Recognition of Retain Images Based on Association Rules

Raniah Ali Mustafa[1], Zahraa Salah Dhaief[2], Haitham Salman Chyad[3] and

Amal Abdulbaqi Maryoosh[4]

Research Scholar [1-4]

Department of Computer Science

College of Education

Baghdad

Iraq

_____

ABSTRACT

The objective of feature extraction is the transformation of input data into a set of features. Features are distinctive properties of input patterns that help in differentiating between the categories of input patterns. The main idea of the proposed system depends on the feature extraction where the system uses wavelet transforms. In this paper apply colour information (RGB colour space) in a specific form of two-level discrete wavelet transforms (DWT) of Retain image; obtain seven bands of texture features are extracted from wavelet coefficients. In the paper, the algorithm depends on the idea of extracting features from retaining images person to extract association rules between these features retina images of person recognition. These rules obtain from using Frequent Pattern (FP-growth) Association Rules (AR) where give for each features code. These rules are id for retina image for the person. The system was tested over a database collected from 30 volunteers, where 15 images for each person were available on the Internet. The used algorithm is Association Rules has been used to extract association rules between features to recognize retain images of person and The achieved training rate was 100% for feature extraction, while the achieved recognition rate for wavelet transform was 82%; an excellent recognition rate and the proposed system implemented in programing language Visual Basic 6.0.

Key Words: *Retain image, discrete wavelet transform, Association Rules, Aprioi, FP-growth.*
_____

1. INTRODUCTION

A pattern is a set of phenomena or concepts or objects where the components of the set are like each other in specific aspects or ways [1]. An individual is a very advanced information system; partly because it has superior pattern recognition ability. A pattern is defined as opposite of a chaos; It is an entity, ambiguous definition, which could be given a name. e.g., a pattern could be a human face, a sound wave, a bar code, a fingerprint image or a handwritten cursive word and other. As shown in Figure (1.1). [2]

In the other meaning pattern is the characterization of each member of the category represents a pattern class. That definition of Pattern class is a category specified through some common features or attributes. The features of a pattern class are the distinguishing attributes common to every patterns belonging to a class. Such features are usually denoted to as interesting features. The features represent the variation among pattern classes which might be denoted to as the interesting features [3,4].

Pattern recognition (PR) could be described as the classification of entry data to specified classes by the extraction of important attributes or features of the data from a background of irrelevant details [3]. It is described as the study of how machines can make acceptable decisions about the categories of the patterns, learn to discriminate different patterns of interest from them background and observe the environment. Through recognition, the given objects are assigned to a specified category. [1]

Figure 1.1: Example of Patterns: Face, Finger Print, Sound Wave, Tree, Character Image and Bar Code

Pattern recognition field has become interesting recently because of rising applications that aren't just challenging but also requesting more computation (view Table (1.1)). These applications include biometrics (personal identification based on different bodily attributes such as fingerprints and face), financial forecasting, document classification (effectively searching text documents), retrieval and organization of multimedia databases, and data mining (recognizing a pattern, e.g., an outlier in millions of multidimensional patterns, or correlation). [5]

Table 1.1: Pattern recognition applications

Problem Domain	Application	Input Pattern	Pattern Classes
Bioinformatics	Sequence analysis	DNA/Protein Sequence	Known types of genes/patterns
Data Mining	Searching for meaningful patterns	Points in multidimensional space	Compact and well-separated clusters
Documents Classification	Internet search	Text documents	Semantic categories (e.g. business, sports, etc.)
Documents Image Analysis	Reading machine for the blinds	Document image	Alphanumeric characters, words
Industrial Automation	Printed circuit board inspection	Intensity or range image	Defective/ non-defective nature of product
Multimedia Database Retrieval	Internet search	Video clip	Video genres (e.g., action, dialogue, etc.)
Biometric Recognition	Personal identification	Face, iris, fingerprint, retina image etc.	Authorized users for access control
Remote Sensing	Forecasting crop yield	Multispectral image	Land use categories, Growth pattern of crops
Speech Recognition	Telephone directory enquiry without operator assistance	Speech waveform	Spoken words

2. RETINA IMAGE

In each person's eye, the retina is that layer which is made of a complicated network of neural cells at the back of the eye. This part has a unique feature for each retina due to the complicated capillaries that is responsible for providing the retina with the blood [6]. Many studies have concluded that the blood vessel pattern is distinctive pattern for each single retina of the same person. This complication pattern of the retina has contributed in the development of new automatic authentication system for the fast distinguish among individuals [7]. In addition, any identical twins do not have the same patterns of their retina.

General scenario of the method: The technology of retinal scan-based authentication system uses a specific sensor for the data acquisition. In this process, a person is required to peep where the sensor directs on a particular spot in the visual field in which the capturing of characteristics of blood vessel pattern is performed [7].

2.1 Advantages: This biometric authentication technique has obtained a large amount of attention because of the high level of accuracy of samples that can be obtained from the recognition of individuals.

2.2 Disadvantages: It causes kind of user's discomfort due to the hard efforts by the contributors while capturing their retina vessels. However, retina-scan based biometrics has some medical factors that can affect the accuracy of its authentication process such as high blood pressure [6]. Also, different papers have addressed other conditions that can decrease the performance of retina scanning like wearing glasses, lenses [8].

3. WAVELETS IN COLOR IMAGES

In the last decade, wavelets have become very popular, and new interest is rising in this topic. The main reason is that a complete framework has been recently built [9] in particular for what concerns the construction of wavelet basis and efficient algorithms for its computation. The proposed work is based on the approach of the wavelet decomposition of faces images for the reasons that are explained hereafter. The main characteristic of wavelets (if compared to other transformations) is the possibility to provide a multi resolution analysis of the image in the form of coefficient matrices. Strong arguments for the use of multi resolution decomposition can be found in psycho visual research, which offers evidence that the human visual system processes the images in a multi scale way. Moreover, wavelets provide a spatial and a frequency decomposition of the image at the same time. Wavelets are also very flexible: several basis exist, and one can choose the basis which is more suitable for a given application. We think that this is still an open problem, and up to now only experimental considerations rule the choice of a wavelet form. However, the choice of an appropriate basis can be very helpful.

Computational complexity of wavelets is linear with the number (N) of computed coefficients $O(N)$ while other transformations, also in their fast implementation, lead to $N{\times}log_2(N)$ Thus, wavelets are adapted also for dedicated hardware design (Discrete wavelet Transform). If the recognition task has real time computation needs, the possibility of embedding part of the process in hardware is very interesting, like in compression tasks [10].

