INFORMATION CLUSTERING SYSTEM OF COMMERCIALLY IMPORTANT FISHES OCCURRING IN THAILAND

JANKET PENGPANIT

A THEMATIC PAPER SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE (TECHNOLOGY OF INFORMATION SYSTEM MANAGEMENT) FACULTY OF GRADUATE STUDIES MAHIDOL UNIVERSITY 2012

Thematic Paper
entitled

INFORMATION CLUSTERING SYSTEM OF COMMERCIALLY IMPORTANT FISHES OCCURRING IN THAILAND

…….
Ms. Janket Pengpanit
Candidate

…….
Lect. Supaporn Kiattisin,
Ph.D. (Electrical and Computer Engineering)
Major advisor

…….
Lect. Waranyu Wongseree,
Ph.D. (Electrical Engineering)
Co-advisor

…….
Asst. Prof. Auemphorn Mutchimwong, Ph.D.
Acting Dean
Faculty of Graduate Studies
Mahidol University

…….
Lect. Supaporn Kiattisin, Ph.D.
Program Director
Master of Science Program in Technology of Information System Management
Faculty of Engineering
Mahidol University

Thematic Paper
entitled

INFORMATION CLUSTERING SYSTEM OF COMMERCIALLY IMPORTANT FISHES OCCURRING IN THAILAND

was submitted to the Faculty of Graduate Studies, Mahidol University
for the degree of Master of Science (Technology of Information System Management)
on
May 10, 2012

…….
Ms. Janket Pengpanit
Candidate

…….
Asst. Prof. Adisorn Leelasantitham, Ph.D.
Chair

…….
Lect. Supaporn Kiattisin, Ph.D.
Member

…….
Lect. Piyasawat Navaratana, Ph.D.
Member

…….
Lect. Waranyu Wongseree, Ph.D.
Member

…….
Asst. Prof. Auemphorn Mutchimwong, Ph.D.
Acting Dean
Faculty of Graduate Studies
Mahidol University

…….
Lect. Worawit Israngkul,
M.S. (Technical Management)
Dean
Faculty of Engineering
Mahidol University

ACKNOWLEDGEMENTS

This thesis would not be success without advices and encouragements from many people. I would like to express my gratitude to my gratitude to my advisor, Dr. Supaporn Kiattisin, and co-advisor, Dr. Waranyu Wongseree. They was suggested me on a research of this topic and necessary knowledge on topic of contractual model. They was gave me a good advices and taught me about how to thinking process.

The indispensable group was my family, my elder sister and my best friends who always stayed beside and encourage me when I was under pressure and discouraged. Even they do not know deeply in this topic, but they tend to support me as much as possible. Finally, I appreciated my boss and colleagues who encouraged me to succeed this thesis

Janket Pengpanit

INFORMATION CLUSTERING SYSTEM OF COMMERCIALLY IMPORTANT FISHES OCCURRING IN THAILAND

JUNKET PENGPANIT 5036638 EGTI/M

M.Sc. (TECHNOLOGY OF INFORMATION SYSTEM MANAGEMENT)

THEMATIC PAPER ADVISORY COMMITTEE : SUPAPORN KIATTISIN, Ph.D., WARAUYU WONGSEREE, Ph.D.

ABSTRACT

The purpose of this study was to find matches for clustering the commercial fish picture in Thailand. Image processing and data mining were applied to the commercial fish database online. MATLAB was one of the tools for transforming the packet. It analyzed grouping and relationships between same and different fish shapes. Clustering used two methods: Hierarchical Clustering Method and K-Mean Clustering Method. Hierarchical Clustering Method showed a tree chart, a Dendogram, which presented the relationship of fish shapes having similar characteristics by cluster distance. K-Mean Clustering also presented similar clustering extract data for grouping. It divided fish data between same and different groups. The results of the analysis showed that the K-Mean Clustering Method was better than the Hierarchical Clustering Method. It was suitable for mass data and same-group cases even if fish shape data were similar, but it could accurately group fish type.

