\mathbf{H} &= \mathbf{H}_I + \mathbf{H}_U + \mathbf{H}_{I \cap U} \\ \psi(\mathbf{R}) &= \psi(\mathbf{R}_I) + \psi(\mathbf{R}_U) + \psi(\mathbf{R}_{I \cap U})\end{aligned}\tag{5-21}$$

Figure 5-6 shows the information contents in both the object and in the background. It is clear from the figure that the information contents (the colour distribution in this case) of the background have more significance than the object.

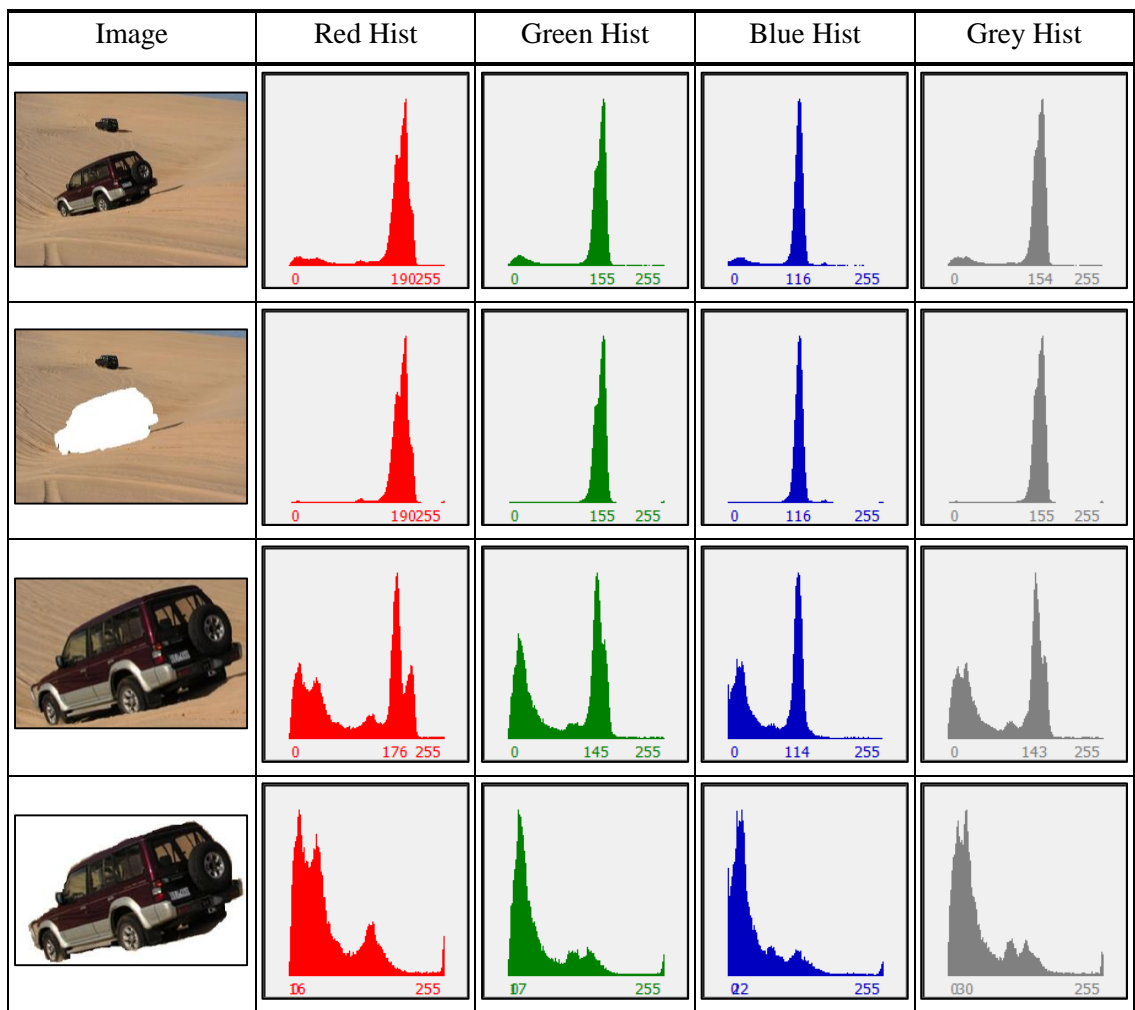


Figure 5-6: Histogram for the bands of an image.

Figure 5-7 illustrates the histogram components of the grey histogram. The figure shows the two components of the grey histogram of the image. From the figure, it is clear that the background component has higher values in each bin; this will give the background component more significance than the object component in the features or measure extraction.



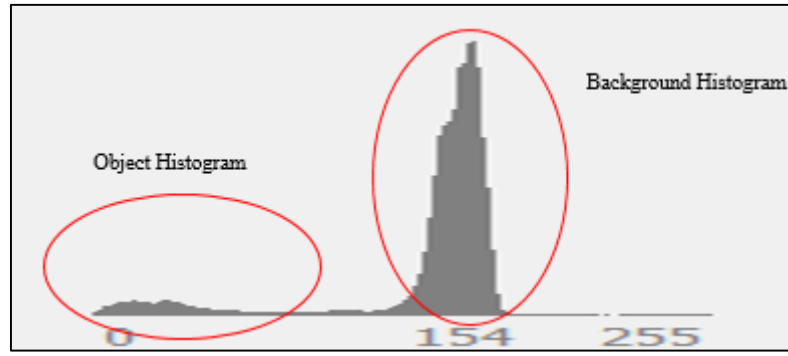


Figure 5-7: Histogram components.

For instance, let us assume that the colour values in an image are represented by the set  $\mathbb{I}$ , which represents the universal set, such that:

$$\mathbb{I} = \{x_i : 1 \leq i \leq W \times H\} \quad 5-22$$

Where  $x_i$  is the value of the pixel colour of the image considering that the image pixels are arranged to be a one-dimensional vector, using the following formula:

$$\begin{aligned} x_k &= I(i, j) \\ k &= (i - 1) \times W + j \end{aligned} \quad 5-23$$

where  $W$  and  $H$  are the width and height of the image.

The set of the pixels is divided into two components, object  $\mathbb{I}_S$  and background  $\mathbb{I}_B$ :

$$\begin{aligned} \mathbb{I} &= \mathbb{I}_S \cup \mathbb{I}_B \\ \mathbb{I}_S &= \{x_{Si}\} \\ \mathbb{I}_B &= \{x_{Bi}\} \end{aligned} \quad 5-24$$

Considering that the object is isolated from the background and the borders are well-defined, i.e.  $\mathbb{I}_S \cap \mathbb{I}_B = \emptyset$ , the cardinality of the universal set is equal to the sum of the cardinality of the two components, i.e.

$$\begin{aligned} |\mathbb{I}| &= |\mathbb{I}_S| + |\mathbb{I}_B| + |\mathbb{I}_S \cap \mathbb{I}_B| \\ |\mathbb{I}_S \cap \mathbb{I}_B| &= 0 \\ |\mathbb{I}| &= |\mathbb{I}_S| + |\mathbb{I}_B| \end{aligned} \quad 5-25$$

Based on the definition of the histogram, we shall define the function  $H(x)$  as the number of occurrences of the variable  $x$  as follows:

$$H(x_i) = P(x_i) \quad 5-26$$

where  $P(x_i)$  is the probability of occurrence of the bin  $x_i$ .  $P(x_i)$  contains the two components mentioned above i.e.

$$H(x_i) = H_B(x_i) + H_S(x_i) + H_{SB}(x_i) \quad 5-27$$

where  $H_B(x_i)$  and  $H_S(x_i)$  are the probability of occurrence of the variable  $x_i$  in the background and in the object regions respectively.  $H_{SB}$  represents the common intensities or colours shared by the object and the background, i.e. the same colour being in both the object region and in the background. After isolating the two regions, the effect of  $H_{SB}$  will be removed. The histogram for the salient region is given by:

$$\begin{aligned} H_S(x_i) &= H(x_i) - H_B(x_i) ; \forall x_i \in \mathbb{I}_S \\ H_S(x_i) &= 0 ; \forall x_i \in \mathbb{I}_B \end{aligned} \quad 5-28$$

In the case of handling the information contents of the image as a whole, the features are extracted from the histogram of the entire image, and the expectation value of the histogram can be extracted as follows:

$$\begin{aligned} E(x) &= \int_{-\infty}^{\infty} x \cdot H(x) dx \\ E(x) &= \int_{-\infty}^{\infty} x \cdot (H_B(x) + H_S(x)) dx \\ E(x) &= \int_{-\infty}^{\infty} x \cdot H_B(x) dx + \int_{-\infty}^{\infty} x \cdot H_S(x) dx \\ E(x) &= E(x \in \mathbb{I}_B) + E(x \in \mathbb{I}_S) \end{aligned} \quad 5-29$$

Assuming that after extracting the salient object, the values of  $x \in \mathbb{I}_B$  are zero, then the expected value will be only for the object. This will dramatically improve the comparison process.

## 5.5 Distance Measures

The distance measure between the query image histogram  $H^Q(x)$  and the  $i^{th}$  image in the database  $H_i^D(x)$  is given by:

$$D(H^Q(x), H_i^D(x)) = D(H_B^Q(x), H_B^D(x)) + D(H_S^Q(x), H_S^D(x)) \quad 5-30$$

By considering only the object in the identification and retrieval process, the background effect shall be considered to be zero, making the results very much more accurate than those obtained with inclusion of the background in the retrieval process.

In the above discussion the image was considered to be crisp, which is not the actual case as we have ignored the common features between the objects and the background. These features are not only due to the incertitude of the borders, but also due to the nature of the colour and texture distribution between them both. Because of that, we need to consider fuzzy logic in defining the regions in the image as follows.

### ***Fuzzy-Distance Measure***

Due to the imprecise nature of the image, and since the comparison technique involves image segmentation, a fuzzy approach is suitable for use here. There are many reasons behind considering a fuzzy approach, of which some are:

- 1- the uncertainty nature of the image, as discussed earlier;
- 2- the borders between the object and the background are not well-defined;
- 3- the homogeneity of the image which leads to difficulties in separating the objects information from the background information.

We shall here define the image as the universal set  $\mathcal{I}$ , which contains three main parts, object  $\alpha$ , background  $\beta$ , and the pixels that are common between the object and the background  $\gamma$ . The definition of each membership function is given below:

$\alpha(x): \mathcal{I} \rightarrow [0,1] \forall x \in \mathcal{I}$  is the membership function which measures how the variable  $x$  is possessive to the object,  $\beta(x): \mathcal{I} \rightarrow [0,1] \forall x \in \mathcal{I}$  is the background membership function, and  $\gamma(x): \mathcal{I} \rightarrow [0,1] \forall x \in \mathcal{I}$  is the common pixels membership function. Each membership function will produce a set that contains the corresponding elements and is a subset of the universal set  $\mathcal{I}$  i.e. we will have the sets  $\alpha \subset \mathcal{I}$ ,  $\beta \subset \mathcal{I}$ , and  $\gamma \subset \mathcal{I}$ .

We shall define here the distance between the images  $\mathcal{D}$  as a fuzzy distance which is given below:

$$\mathcal{D}(v, u): \mathcal{P}^2 \rightarrow R \quad 5-31$$

where  $\mathcal{P}$  is the power set of  $\mathcal{I}$  and  $v, u \in \mathcal{P}$ .

In order to consider the distance  $\mathcal{D}(v, u)$  as a metric it needs to satisfy the following criteria:

- 1- reflexivity:  $\mathcal{D}(v, v) = 0, \forall v \in \mathcal{P}$ ,
- 2- separability:  $\mathcal{D}(v, u) = 0 \Rightarrow v = u, \forall (v, u) \in \mathcal{P}^2$ ,
- 3- symmetry:  $\mathcal{D}(v, u) = \mathcal{D}(u, v), \forall (v, u) \in \mathcal{P}^2$ , and
- 4- triangular inequality  $\mathcal{D}(v, u) \leq \mathcal{D}(v, \partial) + \mathcal{D}(\partial, u), \forall (v, u, \partial) \in \mathcal{P}^3$ .

To make the comparison reasonable, we need to compare the features rather than the elements' values, therefore we shall define the feature extractor  $\psi(.): \mathcal{P} \rightarrow R^n$ , where  $n$  is the number of features extracted using the given feature extractor.

The feature sets shall inherit the imprecision from the fuzzy sets that represent the regions given above. Therefore, applying the feature extractor on each set will produce a new fuzzy set which contains the features extracted from each region as given below:

$$\begin{aligned} \mathcal{J}' &= \psi(\mathcal{J}) \\ \alpha' &= \psi(\alpha) \\ \beta' &= \psi(\beta) \\ \gamma' &= \psi(\gamma) \end{aligned} \tag{5-32}$$

Where  $\mathcal{J}'$ ,  $\alpha'$ ,  $\beta'$ , and  $\gamma'$  are fuzzy sets containing the features extracted respectively from  $\mathcal{J}$ ,  $\alpha$ ,  $\beta$ , and  $\gamma$ .

In order to find the best match for the query image we shall use the three sets in the comparison process and find the minimum distance between these sets and the corresponding sets in the database images. If we consider the Minkowski distance as the metric measure, then the distance between two fuzzy sets is given by:

$$\begin{aligned} \mathcal{D}^{(p)}(v, u) &= \left( \int_0^n |v(\tau) - u(\tau)|^p d\tau \right)^{1/p} \\ \mathcal{D}^{(\infty)}(v, u) &= \sup_{\tau} |v(\tau) - u(\tau)| \end{aligned} \tag{5-33}$$

For discrete finite case, the above equation becomes:

$$\mathcal{D}^{(p)}(v, u) = \left( \sum_{\tau=1}^n |v(\tau) - u(\tau)|^p \right)^{1/p} \tag{5-34}$$

$$\mathcal{D}^{(\infty)}(v, u) = \max_{\tau} |v(\tau) - u(\tau)|$$

In the above equation, one of the features may dominate others since the features might be in different ranges, thus it can be normalized to be as follows:

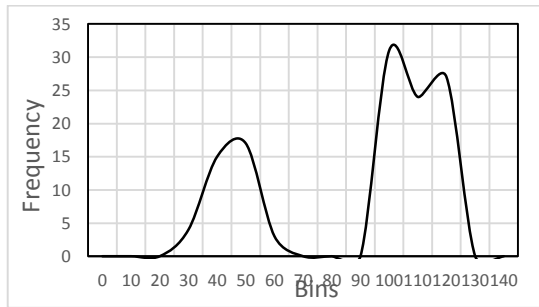
$$\mathcal{D}^{(p)}(v, u) = \left( \sum_{\tau=1}^n \frac{|v(\tau) - u(\tau)|^p}{|v(\tau) + u(\tau)|^p} \right)^{1/p} \quad 5-35$$

In crisp case, the histogram is divided into two parts as shown in Figure 5-7, which is certainly not very accurate since the object contains some information from the background and vice versa, i.e. some of the object information shall be lost and included in the background, as illustrated in the following numerical example.

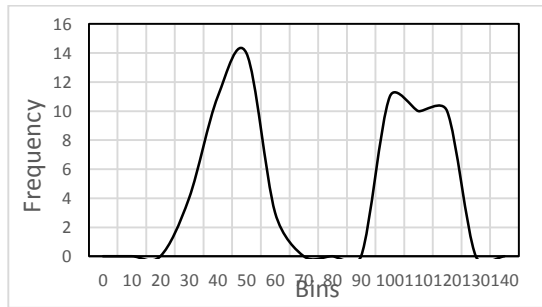
Consider the small image shown in Figure 5-8, which represents a simple image with the object marked in red. In the same figure, (b) shows the histogram of the image and (c) shows the histogram of the region that contains the object.