In *classical wavelet decomposition*, the image is split into an approximation and details images. The approximation is then split itself into a second-level approximation and details. For a *n*-level decomposition, the signal is decomposed in the following way: [3]

$$A_n = \left[H_x * \left[H_y * A_{n-1} \right]_{\downarrow 2,1} \right]_{\downarrow 1,2}$$

$$D_{n1} = \left[H_x * \left[G_y * A_{n-1} \right]_{\downarrow 2,1} \right]_{\downarrow 1,2}$$

$$D_{n2} = \left[G_x * \left[H_y * A_{n-1} \right]_{\downarrow 2,1} \right]_{\downarrow 1,2}$$

$$D_{n3} = \left[G_x * \left[G_y * A_{n-1} \right]_{\downarrow 2,1} \right]_{\downarrow 1,2}$$

where $\downarrow$ denotes the convolution operator, $\downarrow2,1$ ($\downarrow1,2$) sub-sampling along the rows (columns) and $A_0 = I(x,y)$ is the original image. A_n is obtained by low pass filtering H_x and is the approximation image at scale n . The details images D_{ni} are obtained by high pass filtering G_x in a specific direction and thus contain directional detail information at scale n. The original image I is thus represented by a set of sub images at several scales; { A_n, D_{ni}}.

The wavelet packet decomposition, is a generalization of the classical wavelet decomposition. In that case, the details as well as the approximations can be split. This results in a wavelet decomposition tree.

4. HAAR BASIS FILTER

The Haar basis filter contains of Low pass filter (LPF) and high pass filter (HPF) are represented in the following equations (1)-(2):

$$\text{HPF: } \frac{1}{\sqrt{2}} [1 \quad -1], \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots . (1)$$

$$\text{LPF: } \frac{1}{\sqrt{2}} [1 \quad 1], \dots \dots \dots \dots \dots \dots \dots \dots . \dots \dots \dots (2)$$

The LP and HP filters are so termed the decomposition filters for they distinct decompose the picture or the picture down to detailed and approximation. The convolution with the band pass filter in a definite direction results in so termed details picture and the convolution with the LPF outcomes in a so termed approximation picture.

LL band (approximation band) is the outcome of implementing LPF in vertical and horizontal directions and its filter gain as follows:

$$LL = \frac{1}{\sqrt{2}} (1 \quad 1)^t . \frac{1}{\sqrt{2}} (1 \quad 1) = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \dots \dots \dots \dots \dots \dots (3)$$

LH band (detail band) is the effect of applying horizontal LPF and vertical HPF and its filter is derived as follows:

$$LH = \frac{1}{\sqrt{2}} (1 \quad 1)^t . \frac{1}{\sqrt{2}} (1 \quad -1) = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}, \dots \dots \dots . \dots (4)$$

HL band (detail band) is the result of implementing vertical HPF and horizontal LPF and its filter is gain as follows:

$$HL = \frac{1}{\sqrt{2}} (1 \quad -1)^t . \frac{1}{\sqrt{2}} (1 \quad 1) = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}, \dots \dots \dots \dots (5)$$

HH band (detail band) is the outcome of implementing vertical and horizontal HPF and its filter is gain as follows:

$$HH = \frac{1}{\sqrt{2}} (1 \quad -1)^t . \frac{1}{\sqrt{2}} (1 \quad -1) = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \dots \dots \dots (6)$$

The transformed bands (LL scaling bands and three wavelet bands HH HL, and LH) can be gained by implementing above filters to every (2×2) adjacent pixel of the entire picture pixel. This method is named Fast Mallat transform algorithm [11].
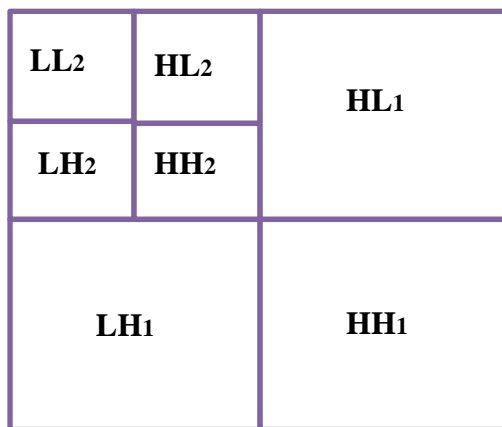

Figure 1.2: Two level Wavelet Sub Band Decomposition

5. DISCRETE WAVELET TRANSFORM DECOMPOSITION

The DWT for 2-dimensional images $x[m,n]$ can be identically defined through applying 1-dimensional DWT to every dimension m and n independently: $DWT_n \left[DWT_m [x[m,n]] \right]$. 2-dimensional WT decomposes a image to "subbands" which are centralized in orientation and frequency. WT is formed through passing the image by sequences of filter bank phases. One phase is demonstrated in Figure 1.3. In which an image is initially filtered in the horizontal direction [11, 12].
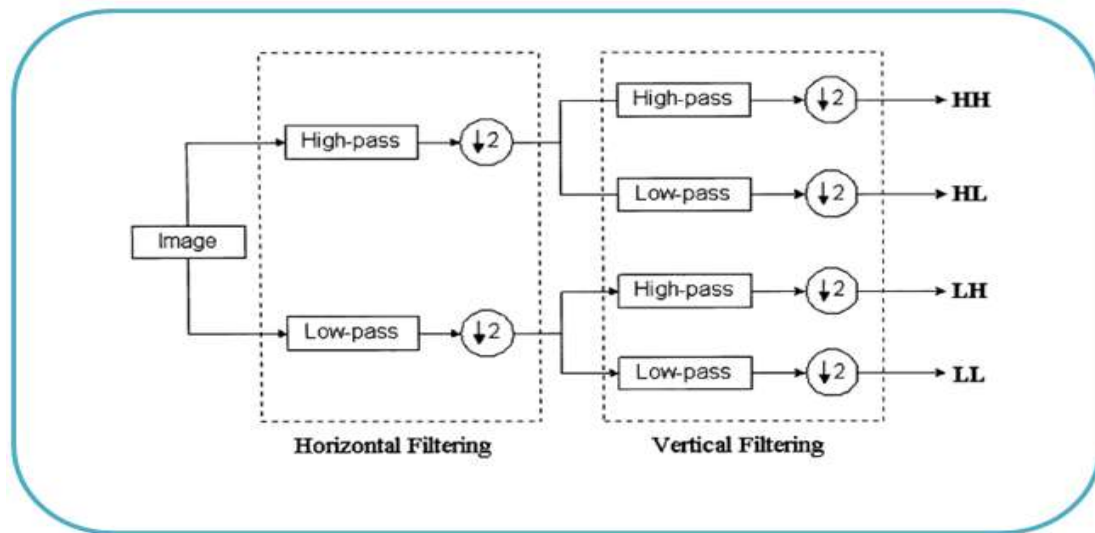
Figure 1.3: Illustration of 2-D Wavelet Transforms

