

# **The Implementation of Cluster of Identification of Data**

By

**Zafirah Muhamad Zulkapley**

A project dissertation submitted in partial fulfilment of  
the requirement for the  
Bachelor of Technology (Hons)  
(Information Communication Technology)

JANUARY 2007

Universiti Teknologi PETRONAS  
Bandar Sri Iskandar  
31750 Tronoh  
Perak Darul Ridzuan

t

QA

76.73

.B3

Z11

1) Visual programming (computer science)  
2) BASIC (computer program language)

# **CERTIFICATION OF APPROVAL**

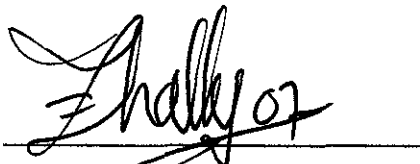
## **The Implementation of Cluster Identification of Data**

by

Zafirah Muhamad Zulkapley

A project dissertation submitted to the  
Computer and Information Sciences Programme  
Universiti Teknologi PETRONAS  
in partial fulfilment of the requirement for the  
BACHELOR OF TECHNOLOGY (Hons)  
(INFORMATION COMMUNICATION TECHNOLOGY)

Approved by,

A handwritten signature in black ink, appearing to read 'Jalil Ahmad', is written over a horizontal line.

(Mr. Jalil Ahmad)

UNIVERSITI TEKNOLOGI PETRONAS

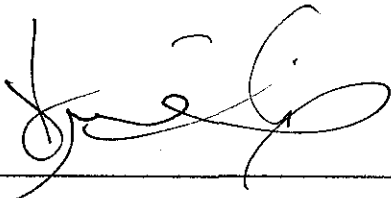
TRONOH, PERAK

January, 2007

## **CERTIFICATION OF ORIGINALITY**

January 2007

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.



---

**ZAFIRAH MUHAMAD ZULKAPLEY**

## **ABSTRACT**

In this paper, we combine few transformations to create a single unique transformation based on Threshold, Edge Detector, Simple Skeleton and Hough Line Transformation with information-theoretic-based criteria for unsupervised hierarchical image-set clustering. The continuous image modeling is based on mixture of Gaussian densities. The unsupervised image-set clustering is based on a generalized version of a recently introduced information-theoretic principle, the information bottleneck principle. Images are clustered such that the mutual information between the clusters and the image content is maximally preserved. Experimental results demonstrate the pattern of the image skeleton. Information theoretic tools are used to evaluate cluster quality. Particular emphasis is placed on the application of the clustering for efficient image search and verification. The application is very suit to offer authentic-looking counterfeit checks.

## **ACKNOWLEDGEMENTS**

**Bismillahirrahmannirrahim**

**Alhamdulillah, thanks to the Almighty for His blessings, this Final Year Project is successfully accomplished.**

**I would like to acknowledge the contributions of the following groups and individuals to the development of my presentation:**

**Mr. Jale Ahmad**

**My utmost gratitude goes to my FYP supervisor, Mr. Jale Ahmad, who has been excellent in providing the information and guidance to me regarding the development of this project. His leadership, supervision and guidance have led to the success of this project.**

**Ms. Aliza Sarlan, FYP Committee, ICT/BIS lecturers and Mr.Adzfulredza**

**Many thanks and appreciations go to Ms. Aliza Sarlan for conducting the Final Year Project in UTP. In addition, thanks to the FYP committee for providing me direction and guidance throughout the entire project. To ICT / BIS lecturers, who have helped me in direct or indirect ways, thank you. The greatest appreciations to the evaluators that helped me in improving this project, recognizing both strengths and weakness. To Mr.Adzfulredza, thank you for giving me information I needed.**

**My parents; Ir.Muhamad Zulkapley Kadri and Ir. Zaharah Mohd Yusof**

**My deepest gratitude goes to them for their full encouragement and support,for their guidance and faith.**

**My sisters, Zanariyah and Zubaidah**

**Especially to my elder sister, Zanariyah, who has been there for me when I'm on the verge of nervous breakdown, who drove all the way from Kuala Lumpur to Tronoh at night just to be there for me and stayed at my side until I finish my project.**

**Mr. Ahmad Shah and Ms. Zanira Zahani**

Many thanks to Mr. Ahmad Shah for the never ending support. Ms. Zanira, thank you for taking the place of my sister, when I needed her.

**Mr. Noor Haizad Said**

Without your great assistance and patience, this project would not be a success

Last but not least, to any other person(s) not mentioned here, I would like to say " thank you " for your support and your direct or indirectly involvement until the completion of this project.

## TABLE OF CONTENTS

<b>TITLE</b> .....	<b>i</b>
<b>CERTIFICATION OF APPROVAL</b> .....	<b>ii</b>
<b>CERTIFICATION OF ORIGINALITY</b> .....	<b>iii</b>
<b>ABSTRACT</b> .....	<b>iv</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>v</b>
<b>TABLE OF CONTENTS</b> .....	<b>vii</b>
<b>LIST OF FIGURES</b> .....	<b>ix</b>
<b>LIST OF TABLES</b> .....	<b>x</b>
<b>ABBREVIATIONS</b> .....	<b>xi</b>
<b>CHAPTER 1: INTRODUCTION</b> .....	<b>1</b>
1.1 Background of Study.....	3
1.2 Problem Statement.....	5
1.3 Objectives of Study .....	5
1.4 Scope of Study.....	6
<b>CHAPTER 2: LITERATURE REVIEW AND THEORY</b> .....	<b>7</b>
2.1 Image Processing Walkthrough.....	7
2.2 Introduction to Image Processing .....	11
2.3 Digital Image Definitions. ....	12
2.3.1 Common Values .....	14
2.4 Data Segmentation.....	15
2.5 Segmentation Houghing Processor.....	16
<b>CHAPTER 3: METHODOLOGY/PROJECT WORK</b> .....	<b>17</b>
3.1 Methodology.....	17
3.2 Project Steps and Phase Identification .....	18

3.3 System Design .....	20
3.3.1 Creating the algorithm .....	20
3.3.1.1 Image Segmentation .....	21
3.3.1.2 Wavelet Analysis .....	22
3.3.1.2.1 Discrete Wavelet Transform.....	22
3.3.1.3 Clustering .....	23
3.4 Tools.....	24
3.4.1 Microsoft Visual Studio .NET 2005.....	24
3.4.2 Adobe Photoshop CS and Image Ready.....	24
3.4.3 MatLab7 .....	25
3.5 FiraImage.....	26
3.5.1 Functional Requirements.....	26
3.5.2 Non-Functional Requirements.....	26
3.5.3 Wavelet Algorithm Module.....	27
3.5.4 Clustering Module .....	28
3.5.5 Image Module.....	28
3.5.6 Filters Module .....	28
3.5.7 Texture Module .....	28
3.6 Process Flow For Prototype User Experience .....	29
3.7 Image Processing to Identify Image Genuinity .....	31
3.8 Development Plan .....	31
<b>CHAPTER 4: RESULTS AND DISCUSSION .....</b>	<b>32</b>
4.1 System Acceptance Test.....	32
4.1.1 Test execution confirmation checklist.....	32
4.2 Discussions .....	33
<b>CHAPTER 5: CONCLUSION AND RECOMENDATION.....</b>	<b>34</b>
5.1 Conclusion.....	34
5.2 Recommendation.....	35
<b>REFERENCES .....</b>	<b>36</b>
<b>APPENDICES.....</b>	<b>38</b>



