

PAPER • OPEN ACCESS

Feature Extraction Methods: A Review

To cite this article: Wamidh K. Mutlag *et al* 2020 *J. Phys.: Conf. Ser.* **1591** 012028

View the [article online](#) for updates and enhancements.

You may also like

- [Image Classification based on CBIR](#)
Shaker K. Ali and Salah A. Sadoon
- [Strongest Coronal Magnetic Fields in Solar Cycles 23 and 24: Probing, Statistics, and Implications](#)
Viktor V. Fedenev, Sergey A. Anfinogentov and Gregory D. Fleishman
- [Intra-tumoural heterogeneity characterization through texture and colour analysis for differentiation of non-small cell lung carcinoma subtypes](#)
Yuan Ma, Wei Feng, Zhiyuan Wu et al.



UNITED THROUGH SCIENCE & TECHNOLOGY

 **The Electrochemical Society**
Advancing solid state & electrochemical science & technology

**248th
ECS Meeting**
Chicago, IL
October 12-16, 2025
Hilton Chicago

**Science +
Technology +
YOU!**

**Register by
September 22
to save \$\$**

REGISTER NOW

Feature Extraction Methods: A Review

Wamidh K. Mutlag¹, Shaker K. Ali², Zahoor M. Aydam³ and Bahaa H.Taher⁴

¹Al shatrah Technical Institute, Southern Technical University, Iraq.

^{2,3}Computer sciences and mathematics college, University of Thi_Qar, Thi_Qar, Iraq.

⁴Information Technology college, University of Sumer, Thi_Qar, Iraq.

¹wamid.almuhaysen@stu.edu.iq, ² shaker@utq.edu.iq, ³ zahoormosad@gmail.com and ⁴ ghrabiuk@gmail.com

Abstract. Feature extraction is the main core in diagnosis, classification, clustering, recognition, and detection. Many researchers may be interested in choosing suitable features that used in the applications. In this paper, the most important features methods are collected, and explained each one. The features in this paper are divided into four groups; Geometric features, Statistical features, Texture features, and Color features. It explains the methodology of each method, its equations, and application. In this paper, we made a comparison among them by using two types of image, one type for face images (163 images divided into 113 for training and 50 for testing) and the other for plant images (130 images divided into 100 for training and 30 for testing) to test the features in geometric and textures. Each type of image group shows that each type of images may be used suitable features may differ from other types.

Keywords: Geometric features, statistical features, Textures features and Color features.

1.Introduction

In image processing technology, whether it is binary, colored or gray. Image processing may be performed by extracting features for identification, classification, diagnosis, classification, clustering, recognition and detection. Feature extraction method are utilized to obtain much information as possible of image. The selection and effectiveness of feature chosen and extraction are a major challenge now [1]. Many methods used to extract features, which may depend on Geometric features, Statistical features, Texture features, and Color features. Each main type of feature divided into many subdivided types such as Color features divided into three types (Color moment, Color histogram and Average RGB) [2]. Figure 1 shows the most important features methods.