KEY WORDS :COMMERIAL FISHS / CLUSTERING / IMAGE PROCESSING / DENDOGRAM / MATLAB

62 pages

ระบบการจัดการกลุ่มข้อมูลปลาที่มีความสำคัญในเชิงพาณิชย์ของประเทศไทยด้วยวิธีคลัสเตอร์ริ่ง
INFORMATION CLUSTERING SYSTEM OF COMMERCIALLY IMPORTANT FISHES OCCURRING IN THAILAND

จันทร์เกตุ เพ็งพานิช 5036638 EGTI/M

วท.ม. (เทคโนโลยีการจัดการระบบสารสนเทศ)

คณะกรรมการที่ปรึกษาวิทยานิพนธ์: สุภาพร เกียรติสิน, Ph.D., วรัญญู วงษ์เสรี, Ph.D.

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาข้อมูลจากรูปภาพของปลาเศรษฐกิจในประเทศไทย โดยใช้เทคนิคด้านการประมวลผลภาพ(Image processing) และทำเหมืองข้อมูลมาประยุกต์ใช้กับข้อมูลเกี่ยวกับปลาเศรษฐกิจจากฐานข้อมูลปลาออนไลน์ โดยใช้โปรแกรม MATLAB เป็นเครื่องมือในการแปลงรูปภาพให้เป็นกลุ่มข้อมูล และทำการวิเคราะห์เพื่อจัดกลุ่มและแสดงความสัมพันธ์ของข้อมูลรูปร่างปลาที่มีความคล้าย และที่มีความแตกต่างกันได้ โดยแบ่งแบบคลัสเตอริ่ง(Clustering) ออกมาสองวิธี คือ Hierarchical Clustering Method ซึ่งจะแสดงผลในรูปของแผนภูมิต้นไม้ที่ถูกเรียกว่า “Dendogram” ที่จะแสดงความสัมพันธ์ของข้อมูลรูปร่างปลาที่มีลักษณะใกล้เคียงกันจากการวัดระยะห่างของแต่ละคลัสเตอร์(Cluster) และ K-Mean Clustering Method ซึ่งจะแสดงผลการจัดกลุ่มแบบแยกว่าข้อมูลนี้อยู่ในกลุ่มใดบ้าง ทำให้สามารถแยกได้ว่า ข้อมูลรูปร่างปลาแบบนี้มีลักษณะใกล้เคียงกับข้อมูลที่อยู่ในกลุ่มเดียวกัน และมีลักษณะแตกต่างกันหากอยู่ต่างกลุ่ม จากการศึกษาพบว่าหากมีข้อมูลปริมาณมากและมีความคล้ายคลึงกันสูงการแบ่งกลุ่มด้วยวิธี K-Mean Clustering Method จะมีความเหมาะสมกว่าวิธี Hierarchical Clustering Method และแม้ว่าข้อมูลรูปร่างปลาจะมีความคล้ายคลึงกันแต่ยังสามารถนำมาจัดกลุ่มได้

62 หน้า

CONTENTS

CONTENTS (cont.)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (cont.)

CHAPTER I
INTRODUCTION

1.1 Background

Fish have been classified to be the important protein source long ago. In the present, we use the fish's benefits for consumption which is main activities of fisheries. The beautiful fish industrial have been breed for export. Those fish are the important which the fish are ecology and conservative system, also the good natural balancing by observed from the several of fish biological lean to the section plan for prevent the extinguish of fish in time.

The ability of similarity grouping in the important parts to show the intelligence that it is easily for human studied. When they were young, they will be firstly adapted in growing up. So, it is quite hard work, if let computer do the same processing of human being's learn.

For this research, the researcher will use the classification data grouping analyzing in many fields such as artificial intelligence, squeezing compress the data, data mining, image processing, knowledge of equipments, pattern recognition, biological, statistical, customer relationship managements, marketing, medicals, and psychology.

In biological the clustering will be used to grouping all life to be the species or helping in clustering similar gene which useable in biological.

In fisheries have been try to the separation fish from outside character for identify group the commercial fish. The clustering is one idea will be used to help grouping fish.

Many researches were developed tools for supporting in the taxonomy and precisely. Therefore, the researcher chooses two logics: Hierarchical Clustering and K-Mean Analysis, for searching the length by Euclidean method. And, the researcher analyses in the fish shape data.

This research will analyze for searching the result of both grouping. And, then it also is suitable for grouping by fish shape.