The scaling function (LPF) & wavelet function (HPF) are limited pulsation reply filters. In other words, the output at every point relies just on a limited part of the input. The filtered outputs are then down sampled by a factor of 2 (↓2) in the horizontal direction. Those signals are then each filtered through a similar filter pair in the vertical direction. Eventually decomposite the image to 4 sub bands indicated by (LL, HL, LH, HH). Each of those sub bands could be thought of as a smaller version of the image which represents various image properties. The Low-Low is a coarser approximation to the original image. Low High & High Low records the differences of the image along vertical and horizontal directions, consecutive. High offers the high frequency component of the image. 2-Level decomposition could then be conducted on the Low sub band. Figure 1.4 demonstrates two-Level wavelet decomposition [12].
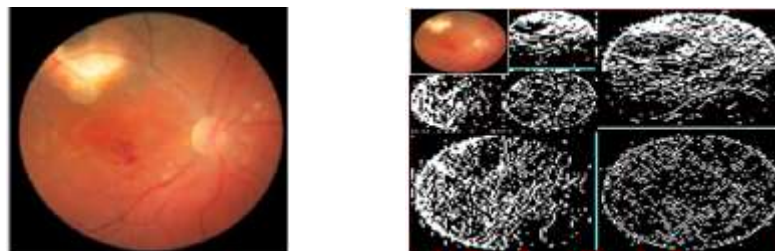

Figure 1.4: (a) Original Image (b) Two-Level Wavelet Decomposition

6. STATISTICAL MEASURES AS DISCRIMINATING FEATURES

The statistical measures, such as Mean, Standard Deviation, Variance, Energy, Entropy is typically useful to representation characteristics or features of data, it is simple and needs minimal calculation load [13, 3].

1. **Mean:** Is a set of intensity values for the elements of image are divided by the number, and is represented by the following equation:-

$$\mu = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} x(i,j), \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots (7)$$

2. **Standard Deviation:** The Standard Deviation is the most utilized indicator of variability and is a measurement concerning to the average distance of the scores from their mean value. It is computed on the basis of the positive square root of the variance, and is represented by the following equation:-

$$\sigma = \sqrt{\frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (x(i,j) - \mu)^2} \ , \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots (8)$$

The standard deviation is significant in detecting the details contented in an image.

3. **Variance:** Is the average of the total square deviations values for the arithmetic mean .In fact it is almost identical to the standard deviation, and is represented the following equation:-

$$\sigma^2 = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(|x(i,j)| - \mu \right)^2, \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (9)$$

4. **Energy:** Is intended power to distribute information in the image, which is usually the factor that image is approaching compressibility, and is represented by the following equation:-

$$ENE = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} x^2(i,j), \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (10)$$

5. **Entropy:** Numerical value representing the entropy of the gray image. Entropy is a statistical measure of the random that could be utilized to describe the texture information of the input image, and is represented by the following equation:-

$$ENT = - \sum_{i=1}^{M} \sum_{j=1}^{N} (x(i,j) \log_2 (x(i,j)), \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (11)$$

where N and M are size of row and column of the x (i, j), a wavelet coefficient.

7. ASSOCIATION RULES TECHNIQUE

Association rules are one of the most researched areas of data mining and have recently received much attention from the database community. They have proven to be quite useful in the marketing and retail communities as well as other more diverse fields [14]. Association rules, first introduced in 1993 are used to identify relationships among a set of items in a database. These relationships are not based on inherent properties of the data themselves, but rather based on co-occurrence of the data items [14].

7.1 APRIOI ALGORITHM

This algorithm is one of the algorithms that have the influence on frequencies item sets for association rules (AR). It was proposed through Srikant and Agrawal [15].The algorithm of names relies on the truth that the algorithm utilizes prior knowledge of frequencies item set characteristics. Algorithm which utilizes a repeated way also it is identified as a level-wise (L-w) seeking, where k-item sets are utilized for exploration of (k+1)-item sets. Initially, the group of frequencies of 1-itemsetsis found out. This set is indicated throughL1. L1 is utilized to find outL2, the set of frequencies of 2-itemsets that is utilized to find outL3, and other, even when no more frequent k-item sets could be discovered. To find out of every L_k needs one entire scan of the DB [16].

7.2 FP-GROWTH ALGORITHM

Recently, an FP-tree based frequent pattern mining method, called FP- growth, developed by Han et al achieves high efficiency, compared with an Apriori- like approach. The FP-growth method adopts the divide-and-conquer strategy, uses only two full I/O scans of the database, and avoids iterative candidate generation [17]. FP-growth is used for mining the complete set of frequent patterns by pattern fragment growth. Efficiency of mining is achieved with three techniques [18]:

1. A large database is compressed into a highly condensed, much smaller data structure, which avoids costly, repeated database scans.

2. FP-tree- based mining adopts a pattern fragment growth method to avoid the costly generation of a large number of candidate sets.

3. A partitioning-based divide-and-conquer method is used to dramatically reduce the search space.
   The frequent Pattern mining problem can be formally defined as follows. Let I = {i1, i2, ..., im} be a set of items, and DB be a transaction database, where each transaction T is a set of items and T ⊆ I. A unique identifier, called TID, is assigned with each transaction. A transaction T contains a pattern P, a set of items in I, if P ⊆ T. The support of a pattern P is the number of transactions containing P in DB. We say that P is a frequent pattern if P is less than a predefined minimum support threshold ξ, frequent pattern mining which consists of two steps [19]:

1. Construct a compact data structure, frequent pattern tree (FP-tree), which can store more information in less space.

2. Develop an FP-tree based pattern growth (FP-growth) method to uncover all frequent patterns recursively.

   Apriori algorithm may still suffer from the following two nontrivial costs [20, 21]:

1. It may still need to generate a huge number of candidate sets. For example, if there are 10^4 frequent 1-itemsets, the Apriori algorithm will need to generate more than 10^7 candidate 2-itemsets.

2. It may need to repeatedly scan the whole database and check a large set of candidates by pattern matching. It is costly to go over each transaction in the database to determine the support of the candidate item sets.