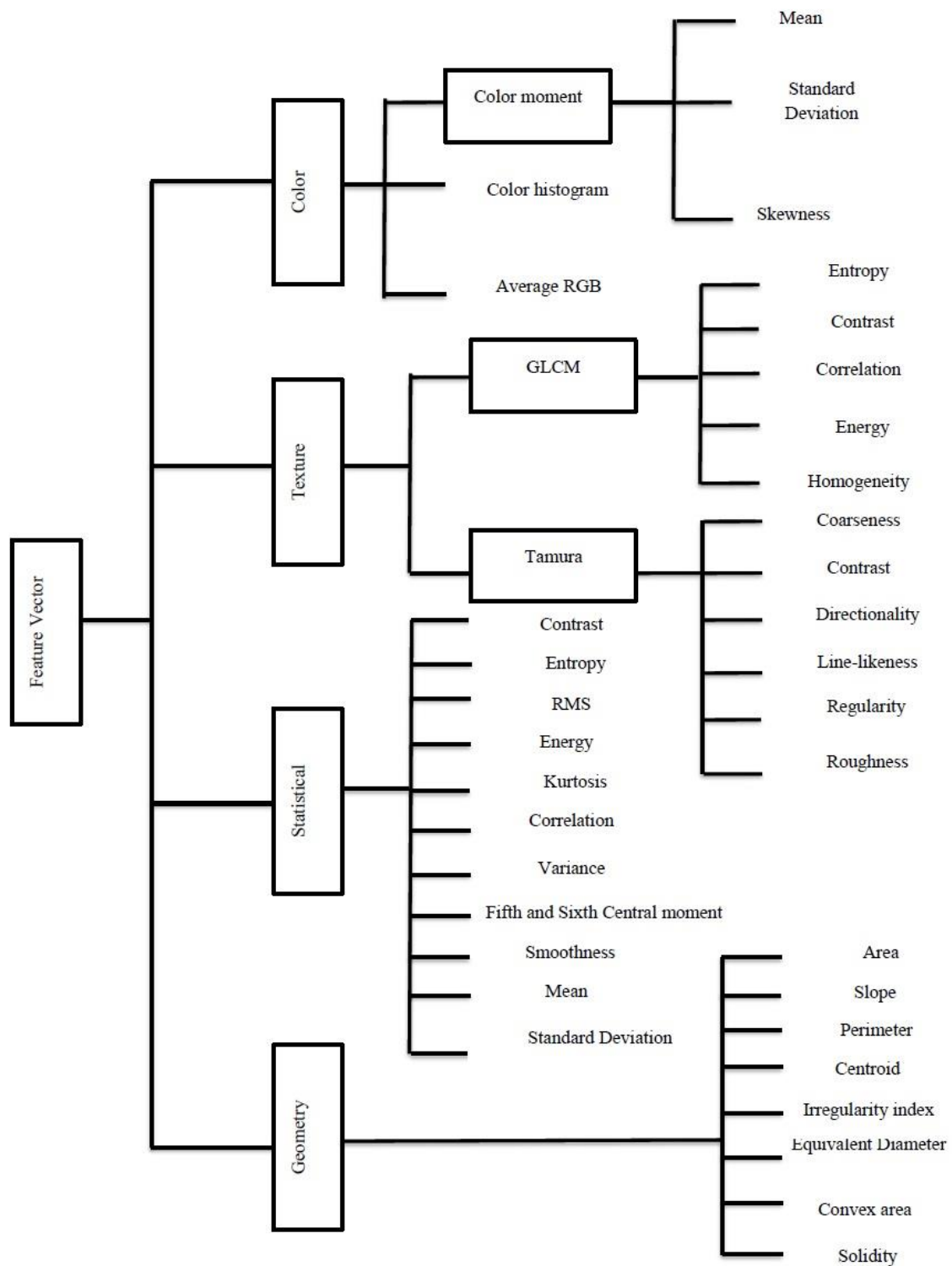


Figure 1. Feature Extraction Methods

2. Feature Extraction

The features can be divided into four types; Geometric features, Statistical features, Texture features, and Color features. Where each type are further divided into sub-types as following:

2.1. Color Features

The color features are divided into three types (Color moment, Color histogram and Average RGB) [3]

2.1.1. Color Moments:

It can be define as scales that can differentiate images based on their own color feature. The color Moments in the image interpreted as the probability distribution. The three-color moments (Mean, Standard Deviation and Skewness).