1.2 Objectives of the Study

1.2.1. To design and analysis the fish image with K-Mean Clustering and Hierarchical Clustering.

1.2.2. To comparison the result of K-Mean Clustering and Hierarchical Clustering.

1.3 Scope of Study

1.3.1. Requirement clustering of image data 100 image samples.

1.3.2. Study the picture on the side view of the fish image.

1.3.3. Study of economic fish in Thailand.

1.3.4. The study must cover

1.3.4.1 The family list of commercial fish in Thailand

1.3.4.2 The picture is Jpeg or Gif type

1.4 Expected Results

1.4.1. The ability grouping fish shape in both K-Mean Clustering and Hierarchical Clustering.

1.4.2. The ability to compare that which way is proper for grouping fish shape between K-Mean Clustering and Hierarchical Clustering.

CHAPTER II

LITERATURE REVIEW

The chapter research methodology represents base knowledge for analyzing: classify characteristic fish part, and technical reviews

2.1 Characteristic Fish Classifications

The fish have spread widely in the variety environment. Thus, evaluation of fish related with shape because it is important for fish living in environment differently. Fish shape characteristics is classified. As follow:

1. *Fusiform shape* is similar as Torpedo-shape. In the front view is circle or slightly oval characteristics. In front of fish is thick and tapering to tail. This fish shape group generally that they swim quickly, like living surface water or in middle of water area (i.e. mackerel, long tail tuna, Spanish mackerel, and shark).

2. *Arrow shape* is similar as spindle but long body such as gar.

3. *Globiform shape* is similar as ball and slowly swims such as puffer fish which kind of these fish are prickle for protecting themselves.

4. Angguiliform, Snake-shape, Serpentile-shape is close to rarely slender shape as snake. The cross section view is circle and flat down to them tails. Swimming movement mostly uses body meat such as general eel.

5. Filifrom shape is the slender characteristics and small as yarn. The example of these fish group: snipe eel

6. Trachipterform shape is the long and flat as ribbon. These fish do not clever for swimming, and trunk muscles and fin for moment such as cutlass fish, wolf herring

7. Cylindrical shape in front view is circle or quit round.

8. Compressed shape - this shape is the flat slide characteristics in front views such as Nile tilapia, silver barb

9. Depressed shape - the characteristic is flattened. If you see in these fish cross view look shape toward the flat bottom such as ray

Moreover, some kind of fish is different characteristics such as seahorses, pipe fish, and butterfly fish.

2.2 Technical Reviews

Clustering Analysis is technical for classify or identify each other group (e.g. human, animal, thing etc.). It is divided into 2 group above up. Same case group is similar characteristics and other case group is different characteristics ("Classify variable groups by Cluster Analysis by Sompode,"2010).

2.2.1 Cluster Analysis Techniques

Cluster Analysis techniques is consisted two techniques: Hierarchical Cluster Analysis and K-Means Cluster Analysis. First, Hierarchical Cluster Analysis was used to classify in same similar group. Second, K-Means Cluster Analysis was techniques for classify with sub-group. Almost, it was used with many cases by the amount of group defined what cluster wants. For example, identify as K group, K-Mean works iteration repeatedly by each time which it build be combination of case until stay in a group. The process will choose case that distance from mean value of the fewer groups. And, it is calculated means again. It repeats do again until means are not being changed or re cycle completely.

Kind of variables were used in K-Mean Clustering. These variables must be quantitative or interval scale or ration scale. And, they cannot use with data which it is frequency value or binary as Hierarchical technique.

It is showed the results as hierarchy of number level that they are participate together. It is called "Dendrogram" in tree structure each node represents a cluster.

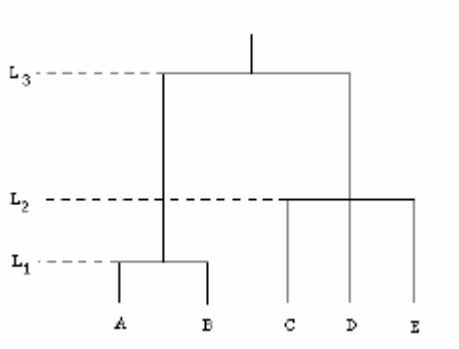


Figure 2.1 Object group in K-Means Clustering

In above figure 2.1 shows object group in K-Means clustering level of the objects. The group of object {A, B, C, D, E} Cluster in level (L1) consisted {A, B}, {C}, {D}, {E}. Then, in level 2 (L2) also concluded {A, B} and {C, D, E}. And finally, in level 3 (L3) is {A, B, C, D, E} which each level can defines set of class. Thus, it cans be moved lower level to be higher level.