8. Normalization Features

An important step between feature extraction and distance computation is feature normalization. Complex image database retrieval systems use features that are generated by many different feature extraction algorithms with different kinds of sources. These feature vectors usually exist in a very high dimensional space. Not all of these features have the same range [22]. The normalization of feature (*f*) is performed using the following equation (12):

$$f\,\mathrm{norm} = \begin{cases} 0 & \text{if } f \leq \text{ fmin} \\ \dfrac{f - \text{fmin}}{\text{fmax} - \text{fmin}} & \text{if fmin} < f < \text{fmax} \\ 1 & \text{if } f \geq \text{fmax} \end{cases} \quad, \cdots\cdots\cdots (12)$$

where *fmin* and *fmax* are *min* and *max* feature values found over all image samples listed in the image database, *form* is the normalized value. The applied normalization process maps the extracted feature values to the range between [0, 1]. [11]

9. THE PROPOSED SYSTEM

The main idea for proposed system that using retain image and then feature extraction using algorithm depends on wavelet transforms where using colour information (RGB colour space) in a specific form of discrete wavelet transform (DWT); then texture features are extracted from wavelet coefficients. Every retina image of person is described by a subset of band filtered images containing wavelet coefficients. These coefficients characterize the retina texture and a set of simple measures such as (Entropy, Energy, Standard derivation, Variance and Mean) to allow the user form compact and meaningful feature vectors. These features are used to extract association rules (Fp-growth) which describe all retain images in data mining term. The proposed system passes through two main phases the training phase and testing phase. The two phases share the two main modules of the system (pre-processing module and feature extraction module). The proposed system is done in following phases:

1. Training Phase   2. Testing Phase

Retina image preparation   Retina image preparation

Retina image pre-processing   Retina image pre-processing

Feature Extraction   Feature Extraction

Training Database Features   Training Database Features

Extract Association rules   Extract Association rules

                                                       Recognition Process

                                                       Decision Result

In this paper, each phases in the proposed system will explained with the results. The figure 1.5 illustrates the each phases for the proposed system.
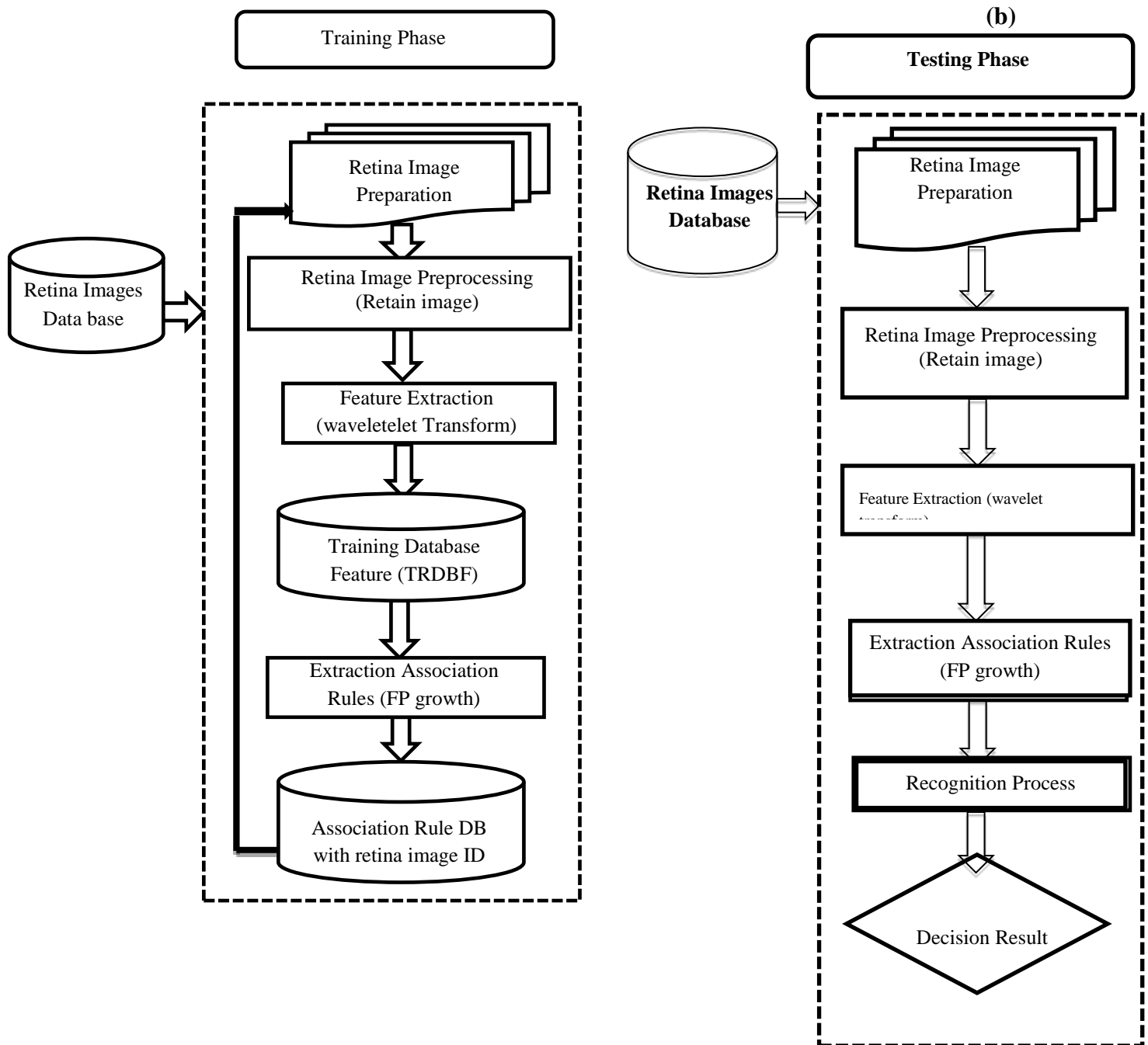
Figure 1.5: The proposed system for Retina Image (a) Training Phase (b) Testing Phase

9.1 READ IMAGE

In this phase the color retina image is loaded from retina image data set. In retina image every pixel has 3 colors (RGB) a color in the RGB color model is depicted by identifying how much of each of the red, green and blue is included in the color expressed as an RGB triple (r, g, b) every component of which can differ from 0 to 255.