100	110	120	100	110	120	100	120	110	100	120
110	110	120	100	110	120	100	120	110	100	120
120	120	50	60	50	50	40	100	110	120	100
100	100	60	40	50	60	50	110	40	50	100
40	110	50	40	30	40	30	120	110	120	120
120	120	50	110	100	50	40	100	110	40	100
100	100	40	30	40	50	40	110	100	120	120
50	120	50	100	40	50	30	110	100	40	110
100	100	50	50	40	50	40	110	50	120	100
110	120	100	120	100	110	110	120	100	110	100
100	110	120	110	120	120	100	120	100	110	100

(a)



(b)



(c)

Figure 5-8: Example of simple image and its histogram, (a) image pixels' values, (b) histogram for the entire image, (c) histogram of the region surrounding the object.

In the image given in Figure 5-8, it is clear that some pixels in the object have values similar to the background pixels and vice versa. Therefore, by cutting the histogram into two parts, one for the object and one for the background, the results will not be accurate as shown in Figure 5-9.

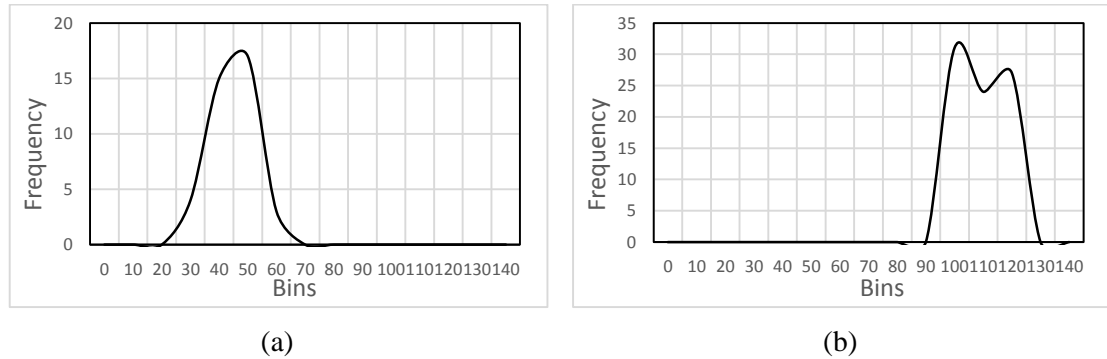


Figure 5-9: Separating the histogram into (a) object histogram and (b) background histogram.

In fact, the histograms corresponding to the object and background, which can be obtained by cutting the image itself, should be as follows:

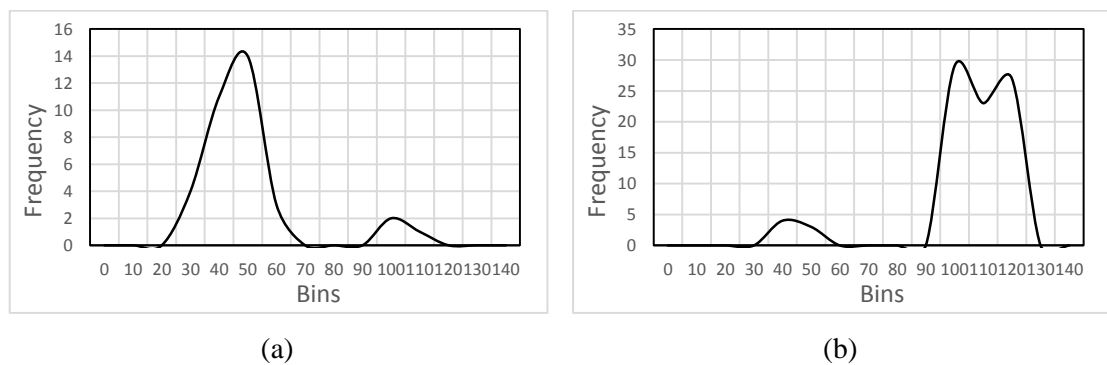


Figure 5-10: Actual histogram for (a) object and (b) background.

From Figure 5-9 and Figure 5-10, it is clear that, although the histograms corresponding to object and background are similar to some extent they are not the same, and this can also be seen by comparing the statistical measures as given in Table 5-5.

Table 5-5: Comparing the Histogram cut with the actual object cut.

Case		Entropy	Mean	Stdv.	Skew
Figure 5-9 Hist. cut	Object-1	0.328759039	2.6	5.590808784	2.21761407
	BG-1	0.138308224	5.466666667	11.39465205	1.728986056
Figure 5-10 Object cut	Object-2	0.583693249	2.333333333	4.353433237	2.124782013
	BG-2	0.329328286	5.733333333	10.79329598	1.668257142

Figure 5-11 visualize the difference in the measures in both cases mentioned above.

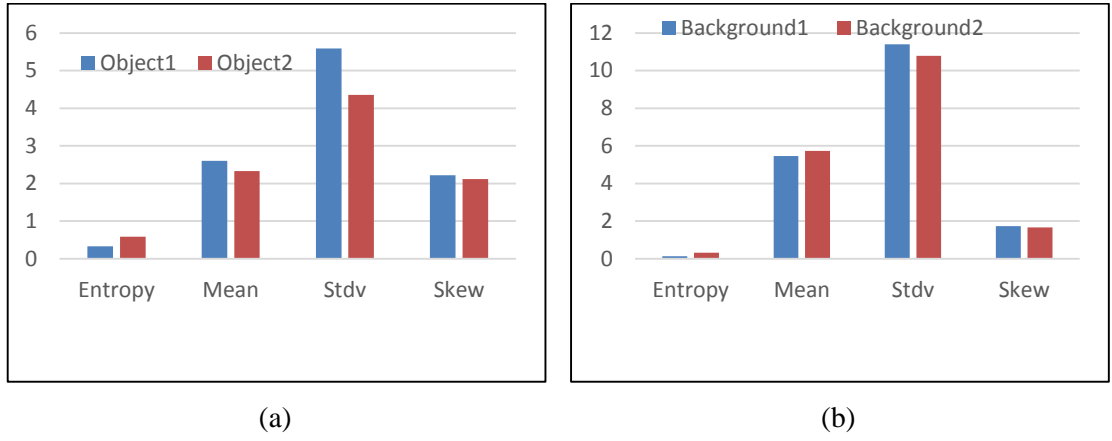


Figure 5-11: Comparing the measures of the histogram cut and object cut, (a) object, (b) background.

Figure 5-12 shows the membership functions for both object and background.

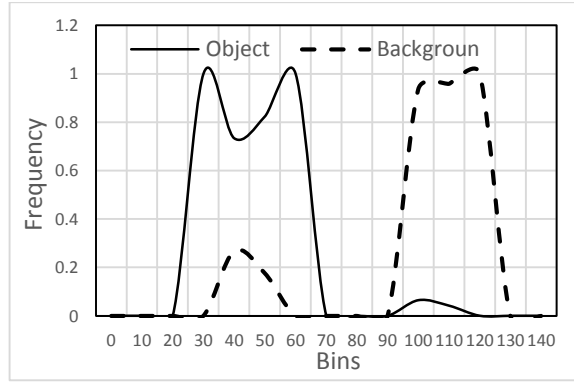


Figure 5-12: Object and background features membership functions.

In the first case, with the histogram cut, we can use the  $\alpha$ -cut to separate the object from the background, i.e.

$$\begin{aligned}\alpha' &= {}^T J' \\ \beta' &= J' - \alpha'\end{aligned}\tag{5-36}$$

where  ${}^T J'$  is the  $\alpha$ -cut of the features set  $J'$  at threshold  $T$ .

In the object cut, which is more accurate than the histogram cut, the features set can be extracted as follows:

$$\begin{aligned}\alpha' &= \mathcal{F}(\alpha(x).J(x)) \\ \beta' &= \mathcal{F}(\beta(x).J(x))\end{aligned}\tag{5-37}$$

In order to compare the contents of two different images, we shall use the features sets given in Eqn. 5-37. The similarity between two objects can be found by calculating the

minimum distance between the object feature sets in the query image, against other images in the database.

Based on the discussion above, one can identify three possible ways of performing the retrieval process. The first one is by comparing the features extracted from the image as a whole; the second retrieval process uses the region of the identification process, and finally the use of the object only in the identification process. We shall refer to these three possible ways as:

- 1- Entire Image Identification (EII).
- 2- Region-Based Image Identification (RBII).
- 3- Object-Based Image Identification (OBII).

In order to study the efficiency of the three aforementioned ways we shall carry out a comparison among them, first by calculating the dissimilarity between two images with same object but with different background and colour distribution and second, with images with different objects but with similar background. Table 5-6 shows the images that have been used in the comparison and the results obtained.

Table 5-6: the distance between pair of image using EII, RBII, and OBII.





Case	Images (1)	Images (2)	EII	RBII	OBII
1			213	309	30
2			12	99	150

Figure 5-13 shows the graph of the distances given in the above table. It is clear from this figure and Table 5-6 that in the first case, the two images are related to each other since both contain the same object, but the distance between them using EII is very large. This is because of the effect of the background of the object, thus, and based on the machine retrieving process, they will not be marked as similar. On the other hand, the



two images in the second case are different but the machine considered them as similar and the distance between them is small. By using RBII, the same results might be obtained but in the second case RBII considered the two images as being not very similar, as was the case with EII.

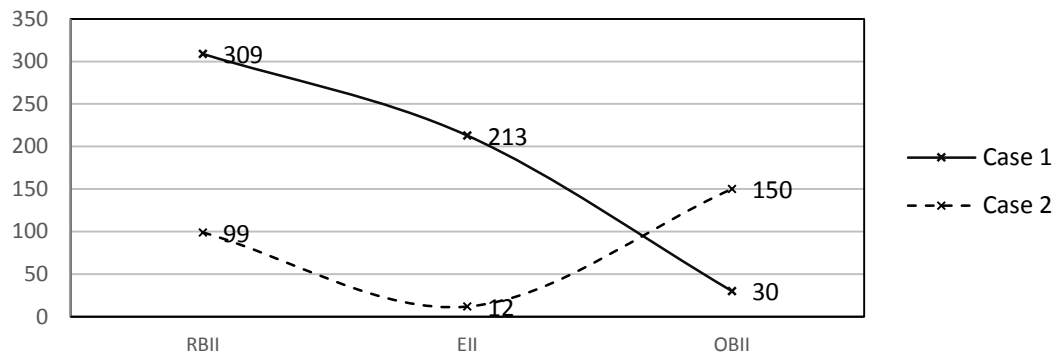


Figure 5-13: Distance measures for the three possible cases, RBII, EII, and OBII

When applying the OBII, the results were reasonable, since the images in case (1) are more similar than the images in case (2). This is clear from the distance between the images, and is because we have ignored the background and focused only on the objects in the images.

## 5.6 Evaluation of Image Retrieval Techniques

To evaluate the results of the three ways of image retrieving, several evaluation measures can be used; one of them is the Precision and Recall Curves. These curves give good visual representation to the result obtained by extracting the values of precision and recall and drawing the relation between them. Figure 5-14 shows the precision-recall graphs, from which one can notice that the EII gave some irrelevant results due to the effect of the background, which dominate the features of the object itself. Similar results might be obtained from applying PRII but with less effect from the background. The best results are obtained by applying OBII in which only the object is compared.

In order to make the evaluation more reasonable, we shall suggest a weighted efficiency evaluation measure (WEEM). In this method, the order of the retrieved images is considered in the evaluation process. In this evaluation measure, images retrieved first will have higher weight than the ones retrieved later.

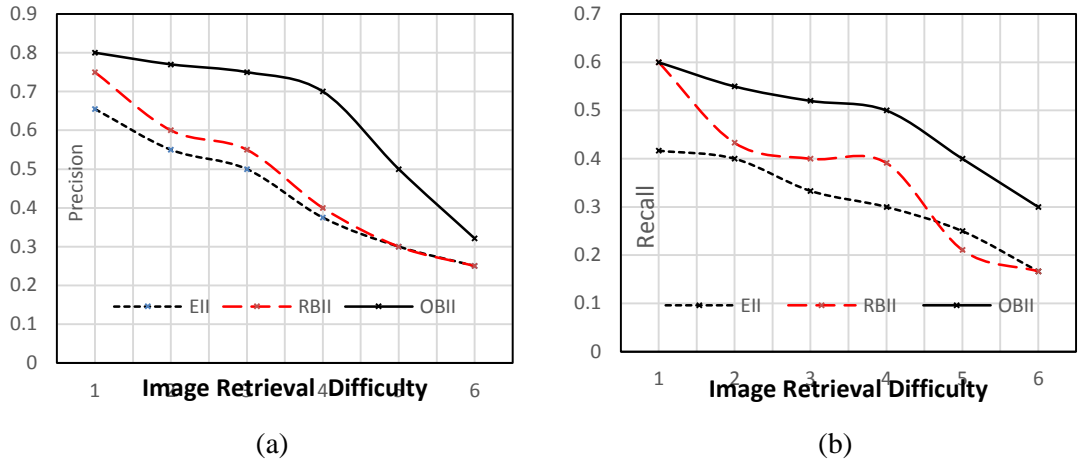


Figure 5-14: Evaluation of the three retrieval ways, (a) Precision, (b) Recall.

The efficiency evaluation measure (EEM) can be calculated by dividing the number of relevant retrieved images by the number of retrieved images, i.e.

$$EEM = \frac{N_{RR}}{N_T} \quad 5-38$$

where EEM is the efficiency evaluation measure,  $N_{RR}$  is the number of relative retrieved images, and  $N_T$  is the number of retrieved images in the ideal case in which all the images retrieved are correct. By considering the order of the retrieved images, EEM is modified by adding weight to the number of retrieved images based on the retrieval order, i.e.

$$N_{RR} = \sum_{i=1}^N K(N - i + 1)$$

$$K = \begin{cases} 0 & \text{for irrelevant images} \\ 1 & \text{for relevant images} \end{cases}$$

$$N_T = \sum_{i=1}^N (N - i + 1)$$

$$WEEM = \frac{\sum_{i=1}^N K(N - i + 1)}{\sum_{i=1}^N (N - i + 1)} \quad 5-39$$

By applying the measure given above, the efficiency of the retrieval process is given in Figure 5-15. In this figure it is shown that OBII efficiency is far better than that of EII and RBII.

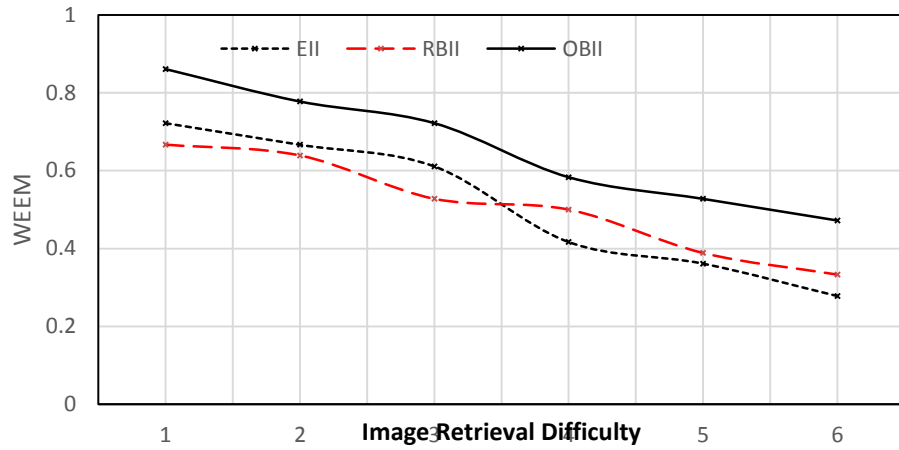


Figure 5-15: Retrieving evaluation using WEEM.