2.3 K-means Clustering

K-means basic step is simply process. The first step, determine number of cluster K and be assume the centroid or center of these clusters. It cans take any random objects as the initial centroids or the first K objects in sequence can also serve as the initial centroids.

Then the K means algorithm will do the three steps below until convergence

Iterate until *stable* (= no object move group) shows below in figure as follows:

1. Determining the centroid coordinate
2. Determining the distance of each object to the centroids
3. Grouping the object based on minimum distance

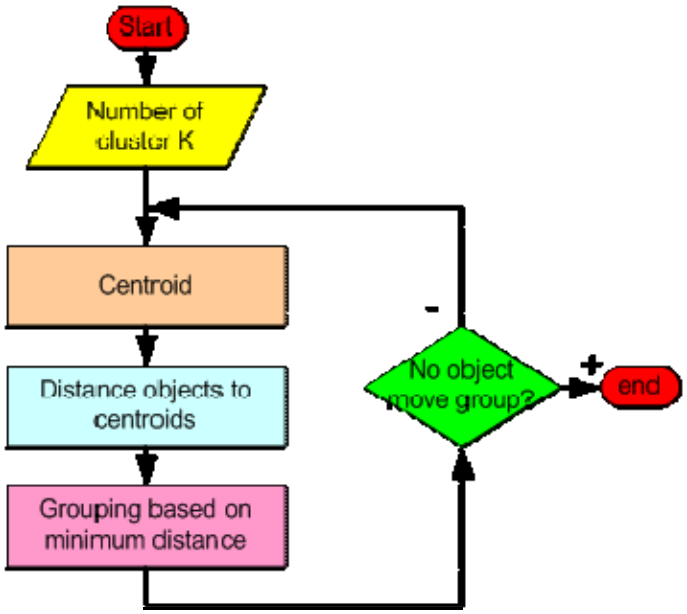


Figure 2.2 K-Means Clustering process

K-means clustering analysis in Figure 2.3, it represents in process of partitions analyzing. First, It was defined k = 2, Blue as group 1, Dark blue as group 2 and Red as Centroid. Then, determining the centroid coordinate and partitioned into two groups as Blue (Group 1) and Dark blue (group 2). Next, determining the distance of each object to the centroids in figure will see Red points as centroids. And, it cans group based on minimum distance. Each record data will be comparing with centroid for assign by using distance for finding what record data is the nearest. Finally, looping until record clustering not is changed or maximum loop fully.

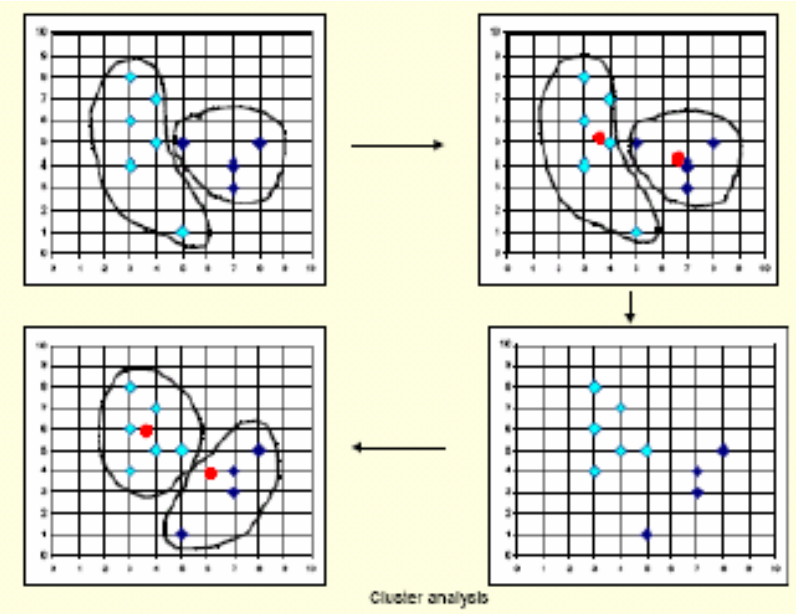


Figure 2.3 K-mean process sampling

Presume if you have several objects (4 types of medicines) and each object have two attributes/features as shown in Table 2.1 as below. Our goal is to group these objects into K=2 group of medicine based on the two features (pH and weight index).