2.1.1.1. Mean

The mean can be defined as the mean color value in the image, which may define in Equation 1

$$M_j = \sum_m \frac{1}{M} P_{ji} \quad (1)$$

2.1.1.2. Standard Deviation (STD)

The standard deviation is the square root of the distribution variation, Equation 2 explain the format of STD

$$\sigma_j = \sqrt{\frac{1}{M} \sum_{i=1}^M (P_{ji} - M_j)^2} \quad (2)$$

2.1.1.3. Skewness

Interpret the deviation as a measure of the degree of asymmetry in the distribution [2]

$$S_j = \sqrt[3]{\frac{1}{M} \sum_{i=1}^M (P_{ji} - M_j)^3} \quad (3)$$

2.1.2. Color Histogram

Color is the most common characteristic and widely used because of its intuitive compared to other qualities and more important information, the ease of extraction from the image and the histogram distributes colors using a set of boxes.

2.1.3. Average RGB

The goal of using this feature is for image filtering when using various features. The second reason for choosing this feature because using a small number of data to represent vector parameters [4].

2.2. Texture Features

Texture is the most important feature for many types of images that appear everywhere in nature such as medical images and sensor images and so on. The texture defined as a superficial manifestation of the human visual systems of natural objects. It is easy to recognize by everyone, but it is difficult to determine the texture in matrix, but it occurs in an area of matrix that analyzed texture by quantitative and qualitative analysis. In this paper, two types of texture features are discussed (GLCM and Tamura)[5].

2.2.1. Gray Level Co-occurrence Matrices (GLCM)

A histogram to measure gray values that occur at a given offset on an image. Used to extract texture from a broken tissue image [6]. Those are five various texture features specific by GLCM Entropy, Contrast, Correlation, Energy and Homogeneity [7].

2.2.2.1 Entropy

A statistical measure of randomness be utilized to distinguish the texture of an input image.

$$\text{Entropy} = - \sum \sum q(i,j) \log q(i,j) \quad (4)$$

Where q is the number of gray-level co-occurrence matrices in GLCM

2.2.2.2 Contrast:

Calculates the density contrast between pixels and adjacent pixels to whole image. Equation 5 explain the contrast.

$$\text{contrast} = \sum (i, j)^2 q(i, j) \quad (5)$$

Where, $q(i, j)$ = pixel at location (i, j) .

2.2.2.3 Correlation

The function of this scale is to measure the probability of the specified of the specified pixel pairs as in Equation 6 [11].

$$\text{correlation} = \frac{\sum_{i=0}^{M-1} \sum_{j=0}^{M-1} (i - n_i)(j - n_j) q(i, j)}{\sigma_i \sigma_j} \quad (6)$$

2.2.2.4 Energy:

Is the summation of squared elements in the GLCM .It is also known as the angular second moment or uniformity.

$$\text{Energy} = \sum \sum q(i, j)^2 \quad (7)$$

2.2.2.5 Homogeneity

It used to measure the approximation of the distribution of elements in the GLCM to the GLCM diagonal, which define in Equation 8.

$$\text{Homogeneity} = \sum_{i, j} \frac{q(j, i)}{1 + |j - i|} \quad (8)$$

2.2.2. Tamura

A description of the quantitative analysis Tamura provided six properties and gave a common description on all types of images texture. Those six various texture features specific by Tamura Contrast, Directionality, Coarseness, Roughness, Line-Likeness and Regularity [8][5].

2.2.2.1. Coarseness

Essentially connect to the distance in the gray levels of spatial changes, which implicitly linked to the size of primeval elements formation the texture. It's have the directly connection to scale and duplication averages and maximum primary texture feature. An image will include iterative textures pattern at different scales, Coarseness tries to find the largest size in which the tissue is present, even in the case of a smaller tissue, as in Equation 9.

$$R_M(x, y) = \sum_{i=x-2^{M-1}-1}^{x+2^{M-1}-1} \sum_{j=y-2^{M-1}-1}^{y+2^{M-1}-1} \frac{F(i, j)}{2^{2M}} \quad (9)$$

Where, $2^M \times 2^M$ size is the average of neighborhood.

$$S_{M,h}(x, y) = |R_M(x + 2^{M-1}, y) - R_M(x - 2^{M-1}, y)| \quad (10)$$

Equation 10 calculates the difference between pair of averages corresponding to non-overlapping neighborhoods.