## 5.7 Object Identification Experimental Results

The above system was tested against the new dataset which we constructed for this purpose and the obtained results were compared with those obtained using traditional techniques. The three cases, EII, RBII, and OBII, have been tested and compared.

Figure 5-16 shows the graphs of the precision, recall, and the WEEM measures. From the graphs, it is clear that the presence of the background affects the results drastically, thus the less background in the image, the better the retrieval efficiency is.

Instead of retrieving images similar to the query image from the database, one could identify the object itself and tag it, and then we only need to search from images with a similar tag. For this reason, we have created a new database which contains images that were tagged by humans. These images were used as ground truth data to evaluate the efficiency of our algorithm.

The image content tagging was applied on the constructed dataset: different objects have been defined and stored in a database. The algorithm receives an image from the user, extracts the salient objects from it, and then it tags the salient objects. Table 5-7 shows the results obtained from applying the aforementioned algorithm on a set of images containing 200 images with different objects. The low ratio of the correct tagging for the human is because the features of the human are changed due to different factors such as the clothes, while the similarity between the stop sign and the red flowers has reduced the accuracy a little.

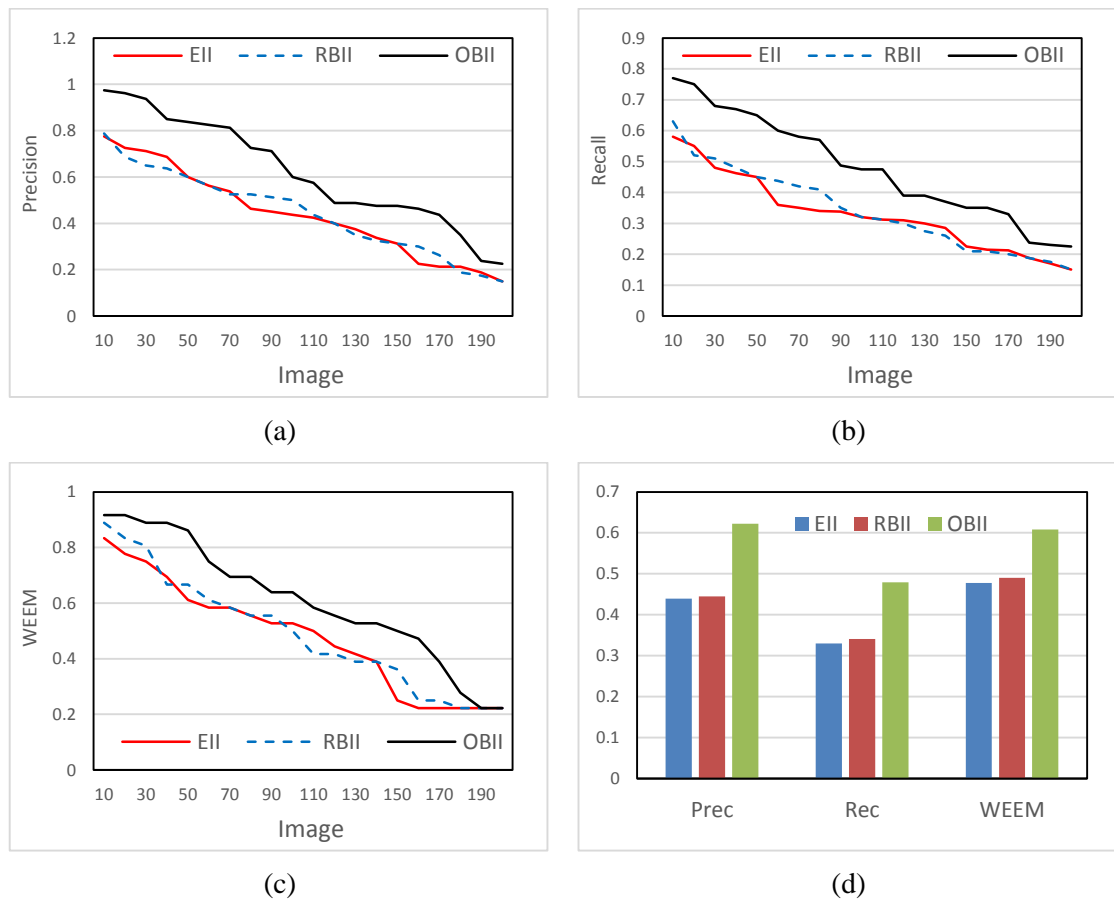













Figure 5-16: Evaluation and comparison among EII, RBII, and OBI, (a) precision, (b) recall, (c) WEEM, (d) average measures.

Table 5-7: The results of tagging the contents of the image.

Object	Cow	Human	Red Flower	Yellow Flower	Traffic Sign	Brown Horse	Average
% Correct	95	87	90	93	90	89	90.67

Table 5-8 shows a sample of the results obtained from applying the saliency-based image auto tagging. It is clear from the results that the background has been removed and only the salient objects have been extracted and tagged. The background will be identified in the background identification phase. In case (1), the white cow has been identified as human for many reasons; first, the object is very small and it contains part of the background, and second, ‘white cow’ is not defined in the object database, so the closest object to its features was ‘human’.

Table 5-8: Sample of the results obtained from applying the auto-tagging algorithm.

Case	Image	Salient objects / tag			
1					
		human	Cow		
2					
		Yellow Rose	Yellow Rose		
3					
		Human	Human	Human	Ball

## 5.8 Background Identification

Background identification is important and its importance is not very much less than that of object identification. For humans, if they want to describe a scene, the background is part of the description, such as a bird in the sky or a bird in a field, etc. The most important feature of background is its regularity. Usually backgrounds are of regular and repeated context: clouds, grass, brick walls, sands, etc. are examples of possible backgrounds. The regular and repetitious nature of the backgrounds are good reasons, in addition to others to be discussed soon, that make texture identification a perfect descriptor for background.

Texture analysis was discussed in Section 5.3. In this Section, we shall discuss how to utilize texture analysis in background identification. In addition, we shall introduce an algorithm into the identification process using Artificial Neural Networks (ANNs) and GLCM. The main advantage of using ANNs is that they can be trained on one set of inputs and thereafter can be used to identify other inputs.

Supervised learning can be used in the training process since both input and output for the dataset are available. The main limitation of ANNs is the need to retrain them with every new input, i.e. they need to forget all that they learned and start learning again from zero. This limits the application of ANNs from identifying images of different natures. Thus, ANNs were not adopted in objects identification since with every new object they need to be retrained.

### **5.8.1 Artificial Neural Network Based Background Identification**

ANNs have been adopted in many image processing applications and in pattern recognition due to their ability to be trained. Appendix B contains the necessary theory and background of the Neural Network.

Different ANNs have been used in the field of image processing and identification. The following is a survey of the state-of-the-art research in the field of image processing and neural networks. Self-Organising Maps ANN was used by Laaksonen et al. (2001); they have used it in implementing the relevance feedback phase of their retrieval algorithm [202]. Hybrid Neural Networks were used in image classification which is used in the image retrieval process by Tsai et al. in 2003 [203]. Relevance feedback has been widely used in Neural Network-based image retrieval systems as in ref. [202], [204], [205], [206] and [207]. Many systems have utilized Radial Basis Function Network to improve the performance of Neural Networks in retrieving images [208], [207]. Lee & Yoo, (2001) introduced a Neural Network-Based CBIR system and a Human Computer Interaction approach to CBIR using the Radial Basis Function (RBF) network [208]. Nematipour et al. in 2011, proposed what they called Enhanced Radial Basis Function Network and Relevance Feedback to design an effective image retrieval mechanism in CBIR [207]. Liu and Mio proposed in 2007 an algorithm that uses spectral histogram features to describe the spatial relationship among pixels as input to their neural network to organize images for CBIR [209]. A Neuro-Fuzzy approach was used with the 2-D wavelet Transform by Balamurugan and Anandhakumar in 2009 as a tool for clustering the images in CBIR techniques [210]. Cubic – Splines Neural Networks have been used to reduce the gap between high level concepts and low level visual features that are used in image retrieval [211] [212]. The application of multilayer neural networks was suggested by Rao et al. in 2010 for automatic image retrieval [213].

From studying the published researches about the use of ANNs in image contents identification, one can note that most of the proposed algorithms were utilizing neural networks in identifying images that belong to similar domains. In other words, the ANN is useful in applications with images datasets that belong to the same domain such as medical, MRI, etc. Applications such as cloud identification and classification [214] [215], MRI [216] [217], tumour identification [218] [219] [220], rock identification [221], fabric defects identification [222], and other applications are examples of the use of ANN in image contents identification. In all the applications listed above, one may notice that the images dataset have relevance to each other i.e. they are from same domain.

Using neural networks in image identification is not always feasible in all applications, several constraints should be considered when using them. First, it is not practical to use the entire image as an input to the ANN; in such cases, the number of inputs would be huge since the number of input nodes is equal to the number of pixels in the image. The second constraint is that they are not practical for identifying different images with no common features among them, although they are very useful in identifying images with common features, such as in tumour identification, pattern recognition, and face recognition since most of the images are with the same nature and having common features. Thus, Neural Networks are more suitable for matching than identification, e.g. searching for matched faces with different poses from the query face.

Based on the above discussion, we shall use neural network in identifying the background since there are a limited number of backgrounds such as sky, cloudy sky, bricks wall, etc. The neural network will be trained to identify the backgrounds and return the class to which a background belongs.

By studying the nature of backgrounds, it was noticed that the texture feature could work perfectly with them since most backgrounds are textured, such as grass or cloudy sky. The neural network will be trained using different texture samples and then it can be used for background identification tasks.

## 5.8.2 Background Identification Features

Several features, such as, texture and colour, can be used in background identification. Colour alone cannot be used, since different backgrounds may produce similar histograms, and texture alone will not give enough description to identify the background, therefore both colour and texture features shall be used in this process.

Colour features are widely used to describe the contents and the nature of images; the most well-known representation for colour features is the histogram which was discussed earlier. Since, different images may have similar histograms, such as, the sky and the sea, or tree leaves and grass, thus, texture can be used to solve such problems. When referring to the description of the image's texture, one usually adopts texture's statistical feature and structural feature, as well as the features based on frequency domain (spectral) [38], [65]. Although texture is not well-defined like the colour feature, it gives a good description for the contents in the image like cloud, trees, bricks, and fabric.

Texture features can be obtained using the Gabor filter, wavelet transform, and local statistics measures. Our main interest is in the statistical approach which includes many techniques, the most well-known ones are Moments of Intensity and Cooccurrence Matrix. The calculation of the moments of intensity is similar to colour histogram moments calculations. For colour histograms, the colour distribution is used while here, the intensity histogram is used.

GLCMs , which were discussed in Sec 5.3, are used to describe the texture since they measure the relation between adjacent pixels displaced by a certain distance. Due to the symmetric nature of GLCMs it was noticed that the opposite angular displacements such as  $0^\circ$  and  $180^\circ$  will produce transposed GLCMs, or more general  $G(x, y, \delta, \theta) = G^T(x, y, \delta, \theta + 180^\circ)$ . Thus it is not feasible to calculate the GLCMs for all possible angles. Thus, angles such as  $0^\circ, 45^\circ, 90^\circ, 135^\circ$  are sufficient to describe the texture.

From the above discussion one can determine that for each texture image there are GLCM prints corresponding to that image. The shape and size of the GLCMs are varied with the variation of the texture brightness, coarseness, regularities and other features. In this work, we shall adopt two properties of the texture, coarseness and regularity, and the rest of the properties can be concluded from these properties.




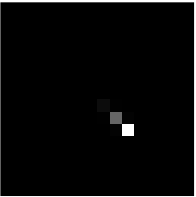
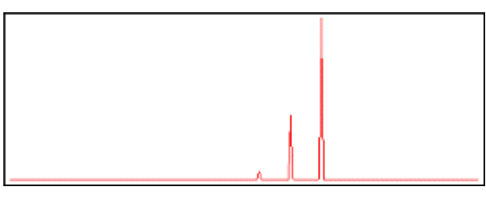

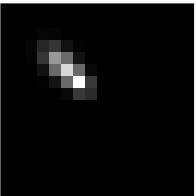
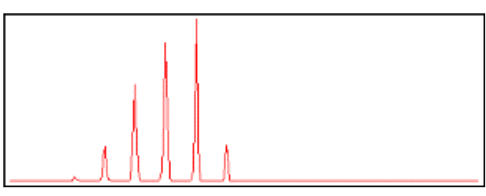
### Coarseness of texture

It was noticed that the shape of the GLCMs is dependent upon the coarseness of the texture; this is because in the case of fine texture, the GLCMs extraction process will produce GLCMs with a clear spot. For instance, assume that we are calculating  $G(x, y, \delta_x, 0)$  which means only the horizontal relationship between the adjacent pixels shall be considered. For coarse texture, if the width of the textures is  $w$ , then if  $\delta_x$  is less than  $w$ , the value of  $g$  at some grey level  $a$  with respect to itself will be higher, i.e.  $g(a, a)$  will be higher than it is in the case of fine texture, in which the ratio of  $\frac{\delta_x}{w}$  is higher than that for coarse texture. In other words, the number of pixels at which  $I(x, y) = I(x + \delta_x, y + \delta_y)$  is higher in coarse texture than in fine texture. Thus, the width of the GLCMs shall be considered as a measure of the coarseness of the texture.

Table 5-9 shows examples of fine and coarse texture. From this table it is clear that the GLCM is concentrated in one region (spot) for fine texture, while its size is larger, and there are some values corresponding to the same grey values, for coarse texture.

The location of the spot in the 2D GLCM or the centre in 1D GLCM has been affected by the brightness of the image, which may affect the result slightly. This problem can be overcome by shifting the centre to a fixed location.

Table 5-9: The effect of coarseness on GLCM.

	Texture	2D GLCM	1D GLCM
Fine Texture			
Coarse Texture			

By testing the most commonly used measures given above, it was noticed that only few measures change in direct relation to the coarseness of the texture.

From Figure 5-17, it is clear that the contrast ( $\mathcal{C}$ ), entropy ( $\mathcal{E}$ ), 1D standard deviation ( $\sigma_{1D}$ ), and the range ( $\rho$ ) are increased with the increase of the coarseness, while other measures change in an irregular way with the change of the coarseness. Therefore, we shall adopt the above four measures in identifying the coarseness of the texture.