## LIST OF FIGURES

Figure 1.1 : Hough Line Transformation .....	3
Figure 1.2 : Comparison of Transformation of Original to Forged Cash Notes .....	4
Figure 2.1 : Various Geometries for Simple Skeleton .....	9
Figure 2.2 : Hough Transform Result.....	10
Figure 2.3 : Digitization of continuous image .....	13
Figure 3.1 : Project Methodology .....	17
Figure 3.2 : Image Segmentation Performance .....	20
Figure 3.3 : Example of Image Segmentation .....	21
Figure 3.4 : Two- Dimensional discrete wavelet transform .....	22
Figure 3.5 : Simple Graphical Example .....	23
Figure 3.6 : System Architecture for Prototype.....	27
Figure 3.7 : FiraImage Basic Interface .....	29
Figure 3.8 : Image Transformation.....	30
Figure 3.9 : Menu based system .....	30
Figure 3.10 : Transformation .....	31
Figure 3.11 : Development Plan.....	31
Figure 4.1 : Result of Transformation for Original Cash Notes .....	33
Figure 4.2 : Result of Transformation for Forged Cash Notes .....	33

## **LIST OF TABLES**

Table 1: Common Value of Digital Image Parameters .....	14
Table 2: System Acceptance Test .....	32

## **ABBREVIATIONS**

<b>UTP</b>	<b>Universiti Teknologi PETRONAS</b>
<b>FYP</b>	<b>Final Year Project</b>
<b>ICT</b>	<b>Information and Communication Technology</b>
<b>BIS</b>	<b>Business Information System</b>

# **CHAPTER 1**

## **INTRODUCTION**

This project and research are mainly on getting the better understanding about digital image processing technique and how this technology can be implemented and developed in various ways. When we talked about image processing, not many know on how much image processing has been an important technology contributed to the human being i.e health, safety and better lifestyle.

Digital image processing allows the use of much more complex algorithms for image processing, and hence can offer both more sophisticated performance at simple tasks, and the implementation of methods which would be impossible by analog means. In particular, digital image processing is the only practical technology for:

- Classification
- Feature extraction
- Pattern recognition
- Projection
- Multi-scale signal analysis

From above mentioned usage, several applications have been developed to improve the human being way of life e.g.:

- Photography and printing
- Cash Notes Fraud Protection
- Satellite image processing
- Medical image processing
- Face detection, feature detection, face identification
- Microscope image processing
- Car barrier detection

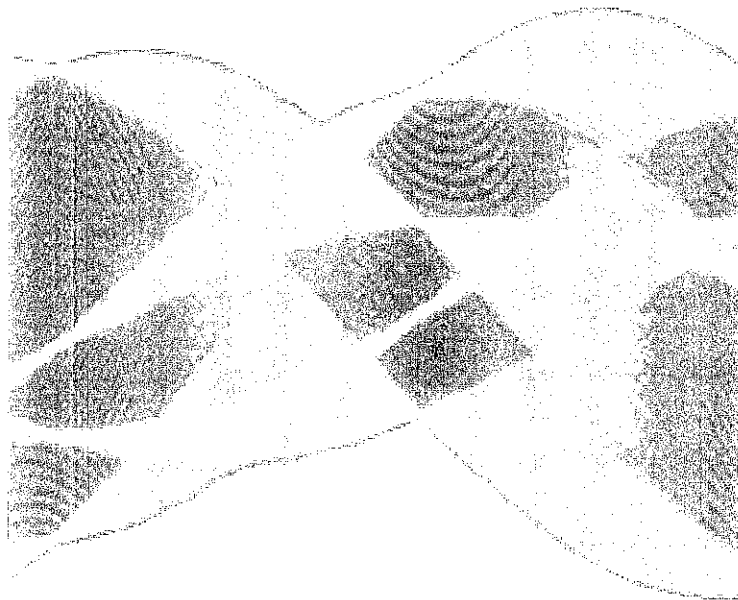
In the project, we will develop a prototype for identifying the number of segments/object nodes and use the result to verify the validity of certain product/object e.g. Cash Notes.

A set of algorithms will be used for performing the image segmentation, extraction and generate object unique pattern based on Lines Hough Transform.

## 1.1 Background of Study

Advancements in personal computer and document duplication technology are making it easier than ever for criminals to produce authentic-looking counterfeit checks, resulting in substantial check fraud losses. Vectra Bank Colorado's Positive Pay service is designed to cut check fraud losses with early detection and return of unauthorized checks prior to the expiration of Regulation CC (Expedited Funds Availability) deadlines.

Many banks are already implemented several technologies to offer authentic-looking counterfeit checks. Forged documents are often used to conceal other thefts; banks tend to count their money meticulously so every penny must be accounted for. A document claiming that a sum of money has been borrowed as a loan, withdrawn by an individual depositor or transferred or invested can therefore be valuable to a thief who wishes to conceal the minor detail that the bank's money has in fact been stolen and is now gone. A fast method to identify these forged documents, money or even cheques. Image clustering, segmentation and Hough line transformation.

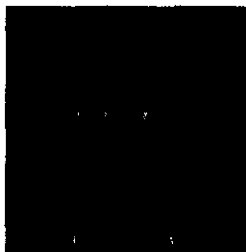


**Figure 1.1: Hough Line Transformation**

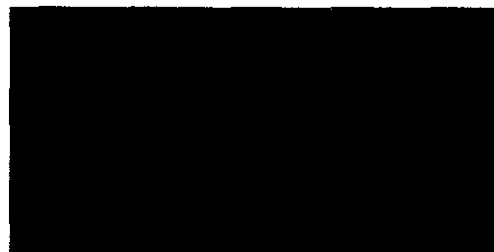
The Hough transform is a technique which can be used to isolate features of a particular shape within an image. Because it requires that the desired features be specified in some parametric form, the classical Hough transform is most commonly used for the detection of regular curves such as lines, circles, ellipses, etc. A *generalized* Hough transform can be employed in applications where a simple analytic description of a feature(s) is not possible. Due to the computational complexity of the generalized Hough algorithm, we restrict the main focus of this discussion to the classical Hough transform. Despite its domain restrictions, the classical Hough transform (hereafter referred to without the *classical* prefix) retains many applications; as most manufactured parts (and many anatomical parts investigated in medical imagery) contain feature boundaries which can be described by regular curves. The main advantage of the Hough transform technique is that it is tolerant of gaps in feature boundary descriptions and is relatively unaffected by image noise.



Real Note



Forged Note



**Figure 1.2: Comparison of Transformation of Original to Forged Cash Notes**

## **1.2 Problem Statement**

Current clustering techniques do not address all the requirements adequately and concurrently. In details, dealing with large number of dimensions and large number of data items can be problematic because of time complexity. Clustering results are crucially dependent on the distance (measure of similarity) “points” to be clustered. The effectiveness of the method depends on the definition of distance. If an obvious distance measure doesn't exist we must “define” (or create) it, which is not always easy, especially in multi-dimensional spaces. The result of the clustering algorithm (that in many cases can be arbitrary itself) can be interpreted in different ways

## **1.3 Objectives of Study**

The objectives of this project are:

- To develop an algorithm which can identify the number of cluster in data space
- To perform an image segmentation from an image with different texture
- To develop an algorithm to identify the number of cluster from a textured image
- To identify better interpretation of image segmentation result



#### **1.4 Scope of Study**

The scope of the study has been narrowed down to detect forgery on Cash Notes. Currently, the nation is having a major issue about “Black Money”. It has been a national’s mission to curb this problem for it can contribute to an economic downturn tremendously. It is hope that, with this research as a starting point, we could help the government to develop a technology to overcome or at least put a high level of difficulty on cash notes forgery, thus increase the country’s economy.