9.1 RETINA IMAGE PRE-PROCESSING

After reading the color retina image as bmp file format, apply Gaussian filter as expressed in equation (13) is used in this stage is to reduce the existing noise and improve the shape of retina details. The sigma which is equal to (0.8) and kernel size (3x3) have been chosen in the proposed system. Figure (1.6) explains application of Gaussian filter with kernel size (3x3) and sigma value (0.8) with its array value that are produced for case. The smoothing step is processed to achieve better edge extraction results.

$$f(x,y) = \frac{1}{2\pi\sigma^2} exp\left(-\frac{x^2+y^2}{2\sigma 2}\right), \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. \ldots\ldots (13)$$

Sigma
Kernel Size 3x3
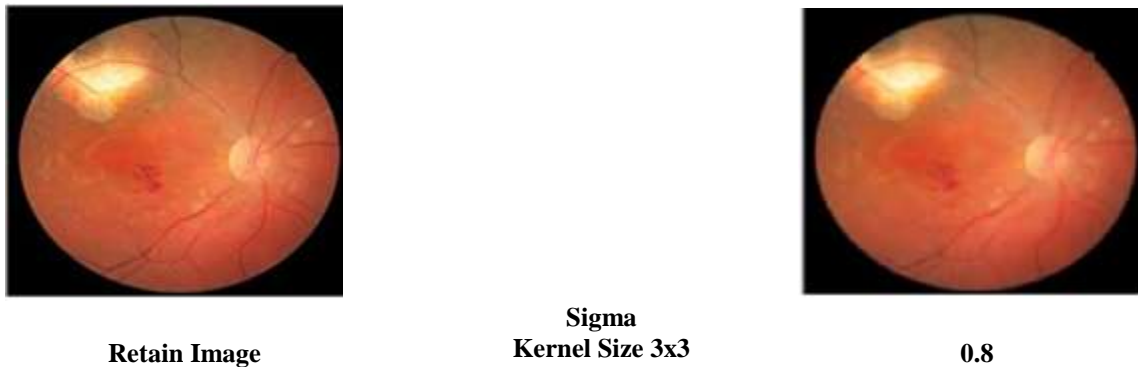
0.8

Retain Image

Figure 1.6: The applying Gaussian filter with sign.

9.3 FEATURE EXTRACTION BASED ON DISCRETEWAVELET TRANSFORM STAGE

In this phases, retina image will passing several sub stage for the purpose of extracted texture features, which uses for extracted association rules for retina image. The proposed system uses two collections of images (Red and green bands color sub images) which are entered next stage to be transformed. Figure (1.7) explains images preprocess before entering the next phase.
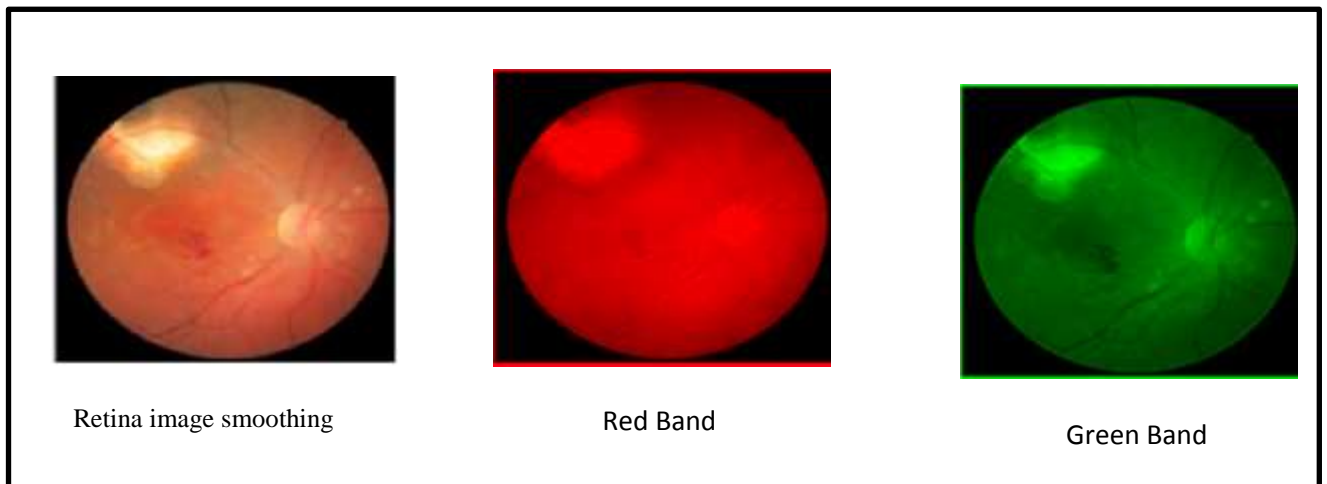

Retina image smoothing

Red Band

Green Band

Figure 1.7: Images Preprocess for Two Bands

### 9.4 Statistical Measures for Feature Extraction

In this paper, we will compute statistical measures for each bands (LH1, HL1, HH1, LL2, HL2, LH2 and HH2 bands) are usually useful to represent characteristics or features of data; it is simple and requires less computation load. See Figure (1.8) Statistical measures adopted by:

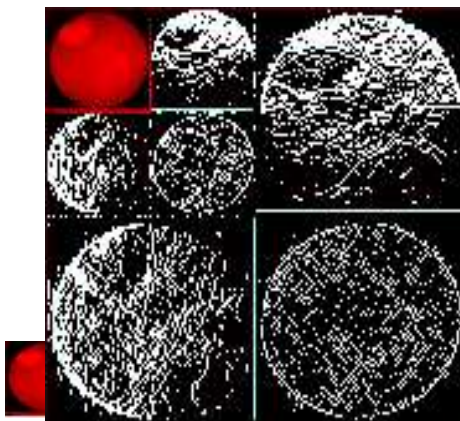   a.   Energy b. Entropy c. Standard Deviation d. Variance e. Mean