2.2.2.2. Contrast

Measurement allocation of gray levels that change for extent is its distribution into black or white. Determine the contrast, use the central moments of the fourth order of gray levels and the second order

$$\text{Contrast} = \sigma / \alpha_4 \quad (11)$$

$$\alpha_4 = N_4 / \sigma^4$$

where, N_4 is the fourth moment about the mean and σ^2 is the variance. $m=1/4$ to give the closest value according to Tamura.

2.2.2.3. Directionality

Directionality of an image is measured by which the frequency of local edges directed against directional angles distributed. It is a global property over a region. This feature cannot distinguish between trends or patterns but measures the overall degree of directivity in an image by directivity. The most important feature among Tamura features through the matrix is to distinguish between one image and another in the consistency of the region.

$$\text{Directionality} = 1 - rm_{peaks} \sum_{p=1}^{m_{peaks}} \sum_{b \in w_p} (b - b_p)^2 H_{\text{directionality}}(b) \quad (12)$$

Where

m_p : number of peaks

b_p : is the position of the peak

w_p : is the range of the angles attributed to the P th peak

r : denotes a normalizing factor related to quantizing levels of the angles b and b denotes quantized directional angle

$H_{\text{directionality}}$: is the histogram of quantized direction values,

b : is constructed by counting number of the edge pixels with the corresponding directional angles.

2.2.2.4. Line-Likeness

Line-likeness refers only the shape of the texture primitives. A line-like texture has straight or wave like primitives whose orientation may not be fixed. Often the line-like texture is simultaneously directional. Line-likeness (flin) can be computed as follows

$$flin = \sum \sum PDD(i, j) n_j m_i \cos \left[\frac{2\pi n}{w} (i - j) \right] \sum \sum PDD(i, j) n_j m_i \quad (13)$$

Where $PDD(i, j)$ is $n \times n$ local direction co-occurrence matrix of points at a distance d .

2.2.2.5. Regularity

Regularity measures a constant pattern or comparable in an image. define in Equation 14.

$$Y_{\text{Regularity}} = 1 - R(C_{CRS} + C_{con} + C_{dir} + C_{lin}) \quad (14)$$

where R is a normalizing factor and C_{xxx} means the standard deviation of f_{xxx}

2.2.2.6. Roughness

Is the sum of the measures of coarseness and contrast

$$\text{Roughness} = \text{Coarseness} + \text{contrast} \quad (15)$$

2.3. Statistical Features

Approaches do not attempt to understand explicitly the hierarchical structure of the texture. Instead, they represent the texture indirectly by the non-deterministic properties that govern the distributions and relationships between the grey levels of an image. eleven types of texture features are discussed (Contrast, Entropy, RMS, Energy, Kurtosis, Correlation, Variance, Fifth and sixth central moment, Smoothness, Mean and standard deviation) [9-10].

2.3.1. Contrast

Contrast is a measure of intensity or gray-level variations between the reference pixel and its neighbor. Contrast is determined by the brightness of the object color, and other objects within the same display area. define in Equation 16

$$F = \sum_m^{M_g-1} m^2 \left[\sum_{i=0}^{M_g-1} \sum_{j=0}^{M_g-1} q_{d,\theta}(i, j) \right] \quad (16)$$

Where $m=|i-j|$ When i and j are equal, the cell is on the diagonal and $i - j = 0$. These values appear pixels entirely comparable to their neighbor, so there are specific a weight of 0. If i and j differ by 1, there is a small contrast, and the weight is 1. If i and j differ by 2, the contrast is increasing and the weight is 4. The weights continue to increase exponentially as $(i - j)$ increases.

2.3.2. Entropy

It is used to measure the system disturbance in the physics of thermodynamics. Entropy measurement is an ideal way to measure the level of unstable signal disturbance as well as to measure the amount of information contained in the event. define in Equation 17

$$\text{Entropy} = -\sum (P * \log (P)) \quad (17)$$

Where: P:probability vector.