Table 2.1 Medicine group of objects sample

Object	**attribute 1 (X): weight index**	**attribute 2 (Y): pH**
Medicine A	1	1
Medicine B	2	1
Medicine C	4	3
Medicine D	5	4

Each medicine shows one point with two attributes (X, Y) that we can represent it as coordinate in an attribute space as shown in the figure below.

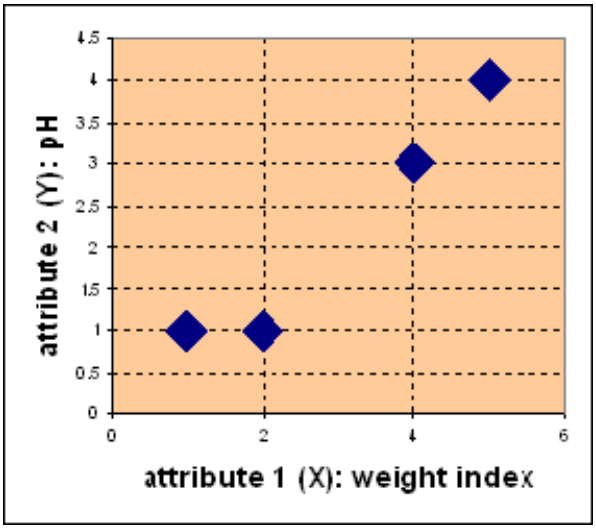


Figure 2.4 Coordinating in an attribute space

1. *Initial value of centroids* : Suppose we use medicine A and medicine B as the first centroids. Let c_1 and c_2 denote the coordinate of the centroids, then $c_1 = (1, 1)$ and $c_2 = (2, 1)$

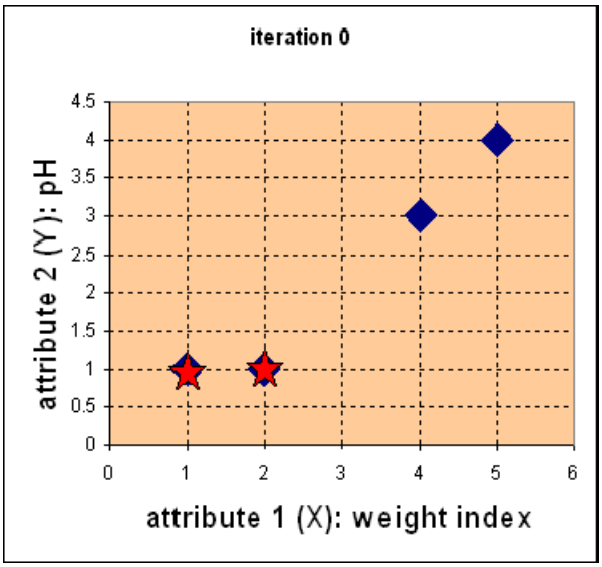


Figure 2.5 Initial value of centroids

2. *Objects-Centroids distance:* we calculate the distance between cluster centroid to each object. Let us use Euclidean distance, then we have distance matrix at iteration 0 is

$$D^0 = \begin{bmatrix} 0 & 1 & 3.61 & 5 \\ 1 & 0 & 2.83 & 4.24 \end{bmatrix} \begin{matrix} c_1 = (1,1) & group-1 \\ c_2 = (2,1) & group-2 \end{matrix}$$

$$\begin{matrix} A & B & C & D \end{matrix}$$

$$\begin{bmatrix} 1 & 2 & 4 & 5 \\ 1 & 1 & 3 & 4 \end{bmatrix} \begin{matrix} X \\ Y \end{matrix}$$

Figure 2.6 Objects-Centroids distance

Each column in the distance matrix symbolizes the object. The first row of the distance matrix corresponds to the distance of each object to the first centroid and the second row is the distance of each object to the second centroid. For example, distance from medicine C = (4, 3) to the first centroid as:

$c_1 = (1, 1)$ is $\sqrt{(4-1)^2+(3-1)^2} = 3.61$

And, its distance to the second centroid below as:

$c_2 = (2, 1)$ is $\sqrt{(4-2)^2+(3-1)^2} = 2.83$, etc.