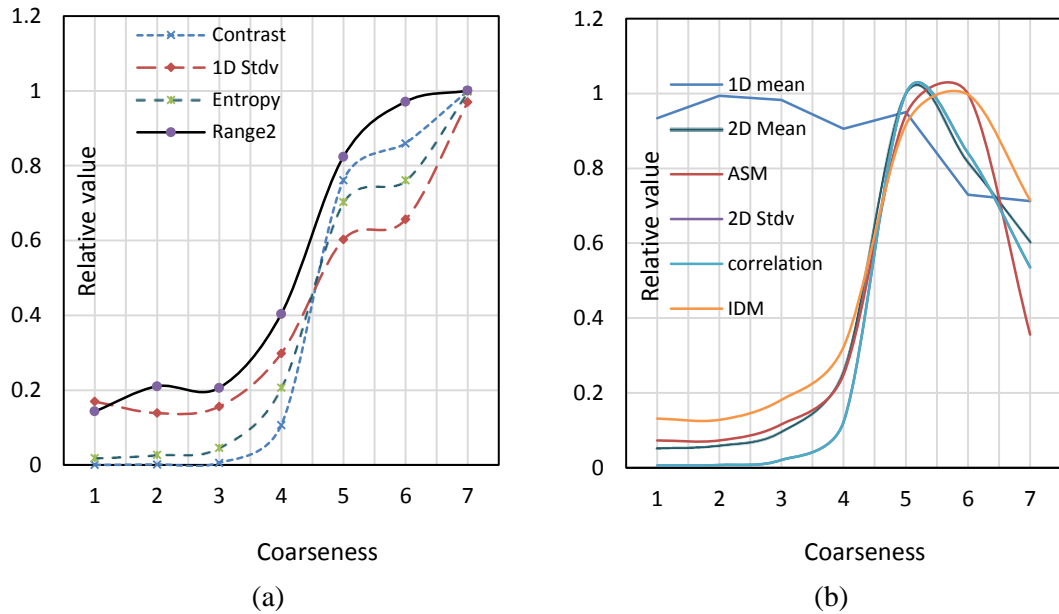


Figure 5-17: the variation of the statistical measures with respect to coarseness (a) Measures that change directly with coarseness, (b) measures that do not change directly with coarseness.

### ***Regularity and Irregularity of texture***

Regularity is another important property of texture that can affect the shape of the GLCM. Regularity measures the repetition of the texture in the image, e.g. grass is a regular texture, since the same texture is repeated continuously, whilst cloudy sky is irregular. The irregularity effect on the GLCMs is characterized by having more than one spot in the GLCM, one spot for each part, e.g. in the case of cloudy sky there are two spots; one for the blue part and one for the clouds. Thus, the GLCM shall be divided into more than one part i.e.  $G = G_1 + G_2 + \dots + G_m$ . Where  $G_i$  for  $i = 1, 2, \dots, m$  are the GLCM corresponding to different parts of the texture, and each one will produce its own spot since they are calculated for a specific range of grey levels. Figure 5-18 shows the effect of irregularity on GLCMs; from the figure, it is obvious that there are two major peaks for the 1D GLCMs, one for the clouds and one for the blue background; and two spots in the 2D GLCM for the same reason.

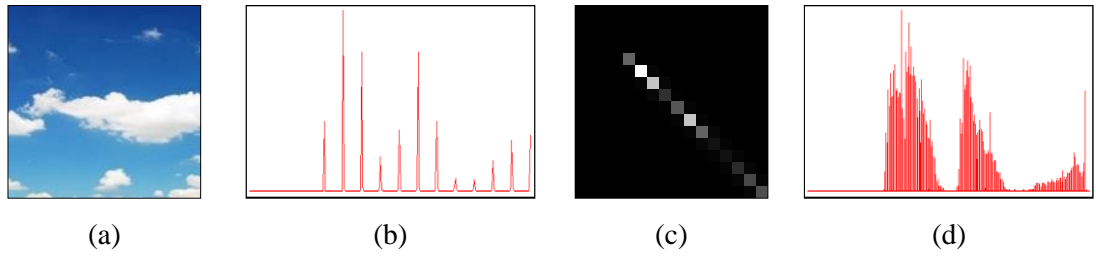


Figure 5-18: The effect of irregularity on the GLCM, (a) original image, (b) 1D GLCM 16 grey levels, (c) 2D GLCM 16 grey levels, (d) 1D GLCM 256 grey levels.

To measure the texture features we shall use different measures that can be extracted from 1D GLCMs, such as the mean  $\mu$ , standard deviation  $\sigma$ , range  $\rho$ , number of peaks (tops)  $\tau$ , and average of the tops  $\alpha$ . Each one of the measures is affected by the properties of the texture, e.g. the coarseness affects the range, and the regularity increases the number of peaks. As shown in Figure 5-19, only the range and the number of tops change with the increase in the irregularity of the texture, while all other measures change randomly and in irregular ways.

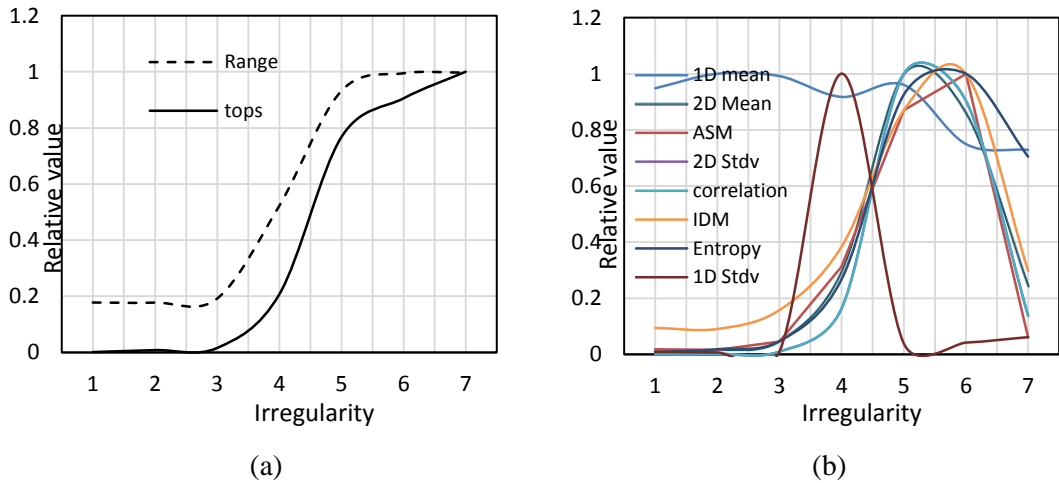


Figure 5-19: The variation of the statistical measures with respect to regularity (a) measures that change directly with regularity, (b) measures that do not change directly with regularity.

To test the GLCM as a measure of texture similarity, different numbers of grey levels such as 8, 16, 32 and 256 have been used, from which GLCM with sizes of  $8 \times 8$ ,  $16 \times 16$ ,  $32 \times 32$  and  $256 \times 256$  have been obtained. There should be some kind of optimization between the information in the image and the size of the GLCM. Selecting 256 grey levels will result in having a very large GLCM ( $256 \times 256$ ), which will give a vector of size of (65536). This means a neural network with more than six thousand input neurons if we want to use this vector as an input to an ANN. Selecting a lower

number of grey levels may degrade the information in the image, but it will reduce the number of inputs to the ANN.

The use of 16 grey levels gave good results since the degradation in the texture is small and the produced GLCMs give sufficient description, as shown in Figure 5-20, which shows a comparison between 256 and 16 grey levels GLCMs.

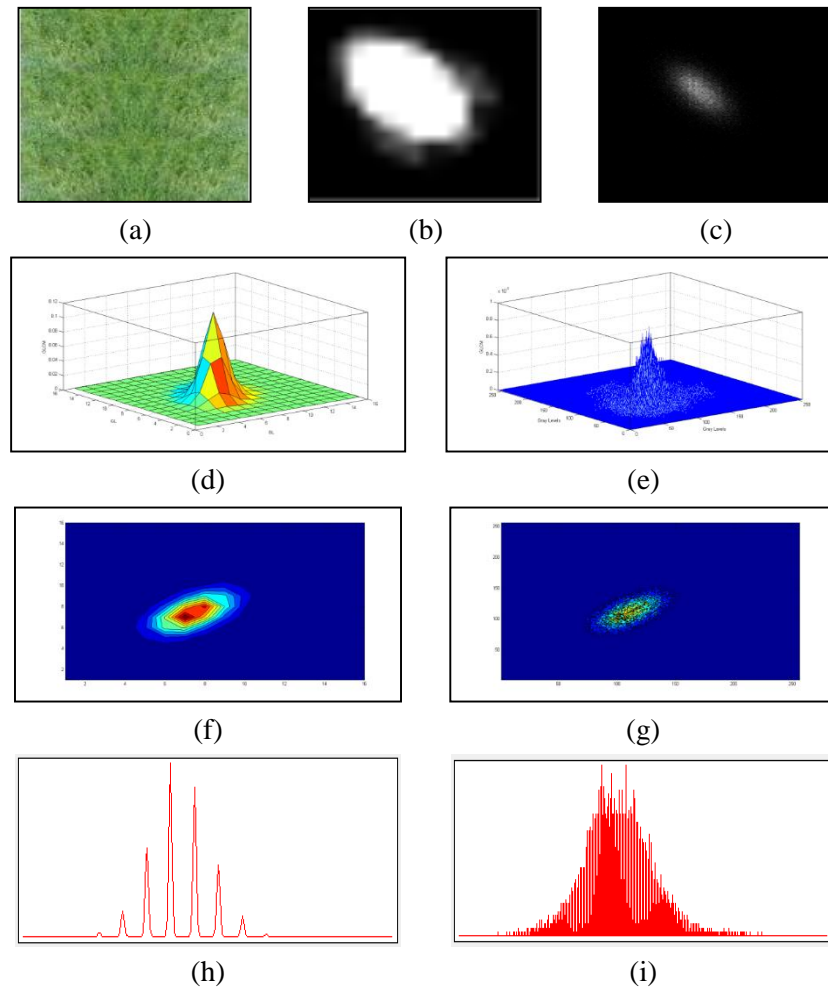


Figure 5-20: GLCM, (a) original image, (b) 16 grey level GLCM, (c) 256 grey level GLCM, (d) 16 grey level 2-D GLCM histogram, (e) 256 grey level 2-D GLCM histogram, (f) 16 grey level 3-D GLCM contour, (g) 256 grey level 3-D GLCM contour, (h) 16 grey level 2-D GLCM histogram, (i) 256 grey level 2-D GLCM histogram.

### 5.8.3 Neural-Fuzzy-Based Background Identification

From the above discussion, one may notice that there is a major limitation with neural networks, which is their need to be retrained with every new image type, so they fail to be useful for general image recognition. However, they work perfectly with images of the same nature or which are related to the same domain such as medical, x-ray, MRI,

fingerprints, etc. The reason that they work well with such kinds of images is that they are usually used to identify abnormal cases in a set of similar images such as identifying the tumour in case of medical images. For example, a set of brain images maybe fed to the network and the network may then be requested to identify any abnormal cases. This process is not suitable in the case of identifying the contents of different images with different natures. In the case of images with a different nature, one needs to train the neural network to identify some images, but if a new image needs to be identified then the new image must be added to the training images set and the network has to be retrained from the beginning.

In order to overcome the abovementioned problem and to design an algorithm inspired by the human perception system, we shall consider using a bank of neural networks in addition to a Fuzzy interfacing system. The neural network will be used to identify the features of the images and not the images themselves. The proposed algorithm has been built upon the following assumptions,

- 1- No frequent retraining for the NN should need to be performed,
- 2- Identifying texture coarseness and regularity in addition to colour using NN,
- 3- Using Fuzzy intelligence to combine the results obtained from the NN to decide the texture type.

The role of the ANN in the proposed algorithm is to identify the background properties such as coarseness, regularity, and colour. We shall use a bank of ANNs consisting of three ANNs; the first one is to identify the coarseness, the second is to identify the regularity and the last one is dedicated to identifying the colour. The inputs to the first two ANNs are the measures of the reduced 1D GLCMs. The idea of using the reduced GLCMs and not the original one is to reduce the complexity of calculating the measures of the input vectors. The reduction in GLCM will not affect the result significantly since the GLCMs will maintain the correspondence with the texture.

The input to the third NN is the colour components of the texture image. The Hue Saturation and Value (HSV) colour system is used in identifying the colour since it is analogous to HVS and Munsell Well. It is possible to obtain most of the Fuzzy Standard Colour (FSC) using HSV. In FSC, we aim at finding a way to describe the colour in a

way similar to a human's description, which means, to give a possessive value for the colour, which might be represented by a Fuzzy membership function. For example, we may describe the colour as Reddish Brown, which means the brown membership value is higher than that for red, but there is still a value for red.

Figure 5-21 shows the block diagram of the proposed algorithm. ANN1 is used to identify the coarseness, ANN2 is used to identify the regularity, and ANN3 is used to identify the colours.

The FIS consists of the linguistic database which contains linguistic variables such as Colour= {Green, Brown, Blue, etc.}, Coarseness= {very fine, fine, course, very Coarse}, Regularity= {very regular, regular, irregular, very regular} and finally texture = {field grass, beach sand, desert, sky, cloudy sky, etc.}. In addition to the linguistic database, a rule base should be defined as well. Rules such as “**IF** the coarseness **IS** smooth and the Regularity **IS** very regular and The Colour **IS** light brown **THEN** the texture **IS** Desert Sand” can be used to identify the texture.

The above algorithm has been applied on different sets of backgrounds with different textures. The standard colours that were used to train the neural network were obtained from different users feedbacks with different backgrounds. The standard colours set contained colours like red, brown, green, blue, yellow, white, black, orange, and grey.

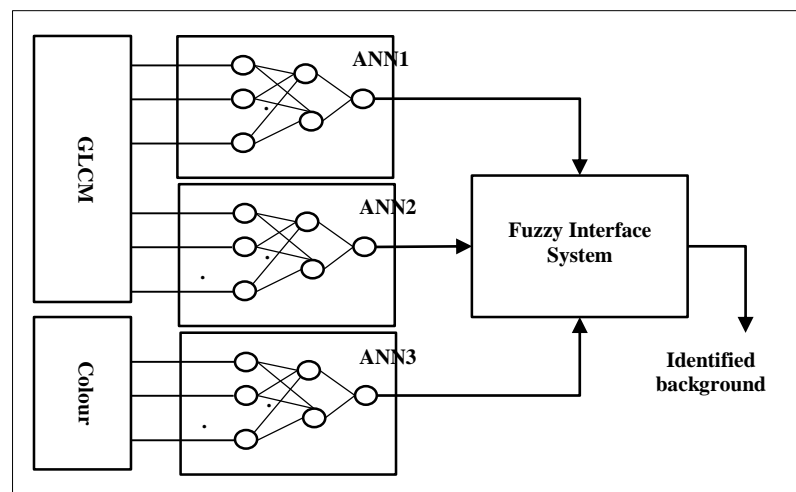


Figure 5-21: The proposed algorithm block diagram.

The colour experimental results are shown in Figure 5-22, in which different colours have been labelled based on the Hue value with a constant saturation and value.

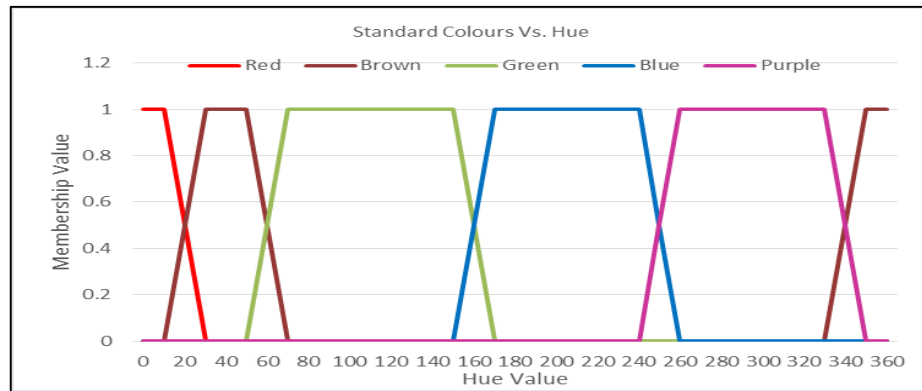


Figure 5-22: FSC vs. Hue, Saturation=0.5 and Value = 0.5.