2.3.3. RMS(Root mean square error)

The RMS value Gradually increase the value with the development of error However, the unable to provide the information Special of incipient fault stage while The value increases gradually as the error develops as define in Equation 18

$$R = \sqrt{\frac{1}{M} \sum_{j=1}^M |y_j|^2} \quad (18)$$

2.3.4. Energy

It is utilize to describe a measurement of information while Perform an operation under a probability frame this as (maximum a priori) assessment in coupling with Markov Random domain. Sometimes the energy can be a positive measure to maximize and sometimes it is a negative measure to minimize. It is specific by mean of [19]

$$F = \sum_j \sum_i q(j,i)^2 \quad (19)$$

2.3.5. Kurtosis

It Measures the stability of the distribution, which relates to the normal distribution

$$\text{Kurtosis} = \sum_{j=1}^M \sum_{i=1}^N \frac{(q(j,i)-m)^4}{(MN)\sigma^4} \quad (20)$$

2.3.6. Correlation

Correlation is the basic process used to extract the information from the image as shown in the following Equation 21

$$\text{Correlation} = \frac{\sum_i \sum_j q(i,j) - M_x M_y}{\sigma_x \sigma_y} \quad (21)$$

2.3.7. Variance

The variance is define as the mean of the signal square and is calculated after the mean value is removed, define in Equation 22

$$\sigma^2 = \frac{1}{q} \sum_{i=1}^q (Y_j - M)^2 \quad (22)$$

Where: σ =Variance, q =no of samples, Y_j =input heart signal μ = mean

2.3.8. Fifth and sixth central moment:

That give the deviation about average. Fifth central moment,

$$= \sum_{j=1}^M \sum_{i=1}^N \frac{(q(j,i) - m)^5}{(MN)\sigma^5} \quad (23)$$

Sixth central moment

$$= \sum_{j=1}^M \sum_{i=1}^N \frac{(q(j,i)-m)^6}{(MN)\sigma^6} \quad (24)$$

2.3.9. Smoothness

Comparative smoothness, Q is a measurement of gray level disparity that which can used to create relative smoothness recipes. The smoothness is specific by Equation 25

$$Q = 1 - \frac{1}{1 + \sigma^2} \quad (25)$$

Where, σ is the standard deviation of the image

2.3.10. Mean

Calculates the average values in the image

$$\text{Mean} = \sum_{i=1}^r \sum_{j=1}^t \frac{q(i,j)}{rt} \quad (26)$$

Where $q(i, j)$ is the intensity value of the pixel at the point (i, j) . The image is of r by t size.

2.3.11. standard deviation

Calculates the mean distance between the pixel value and the mean where the low standard deviation value indicates that there is less deviation of the pixels from the mean and the higher value indicates the high contrast, define in Equation 27

$$\sigma = \sqrt{\frac{\sum_{i=1}^r \sum_{j=1}^t (q(i,j) - m)^2}{rt}} \quad (27)$$

2.4. Geometry Features

There are eight types of geometry features as following :

2.4.1. Area

Is the extension of shapes, and it is different from the perimeter. Where the linked inside the shape, there are many known formulas for simple forms such as triangles, rectangles, and circles. Using these formulas, any polygon area can be calculated by dividing the polygon into triangles or circles to obtain curved shapes with borders and then collected after the calculation of their areas and when the polygon is irregular can polygon area is calculated by equation Gauss trapezoidal and described as in the following Equation(28)[11]

$$A = \frac{1}{2} \sum_{i=0}^{m-1} (p_i * q_{i+1}) - (p_{i+1} * q_i) \quad (28)$$