3. *Objects clustering:* The analysts assign each object based on the minimum distance. Thus, medicine A is assigned to group 1, medicine B to group 2, medicine C to group 2 and medicine D to group 2. The element of Group matrix below is 1 if and only if the object is assigned to that group.

$$G^0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} \begin{matrix} group-1 \\ group-2 \end{matrix}$$

$$\begin{matrix} A & B & C & D \end{matrix}$$

Figure 2.7 Objects clustering

4. *Iteration-1, determine centroids*: Knowing the members of each group, now we compute the new centroid of each group based on these new memberships. Group 1 only has one member thus the centroid remains in $c_1 = (1,1)$. Group 2 now

has three members, thus the centroid is the average coordinate among the three members: $c_2 = \left(\frac{2+4+5}{3}, \frac{1+3+4}{3}\right) = (4, 2\frac{2}{3})$.

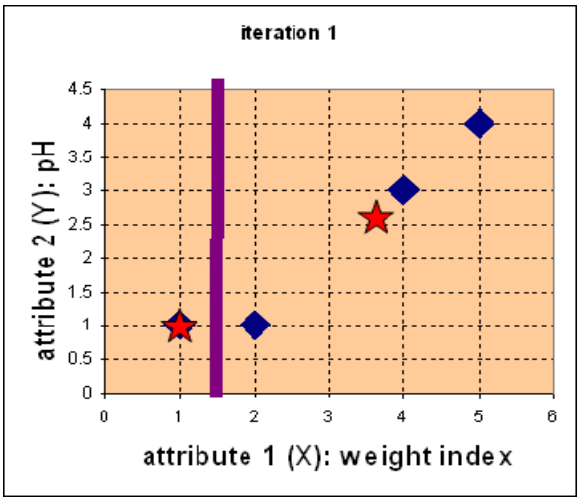


Figure 2.8 Iteration-1, determine centroids

5. *Iteration-1, Objects-Centroids distances* : The next step is to compute the distance of all objects to the new centroids. Similar to step 2, we have distance matrix at iteration 1 is

$$D^1 = \begin{bmatrix} 0 & 1 & 3.61 & 5 \\ 3.14 & 2.36 & 0.47 & 1.89 \end{bmatrix} \begin{matrix} c_1 = (1,1) & group-1 \\ c_2 = (\frac{11}{3}, \frac{8}{3}) & group-2 \end{matrix}$$

$$\begin{matrix} A & B & C & D \end{matrix}$$

$$\begin{bmatrix} 1 & 2 & 4 & 5 \\ 1 & 1 & 3 & 4 \end{bmatrix} \begin{matrix} X \\ Y \end{matrix}$$

Figure 2.9 Iteration-1, determine centroids

6. *Iteration-1, Objects clustering:* Similar to step 3, we assign each object based on the minimum distance. Based on the new distance matrix, we move the medicine B to Group 1 while all the other objects remain. The Group matrix is shown below.

$$G^1 = \begin{matrix} \\ group-1 \\ group-2 \end{matrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{matrix} \\ \\ A \quad B \quad C \quad D \end{matrix}$$

Figure 2.10 Iteration-1, determine centroids

7. *Iteration 2, determine centroids:* Now we repeat step 4 to calculate the new centroids coordinate based on the clustering of previous iteration. Group1 and group 2 both has two members, thus the new centroids are equitation as below:

$c_1 = (\frac{1+2}{2}, \frac{1+1}{2}) = (1\frac{1}{2}, 1)$ and $c_2 = (\frac{4+5}{2}, \frac{3+4}{2}) = (4\frac{1}{2}, 3\frac{1}{2})$.