Some other FSC can be obtained by changing the values of the saturation and value. The names of the FSC along with the HSV values are used to train the neural network ANN3, and then this is used to identify other colours. ANN3 has a number of outputs equal to the number of FSC used. Thus the output corresponding to a specific colour is considered as the membership value for that colour, e.g. if the output of the red colour is 0.7 and that of the brown is 0.3, that means the colour is brownish red.

In the same way, the output of ANN1 will give how coarse the texture is, e.g. if the output is 0.9 this means very coarse; 0.7 means coarse, and 0.2 means fine. In addition, ANN2 gives how regular the texture is. The output of the neural networks then is used to define the type of the texture in FIS.

To test the output of the ANNs, consider Table 5-10, in which different images have been tested. The coarseness ranges from (−1) smooth to (1) coarse. Any value between these two values is a measure of how coarse the image is. Similarly, the irregularity ranges from (−1) regular to (1) irregular.

Table 5-10: Sample examples on the ANNs outputs.

Image						
Av. Col.						
ANN o/p	0.98 Brown	0.99 Brown	0.95 Brown	0.93 Green	0.75 Green	0.87 Blue
Coarseness	-0.12	-0.85	0.71	0.20	0.039	-0.06
Irregularity	0.21	-0.89	0.31	-0.20	0.82	0.67

## 5.9 Background Identification Experimental Results



The Neuro-Fuzzy system was used in this phase to identify the background, the ANNs details are given in Table 5-11. The table gives information such as the number of neurons and time elapsed training each ANN.

Table 5-11: Neural Networks details.

NN	Number of Neurons in			Training Images	Epochs	Elapsed Time (sec)	Min Error
	input	hidden	output				
NN1	5	5	1	20	1300	0.2	0.099
NN2	5	5	1	20	25000	2.3	0.099
NN3	3	3	10	20	11300	2.5	0.099

Table 5-12 shows an example of the results obtained from applying the proposed algorithm:

Table 5-12: measures obtained from applying the proposed algorithm.



#	Image	$\mu$	$\sigma$	$\rho$	$\tau$	$\alpha$	Colour
1		104	23	151	1	1	0.6 Green 0.1 Brown
2		137	56	187	3	0.64	0.9 Blue 0.002 Green

From Table 5-12 one can notice that in case 1,  $\sigma$  is smaller than that in case 2, which means that the first texture has higher coarseness than the second one. The value of  $\rho$  in (1) is higher than that in (2) for the same reason. The number of peaks in (1) is 1 while in (2) is 3 which means that the first image is regular and the second one has less regularity.

By using the trained neural networks and the FIS, the results shown in Table 5-13 have been obtained.



Table 5-13: Results obtained from applying the proposed algorithm

#	Image	Coarseness	Regularity	Colour
1		0.87 Coarse	0.78 Regular	0.6 Green 0.1 Brown
2		0.73 Fine	0.68 Irregular	0.9 Blue 0.002 Green

Based on the results in the above table, the description of the image in (1) will be a medium coarseness, green grass, and the image in (2) is day, blue, partially cloudy sky.

The algorithm was tested using 250 different background images and the results are shown in Table 5-14.

Table 5-14: The percentage of correct results.

Colour	Regularity	Coarseness	Class
88%	84%	91%	85%

It is noticed from the obtained results that the algorithm worked perfectly on backgrounds with textures such as grass, sand, trees leaves, sea, clouds, stones, and many others. The percentages of wrongly classified backgrounds were in fine clouds and in bricks. In the first case the fine clouds were similar to sea and in the case of bricks, the texture of the brick itself was considered and not the texture of the wall.

Another important merit of the proposed algorithm is that, due to the use of Fuzzy intelligence, it is possible to have more than one class for the same image, for example the image shown in Table 5-15.

Table 5-15: A texture may belong to more than one class.

Image	Colour	Coarseness	Regularity	Class
	0.99996 Brown	0.12	0.9	Sand
	0.99 Brown	0.98150	0.76	Gravel
	0.9933 Brown	0.44	0.78	0.4 Coarse Sand 0.6 Fine Gravel

## 5.10 Conclusions

In this chapter, we have studied the main two features that are used as image similarity measures, which are the colour and texture features. These features were used to support our efforts in identifying the image contents based on the salient regions or the object they contain. A saliency based image contents identification algorithm was developed which used the principle of region and object saliency to identify the contents of an image. The developed method was compared with traditional techniques and it was shown that the obtained results are more reasonable than those obtained from traditional techniques.

A Neuro-Fuzzy Background Identification Algorithm has been developed as well. By studying the nature of the backgrounds, it was found that the texture and the colour feature could be used efficiently to recognize the background. Many features, and how they are affected by the coarseness and regularity of the background, have been studied. The proposed system has overcome the neural nets retraining problem by including a fuzzy interfacing system.

# Chapter 6

## Discussion and Conclusions

### 6.1 Introduction

This Chapter presents the general conclusions and discussions based on the obtained results. It also summarizes and reviews the contents of each Chapter in the dissertation. As the results have been discussed in detail separately in the concerned Chapters, the discussion here shall be more general. In addition, the general conclusions about the findings shall be derived in the second part of the Chapter. Finally, just as we presented the intended contributions in Chapter 1, the achieved contributions shall be discussed in detail here.

### 6.2 General Review

In this section, a general review of the main topics discussed in the thesis shall be introduced. The thesis consists of six chapters and covered the theoretical and empirical parts of the work. In Chapter 1, a general introduction was presented and different problems with the available machine vision techniques were addressed. The main problem, the one which is the main concern of this thesis was identified. The motivations for this research have been discussed as well, in addition to the feasibility of some techniques such as computational intelligence. The intended solutions for the main problem were introduced in terms of intended contributions to the field of research.

In Chapter 2, the main concern was to present the necessary theoretical background and review important techniques useful in deriving the proposed algorithms and theory.

Image processing, feature extraction, segmentation and clustering were presented in this chapter since they are utilized in different stages of the proposed algorithms development.

In addition, computational intelligence techniques such as adaptive learning and fuzzy logic were discussed since they are important in developing an image identification system which is inspired by the human vision system.

By studying the human vision system and the machine vision system, the main strengths of the human vision system have been identified. The utilization of these strengths in improving machine vision systems has been discussed. Human attention phases were discussed since they have been selected to be used in the improvement process.

In Chapter 3, the data collection and analysis was presented, different datasets both for saliency identification and for image retrieving were presented and discussed in details. Some necessary algorithms for extracting ground truth data were presented as well and the obtained results were analysed.

The main contributions of the dissertation were presented in Chapters 4, and 5. In Chapter 4, saliency extraction techniques, as a representation of human attention, were analysed and discussed. By analysing the existing algorithms, most of their drawbacks and weaknesses were identified and therefore we were able to design a novel saliency extraction technique that utilises the principles of human attention. In the proposed technique, we have introduced irregularity as a new approach for saliency identification. Irregularity approach was built based on the fact that usually human vision is attracted by irregular regions, such as a ball in a playing field. Based on this definition, we have developed new saliency extraction algorithms, which utilize both local and global features of the regions to mark them as salient or non-salient. The necessary theory, measures and mathematical formation for the new approach were derived and new algorithms developed. The proposed algorithms were tested against most of the existing saliency extraction algorithms using various datasets and it was found that our saliency extraction algorithms give very satisfactory results as compared to existing methods and the ground truth data.

Other techniques such as thresholding and filtering were needed in the proposed saliency extraction algorithms. Thresholding was studied thoroughly and a thresholding technique

proposed. The proposed technique uses the principles of bimodality of histogram to isolate the regions from each other and fuzzy logic to resolve the issue of incertitude of the image.

In order to form salient regions from the salient points, we had to use clustering algorithms. Most of the existing clustering algorithms have been studied and summarised and a blob-based clustering algorithm developed. The developed clustering algorithm is suitable for clustering gaze points and salient points since it assumes the point to be a small region surrounding the point. This was also derived based on the nature of the human vision system, that is, the human does not look at a particular point but he looks at the surrounding small region instead. Based on this fact, blobs principles and region merging techniques have been utilized in the developed clustering technique. The algorithm is iterative and the blobs grow with iteration: every overlapped blobs are merged together to form a new larger blob.

Finally, in this chapter, we have studied most of the existing saliency evaluation techniques, and suitable evaluation techniques have been developed which are useful with applications such as our application in this work. The evaluation technique considers the nature of the result needed to be compared with the ground truth data. Some techniques have considered the point distribution in the image and some others have considered the region instead of points. The pros and cons for each method have been studied and discussed.

Image contents identification was the focus of Chapter 5, where most of the image identification techniques and similarity measures approaches have been analysed to develop an algorithms that can overcome the existing methods' limitations. In this system, only the salient objects are identified and used in image contents description. Since the image by its nature is not crisp and contains some incertitude, both geometrically and spatially, then fuzzy logic has been considered in most of the comparison methods in this investigation. In object identification, we have considered fuzzy membership functions principles to form sets of features for the object and the background. In addition, background identification has been discussed as well and a neuro-fuzzy background identification system was proposed. The proposed system uses

the image description which can be extracted from the colour and texture features to describe the background and hence to identify it.

Again, and in order to benchmark the proposed algorithm with other existing algorithms, most of the available retrieving evaluation methods were studied. The merits and demerits of each method were analysed and discussed. An algorithm for retrieval evaluation has been developed. This algorithm gives importance (rank) to the image based on the retrieval sequence, i.e. the image retrieved first will be ranked higher than the next image, and so on.

### **6.3 Academic Contributions**

In the following, we shall discuss the main contributions that the thesis has achieved:

- 1- Saliency definition and extraction: Human attention has been reviewed and studied in order to utilize it in improving image contents identification. Human attention has been simulated in salient point extraction, which was used to highlight the important objects in the images. In this context, a new approach for saliency extraction has been suggested based on the irregularity of the regions with respect to the image. A region is considered as salient region when it is irregular in comparison to other regions. In addition, we have suggested a new measure, dependent upon the statistical measure, and the necessary theoretical derivations have been derived and proved.
- 2- Gaze and salient points clustering: Clustering is another important challenge we need to handle as it is used to convert the salient points into salient regions. The proposed technique utilises the principle of image formation in human eyes and the blobs technique. The developed algorithm is fully automatic, which means it does not need any parameters to be specified, such as number of clusters or threshold value, as is required by other clustering techniques. Instead, we have defined a stopping criterion which can be used to terminate the clustering iterations.
- 3- Automatic thresholding: Thresholding is an important factor in most of the image processing algorithms; therefore, it was a crucial issue to develop an automatic thresholding technique that can be used in different image processing algorithms.

A thresholding technique, which utilizes the fuzziness nature of the image and the bimodality of the histogram, was developed. The necessary theory of the proposed algorithm was derived, discussed and compared with other existing techniques. The basic idea of the proposed technique is to represent the histogram with two membership functions and divide the image into four regions rather than two. This approach has been applied to identify the salient and non-salient regions in the saliency map.

- 4- Image Identification: In this work, we have introduced a competitive image contents identification algorithm that identifies the objects in the image as well as the background. Therefore, two identification algorithms have been developed, objects identification and background identification. The objects identification algorithm has been developed to identify the salient objects only and a Saliency Based Image Retrieval (SBIR) technique has been developed. The algorithm has used common features such as colour and texture to describe the salient object in an image. In addition, an image background identification algorithm has been developed which utilises the computational intelligence and colour-texture features in identifying the contents of an image. The problem of neural network retraining has been considered and solved by using a new neural-fuzzy structure.
- 5- Evaluation: In order to evaluate the proposed algorithms, we have studied the available evaluation techniques, and in some cases, we had to develop evaluation algorithms suitable for the application at hand. Two evaluation techniques have been developed; the first one is for evaluating the saliency extraction results by comparing them with the available algorithms and with the ground truth data. Image retrieval evaluation is the second developed evaluation technique. It was noticed that most of the available evaluation methods do not consider the order of the retrieved data, and usually they find a percentage of correctly retrieved images to images in the database. In fact, the order of the retrieved images is important, thus we have given a weight to every retrieved image based on its order in the retrieval process e.g. the first image will get the highest weight then the next one will get the next highest, and so on.

## 6.4 Conclusions

This work aimed to design an Image retrieval system that is inspired by the human visual system to improve overall performance and to reduce the gap between the human and machine visual systems. It is well accepted that the human visual system evolved over millions of years to be perfect for scene interpreting and recognition. Therefore, using features from this vision system and applying them in a machine vision system shall drastically improve the results.

In this work, we have investigated and reviewed most of the state-of-the-art techniques in different fields such as human attention, saliency extraction, CBIR, and image identification. The main concern of this work was to add some semantic capabilities to the machine vision techniques and to design an image retrieving system inspired by the human vision system, therefore the human vision system was studied and some of its functionalities simulated and implemented. In addition, algorithms' automation was one of the concerns of this work as it was aimed to design fully automatic algorithms that do not need any human intervention or parameter settings; therefore, many algorithms have been developed to achieve this automation.

The proposed algorithms were supported with the necessary theoretical concepts and mathematical modelling and proofs. In addition, they were applied on standard datasets and compared with other existing algorithms both qualitatively and quantitatively. The benchmarking and comparisons showed the feasibility of the proposed algorithms as discussed thoroughly in each chapter. In some cases, there was a need to adopt datasets other than the standard datasets, as they were too simple and did not give any impression about the efficiency of the algorithms. In such cases, we constructed new datasets and tested the existing algorithms against our dataset, as discussed in the image identification datasets.

The importance of this work lies in developing a system that can label the objects in an image and describe the image based on these labels in addition to the background description. Furthermore, with this system, it is possible to search for an object in the images that are stored in a database even if they contain more than one object; this feature was not available in most of the visual search techniques without human intervention.



Identifying and labelling the salient objects only is another improvement that have speeded up the search process. In most of the existing SBIR systems, the authors use the saliency as another feature that can be used in matching the query image with other similar images, with different pose, for example. This way is useful in performing matching process rather than identification process and it is widely applicable in specialised applications such as fingerprints matching and satellite and urban imageries matching. In our approach, the saliency has been used to extract the important object in an image as a whole and isolate it from the background and then identify and label it.

The knowledge database or the dictionary, which contains the features and labels corresponding to the objects, was needed to label the salient objects after extracting their features. This will convert the search process from features distance measure process into text search process. This conversion shall improve the search speed as it does not need any distance calculation with the features of the images in the image database, which is faster than traditional features comparison search. The knowledge database is very much smaller in size than the images database which is unlimited as the number increases with every new image. Hence, the search time complexity of the proposed algorithm is less than that for the traditional search algorithms as they both follow linear time complexity but the number of search process needed in the proposed algorithm is very much less than those required in the traditional search algorithms.

In the case of background identification, the time complexity of the identification is constant as no search process is needed and background is identified directly by the system. However, the neural network training consumes time for training, but the training process is needed only once.

The saliency extraction algorithm showed competitive results in segmenting the image in a more reasonable way in contrast to traditional segmentation approaches which segment the image into small pieces. For example, in a colour-based traditional segmentation algorithm, the car might be divided into small parts as it contains different colours, while in our saliency extraction algorithm, we shall utilise the structure, the intensity, and location of the objects to form single object which can be identified as a whole. Although, extracting the salient objects may add extra calculation burden to the process, but it is performed only once for each new image. Given that after the auto-

labelling shall be performed only on the knowledge database and no need to search for the entire image database.