## **CHAPTER 2**

### **LITERATURE REVIEW AND THEORY**

#### **2.1 Image Processing Walkthrough**

A few decades ago, image processing was done largely in the analog domain, chiefly by optical devices. These optical methods are still essential to applications such as holography because they inherently parallel; however, due to significant increase in computer speed, these techniques are increasingly being replaced by digital image processing methods.

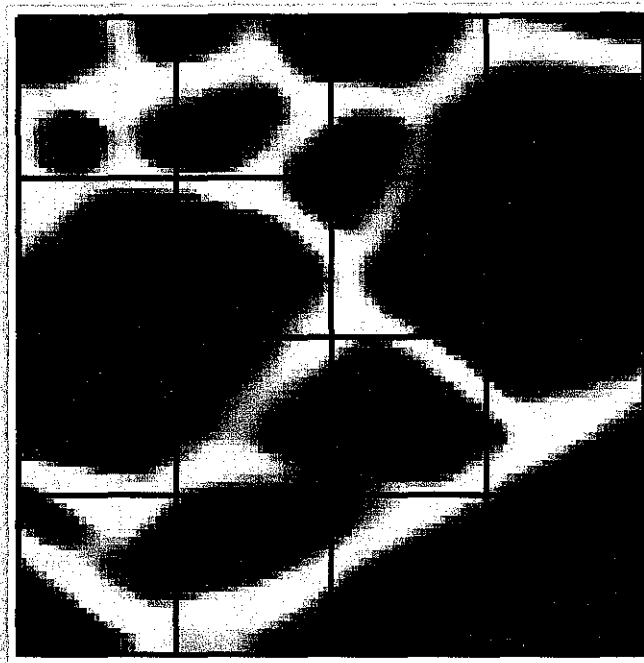
Digital image processing techniques are generally more versatile, reliable and accurate; they have the additional benefit of being easier to implement than their analog counterparts. Specialized hardware is still used for digital image processing: computer architectures based on pipelining have been the most commercially successful. There are also many massively parallel architectures that have been developed for the purpose. Today, hardware solutions are commonly used in video processing systems. However, commercial image processing tasks are more commonly done by software running on conventional personal computers.

Many of the techniques of digital image processing, or digital picture processing as it was often called, were developed in the 1960s at the Jet Propulsion Laboratory, MIT, Bell Labs, University of Maryland, and a few other places, with application to satellite imagery, wirephoto standards conversion, medical imaging, videophone, character recognition, and photo enhancement. But the cost of processing was fairly high with the computing equipment of that era. In the 1970s, digital image processing proliferated, when cheaper computers and dedicated hardware became available. Images could then be processed in real time, for some dedicated problems such as television standards

conversion. As general-purpose computers became faster, they started to take over the role of dedicated hardware for all but the most specialized and compute-intensive operations.

With the fast computers and signal processors available in the 2000s, digital image processing has become the most common form of image processing, and is generally used because it is not only the most versatile method, but also the cheapest.

Digital image processing allows the use of much more complex algorithms for image processing, and hence can offer both more sophisticated performance at simple tasks, and the implementation of methods which would be impossible by analog means. In this project, a few complex algorithms are used for example Hough transform, Simple Skeleton, and Edges.



**Figure 2.1: Various Geometries for Simple Skeleton**

Simple Skeleton creates a skeleton by converting the pixels of the Microstructure directly into elements. If `skeleton_geometry` is `QuadSkeleton`, then each pixel leads to one quadrilateral element. If `skeleton_geometry` is `TriSkeleton`, each pixel will be split into two triangular elements, with the direction of the split determined by the `TriSkeleton arrangement` parameter. The various geometries are illustrated in Figure 2.1.



**Figure 2.2: Hough Transform Result**

The Hough transform is a technique which can be used to isolate features of a particular shape within an image. Because it requires that the desired features be specified in some parametric form, the *classical* Hough transform is most commonly used for the detection of regular curves such as lines, circles, ellipses, *etc.* A *generalized* Hough transform can be employed in applications where a simple analytic description of a feature(s) is not possible. Due to the computational complexity of the generalized Hough algorithm, we restrict the main focus of this discussion to the classical Hough transform. Despite its domain restrictions, the classical Hough transform (hereafter referred to without the *classical* prefix) retains many applications; as most manufactured parts (and many anatomical parts investigated in medical imagery) contain feature boundaries which can be described by regular curves. The main advantage of the Hough transform technique is that it is tolerant of gaps in feature boundary descriptions and is relatively unaffected by image noise.

## 2.2 Introduction to Image Processing

Modern digital technology has made it possible to manipulate multi-dimensional signals with systems that range from simple digital circuits to advanced parallel computers. The goal of this manipulation can be divided into three categories:

\* Image Processing *image in -> image out*

\* Image Analysis *image in -> measurements out*

\* Image Understanding *image in -> high-level description out*

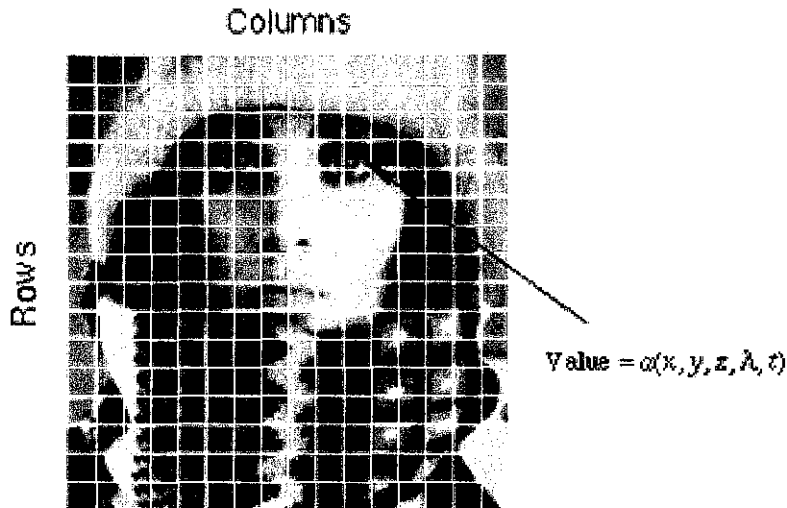
Let's begin with certain basic definitions. An image defined in the "real world" is considered to be a function of two real variables, for example,  $a(x,y)$  with  $a$  as the amplitude (e.g. brightness) of the image at the *real* coordinate position  $(x,y)$ . An image may be considered to contain sub-images sometimes referred to as *regions-of-interest*, *ROIs*, or simply *regions*. This concept reflects the fact that images frequently contain collections of objects each of which can be the basis for a region. In a sophisticated image processing system it should be possible to apply specific image processing operations to selected regions. Thus one part of an image (region) might be processed to suppress motion blur while another part might be processed to improve color rendition.

The amplitudes of a given image will almost always be either real numbers or integer numbers. The latter is usually a result of a quantization process that converts a continuous range (say, between 0 and 100%) to a discrete number of levels. In certain image-forming processes, however, the signal may involve photon counting which implies that the amplitude would be inherently quantized. In other image forming procedures, such as magnetic resonance imaging, the direct physical measurement yields a complex number in the form of a real magnitude and a real phase.

### 2.3 Digital Image Definitions

A digital image  $a[m,n]$  described in a 2D discrete space is derived from an analog image  $a(x,y)$  in a 2D continuous space through a *sampling* process that is frequently referred to as digitization. The mathematics of that sampling process will be described in Section 5. The effect of digitization is shown in Figure 1.