Where: m : number of points, p_i : X axis coordinates, q_i : y axis coordinates

2.4.2. Slope

The straight line is a set of points, that which has a fixed slope between any two points. The slope of the straight line is usually determine by the value of the ratio of vertical change to horizontal variation. The slope usually describes the slope of the two-point line. The parallel line of the x-axis is define as the horizontal line, Zero. The parallel line of the y-axis known as the vertical line, and its slope always has an undefined value. The parallel two lines always have slope equal [12]. This is described by the following Equation (29)

$$\text{Slope} = \frac{pz - p_0}{qz - q_0} \quad (29)$$

2.4.3. Perimeter

It is the length of the line that surrounds of two-dimensional shapes such as; the circle, square, rectangle or irregular shapes. The perimeter can be calculated as in Equation (30) if the shape is equilateral while the equation (31) calculates the perimeter if the shape is ribbing inequilaterally [13].

$$\text{Per} = n * (x) \quad (30)$$

$$\text{Per} = \sum_{i=0}^{n-1} x_i$$

Where: n is number of ribs, x :length of the rib

$$(31)$$

2.4.4. Centroid

The centroid is a fixed point in the object where the lines pass through this point, which represents the weight of the object. The centroid is different from each other in terms of form or acclimatization and thus determine the status of a centroid related to this difference. The centroid can calculate according to the following Equation (32) [14].

$$x_o = \frac{\sum x_{oi} A_i}{\sum A_i}, \quad y_o = \frac{\sum y_{oi} A_i}{\sum A_i} \quad (32)$$

Where: x_o : the x- axis value when center point of shape

y_o : the y- axis value when center point of shape

x_{oi} : The distance at which the center of the shape be far from the junction point of the axes on axis (x)

y_{oi} : The distance at which the center of the shape be far from the junction point of the axes on axis (y)

A_i : area the shape.

2.4.5. Irregularity Index[15]:

The boundaries of irregular shapes is calculated by the Equation(33)

$$L = \frac{4\pi \cdot A}{\text{per}} \quad (33)$$

Where: A:is area, Per: is perimeter

The metric value irregularity index (L) is equal to one only for circle and it is < 1 for any other shape.

2.4.6. Equivalent Diameter [16]:

Numerical that determine the diameter of the circle together with the same region as the area. It is calculate as in the following Equation (34).

$$Q_{\text{diameter}} = \sqrt{\frac{4 \cdot A}{\pi}} \quad (34)$$

Where: A: is area

2.4.7. Convex Area

The closed convex or convex represents the set X of points in the Euclidean level the smallest convex set contains X. For example, when X you are a limited subset of the plane. The convex area is the number of pixels in the convex image. The size of the square surrounding the area. Where the bounding box is a convex hull[17].

2.4.8. Solidity

Calculates the pixel ratio in a convex hull located in the area[18]

$$\text{Solidity} = \frac{A}{\text{Convex Area}} \quad (35)$$

Where: A: is area

3. Comparison and analysis

In this section, discuss the comparative results of features extraction and database used for general purposes including about 280 images by two categories with the KNN algorithm for classification as show in Table.1.

4. Discussion

In this paper, four main types and subtypes for each type of features extraction are collected, each feature extraction is necessary for special application anyway. The similarity measurement of efficiency conclusion includes TP values and TN values [19]. The equation (36) shows the accuracy four main types of extraction features and subtypes

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FP} + \text{FN}} \quad (36)$$

5. Conclusions

This section displays the database including two different categories and accuracy are gain from this review .Table 1 portrays the features and the database used

Table 1. Comparison of Features Extraction

Database Contents	Number of Samples	Type of features used	Accuracy	
			Training	Testing
Face	Training:113 Face images Testing :50 Face images	Geometry	100%	98%
		Statistic	99.5%	96%
		Color	98.8%	96%
		Texture	97%	95%
Plant	Training:100 Plant images Testing :30 Plant images	Geometry	100%	93%
		Statistic	100%	95%
		Color	100%	97%
		Texture	100%	98%

References

- [1] Ethem Alpaydin 2014 *Introduction to Machine Learning* (MIT Press).
- [2] Kavya R and Harisha 2015 *Feature Extraction Technique for Robust and Fast Visual Tracking: A Typical Review* (International Journal of Emerging Engineering Research and Technology Vol 3 Issue 1) PP 98-104,
- [3] S.R. Kodituwakku and S.Selvarajah 2011 *Comparison of Color Features for Image Retrieval* (Indian Journal of Computer Science and Engineering Vol 1, No 3) pp 207-211.