The result is a vector with (28) for energy values E = (El, E2 ,............. E28), (28) for entropy values En = (Enl, En2, ... En28), (28) for standard deviation values ST = (STl, ST2 ,............. ST28), (28) for variance values V = (Vl, V2 ,............. V28), and (28) for Mean values M = (Ml, M2 ,............. M28). Therefore will obtain (140) features for retina image
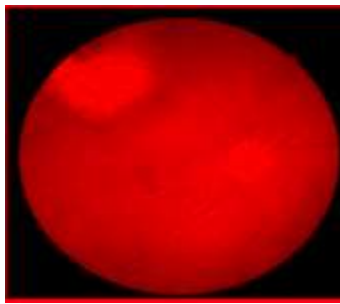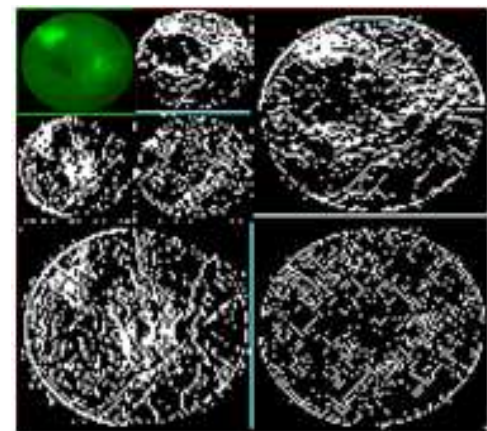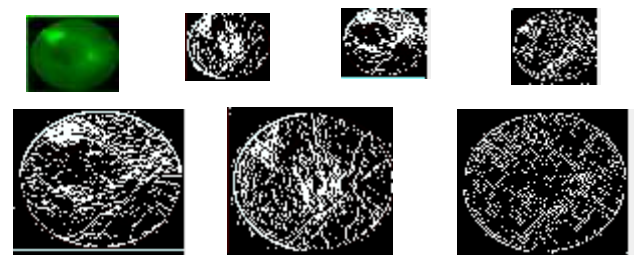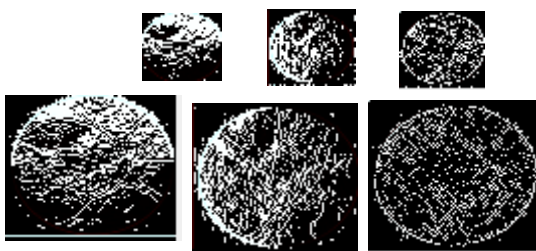
Red Band
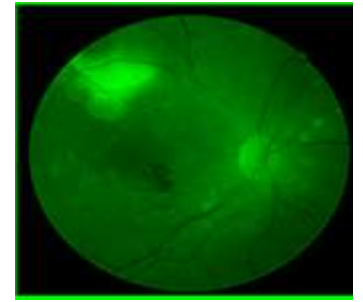
Green Band

Two Level Wavelet Transform (Red Band)

Two Level Wavelet Transform (Green Band)

LH1, HL1, HH1, LL2, HL2, LH2 and HH2 bands

Figure 1.8: Two Level Wavelet Decomposition of retina Area

9.5 Training Database Features (TRDBF)

The extracted features from wavelet coefficients by applied a set of simple measures are stored in database which is called (TRDBF). This database includes (5) attributes for features such as (Energy, Entropy, Standard deviation, Variance, Mean,)

Table 1.2: Result of Extracted Features for Some Sample of retina Database (Before Normalization)

Class	Sub-Image	Energy	Entropy	Standard deviation	Variance	Mean
C1	GLH1	19223.5084	3.33161522508345	6768.16378541975	343560307697.005	579457
	GHL1	27434.3874666667	4.48460537319134	9751.85476060414	713240034539.383	834906
	GHH1	23278.4454666667	4.12806852329301	8274.06637193801	513451307454.266	708385
	GLL2	19695.6329333333	7.16092660258417	12044.1489274552	1087961425400.4	1031161
	GLH2	15382.3124	2.57495374238022	5367.00907680893	216035898229.121	459497
	GHL2	25969.0628	3.88238612268885	9133.19052835017	625613769203.59	781939
	GHH2	20456.9948	3.61715891621585	7220.3036552158	390995886551.42	618167
	RLH1	19551.3792333333	3.42920098895599	13689.6435356939	5622190204030.98	2357443
	RHL1	27369.5682666667	4.61752695863448	9762.25012251758	714761455909.459	835796
	RHH1	22215.0331333333	4.10155347111544	15727.9370170439	7421040084363.02	2708450
	RLL2	25396.8313333333	7.41140953408381	13558.8583893326	1378819806164.81	1160843
	RLH2	25865.6770666667	4.48792295538651	18350.3263994003	10102034368935.8	3160042
	RHL2	20530.8684	3.74557858719745	7272.39726914549	396658215302.061	622627
	RHH2	23543.5946666667	4.32253152297974	8404.28872746619	529740517609.614	719534

- **Extracted Features after Normalization:** Then apply normalization frequency on the data base where to obtain the following results.

Table 1.3: Result extracted features from sample retina Database (After Normalization)

Class	Features					
	Sub-Image	Energy	Entropy	Standard deviation	Variance	Mean
C1	GLH1	0.5669	0.3156	0.798	0.7288	0.8328
	GHL1	0.8698	0.3955	1	1	1
	GHH1	0.3459	0.1766	0.641	0.5468	0.7028
	GLL2	0.8309	1	0.631	0.1176	0.2597
	GLH2	0.6772	0.3613	0.2339	0.0317	0.0963
	GHL2	0.9946	0.4223	0.3385	0.0504	0.1393
	GHH2	0.4272	0.242	0.1468	0.0183	0.0604
	RLH1	0.3187	0.1564	0.1079	0.0129	0.0444
	RHL1	0.8784	0.2703	0.2901	0.0414	0.1194
	RHH1	0	0	0	0	0
	RLL2	0.3579	0.9482	0.5143	0.0882	0.2117
	RLH2	0.6552	0.3211	0.2239	0.0301	0.0922
	RHL2	1	0.3948	0.3377	0.0503	0.139
	RHH2	0.4211	0.2155	0.1427	0.0177	0.0588

9.6 Apply FP-growth: After get features will apply FP-growth algorithm where features encoding to find repetition and draw tree then find the rule.

Algorithm (1):   Customized FP-Growth Algorithm

Input: A transactional database DB, minimum support threshold

Output: Its frequent pattern tree, tree structure

Process

Begin

The FP-tree is constructed in the following steps:

Scan the transaction database DB once. Collect the set of frequent items F and their supports. Sort F in support descending order as L, the list of frequent items;

Create the root of an FP-tree, T, and label it as "root";

While Trans in DB do begin:

Select and sort the frequent items in Trans according to the order of L;

Sorted frequent item list = [p | P], where p is the first element and Pis the

remaining list;

Call *insert tree function ([p | P], T);*

Call FP-growth (FP-tree, null).   // The FP_tree is mined

End Process;

Insert   tree function ([p | P], T) :

Begin

If T has a child N such that (N.item-name = p.item-name) then

      Increment N's count by 1;

Else

Create a new node N, and let its count be 1, its parent link be linked to T, and its   node-link be linked to the nodes with the same item-name via the node-link structure;

End If;

If P is nonempty, call insert_tree (P, N) recursively;

End;

FP-growth function (FP-tree, null):

// Procedure FP-growth(Tree, A)

Begin

If Tree contains a single path P then

      for each combination (denoted as B) of the nodes in the path P generate pattern B ∪ A with support=minimum support of nodes in B

Else

    for each ai in the header of the Tree do

     {

Generate pattern B = ai ∪ A with support = ai.support;

Construct B's conditional pattern base by traverse the FP-tree by following the

    link of each frequent item and accumulate all of transformed prefix paths of that item ;

Construct B's conditional FP_tree (Tree B) by accumulate the count for each item

   in the conditional pattern base >= minimum support;

If Tree B ≠ Ø then

    call FP-growth(Tree B, B);

End If;

   }

End If;

  End;

In this paper using FP-growth, where will encode the features to find the frequency and that the advantages of the algorithm is to reduce of scan and then will build the tree and then extract the rules for retrieval retina images person. The table (1.4) illustrates the data base after encoding and table (1.5) illustrate the frequency.