The 2D continuous image  $a(x,y)$  is divided into  $N$  rows and  $M$  columns. The intersection of a row and a column is termed a *pixel*. The value assigned to the integer coordinates  $[m,n]$  with  $\{m=0,1,2,\dots,M-1\}$  and  $\{n=0,1,2,\dots,N-1\}$  is  $a[m,n]$ . In fact, in most cases  $a(x,y)$  is actually a function of many variables including depth ( $z$ ), color ( $\lambda$ ), and time ( $t$ ).



**Figure 2.3: Digitization of a continuous image. The pixel at coordinates  $[m=10, n=3]$  has the integer brightness value 110.**

The image shown in Figure 2.3 has been divided into  $N = 16$  rows and  $M = 16$  columns. The value assigned to every pixel is the average brightness in the pixel rounded to the nearest integer value. The process of representing the amplitude of the 2D signal at a given coordinate as an integer value with  $L$  different gray levels is usually referred to as amplitude quantization or simply *quantization*.

In particular, digital image processing is the only practical technology for:

- Feature extraction
- Pattern recognition
- Multi-scale signal analysis



### 2.3.1 Common Values

There are standard values for the various parameters encountered in digital image processing. These values can be caused by video standards, by algorithmic requirements, or by the desire to keep digital circuitry simple. Table 1 gives some commonly encountered values.

<i>Parameter</i>	<i>Symbol</i>	<i>Typical values</i>
Rows	$N$	256,512,525,625,1024,1035
Columns	$M$	256,512,768,1024,1320
Gray Levels	$L$	2,64,256,1024,4096,16384

**Table 1:** Common values of digital image parameters

Quite frequently we see cases of  $M=N=2^K$  where  $\{K = 8,9,10\}$ . This can be motivated by digital circuitry or by the use of certain algorithms such as the (fast) Fourier transform

The number of distinct gray levels is usually a power of 2, that is,  $L=2^B$  where  $B$  is the number of bits in the binary representation of the brightness levels. When  $B>1$  we speak of a *gray-level image*; when  $B=1$  we speak of a *binary image*. In a binary image there are just two gray levels which can be referred to, for example, as "black" and "white" or "0" and "1".

## 2.4 Data Segmentation

A central problem, called *segmentation*, is to distinguish objects from background. For intensity images (ie, those represented by point-wise intensity levels) four popular approaches are: threshold techniques, edge-based methods, region-based techniques, and connectivity-preserving relaxation methods.

Threshold techniques, which make decisions based on local pixel information, are effective when the intensity levels of the objects fall squarely outside the range of levels in the background. Because spatial information is ignored, however, blurred region boundaries can create havoc.

Edge-based methods center around contour detection: their weakness in connecting together broken contour lines make them, too, prone to failure in the presence of blurring.

A region-based method usually proceeds as follows: the image is partitioned into connected regions by grouping neighboring pixels of similar intensity levels. Adjacent regions are then merged under some criterion involving perhaps homogeneity or sharpness of region boundaries. Over stringent criteria create fragmentation; lenient ones overlook blurred boundaries and over merge. Hybrid techniques using a mix of the methods above are also popular.

## 2.5 Segmentation Houghing Processor

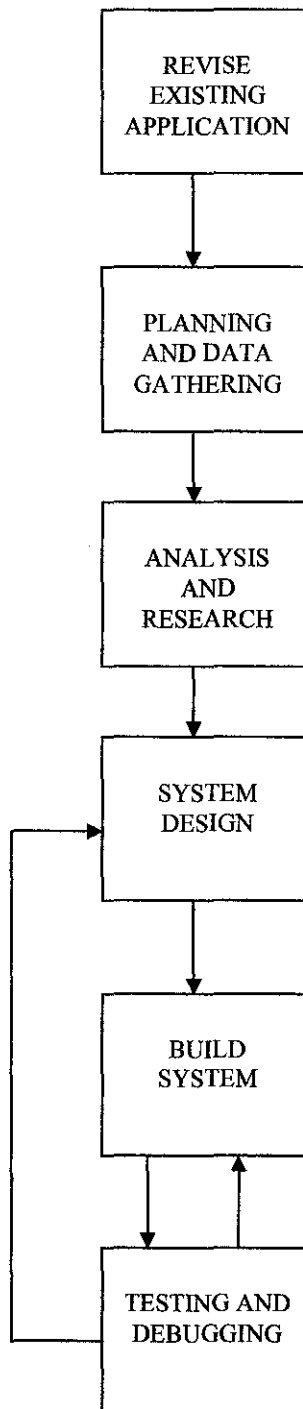
The Hough transform can be used to identify the parameter(s) of a curve which best fits a set of given edge points. This edge description is commonly obtained from a feature detecting operator such as the Roberts Cross, Sobel or Canny edge detector and may be noisy, *i.e.* it may contain multiple edge fragments corresponding to a single whole feature. Furthermore, as the output of an edge detector defines only *where* features are in an image, the work of the Hough transform is to determine both *what* the features are (*i.e.* to detect the feature(s) for which it has a parametric (or other) description) and *how many* of them exist in the image.

In this project, after segmentation, Houghing process will be adapted to identify the curve line for each object in the image. The result of this can be used to compare the original data with the result.

# CHAPTER 3

## METHODOLOGY / PROJECT WORK

### 3.1 Methodology



**Figure 3.1 Project Methodology**

## **3.2 Project Steps and Phase Identification**

As mention in the introduction part, in enhancing and developing this new algorithm, it will involve 6 main phases, which is basically referred to the normal system development life cycle, the 6 main phases are:

### **a) Revise existing application**

- In this phase, analyzing of the existing application is very significant to gather all the ideas needed. From the analysis, it is to acquire the information on the limitation of the system. Is the functionality of the application completely satisfied for the business need?

### **b) Planning**

- From the data gathered on the limitation of the previous application, new ideas should be formed on how to improve the limitation and to make sure the ideas should meet the latest users and business requirements.
- The importance of planning and collect reliable information is to gain as much knowledge and understanding the scope of image processing.

### **c) Analysis**

- Data gathering and fact finding by doing research, meetings, interviews and etc to collect information regarding the new system developments, problems statement, requirements and preferences.

**d) Design**

Plan and design the application on paper. Design application storyboard and user interface. Logical and physical design created such as data flow diagram and ERD diagram of the system.

**e) Build**

- Design the prototype by creating the functioning application and the design of the source code involve the application.

**f) Testing**

- Provide training materials and test the application in the real environment experimenting user several users and data. Test the application compatible with the requirements.

### 3.3 System Design

#### 3.3.1 Creating the algorithm

In developing this project i.e. to develop an algorithm to identify the number of cluster from a textured image, image segmentation needs to be performed. Several approaches proposed to be implemented to perform the image segmentation:

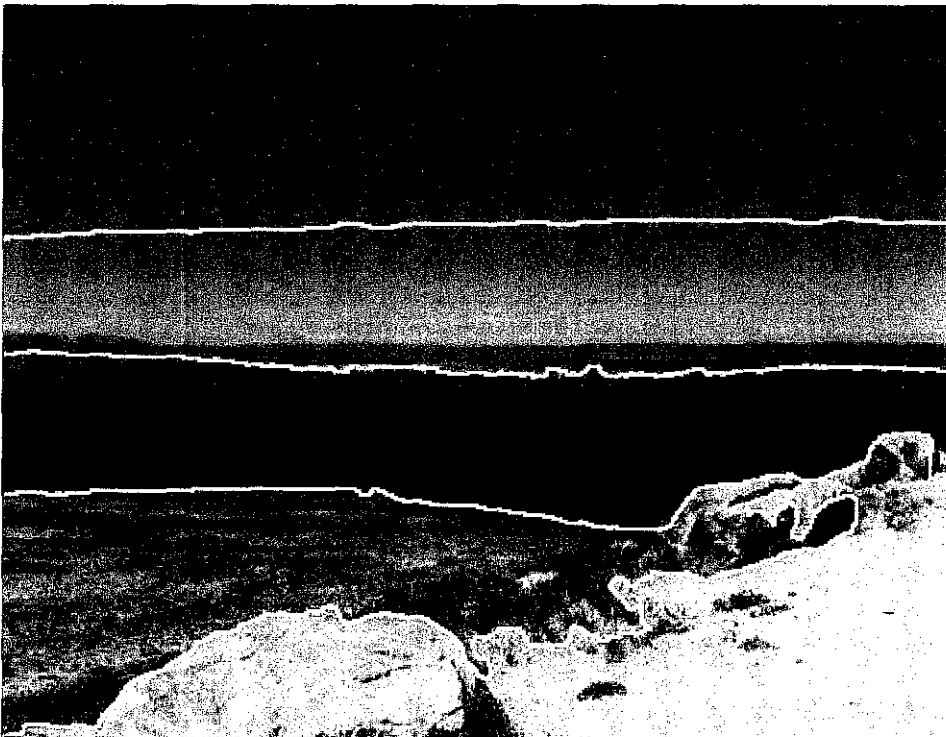
- Wavelet transform
- Clustering



**Figure 3.2: Image Segmentation Performance**

### 3.3.1.1 Image Segmentation

Image segmentation is defined as the partition of a digital image with texture into a set of non overlapping regions whose union is an entire image. It is meant to decompose the image into parts that are meaningful to particular application. The challenges in segmenting an image are when the image is made up of complex texture.



**Figure 3.3: Example of Image Segmentation**

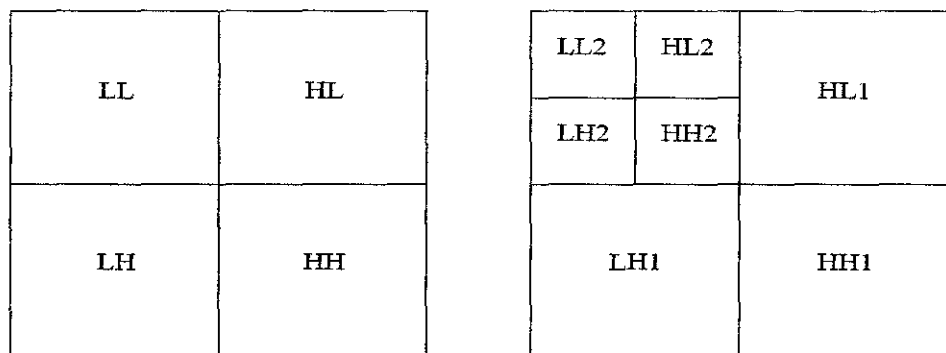


### 3.3.1.2 Wavelet Analysis

Wavelet analysis represents the next logical step: a windowing technique with variable-sized regions. Wavelet analysis allows the use of long time intervals where we want more precise low-frequency information, and shorter regions where we want high-frequency information. In this project, wavelet analysis is used to determine the optimal feature corresponding to each pixel or block of pixels in the image. A technique called, Discrete Wavelet Transform will be used in this project.

#### 3.3.1.2.1 Discrete Wavelet Transform

In discrete wavelet transform, each row of 2D images are filtered by a low-pass filter (Lx) and high-pass filter (Hx). The output is then down-sampled by factor of 2 to produce intermediate images L and H. This process can be continued until the desired number of decomposition levels is reached.



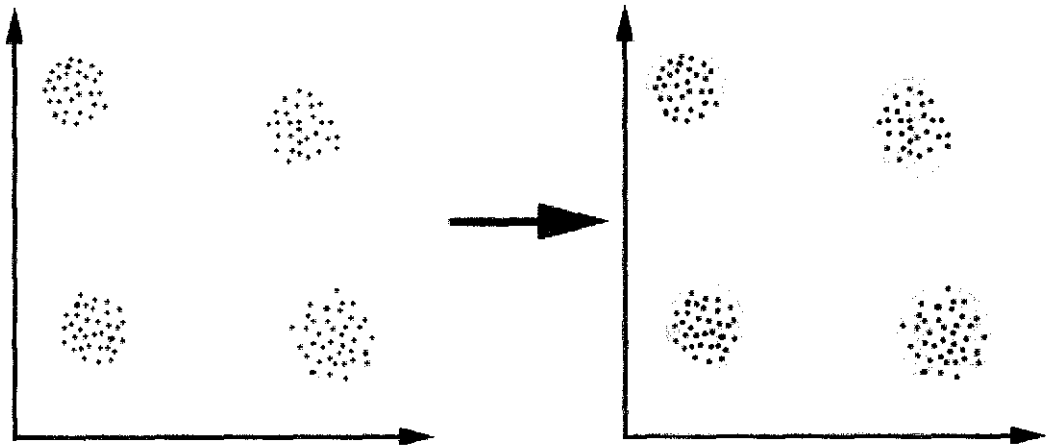
**Figure 3.4:**

**Two-dimensional discrete wavelet transform (one-level transform)**

**Two-dimensional discrete wavelet transform (two-level transform)**

### 3.3.1.3 Clustering

Clustering is the process of organizing objects into groups whose members are similar in some way



**Figure 3.5: Simple Graphical Example**

There are several types of clustering algorithms, which are:

- K-means - *exclusive clustering algorithm*
- Fuzzy C-means - *overlapping clustering algorithm*
- Hierarchical clustering
- Mixture of Gaussians - *probabilistic clustering algorithm*

For this project, K-means algorithms will be used. It's one of the simplest unsupervised learning algorithms that solve the well known clustering problem. It's used to classify a given data set through a certain number of clusters (assume  $k$  clusters) fixed *a priori*. The main idea is to define  $k$ -centroids, one for each cluster.

## **3.4 Tools**

To accomplish the task designing and developing this application, some applications and tools required. Different parts of the prototype need to author by using different tools such as the database, creating the algorithm, the coding to create function and to design the user interface.

### **3.4.1 Microsoft Visual Studio.NET 2005**

Microsoft Visual Studio.NET 2005 is Microsoft's flagship software development product for computer programmers. It centers on an integrated development environment which lets programmers create standalone applications, web sites, web applications, and web services that run on any platforms supported by Microsoft's .NET Framework.

### **3.4.2 Adobe Photoshop CS and Image Ready**

Adobe Photoshop CS and Image Ready is hands down, the most popular program for creating and modifying images. This image authoring tools is used for the design phase of the application. Adobe Photoshop will be used to design the interfaces of the system and used to editing some graphics or pictures which need to be touch to produce good image.

### **3.4.3 MatLab7**

This is the latest version of MatLab. MATLAB is a high-level technical computing language and interactive environment for algorithm development, data visualization, data analysis, and numeric computation.