- [4] Amera H.M alzoubi 2015 *Comparative Analysis of Image Search Algorithm using Average RGB, Local Color Histogram Global Color Histogram and Color Moment HSV*(thesis, Faculty of Computer Science and Information Technology Universiti Tun Hussein Onn Malaysia).
- [5] C. Umamaheswari, Dr. R. Bhavani and Dr. K. Thirunadana Sikamani 2018 *Texture and Color Feature Extraction from Ceramic Tiles for Various Flaws Detection Classification* (International Journal on Future Revolution in Computer Science & Communication Engineering Vol 4 ,Issue. 1)pp 169 – 179.
- [6] M. S. Ahmad, M. S. Naweed, and M. Nisa 2009 *Application of texture analysis in the assessment of chest radiograph* (International Journal of Video & Image Processing and Network Security (IJVIPNS) Vol 9, No 9) pp 291-297.
- [7] .Fritz Albregtsen 2008 *Statistical Texture Measures Computed from Gray Level Co-occurrence Matrices* (Image Processing Laboratory Department of Informatics University of Oslo).
- [8] Peter Howarth and Stefan Ruger 2004 *Evaluation of Texture Features for Content-Based Image Retrieval* (Department of Computing, Imperial College London South Kensington Campus London).
- [9] Monika Sharma, R. B. Dubey and Sujata and S. K. Gupta 2012 *Feature Extraction of Mammograms*(International Journal of Advanced Computer Research Vol.2, No.3)
- [10] Jaspinder Kaur, Nidhi Garg and Daljeet Kaur 2014 *Segmentation and Feature Extraction of Lung Region for the Early Detection of Lung Tumor*(International Journal of Science and Research (IJSR) Vol 3, Issue 6).
- [11] A. H. Stroud 1971 *Approximate Calculation of Multiple Integrals*(Prentice-Hall Inc., Englewood Cliffs, N. J.).
- [12] Christopher Clapham, James Nicholson 2009 *Oxford Concise Dictionary of Mathematics*(OUP oxford).
- [13] Dr Yeap Ban Har, Dr Joseph Yeo, Teh Keng Seng, Loh Cheng Yee, Ivy Chow, Neo Chai Meng and Jacinth Liew 2018 *New Syllabus Mathematics Teacher's Resource Book1*(OXFORD UNIVERSITY Press) 7th edition.
- [14] Dan B. Marghitu and Mihai Dupac 2012 *Advanced Dynamics* (Springer) chapter 2 pp 73-141.
- [15] S. A. Patil and V. R. Udipi 2010 Chest x-ray features extraction for lung cancer classification(Journal of Scientific and Industrial Research Vol 69) pp 271-277.
- [16] R. C. Gozalez and R. E. Woods 2002 *Digital Image Processing Using Matlab*, 2nd ed, Gatesmark (USA) chapter 12 pp 642-654.
- [17] Nitin S. Lingayat and Manoj R. Tarambale 2013 *A Computer Based Feature Extraction of Lung Nodule in Chest X-Ray Image*(International Journal of Bioscience, Biochemistry and Bioinformatics Vol 3, No. 6) .
- [18] K. P. Aarthy and U. S. Ragupathy 2012 *Detection of lung nodule using multiscale wavelets and support vector machine*(International Journal of Soft Computing and Engineering (IJSCE) Vol 2, Issue 3).
- [19] Mariam A. Sheha , Mai S. Mabrouk and Amr Sharawy 2012 *Automatic Detection of Melanoma Skin Cancer using Texture Analysis*(International Journal of Computer Applications Vol 42, No 20)