Table 1.4: Encoding database

Id Set (class)	Item_Set
C1	a,b,c,d,e
C2	e,b,e,d,e
C3	a,b,a,d,d
C4	a,e,d,e
C5	c,b,c,d,c
C6	d,b,d,d,e
C7	a,a,c,d,a
C8	a,b,c,d,e
C9	a,b,c,b
C10	a,b,d,e

Table 1.5: Frequency Item set (First step of FP-growth algorithm)

Item ID	Support Count
A	10
B	8
C	7
D	12
E	9

Then will build the tree and then extract the rules for recognition retina images person. The table (1.6) illustrates rules (id) will used for recognition retina images person.

Table 1.6: Result Association Rules Extracted database (FIAEDB) and their Confidence from sample retina images person

Class	Association Rules			
	Symbol	**Equation**	**Relation**	**Confidence**
	C26	A,C,/A,	0.1094/0.1094	A-->C
C1	C26	B,G,/B,	0.1094/0.1094	B-->G
	C26	C,D,/D,	0.1094/0.1094	D-->C
	C26	E,G,/E,	0.1094/0.1094	E-->G

9.7 Evaluation Criteria

The evaluating performance of retina recognition system is calculated by two measures called False Alarm Rate (FAR) and Recognition Rate (RR). The formula for calculating these measures are given as in equations (14) and (15) respectively.

1. **FAR**: is defined as the ratio between the numbers of false recognition decision to the total number of attempts.

$$FAR = \frac{\text{Number of false recognition attempts}}{\text{Total number of attempts}} * 100, \dots \dots \dots \dots \dots \dots \dots \dots \dots (14)$$

2. **Recognition Rate**: is defined as the ratio between the numbers of correct recognition decision to the total number of attempts.

$$RR = \frac{\text{Number of correct attempts}}{\text{Total number of attempts}} * 100, \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots (15)$$

10. Retina Image Dataset

The retinal images are considered as the raw material to be enhanced, segmented and evaluated. Image datasets are normally accompanied with a ground truth which acts as a benchmark for comparing and evaluating the achieved experimental results using the true results provided usually by medical experts (i.e. ophthalmologists) on an image dataset. In the paper use images retina is dataset standard and size (200,250) and bitmap format .In the figure (1.9) show some sample retina image.
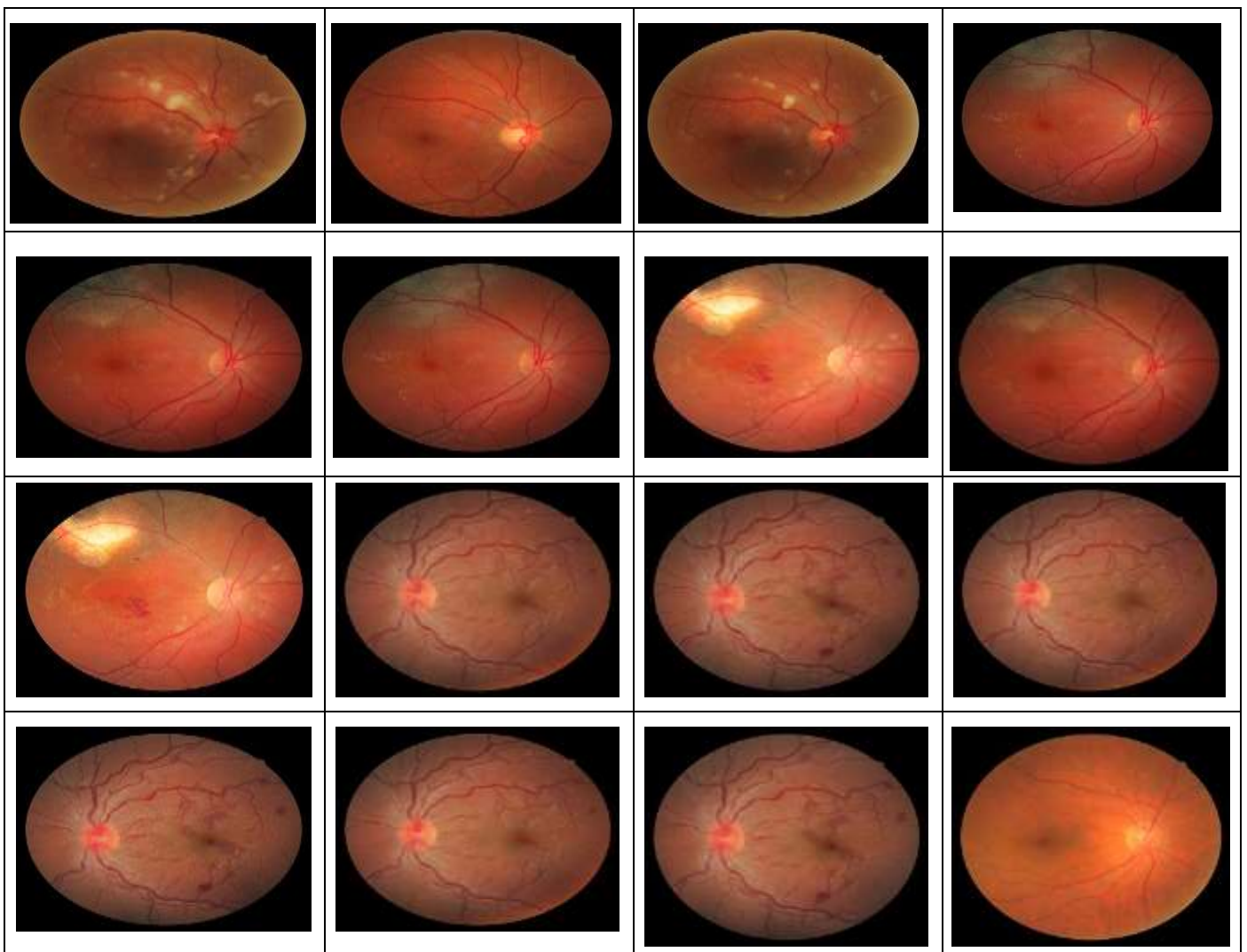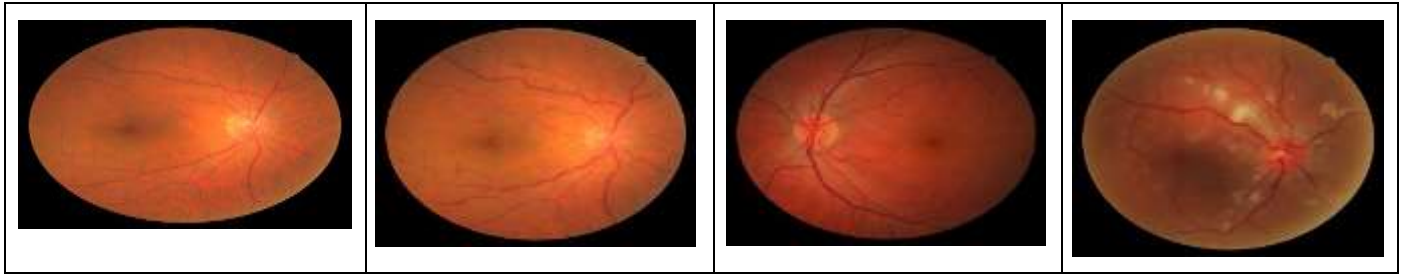
Figure 1.9: Dataset Images