Another significant achievement in this work was that all the algorithms are fully automatic and does not need any manual parameters specification as in most of the existing techniques. For example, in most of the existing algorithms which are suitable for gaze points clustering, some parameters are needed to be specifies such as number of cluster or threshold values. In our approach, the clustering was fully automatic and did not need any parameters to be specified by the user.

## **6.5 Limitation and Further Possible Improvement**

This work opens new horizons for research in this field and is useful for researchers interested in presenting research that considers more than the classic image processing techniques. Various improvements and applications can be suggested in this field.

The main limitation of the proposed saliency algorithm was it did not give high accuracy in the cases when the image contains large regular object and a small irregular background, this was solved previously in the existing approaches by considering the location of the object. They have considered that the salient object should be in the centre, which we have proved that it is not a reliable approach as some salient objects might be anywhere in the image.

As it was discussed in different occasions, we need to use two databases in the search process, one for the objects, which we have referred to as object database (or dictionary), and one for the images which was referred to as image database. Although it was shown that the number of search process in our approach is less than that is needed for traditional approaches, but still, there is a possibility to improve the algorithm by using different search algorithms such as tree or binary search instead of the sequential search that we have adopted. Another improvement relating to the image database is possible. As the database might become very large, since the database grows with every new entry. Then big data principles can be applicable to improve the efficiency of the image retrieval process.

As we have suggested, computation intelligence may be used in extracting the regions of interest from the salient points that were obtained from applying the saliency extraction

algorithm. The main limitation of this approach was that the neural networks need to be trained for each new image, so considering other artificial intelligence approaches can improve the performance. Techniques such as the support vector machine (SVM) and Swarm-based neural networks could be utilised for this purpose.

For testing reasons we have adopted texture and colour features only, though, other features such as shape features can be tested as well. In addition, the proposed system can be integrated with other image identification approaches and study the possible improvement.

Even though it was shown that the object isolation has given very satisfactory results but still, there is some small parts from the background. The process of isolating the object in more accurate approach may improve the identification process.

## References

- [1] H. Yang and X. Zhou, "Research of content based image retrieval technology," in *Proceedings of the third international symposium on electronic commerce and security workshops (ISECS '10)*, Guangzhou, China, 2010.
- [2] L. Itti, C. Koch and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254-1259, 1998.
- [3] X. Hu, H. Wu, Y. Zhang and L. Sun, "Flower image retrieval based on saliency map," in *Proceedings of the international symposium on computer, consumer and control (IS3C)*, Taichung, China, June 2014.
- [4] Z. Zdziarski and R. Dahyot, "Feature selection using visual saliency for content-based image retrieval," in *Proceedings of IET Irish signals and systems conference (ISSC 2012)*, Maynooth, Ireland, June, 2012.
- [5] G. Ozyer and F. Vural, "An attention-based image retrieval system," in *Proceedings of the 10th international conference on machine learning and applications and workshops (ICMLA)*, Honolulu, USA, Dec. 2011.
- [6] R. de Carvalho Soares, I. da Silva and D. Guliato, "Spatial locality weighting of features using saliency map with a bag-of-visual-words approach," in *Proceedings of IEEE 24th international conference on tools with artificial intelligence (ICTAI)*, Athens, Greece, Nov. 2012.
- [7] B. Wang, X. Zhang, M. Wang and P. Zhao, "Saliency distinguishing and applications to semantics extraction and retrieval of natural image," in *proceedings of the international conference on machine learning and cybernetics (ICMLC)*, Qingdao, China, July 2010.
- [8] K. A. Hua, G. A. Shaykhian, R. J. Beil, K. Akpinar and K. Martin, "Saliency-based CBIR system for exploring lunar surface imagery," in *Proceedings of the 121th. ASEE annual conference and exhibition*, Indianapolis, USA, 2014.
- [9] S. Wan , P. Jin and L. Yue, "An approach for image retrieval based on visual saliency," in *Proceedings of the international conference on image analysis and signal processing*, Taizhou, China, April, 2009.
- [10] Y. Lei , . Z. Shi , X. Jiang and Q. Li , "Image retrieval based on color saliency histogram," in *Proceedings of the international symposium on computer network and multimedia technology*, Wuhan, China, Jan. 2009.
- [11] H. Chen and R. Wang, "Image auto-annotation and retrieval using saliency region detecting and segmentation algorithm," in *Proceedings of the fourth international conference on digital home (ICDH)*, Guangzhou, China, Nov. 2012.
- [12] Z. Liang, H. Fu , Z. Chi and D. Feng, "Image pre-classification based on saliency map for image retrieval," in *Proceedings of the 7th international conference on information, communications and signal processing*, Macau, China, Dec. 2009.
- [13] J. An, S. H. Lee and N. I. Cho, "Content-based image retrieval using color features of salient regions," in *Proceedings of IEEE international conference on image processing (ICIP)*, Paris, France, 2014.
- [14] J. Liu, F. Meng, F. Mu and Z. Yichun, "An improved image retrieval method based on SIFT algorithm and saliency map," in *Proceedings of the 11th international conference on Fuzzy systems and knowledge discovery (FSKD)*, Xiamen, China, Aug. 2014.
- [15] Y. Gao, M. Shi, D. Tao and C. Xu, "Database Saliency for Fast Image Retrieval," *IEEE Transactions on Multimedia*, vol. 17, no. 3, pp. 359-369, 2015.
- [16] E. Giouvanakis and C. Kotropoulos, "Saliency map driven image retrieval combining the bag-of-words model and PLSA," in *Proceedings of the 19th international conference on digital signal processing (DSP)*, Hong Kong, Aug. 2014.
- [17] D. H. Ballard and C. M. Brown, *Computer Vision*, NY: Prentice Hall , 1982.

- [18] D. Hubel, *Eye, Brain, and Vision*, Henry Holt and Company, 1995.
- [19] E. L. van den Broek, *Human-Centered Content-Based Image Retrieval*, Enschede - The Netherlands: Print Partners Ipskamp, 2005.
- [20] A. Desolneux, L. Moisan and J.-M. Morel, *From Gestalt Theory to Image Analysis: A Probabilistic Approach*, New Yourk: Springer, 2008.
- [21] I. Biederman, "Recognition by components: A theory of human image understanding," *Psychological Review*, vol. 94, no. 2, pp. 115-147, 1987.
- [22] D. L. Borges, "3D recognition by parts: A complete solution using parameterized volumetric models," in *Anais do IX SIBGRAPI*, Brasil, 1996.
- [23] D. S. Alexandre and J. M. R. S. Tavares, "Introduction of human perception in visualization," *International Journal of Imaging*, vol. 4, no. A10, pp. 60-70, 2010.
- [24] V. Kadiyala, S. Pinneli, E. C. Larson and D. M. Chandler, "Quantifying the perceived interest of objects in images: Effects of size, location, blur, and contrast," in *Proceedings of human vision and electronic imaging*, San Jose, USA, 2008.
- [25] S. Kim, S. Park and M. Kim, "Central object extraction for object-based image retrieval," in *Proceedings of the 2nd international conference on Image and video 03*, Heidelberg, Germany, 2003.
- [26] P. Kapsalas, K. Rapantzikos, A. Sofou and Y. Avrithis, "Regions of interest for accurate object detection," in *Proceedings of the international workshop on content-based multimedia indexing, CBMI 2008.*, London, UK, 2008.
- [27] L. M. Ward, "Attention," *Scholarpedia*, vol. 3, no. 10, p. 1538, 2008.
- [28] R. M. Ivanoff and J. Klein, "Inhibition of return," *Scholarpedia*, vol. 3, no. 10, p. 3650, 2008.
- [29] Y. Pinto, A. R. v. d. Leij, I. G. Sligte, V. A. F. Lamme and H. S. Scholte, "Bottom-up and top-down attention are independent," *Journal of Vision*, vol. 13, no. 3, pp. 1-14, 2013.
- [30] B. W. Tatler, M. M. Hayhoe, M. F. Land and D. H. Ballard, "Eye guidance in natural vision: Reinterpreting salience," *Journal of Vision*, vol. 11, no. 5, pp. 1-23, 2011.
- [31] T. J. Buschman and E. K. Miller, "Goal-direction and top-down control," *Philosophical Transactions of the Royal Society (Biological Science)*, 29 9 2014. [Online]. Available: [rstb.royalsocietypublishing.org](http://rstb.royalsocietypublishing.org). [Accessed 15 10 2014].
- [32] C. E. Connor, H. E. Egeth and S. Yantis, "Visual attention: Bottom-up versus top-down," *Current Biology*, vol. 14, no. Oct., 2004.
- [33] O. Le Meur, P. L. Callet, D. Barba and D. Thoreau, "A coherent computational approach to model bottom-up visual attention," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 802-817, May 2006.
- [34] T. Kadir and M. Brady, "Scale, saliency and image description," *International Journal of Computer Vision*, vol. 45, no. 2, pp. 83-105, 2001.
- [35] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE Transactions on Image Processing*, vol. 19, no. 1, pp. 185-198, Jan. 2010.
- [36] A. Toet, "Computational versus psychophysical bottom-up image saliency: A comparative evaluation study," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 33, no. 11, pp. 2131-2146, 2011.
- [37] S. Kulkarni, "Content based image retrieval system with hybrid feature set and recently retrieved image library," *International Journal of Computer Applications*, vol. 59, no. 5, pp. 46-55, 2012.
- [38] N. Singhai and S. . K. Shandilya, "A survey on: content based image retrieval systems," *International Journal of Computer Applications*, vol. 4, no. 2, pp. 22-26, July 2010.
- [39] T. . M. Lehmann, M. O. Gülda, C. Thies, B. Plodowski, D. Keysers, B. Ott and H. Schubert, "IRMA – content-based image retrieval in medical applications," in *Proceedings of the 14th world congress on medical informatics*, Amsterdam, Netherlands, 2004.
- [40] D. Heesch, "A survey of browsing models for content based image retrieval," *Multimedia Tools and Applications*, vol. 40, no. 2, pp. 261-284 , 2008.

- [41] G. Jaswal and A. Kaul, "Content based image retrieval: A literature review," in *Proceedings of the national conference on computing, communication and control (CCC-09)*, Himachal Pradesh, India, 2009.
- [42] G. Rafiee, . S. Dlay and W. Woo, "A review of content-based image retrieval," in *Proceedings of the 7th international symposium on communication systems networks and digital signal processing (CSNDSP)*, Newcastle, UK, 2010.
- [43] M. Jain and S. K. Singh, "A survey on: content based image retrieval systems using clustering techniques for large data sets," *International Journal of Managing Information Technology*, vol. 3, no. 4, pp. 23-39, 2011.
- [44] Y. Liu, D. Zhang, G. Lu and W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognition*, no. 40, p. 262–282, 2007.
- [45] R. C. Veltkamp and M. Tanase, "Content-based image retrieval systems: A survey," Av. online at: <http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.104.8806>, Utrecht, The Netherlands, 2000.
- [46] F. Rajam and S. Valli, "A Survey on content based image retrieval," *Life Science Journal*, vol. 10, no. 2, pp. 2475-2487, 2013.
- [47] S. Kaur and V. K. Banga, "Content based image retrieval: Survey and comparison between RGB and HSV model," *International Journal of Engineering Trends and Technology*, vol. 4, no. 4, pp. 575-579, 2013.
- [48] H. B. Kekre, D. Mishra and A. Kariwala, "A survey of CBIR techniques and semantics," *International Journal of Engineering Science and Technology*, vol. 3, no. 5, pp. 4510-4515, 2011.
- [49] M. Puri, "A survey on content based image retrieval," *International Journal of Computer Science & Engineering Technology*, vol. 4, no. 7, pp. 1004-1008, 2013.
- [50] T. Dharani and I. L. Aroquiaraj, "A survey on content based image retrieval," in *Proceedings of the international conference on pattern recognition, informatics and mobile engineering*, Salem, India, 2013.
- [51] G. Pass and R. Zabih, "Histogram refinement for content-based image retrieval," in *Proceedings of IEEE workshop on applications of computer vision*, Paris, France, 1996.
- [52] Y. Liu, X. Chen, C. Zhang and A. Sprague, "An interactive region-based image clustering and retrieval platform," in *Proceedings of IEEE international conference on multimedia and expo*, Toronto, Canada, 2006.
- [53] P. Suhasini, K. S. R. Krishna and I. V. M. Krishna, "CBIR using color histogram processing," *Journal of Theoretical and Applied Information Technology*, vol. 6, no. 1, pp. 116 - 122, 2009.
- [54] P. B. Thawari and N. J. Janwe, "CBIR base on color and texture," *International Journal of Information Technology and Knowledge Management*, vol. 4, no. 1, pp. 129-132, January-June 2011.
- [55] N. Sharma, P. Rawat and J. Singh, "Efficient CBIR using color histogram processing," *International Journal of Signal & Image Processing*, vol. 2, no. 1, 2011.
- [56] M. R. Nazari and E. Fatemizadeh, "A CBIR system for human brain magnetic resonance image indexing," *International Journal of Computer Applications*, vol. 7, no. 4, pp. 33-37, October 2010.
- [57] D. De Ridder, R. Duin, M. Egmont-Petersen, L. van Vliet and P. Verbeek, "Nonlinear image processing using artificial neural networks," *Adv. Imaging Electron Phys.*, vol. 126, pp. 352-450, 2003.
- [58] M. Juneja and P. S. Sandhu, "Performance evaluation of edge detection techniques for images in spatial domain," *International Journal of Computer Theory and Engineering*, vol. 1, no. 5, 2009.
- [59] W. K. Pratt, *Digital Image Processing*, 4th. ed., Hoboken, New Jersey: John Wiley & Sons, Inc., 2007.
- [60] S. Konishi, . A. L. Yuille and J. M. Coughlan, "Statistical edge detection: Learning and evaluating edge cues," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 25, no. 1, pp. 57-74, 2003.
- [61] J. Canny, "A computational approach to edge detection," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679-698, 1986.