Its key features are:

- High-level language for technical computing
- Development environment for managing code, files, and data
- Interactive tools for iterative exploration, design, and problem solving
- Mathematical functions for linear algebra, statistics, Fourier analysis, filtering, optimization, and numerical integration
- 2-D and 3-D graphics functions for visualizing data
- Functions for integrating MATLAB based algorithms with external applications and languages, such as C, C++, Fortran, Java, COM, and Microsoft Excel

## **3.5 FiraImage**

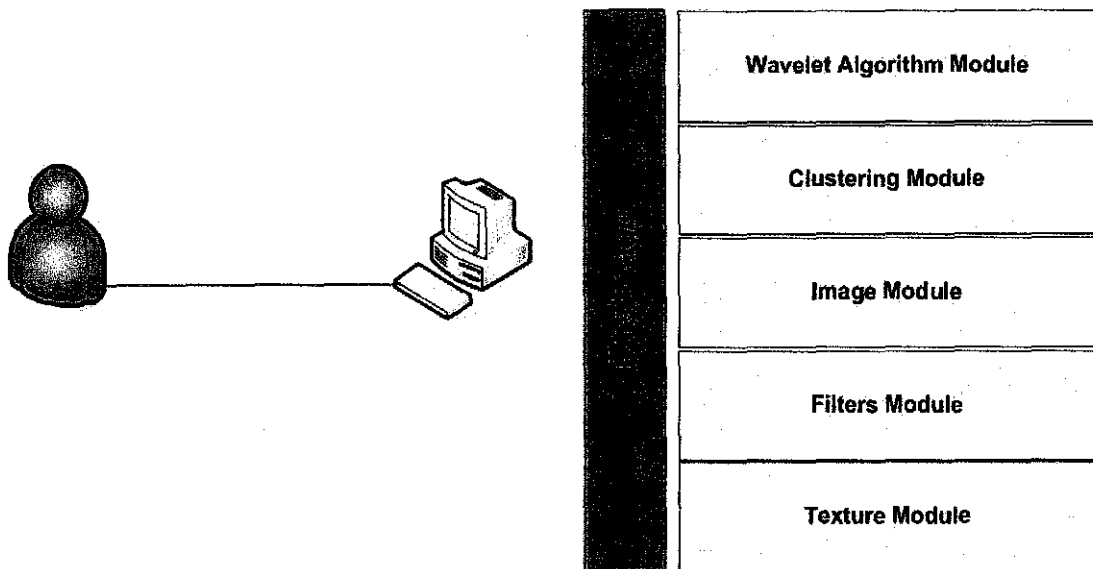
FiraImage System will provide the following functional requirements and non-functional requirements.

### **3.5.1 Functional Requirements**

- To open image file and display on the viewer
- To process image file and store segment info into database
- To list image files into a group of clusters

### **3.5.2 Non-Functional Requirements**

- Image size can be any size
- Image types can be JPG, BMP or GIF



**Figure 3.6: System Architecture for Prototype**

FiraImage system architecture consists of several main modules as shown in Figure 1. The modules are the basis for delivering the result of image clustering and image segmentation.

### **3.5.3 Wavelet Algorithm Module**

Wavelet Algorithm Module provides a windowing technique with variable-sized regions. Wavelet Algorithm Module allows the use of long time intervals where we want more precise low-frequency information, and shorter regions where we want high-frequency information. In this project, wavelet analysis is used to determine the optimal feature corresponding to each pixel or block of pixels in the image. A technique called, Discrete Wavelet Transform will be used in this project.

#### **3.5.4 Clustering Module**

Clustering is the process of organizing objects into groups whose members are similar in some way.

For this project, K-means algorithms will be used. It's one of the simplest unsupervised learning algorithms that solve the well known clustering problem. It's used to classify a given data set through a certain number of clusters (assume k clusters) fixed *a priori*. The main idea is to define k centroids, one for each cluster.

#### **3.5.5 Image Module**

This module handles the management of images including opening image, editing image, and saving images. The image module is also catering the multiple image formats such as JPEG, GIF, BMP etc.

#### **3.5.6 Filters Module**

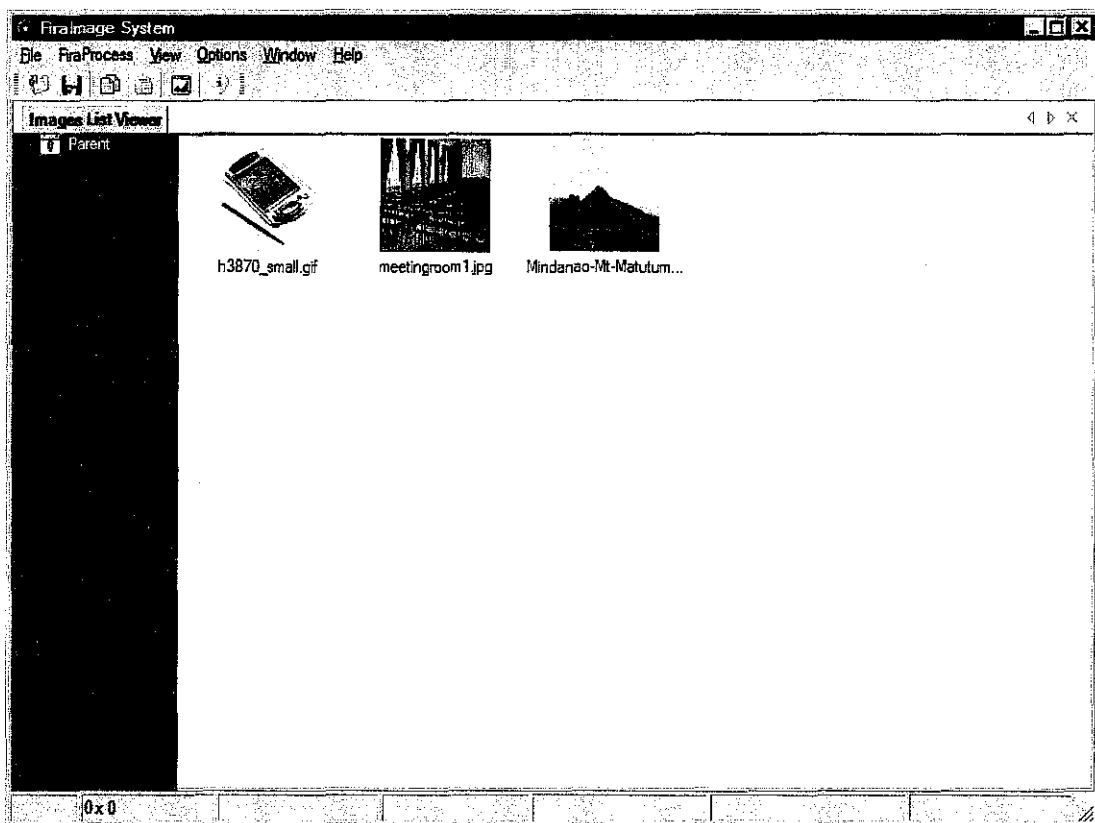
This module manages all the filtering process such as Edge Detector, Convolution and Correlation, HSL Filters, Gaussian Blur etc.

#### **3.5.7 Texture Module**

This module caters for texture reformatting such as Clouds Texture, Labyrinth Texture, Marble Texture, Textile Texture and Wood Texture.

### 3.6 Process Flow For Prototype User Experience

FiraImage System also considers user experience as one of the key points of good application and user-friendly system. A window based application is the suitable method to generate the image clustering and segmentation, as it is the most common way for non-IT literate users. (See Figure 3.7).