11. CONCLUSION

In this paper extract features of retina region operation achieved good result and the algorithm used give good results. Where using two level wavelet transforms to extract the features give the best result in recognition because use two color band (red and green bands). Applying Association FP growth in recognition achieved good result because use wevelet from where extraction. The system proposed achieved good percentage (82%).

REFERENCES

[1] P. Sharma and M. Kaur, "Classification in Pattern Recognition: A Review", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 4, pp. 298-306, April 2013.

[2] M. A. Anjum, "Improved Face Recognition using Image Resolution Reduction and Optimization of Feature Vector", Ph.D. Thesis. Department of Computer Engineering of the National University of Sciences and Technology (NUST) Rawalpindi Pakistan, January, 2008.

[3] A. W. S. Ibrahim, "Face Recognition Using Skin Color and Texture Features", Ph.D. Thesis, Department of Computer Sciences of the University of Technology, 2005.

[4] M. Parasher, S. Sharma, A.K. Sharma and J. P. Gupta, "Anatomy on Pattern Recognition", Indian Journal of Computer Science and Engineering (IJCSE), Vol. 2, No. 3, Jun-Jul 2011.

[5] A. K. Jain, R.P.W. Duin, and J. Mao, " Statistical Pattern Recognition: A Review", IEEE Transactions on Pattern Analysis and Machine Intelligence, VOL. 22, NO. 1, January 2000.

[6] K. Dharavath , F.A.Talukdar and R.H.Laskar , "Study on biometric authentication systems, challenges and future trends: A review",in Computational Intelligence and Computing Research (ICCIC), 2013 IEEE International Conference on , vol., no., pp.1-7, 26-28 Dec.2013.

[7] A.N. Kataria, D.M. Adhyaru and etc., "A survey of automated biometric authentication techniques", in Engineering (NUiCONE), 2013 Nirma University International Conference on , vol., no.,pp.1-6, 28-30 Nov. 2013 .

[8] A.K. Jain, A. Ross, and S.Prabhakar, "An introduction to biometric recognition", in Circuits and Systems for Video Technology, IEEE Transactions on , vol.14, no.1, pp.4-20, Jan. 2004.

[9] "Haar wavelet transform", Web site:
http://www.owlnet.rice.edu/~elec301/projects99/imcomp/Haar.htm

[10] Gundimada, Satyanadh and Asari, Vijayan, "Face Detection Technique Based on Rotation Invariant Wavelet Features", VLSI Systems Laboratory, Department of Electrical and Computer Engineering, Old Dominion Sebastien Marcel and Samy Bengio University, Norfolk, E-mail: {sgund002, vasari}@odu.edu , Web site:

http://csdl.computer.org/comp/proceedings/itcc/2004/2108/02/210820157abs.htm

[11] Raniah Ali Mustafa, Kawther Thabt Saleh and Haitham Salman Chyad," Feature Extraction Based on Wavelet Transform and Moment Invariants for Medical Image", International Journal of Engineering Research and Advanced Technology (IJERAT), E-ISSN : 2454-6135, Volume.4, Issue 8, August -2018.

[12] B.L. Zhang, H. Zhang and S. S. Ge, "Face Recognition by Applying Wavelet Subband Representation and Kernel Associative Memory", IEEE Transactions on Neural Networks, Vol.15, No.1, January 2004.

[13] Z. M. Hussain, "Development of a Face Recognition System Based on   Canonical Correlation Analysis", Ph.D.Thesis, Informatics Institute for Postgraduate Studies at the Iraqi Commission for Computers and informatics, October 2006.

[14] M. H. Dunham, Xiao Y., Gruenwald L., and Hossain Z., "A Survey of the Association Rules", PDF, 2000.

[15] I. H. Witten, E. Frank and M. A. Hall, "Data Mining Practical Machine Learning Tools and Techniques", Elsevier Inc, 2011.

[16]R. Agrawal, T. Imielinski and A. Swami, "Mining Association Rules Between Sets of Items in Large Database", Washington, USA, May 1993.

[17] F. Verhein, "Frequent Pattern Growth (FP-Growth) Algorithm An Introduction", School of Information Technologies, the University of Sydney, Australia, PDF, 2008.

[18] Han J., Pei J., and Yin Y., "Mining Frequent Patterns without Candidate Generation", University of Illinois at Urbana-Champaign, PDF, 2000.

[19]J. Han J. and Yin Y., "Mining frequent patterns without candidate generation", In Proc. of the ACM SIGMOD Conference on Management of data, PDF, 2000.

[20] Mazman S., and Usluel Y., "The Usage of Social Networks in Educational Context", World Academy of Science, Engineering and Technology, PDF, 2009.

[21] Agrawal R. and Srikant R., "Fast Algorithms for Mining Association Rules in Large Databases", Proceedings of the Twentieth International Conference on Very Large Databases, PDF, 1994.
[22] C. Nastar, M. Mitschke, and C. Meilhac," Efficient Query Refinement For Image Retrieval", In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, June 1998