- [62] M. S. Nixon and A. S. Aguado, *Feature Extraction and Image Processing*, Oxford: Newnes A division of Reed Educational and Professional Publishing Ltd, 2002.
- [63] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, New Jersey: Prentice-Hall, Inc., 2002.
- [64] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629-639, 1990.
- [65] R. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786-804, 3 1979.
- [66] R. Haralick and L. Watson, "A facet model for image data," *Computer Vision, Graphics, and Image Process*, no. 15, pp. 113-129, 1981.
- [67] S. O. Elberink and H.-G. Maas, "The use of anisotropic height texture measures for the segmentation of airborne laser scanner data," in *International archives of photogrammetry and remote sensing, Vol. XXXIII, Commission III*, Amsterdam, Holland, 2000.
- [68] B. M. Mehtre, M. S. Kankanhalli and W. F. Lee, "Shape measures for content based image retrieval: A comparison," *Information Processing & Management*, vol. 33, no. 3, pp. 319-337, 1997.
- [69] A. Jain, R. Muthuganapathy and K. Ramani, "Content-based image retrieval using shape and depth from an engineering database," *Lecture Notes in Computer Science*, vol. 4842, pp. 255-264, 2007.
- [70] A. Koschan and M. Abidi, *Digital Color Image Processing*, New Jersey: John Wiley & Sons, Inc., 2008.
- [71] H. Zhou, J. Wu and J. Zhang, *Digital Image Processing Part II*, Ventus Publishing, 2010.
- [72] S. Thilagamani and N. Shanthi, "A Survey on image segmentation through clustering," *International Journal of Research and Reviews in Information Sciences*, vol. 1, no. 1, pp. 14-17, March 2011.
- [73] T. N. Pappas, "An adaptive clustering algorithm for image segmentation," *IEEE Transaction on Signal Processing*, vol. 10, no. 1, pp. 901 - 914, 1992.
- [74] K. Deshmukh and G. N. Shinde, "An adaptive neuro-fuzzy system for color image segmentation," *Journal of Indian Institute of Science*, no. 8, pp. 493-506, Sept.-Oct. 2006.
- [75] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888-905, 2000.
- [76] Z.-K. Huang and K.-W. Chau, "A new image thresholding method based on gaussian mixture model," *Applied Mathematics and Computation*, vol. 205, no. 2, pp. 899-907, 2008.
- [77] H. Al-Rawi and J. J. Stephan, "Histogram-based optimal multiple thresholding using genetic algorithm," in *Proceedings of the 2nd international conference on intelligent knowledge systems*, Istanbul, Turkey, 2005.
- [78] A. E. Savakis, "Adaptive document image thresholding using foreground and background clustering," in *Proceedings of the international conference on image processing (ICIP)*, Chicago, USA, 1998.
- [79] N. Papamarkos, C. Strouthopoulos and I. Andreadis, "Multithresholding of color and gray-level images through a neural network technique," *Image and Vision Computing*, vol. 18, p. 213-222, 2000.
- [80] H. Hamza, E. Smigiel and A. Belaid, "Neural based binarization techniques," in *Proceedings of the eighth international conference on document analysis and recognition (ICDAR'05)*, Washington, USA, 2005.
- [81] G. Rao, C. NagaRaju, L. Reddy and E. Prasad, "A novel thresholding technique for adaptive noise reduction using neural networks," *International Journal of Computer Science and Network Security*, vol. 8, no. 12, 2008.
- [82] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 146-165, 2004.
- [83] F. Tomita, M. Yachida and S. Tsuji, "Detection of homogeneous regions by structural analysis," in *Proceedings of the international joint conference on artificial intelligence*, Stanford, USA, August 1973.
- [84] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Transaction on Neural Network*, vol. 16, no. 3, pp. 645-678, May, 2005.

- [85] A. K. Jain, M. N. Murty and P. J. Flynn, "Data clustering: A review," *ACM Computing Surveys (CSUR) Surveys*, vol. 31, no. 3, pp. 264-323, Sept. 1999.
- [86] P. Rai and S. Singh, "A survey of clustering techniques," *International Journal of Computer Applications*, vol. 7, no. 12, pp. 1-5, 2010.
- [87] P. Tryfos, "Chapter 15: Cluster analysis.," in *Methods for Business Analysis and Forecasting: Text & Cases*, , Wiley, 1998.
- [88] D. Cook and D. F. Swayne, "chapter 5: Cluster Analysis," in *Interactive and Dynamic Graphics for Data Analysis*, Springer, 2007, pp. 103-128.
- [89] R. B. Burns and R. A. Burns , "Chapter 23: Cluster Analysis," in *Business Research Methods and Statistics Using SPSS*, Sage Publications Ltd., 2008, pp. 552-567.
- [90] P.-N. Tan, M. Steinbach and K. Vipin , "Cluster Analysis: Basic Concepts and Algorithms," in *Introduction to Data Mining*, Addison-Wesley Companion, 2006, pp. 488-568.
- [91] G. Milligan and M. Cooper, "Methodology review: Clustering methods," *Applied Psychological Measurement*, vol. 11, no. 4, pp. 329-354, 1987.
- [92] Chowdhury, Nirmalya, C. A. Murthy and S. K. Pal, "Cluster detection using neural networks," in *Proceedings of IEEE international conference on neural networks*, 1995., Perth, Australia, 1995.
- [93] B. Gour, T. K. Bandopadhyaya and S. Sharma, "ART neural network based clustering method produces best quality clusters of fingerprints in comparison to self organizing map and k-means clustering algorithms," in *Proceedings of the international conference on innovations in information technology, IIT 2008*, Al Ain, UAE, 2008.
- [94] G.-Y. Huang, D.-P. Liang, C.-Z. Hu and J.-D. Ren, "An algorithm for clustering heterogeneous data stream with uncertainty," in *Proceedings of the ninth international conference on machine learning and cybernetics*, Qingdao, China, 2010.
- [95] A. Amini, T. Y. Wah, M. Saybani and S. Yazdi, "A study of density-grid based clustering algorithms on data streams," in *Proceedings of the eighth international conference on fuzzy systems and knowledge discovery (FSKD) Volume: 3*, Shanghai, China, 2011.
- [96] A. A. Torn, "Cluster analysis using seed points and density-determined hyperspheres as an aid to global optimization," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 7, no. 8, pp. 610-616, 1977.
- [97] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Transaction on Fuzzy Systems*, vol. 1, no. 2, pp. 98-110, 1993.
- [98] V. S. Tseng and C.-P. Kao, "A novel similarity-based fuzzy clustering algorithm by integrating PCM and mountain method," *IEEE Transaction on Fuzzy Systems*, vol. 15, no. 6, pp. 1188-1196, 2007.
- [99] M.-S. Yang, "A survey of fuzzy clustering," *Mathl. Comput. Modelling*, vol. 18, no. 11, pp. 1-16, 1993.
- [100] Y. Jung, H. Park, D. Du and B. Drake, "A decision criterion for the optimal number of clusters in hierarchical clustering," *Journal of Global Optimization*, no. 25, pp. 91-111, 2003.
- [101] J. A. Hartigan, *Clustering algorithms*, New Yourk: Wiley, 1975.
- [102] T. Pedersen and A. Kulkarni, "Automatic cluster stopping with criterion functions and the gap statistic," in *Proceedings of the 2006 conference of the north American chapter of the association for computational linguistics on human language technology: companion volume: demonstrations*, Morristown, USA,, 2006.
- [103] A. Konar, *Artificial Intelligence and Soft Computing: Behavioral and Cognitive Modeling of the Human Brain*, Boca Raton London New York Washington, D.C.: CRC Press, 1999.
- [104] A. Gelbukh and R. Monroy, "Advances in artificial intelligence applications," *Research on Computer Science*, vol. 17, pp. 151-160, 2005.
- [105] L. V. Fausett, *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*, Prentice-Hall,, 1994.
- [106] G. Lendaris and K. Mathia, "Efficient numerical inversion using multilayer feedforward neural networks," in *Proceedings of the world congress on neural networks (WCNN'96)*, San Diego, California, 1996.



- [107] R. P. Lippmann, "An introduction to computing with neural nets," *IEEE ASSP Magazine*, no. April, pp. 4-22, 1987.
- [108] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338-353, 1965.
- [109] S. K. Pal, "Fuzzy models for image processing and applications," *Proceedings of Indian National Science Academy*, vol. 65, no. 1, pp. 73-90, 1999.
- [110] J. Sun, "MSRA Salient Object Database," Microsoft, 2007. [Online]. Available: [http://research.microsoft.com/en-us/um/people/jiansun/salientobject/salient\\_object.htm](http://research.microsoft.com/en-us/um/people/jiansun/salientobject/salient_object.htm). [Accessed 13 2014].
- [111] "MIT saliency benchmark," Massachusetts Institute of Technology, 2012. [Online]. Available: <http://saliency.mit.edu/>. [Accessed 13 2014].
- [112] R. Achanta, "Saliency detection using maximum symmetric surround," Image and visual representation lab / IVRL, 2013. [Online]. Available: <http://ivrl.epfl.ch/research/saliency/MSSS.html>. [Accessed 13 2014].
- [113] J. Li, M. D. Levine, X. An, X. Xu and H. He, "Visual saliency based on scale-space analysis in the frequency domain," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, , vol. 35, no. 4, pp. 996-1010, 2013.
- [114] J. Li, "Saliency Database," China Science and Technology Network, 2007. [Online]. Available: <http://www.escience.cn/people/jianli/DataBase.html>. [Accessed 13 2014].
- [115] "CV Datasets on the web," [Online]. Available: <http://www.cvpapers.com/datasets.html>. [Accessed 13 2014].
- [116] R. Achanta, S. Hemami, F. Estrada and S. Susstrunk, "Frequency-tuned salient region detection," in *Proceedings of IEEE conference on computer vision and pattern recognition. (CVPR 2009)*, Miami, USA, 2009.
- [117] Y.-F. Ma and H.-J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Proceedings of ACM international conference on multimedia*, Berkeley, CA, USA, 2003.
- [118] J. Harel, C. Koch and P. Perona, "Graph-based visual saliency," *Advances in Neural Information Processing Systems*, vol. 19, pp. 545-552, 2007.
- [119] X. Hou and L. Zhang, "Saliency detection: A spectral residua approach.," in *Proceedings of IEEE conference on computer vision and pattern recognition*, Minneapolis, USA, 2007.
- [120] R. Achanta, F. Estrada, P. Wils and S. SÄusstrunk, "Salient region detection and segmentation," in *Proceedings of the international conference on computer vision systems (ICVS '08)*, Marseille, France, 2008.
- [121] R. Achanta and S. Susstrunk, "Saliency detection using maximum symmetric surround," in *Proceedings of the 17th IEEE international conference on image processing (ICIP)*, , Hong Kong, 2010.
- [122] S. Luo, Z. Liu, L. Li, X. Zou and L. Meur, "Efficient saliency detection using regional color and spatial information," in *Proceedings of IEEE 4th European workshop on visual information processing (EUVIP)*, , Paris, France, 2013.
- [123] R. Margolin, A. Tal and L. Zelnik-Manor, "What makes a patch distinct," in *Proceedings of IEEE conference on computer vision and pattern recognition (CVPR '13)*, Washington, USA, 2013.
- [124] Y. Fang, Z. Chen, W. Lin and C.-W. Lin, "Saliency-based image retargeting in the compressed domain," in *Proceedings of the 19th ACM international conference on multimedia*, New York, USA, 2011.
- [125] J. Wang, "James Z. Wang Research Group," Penn State University, 2004. [Online]. Available: <http://wang.ist.psu.edu/docs/related/>. [Accessed 13 2014].
- [126] D. B. Walther, "Saliency Toolbox," California Institute of Technology, 8 7 2013. [Online]. Available: <http://www.saliencytoolbox.net/>. [Accessed 13 2014].
- [127] R. Achanta, "Saliency maps from 5 state-of-the art methods," Image and visual representation lab (IVRL), 2013. [Online]. Available: <http://ivrl.epfl.ch/page-75955-en.html>. [Accessed 13 2014].
- [128] R. H. S. Achanta, F. Estrada and S. Süssstrunk, "Frequency-tuned salient Region Detection," Image and visual representation lab (IVRL), 2013. [Online]. Available: [http://ivrl.epfl.ch/supplementary\\_material/RK\\_CVPR09/index.html](http://ivrl.epfl.ch/supplementary_material/RK_CVPR09/index.html). [Accessed 13 2014].

- [129] J. Li, "Visual Saliency Based on Scale-Space Analysis in the Frequency Domain," [Online]. Available: <https://sites.google.com/site/jianlinudt/hft>. [Accessed 13 2014].
- [130] R. Achanta and S. Süsstrunk, "Saliency Detection using Maximum Symmetric Surround," Image and visual representation lab (IVRL), 2013. [Online]. Available: <http://ivrl.epfl.ch/page-74626-en.html>. [Accessed 13 2014].
- [131] Y. Fang, "Saliency Detection in the Compressed Domain for Adaptive Image Retargeting," 2012. [Online]. Available: <https://sites.google.com/site/leofangyuming/Home/smap-compressed-domain>. [Accessed 13 2014].
- [132] E. Loupías, N. Sebe, S. Bres and J.-M. Jolion, "Wavelet-based salient points for image retrieval," in *Proceedings of the international conference on image processing*, Vancouver, Canada, 2000.
- [133] Q. Tian, N. Sebe, M. Lew, E. Loupías and T. S. Huang, "Image retrieval using wavelet-based salient points," *Journal of Electronic Imaging*, vol. 10, no. 4, pp. 835-849, 2001.
- [134] H. Song, B. Li and L. Zhang, "Color salient points detection using wavelet," in *Proceedings of the 6th world congress on intelligent control and automation*, Dalian, China, 2006.
- [135] D.-W. Lin and S.-H. Yang, "Wavelet-based salient region extraction," in *Advances in Multimedia Information Processing – PCM 2007. Vol. 4810*, Hong Kong, Springer, 2007, pp. 389-392.
- [136] S. Arivazhagan and R. N. Shebiah, "Object recognition using wavelet based salient points," *The Open Signal Processing Journal*, vol. 2, pp. 14-20, 2009.
- [137] J. Davis and V. Sharma, "Robust background subtraction for person detection in thermal imagery," in *Proceedings of IEEE conference on computer vision and pattern recognition workshops (CVPRW'04)*, NA, 2004.
- [138] H. Zhang and S. A. Goldman, "Image segmentation using salient points-based object templates," in *Proceedings of the 13th international conference on image processing (ICIP)*, Atlanta, USA, 2006.
- [139] C. Schmid and R. Mohr, "Local greyvalue invariants for image retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 5, pp. 530-535, 1997.
- [140] H. P. Morevec, "Towards automatic visual obstacle avoidance," in *Proceedings of the 5th international joint conference on Artificial intelligence*, San Francisco, USA, 1977.
- [141] Z. Zheng, H. Wang and E. K. Teoh, "Analysis of gray level corner detection," *Pattern Recognition Letters*, vol. 20, pp. 149-162, 1999.
- [142] S. M. Smith and J. M. Brady, "SUSAN – a new approach to low level image processing," *International Journal of Computer Vision*, vol. 23, no. 1, pp. 45-78, 1997.
- [143] C. Koch and S. Ullman, "Shifts in selective visual attention: Towards the underlying neural circuitry," *Human Neurobiology*, vol. 4, p. 219-227, 1985.
- [144] N. Bruce, D. Loach and J. Tsotsos, "Visual correlates of fixation selection: A look at the spatial frequency domain," in *Proceedings of IEEE international conference on image processing. ICIP 2007 (Vol 3)*, San Antonio, USA, 2007.
- [145] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proceedings of IEEE conference on computer vision and pattern recognition. CVPR '07.*, Minneapolis, USA, 2007.
- [146] B. Zhou, X. Hou and L. Zhang, "A phase discrepancy analysis of object motion," in *Proceedings of the 10th Asian conference on computer vision ACCV'10*, Queenstown, New Zealand, 2010.
- [147] Y. Fang, W. Lin, B.-S. Lee, C.-T. Lau, Z. Chen and C.-W. Lin, "Bottom-up saliency detection model based on human visual sensitivity and amplitude spectrum," *IEEE Transaction on Multimedia*, vol. 14, no. 1, pp. 187-198, 2012.
- [148] N. Sebe, Q. Tian, E. Loupías, M. Lew and T. Huang, "Evaluation of salient point techniques," *Image and Vision Computing*, vol. 21, pp. 367-377, 2002.
- [149] M. Gide and L. Karam, "Improved foveation- and saliency-based visual attention prediction under a quality assessment task," in *Proceedings of the fourth international workshop on quality of multimedia experience (QoMEX)*, Yarra Valley, USA, 2012.
- [150] A. Borji, D. N. Sihite and L. Itti, "Salient object detection: A benchmark," in *Proceedings of the 12th European conference on computer vision proceedings, Part II*, Florence, Italy, Oct. , 2012.