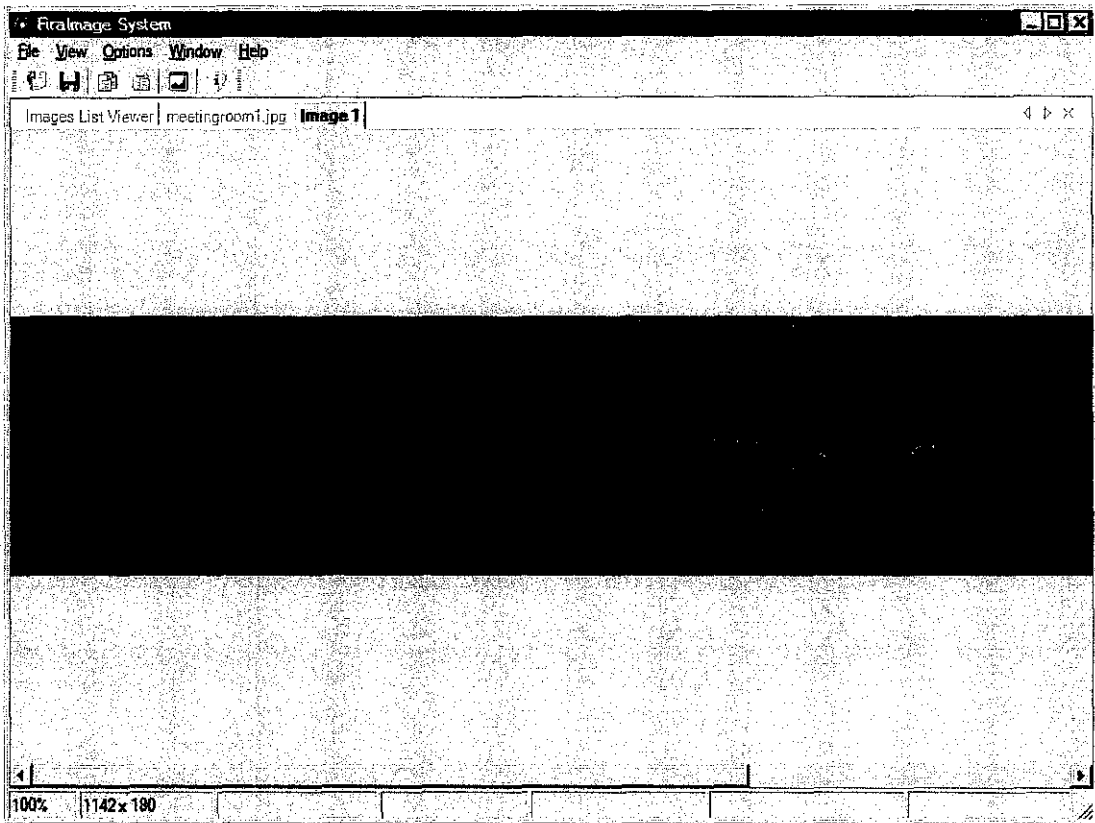


**Figure 3.7: FiraImage Basic Interface**

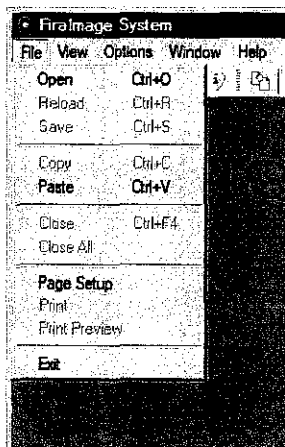
FiraImage System will provide a menu to allow the user to do a few functions:-

- a) Open Image to be filtered, segment, clustered and save
- b) To filter image to allow better segmented operation
- c) To segment image
- d) To cluster image
- e) To save image and cluster data





**Figure 3.8: Image Transformation**



**Figure 3.9: Menu based system**

### 3.7 Image Processing To Identify Image Genuinity

FiraImage will open any image and perform several images processing to allow the identification of image genuinity. The processes are:

```
AForge.Imaging.Filters.Threshold();
AForge.Imaging.Filters.Edges();
AForge.Imaging.Filters.SimpleSkeletonization();
AForge.Imaging.LineHoughTransform();
```

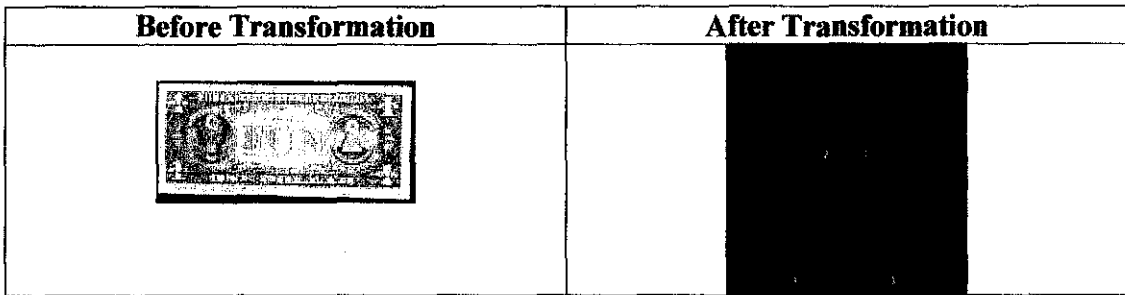


Figure 3.10: Transformation

### 3.8 Development Plan

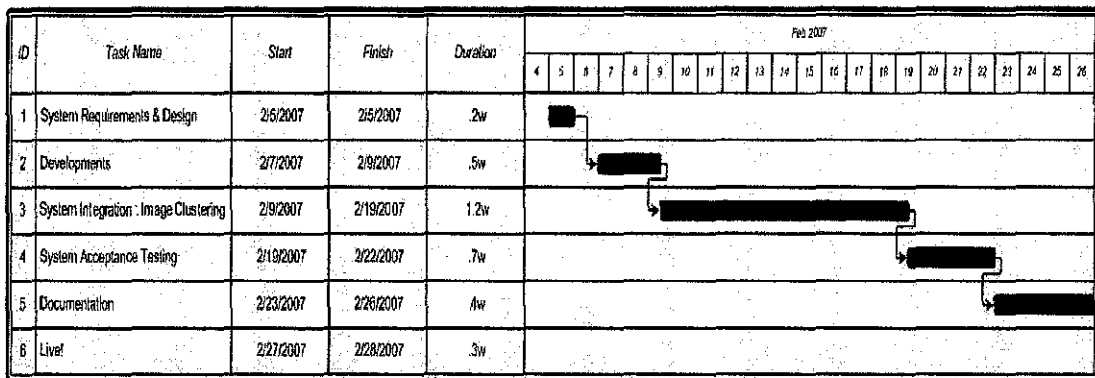


Figure 3.11 : Development Plan

## CHAPTER 4

### RESULTS AND DISCUSSION

#### 4.1 System Acceptance Test

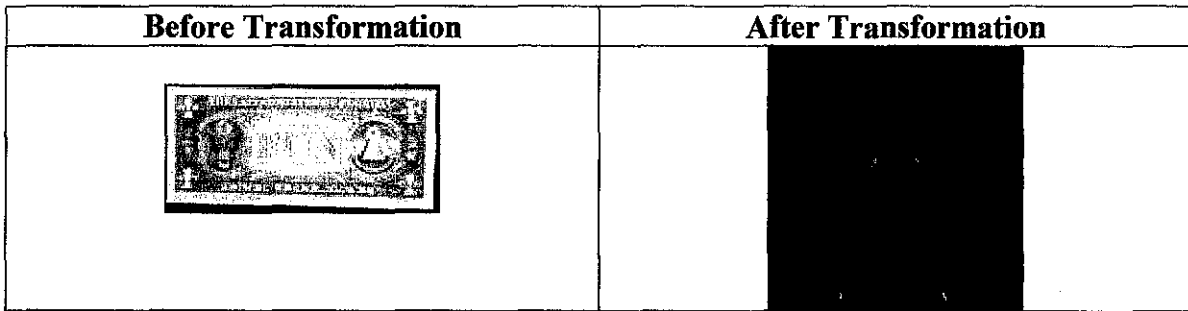
During this stage, we will provide build and test document on a high level could include:

##### 4.1.1 Test execution confirmation checklist

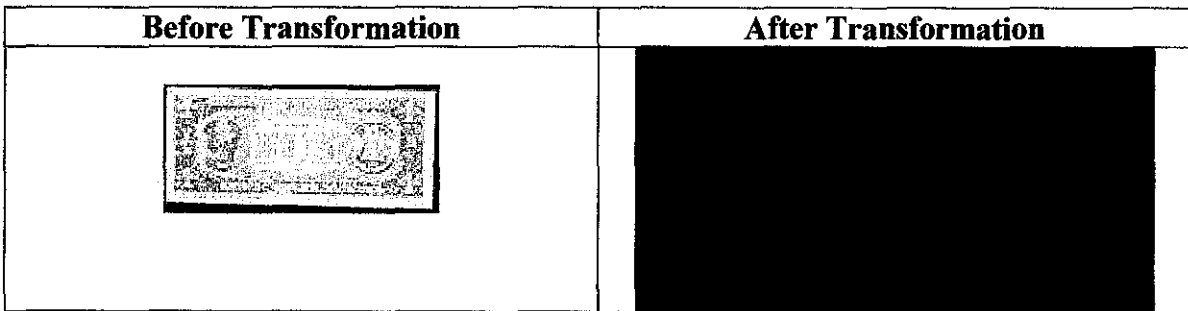
*Outlines the features to be tested and how it fits in the test sequence. It also acts as a summary of test review. Below is basic functional script to be tested.*

<b><i>Functional</i></b>	<b><i>Testing</i></b>	<b><i>Result</i></b>
Open Image	Success opening image	Image displayed.
Filter Image	Filter image from other files.	Only images file displayed.
Process Image	To do a list of transformation to a single image	The image is successfully displayed as shown as in Figure 8.
Save Image	To save image for image comparison.	The image is successfully saved in the storage.

**Table 2: System Acceptance Test**



**Figure 4.1: Result of Transformation for Original Cash Notes**



**Figure 4.2: Result of Transformation for Forged Cash Notes**

## 4.2 Discussion

From the observed result, the two differences between first transformation (Figure 4.1) and the second transformation (Figure 4.2) is the location of the line and the number of strip lines between original cash notes. In reality, the cash notes counterfeit is really difficult to trace. Since the validation of cash notes need robust algorithm and better process, our result can help to improve the current counterfeit measurement. The result is just a starting point of better anti-counterfeit solution.

## **CHAPTER 5**

### **CONCLUSION AND RECOMMENDATION**

#### **5.1 Conclusion**

This study will introduced a new algorithm in processing image as an improvement of current technique. It presented a thorough quantitative and quantitative evaluation of the application. This section will conclude the study and provide recommendation for future work.

It is hope that this new algorithm development, will fulfill all the adequate requirements that are not addressed in current image processing technique, which are, dealing with different types of attributes, discovering clusters with arbitrary shape, minimal requirements for domain knowledge to determine input parameters, ability to deal with noise and outliers, insensitivity to order of input records, high dimensionality and interpretability and usability.

## **5.2 Recommendation**

This project as a large scope and prospects. But due to time constraints, this study has been carried out within a very limited scope. For future study, several recommendations are proposed:

1. To enhance the algorithm to quantify the number of edges and further record the value
2. To use better digital camera for grabbing better resolution of the image
3. To provide more powerful PC processor from processing the image to validation of image genuinity

## REFERENCES

Spectral methods for image clustering, Marco Barreno, Computer Science Division, U.C. Berkeley, May 2004

A Fast Algorithm For Image Deblurring, DL Donoho, October 2004

An Algorithm for Image Clustering and Compression, Metin Kaya, 2005

Unsupervised Image-Set Clustering Using an Information Theoretic Framework, Jacob Goldberger, February 2006

**D. Ballard and C. Brown** *Computer Vision*, Prentice-Hall, 1982, Chap. 4.

**R. Boyle and R. Thomas** *Computer Vision: A First Course*, Blackwell Scientific Publications, 1988, Chap. 5.

**A. Jain** *Fundamentals of Digital Image Processing*, Prentice-Hall, 1989, Chap. 9. **D. Vernon** *Machine Vision*, Prentice-Hall, 1991, Chap. 6

M.F.A.Fauzi, "Texture Automated Segmentation", Content-Based Image Retrieval of Museum Images, August 2004

F. Smeraldi, J. Bigun, W. Gerstner, On the role of Dimensionality in Face Authentication, Proceedings of the 2000 Symposium of the Swedish Society for Automated Image Analysis, 2000

Wikipedia,Data Clustering [http://en.wikipedia.org/wiki/Data\\_Clustering](http://en.wikipedia.org/wiki/Data_Clustering)

Y. Deng, B. S.Manjunath and H.Shin "Color image segmentation", Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR '99, Fort Collins, CO, vol.2, pp.446-51, June 1999

Robert Porter and Nishan Canagarajah,"A Robust Automatic Clustering Scheme For Image Segmentation Using Wavelets,IEEE Transaction on Image Processing, Vol 5, April 1996

Ze-Nian Li and Mark S. Drew, "Fundamentals of Multimedia".

Rafael C. Gonzales, Richard E. Woods, Steven L. Eddins, "Digital Image Processing using MatLab.



---

## **APPENDICES**

### **Data for Experiments**

---

# Data for Experiment I

The screenshot shows the MATLAB workspace with the following variables and their properties:

Name	Size	Bytes	Class
CAN72	32x32	8192	double array
CAD	64x64	32768	double array
CAD2	64x64	32768	double array
CAE	64x64	32768	double array
CAE2	64x64	32768	double array
CAV	64x64	32768	double array
CAV2	64x64	32768	double array
CD	128x128	131072	double array
CD2	128x128	131072	double array
CH	128x128	131072	double array
CH2	128x128	131072	double array
C7	128x128	131072	double array
C72	128x128	131072	double array
I	256x256	65536	uint8 array
i	1x1	8	double array
Ixy	32x32x4	32768	double array
Ixy2	64x64x5	163840	double array
Ixy3	128x128x5	655360	double array
J	1x1	8	double array
cluster	1x1	8	double array
segmentedImage	32x32	8192	double array
segmentedImage2	64x64	32768	double array
segmentedImage3	256x256	524288	double array
selected_peaks	5x2	80	double array

## Data for Experiment II

The screenshot shows the MATLAB workspace window. The title bar reads 'MATLAB'. The menu bar includes 'File', 'Edit', 'View', 'Help', 'Workspace', and 'Help'. The current directory is 'C:\MATLAB95\bin\matlab code'. The workspace contains the following variables:

Name	Size	Bytes	Class
CAAF2	32x32	8192	double array
CAB	64x64	32768	double array
CAB2	64x64	32768	double array
CAR	64x64	32768	double array
CAR2	64x64	32768	double array
CAV	64x64	32768	double array
CAV2	64x64	32768	double array
CD	128x128	131072	double array
CD2	128x128	131072	double array
CH	128x128	131072	double array
CH2	128x128	131072	double array
CV	128x128	131072	double array
CV2	128x128	131072	double array
I	256x256	65536	uint8 array
i	1x1	8	double array
img	32x32x4	32768	double array
img2	64x64x6	196608	double array
img3	128x128x8	1048576	double array
j	1x1	8	double array
cluster	1x1	8	double array
segmentedImage	32x32	8192	double array
segmentedImage2	64x64	32768	double array
segmentedImage3	256x256	524288	double array
selected_peaks	6x5	320	double array

The workspace window also shows 'Workspace' and 'Current Directory' tabs. The Windows taskbar at the bottom shows the Start button and several open applications including MATLAB, a file explorer, and a command prompt.