- [151] T. Judd, F. Durand and A. Torralba, "A benchmark of computational models of saliency to predict human fixations," in *Computer science and artificial intelligence laboratory technical report, MIT-CSAIL-TR-2012-001*, Jan. , 2012.
- [152] Q. Zhao and C. Koch, "Learning a saliency map using fixated locations in natural scenes," *Journal of Vision*, vol. 11(3), no. 9, pp. 1-15, 2011.
- [153] Y. Hu, X. Xie, W.-Y. Ma, L.-T. Chia and D. Rajan, "Salient region detection using weighted feature maps based on the human visual attention model," in *Proceedings of the fifth Pacific rim conference on multimedia*, Tokyo, Japan, 2005.
- [154] W. Einhauser, M. Spain and P. Perona, "Objects predict fixations better than early saliency," *Journal of Vision*, vol. 8(14), no. 18, pp. 1-26, 2008.
- [155] Y. Lin, Y. Y. Tang, B. Fang, Z. Shang, Y. Huang and S. Wang, "A visual-attention model using Earth mover's distance-based saliency measurement and nonlinear feature combination," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 2, pp. 314-328, 2013.
- [156] E. Erdem and A. Erdem, "Visual saliency estimation by nonlinearly integrating features using region covariances," *Journal of Vision*, vol. 13(4), no. 11, pp. 1-20, 2013.
- [157] C. Kim and P. Milanfar, "Visual saliency in noisy images," *Journal of Vision*, vol. 13(4), no. 5, pp. 1-14, 2013.
- [158] J. Stöttinger, A. J. Hanbury and T. Gevers, "Lonely but attractive: Sparse color salient points for object retrieval and categorization," in *Proceedings of CVPR workshops*, Miami, USA, 2009.
- [159] Z. Liu, Y. Xue, H. Yan and Z. Zhang, "Efficient saliency detection based on Gaussian models," *IET Image Processing*, vol. 5, no. 2, pp. 122-131, 2011.
- [160] T. Liu, J. Sun, N.-N. Zheng, X. Tang and H.-Y. Shum, "Learning to detect a salient object," in *Proceedings of IEEE conference on computer vision and pattern recognition*, Minneapolis, USA, 2007.
- [161] M. Mancas, "Image perception: Relative influence of bottom-up and top-down attention," in *Proceedings of the international workshop attention and performance in computer vision in conjunction with the international conference on computer vision systems*, N.A., 2008.
- [162] D. Parkhurst, K. Law and E. Nieb, "Modeling the role of salience in the allocation of overt visual attention," *Vision Research*, no. 42, pp. 107-123, 2002.
- [163] C. M. Masciocchi, S. Mihalas, D. Parkhurst and E. Niebur, "Everyone knows what is interesting: Salient locations which should be fixated," *Journal of Vision*, vol. 9(11), no. 25, pp. 1-22, 2009.
- [164] P. L. Rosin, "A simple method for detecting salient regions," *Pattern Recognition*, vol. 42, pp. 2363-2371, 2009.
- [165] L. Zhang, M. H. Tong, T. K. Marks, H. Shan and G. W. Cottrell, "SUN: A bayesian framework for saliency using natural statistics," *Journal of Vision*, vol. 8(7), no. 32, pp. 1-20, 2008.
- [166] J. Davis and M. Goadrich, "The relationship between precision-recall and roc curves," Technical report #1551, University of Wisconsin Madison, Wisconsin, USA, 2006.
- [167] V. Gopalakrishnan, Y. Hu and D. Rajan, "Random walks on graphs to model saliency in images," in *Proceedings of IEEE conference on computer vision and pattern recognition, CVPR 2009*, Miami, USA, 2009.
- [168] O. Pele and M. Werman, "A linear time histogram metric for improved SIFT matching," in *Proceedings of the 10th European conference on computer vision: Part III (ECCV '08)*, Marseille, France, 2008.
- [169] Y. Lin, B. Fang and Y. Tang, "A computational model for saliency maps by using local entropy," in *Proceedings of the twenty-fourth AAAI conference on artificial intelligence (AAAI-10)*, Atlanta, USA, 2010.
- [170] Y. Rubner, C. Tomasi and L. Guibas, "The earth mover's distance as a metric for image retrieval," *International Journal of Computer Vision*, vol. 40, no. 2, pp. 99-121, 2000.
- [171] R. Peters, A. Iyer, L. Itti and C. Koch, "Components of bottom-up gaze allocation in natural images," *Vision Research*, vol. 45, pp. 2397-2416., 2005.
- [172] Y. S. Choi, A. D. Mosley and L. W. Stark, "String editing analysis of human visual search," *Optometry and Vision Science*, vol. 72, pp. 439-45, 1995.

- [173] C. Schmid, R. Mohr and C. Bauckhage, "Evaluation of interest point detectors," *International Journal of Computer Vision*, vol. 37, no. 2, pp. 151-172, 2000.
- [174] M. Al-Azawi, Y. Yang and H. Istance, "A new gaze points agglomerative clustering algorithm and its application in regions of interest extraction," in *Proceedings of IEEE international advance computing conference (IEEE IACC)*, Gurgaon, India, 2014.
- [175] S. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*, California: California Technical Publishing, 2011.
- [176] H. Gross, *Handbook of Optical Systems: Vol. 4 Survey of Optical Instruments*, Weinheim, Germany: WILEY-VCH, 2008.
- [177] R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer, 2010.
- [178] M. Al-Azawi, Y. Yang and H. Istance, "Irregularity-based image regions identification and evaluation," *Multimedia Applications and Tools*, 08 November 2014.
- [179] M. Al-Azawi, Y. Yang and H. Istance, "Irregularity-based saliency identification and evaluation," in *Proceedings of IEEE international conference on computational intelligence and computing research*, Madurai, India, 2013.
- [180] M. Banerjee and M. K. Kundu, "Edge based features for content based image retrieval," *Pattern Recognition*, no. 36, pp. 2649-2661, 2003.
- [181] R. Rahmani, S. A. Goldman, H. Zhang, S. R. Cholleli and J. E. Fritts, "Localized content based image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. November, no. Special Issue, pp. 1-10, 2008.
- [182] D. Walther and C. Koch, "Modeling attention to salient proto-objects," *Neural Networks*, vol. 19, no. 9, pp. 1395-1407, 2006.
- [183] D. Parkhurst, K. Law and E. Niebur, "Modeling the role of salience in the allocation of overt visual attention," *Vision Research*, vol. 42, no. 1, pp. 107-123, 2002.
- [184] L. Elazary and L. Itti, "Interesting objects are visually salient," *Journal of Vision*, vol. 8(3), no. 3, pp. 1-15, 2008.
- [185] N. D. Bruce, "Features that draw visual attention: an information theoretic perspective," *Neurocomputing*, Vols. 65-66, pp. 125-133, 2005.
- [186] P. Rosin, "A simple method for detecting salient regions," *Pattern Recognition*, vol. 42, no. 11, pp. 2363-2371, 2009.
- [187] H. D. Cheng, J.-R. Chen and J. LI, "Threshold selection based on fuzzy c-partition entropy approach," *Pattern Recognition*, vol. 31, no. 7, pp. 857-887, 1998.
- [188] Q. Wang, Z. Chi and R. Zhao, "Image thresholding by maximizing the index of nonfuzziness of the 2D grayscale histogram," *Computer Vision and Image Understanding*, vol. 85, p. 100-116, 2002.
- [189] W.-B. Tao, J.-W. Tian and J. Liu, "Image segmentation by three-level thresholding based on maximum fuzzy entropy and genetic algorithm," *Pattern Recognition Letters*, no. 24, pp. 3069-3078, 2003.
- [190] Y. Yong, Z. Chongxun and L. Pan, "A novel fuzzy c-means clustering algorithm for image thresholding," *Measurement Science Review*, vol. 4, no. 1, 2004.
- [191] T. Junwei, H. Yongxuan and Y. Yalin, "Fuzzy c-means cluster image segmentation with entropy constraint," in *Proceedings of the 33rd annual conference of the IEEE industrial electronics society (IECON)*, Taipei, Taiwan, Nov. 5-8, 2007.
- [192] H. R. Tizhoosh, "Image thresholding using type II fuzzy sets," *Pattern Recognition*, no. 38, pp. 2363-2372, 2005.
- [193] M. S. Prasad and e. al, "Unsupervised image thresholding using fuzzy measures," *International Journal of Computer Applications*, vol. 27, no. 2, pp. 32-41, 2011.
- [194] M. Al-Azawi and N. K. Ibrahim, "Bimodal histogram based image segmentation using fuzzy logic," in *Proceedings of the national conference on artificial intelligence applications in engineering (NCAIAE12)*, Massana, Oman, 2012.

- [195] Y. An, M. Riaz and J. Park, "CBIR based on adaptive segmentation of HSV color space," in *Proceedings of the 12th international conference on computer modelling and simulation (UKSim)*, Cambridge, UK, 2010.
- [196] W.-T. Chen, W.-C. Liu and M.-S. Chen, "Adaptive color feature extraction based on image color distributions," *IEEE Transactions on Image Processing*, vol. 19, no. 8, pp. 2005-2010, AUGUST 2010.
- [197] S. Siggelkow, *Feature histograms for content-based image retrieval*, PhD Thesis, Albert-Ludwigs University, Freiburg, Germany, 2002.
- [198] G. Heidemann, "Combining spatial and colour information for content based image retrieval," *Computer Vision and Image Understanding*, vol. 94, pp. 234-270, 2004.
- [199] R. M. Haralick, K. Shanmugam and I. Dinstein, "Textural features for image classification," *IEEE Transaction on Systems, Man and Cybernetics*, Vols. SMC-3, no. 6, pp. 610-621, November 1973.
- [200] M. Tsaneva, "Texture features for segmentation of satellite images," *Cybernetics and Information Technologies*, vol. 8, no. 3, pp. 73-85, 2008.
- [201] S. Selvarajah and S. Kodituwakku, "Analysis and comparison of texture features for content based image retrieval," *International Journal of Latest Trends in Computing*, vol. 2, no. 1, March 2011.
- [202] J. Laaksonen, M. Koskela, S. Laakso and E. Oja, "Self-organising maps as a relevance feedback technique in content-based image retrieval," *Pattern Analysis & Applications*, vol. 4, no. 2-3, p. 140-152, 2001.
- [203] C.-F. Tsai, K. McGarry and J. Tait, "Image classification using hybrid neural networks," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, Toronto, Canada, 2003.
- [204] S. Čabarkapa, N. Kojić, . V. Radosavlje and B. Reljin, "Adaptive content-based image retrieval with relevance feedback," in *Proceedings of the international conference on computer as a tool (EUROCON)*, Serbia & Montenegro, Belgrade, 2005.
- [205] N. S. Kojić, S. K. Čabarkapa, G. J. Zajić and B. D. Reljin, "Implementation of neural network in CBIR systems with relevance feedback," *Journal of Automatic Control, University of Belgrad*, vol. 16, pp. 41-45, 2006.
- [206] B. Thomee and M. Lew, "Relevance feedback in content-based image retrieval: promising directions," in *Proceedings of the 13th annual conference of the Advanced School for Computing and Imaging*, Heijen, Netherlands, 2007.
- [207] S. Nematipour, . J. Shanbehzadeh and R. A. Moghadam, "Relevance feedback optimization in content based image retrieval via enhanced radial basis function network," in *Proceeding of the international multiconference of engineers and computer scientists IMECS*, Hong Kong, 2011.
- [208] H. Lee and S. Yoo, "Intelligent image retrieval using neural network," *IEICE Trans. on Information and Systems*, Vols. E84-D, no. 12, pp. 1810-1819, 2001.
- [209] Y. Zhu, X. Liu and W. Mio, "Content-based image categorization and retrieval using neural networks," in *Proceedings of IEEE international conference on multimedia and expo*, Beijing, China, 2007.
- [210] V. Balamurugan and P. Anandhakumar, "Neuro-fuzzy based clustering approach for content based image retrieval using 2D-wavelet transform," *International Journal of Recent Trends in Engineering*, vol. 1, no. 1, pp. 418-424, 2009.
- [211] S. Sadek, A. Al-Hamadi, B. Michaelis and Sayed, "Cubic-splines neural network- based system for image retrieval," in *Proceedings of the 16th IEEE international conference on image processing (ICIP)*, Cairo, Egypt, 2009.
- [212] S. Sadek, A. Al-Hamad, B. Michaelis and U. Sayed, "Image retrieval using cubic splines neural networks," *International Journal of Video & Image Processing and Network Security IJVIPNS*, vol. 9, no. 10, pp. 17-22, 2010.
- [213] P. Rao, E. Vamsidhar, G. V. P. Raju, R. Satapati and K. Varma, "An approach for CBIR system through multi layer neural network," *International Journal of Engineering Science and Technology*, vol. 2, no. 4, pp. 559-563, 2010.

- [214] B. Tian, M. Azimi-Sadjadi, T. Haar and D. Reinke, "Neural network-based cloud classification on satellite imagery using textural features," in *Proceedings of the international conference on image processing*, Santa Barbara, USA, 1997.
- [215] B. Tian, M. Shaikh, M. Azimi-Sadjadi, T. Haar and D. Reinke, "A study of cloud classification with neural networks using spectral and textural features," *IEEE Transactions on Neural Networks*, vol. 10, no. 1, pp. 138-151, 1999.
- [216] M. Torabi, R. Ardekani and E. Fatemizadeh, "Discrimination between alzheimer's disease and control group in MR-images based on texture analysis using artificial neural network," in *Proceedings of the international conference on biomedical and pharmaceutical engineering, ICBPE*, Singapore, 2006..
- [217] J. Zhang, L. Wang and L. Tong, "Feature reduction and texture classification in MRI-texture analysis of multiple sclerosis," in *Proceedings of IEEE/ICME international conference on complex medical engineering (CME)*, Beijing, China, 2007.
- [218] S. Anand, "Segmentation coupled textural feature classification for lung tumor prediction," in *Proceedings of IEEE international conference on communication control and computing technologies (ICCCCT)*, Ramanathapuram, India, 2010.
- [219] D. Joshi, N. Rana and V. Misra, "Classification of brain cancer using artificial neural network," in *Proceedings of the international conference on electronic computer technology (ICECT)*, Kuala Lumpur, Malaysia, 2010.
- [220] D. B. Kadam, S. S. Gade, M. D. Uplane and R. K. Prasad, "An artificial neural network approach for brain tumor detection based on characteristics of GLCM texture features," *International Journal of Innovations in Engineering and Technology*, vol. 2, no. 1, pp. 193-199, 2013.
- [221] L. Haonan, S. Baojin, H. Yaqu, H. Jingfeng and H. Qiongqiong, "Research on identification of coal and waste rock based on GLCM and BP neural network," in *Proceedings of the 2nd international conference on signal processing systems (ICSPS)*, Dalian, China, 2010.
- [222] J. L. Rahejaa, S. Kumar and A. Chaudhary, "Fabric defect detection based on GLCM and Gabor filter: A comparison," *Optik - International Journal for Light and Electron Optics*, vol. 124, no. 23, p. 6469–6474, 2013.
- [223] R. Achanta, "Image and Visual Representation Group IVRG Dataset," 2010. [Online]. Available: <http://ivrg.epfl.ch/research/saliency/MSSS.html>. [Accessed 1 12 2012].
- [224] M. Al-Azawi and N. Ibrahim, "A new edge intersection-based salient points extraction and its application in computer vision," in *Proceedings of the National Conference on Business and IT*, Masanna-Oman, 2014.
- [225] M. Al-Azawi, "Image thresholding using histogram fuzzy approximation," *International Journal of Computer Applications*, vol. 83, no. 9, pp. 36-40, 2013.