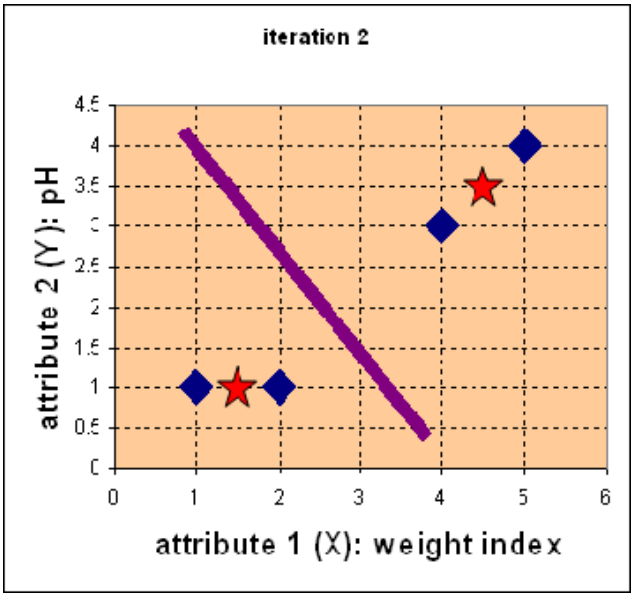


Figure 2.11 Iteration-1, determine centroids

8. *Iteration-2, Objects-Centroids distances:* Repeat step 2 again, we have new distance matrix at iteration 2 as:

$$D^2 = \begin{matrix} \\ c_1 = (1\frac{1}{2}, 1) \quad group-1 \\ c_2 = (4\frac{1}{2}, 3\frac{1}{2}) \quad group-2 \end{matrix} \begin{bmatrix} 0.5 & 0.5 & 3.20 & 4.61 \\ 4.30 & 3.54 & 0.71 & 0.71 \end{bmatrix} \begin{matrix} \\ \\ A \quad B \quad C \quad D \end{matrix}$$

$$\begin{bmatrix} 1 & 2 & 4 & 5 \\ 1 & 1 & 3 & 4 \end{bmatrix} \begin{matrix} X \\ Y \end{matrix}$$

Figure 2.12 Iteration-2, objects-centroids distances

9. *Iteration-2, Objects clustering:* Again, we assign each object based on the minimum distance.

$$G^2 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{matrix} group-1 \\ group-2 \end{matrix}$$
$$A \quad B \quad C \quad D$$

Figure 2.13 Iteration-2, Objects clustering

It was obtain result that $G^2 = G^1$. Comparing the grouping of last iteration and this iteration reveals that the objects does not move group anymore. Thus, the computation of the k-mean clustering has reached its stability and no more iteration is needed. It gets the final grouping as the results

Table 2.2 K-mean clustering comparing between groups

Object	Feature 1 (X): weight index	Feature 2 (Y): pH	Group (result)
Medicine A	1	1	1
Medicine B	2	1	1
Medicine C	4	3	2
Medicine D	5	4	2

Source: (“Numerical Example of K-Means Clustering,” 2012)

Step processing of k means clustering algorithm. It is presented in Figure below.

First, it will be begin with a decision on the value of k = number of clusters. Then, put any initial partition that classifies the data into k clusters. Analysts may assign the training samples randomly, or systematically as the following:

1. Take the first k training sample as single-element clusters
2. Assign each of the remaining (N-k) training samples to the cluster with the nearest centroid. After each assignment, recomputed the centroid of the gaining cluster.

Next, take each sample in sequence and compute its distance from the centroid of each of the clusters. If a sample is not currently in the cluster with the

closest centroid, switch this sample to that cluster and update the centroid of the cluster gaining the new sample and the cluster losing the sample.

And, finally, it will repeat step 3 until convergence is achieved, that is until a pass through the training sample causes no new assignments.

In this case, if the number of data is less than the number of cluster then analysts assign each data as the centroid of the cluster. Each centroid will have a cluster number. If the number of data is bigger than the number of cluster, for each data, analysts calculate the distance to all centroid and get the minimum distance. This data is said belong to the cluster that has minimum distance from this data ("Numerical Example of K-Means Clustering," 2012).

2.3.1 K-means Features and Priority

Similar to other algorithm, K-mean clustering has many weaknesses:

- When the numbers of data are not so many, initial grouping will determine the cluster significantly.
- The number of cluster, K, must be determined beforehand.
- The analysts never know the real cluster, using the same data, if it is inputted in a different order may produce different cluster if the number of data is a few.
- Sensitive to initial condition. Different initial condition may produce different result of cluster. The algorithm may be trapped in the local optimum.
- We never know which attribute contributes more to the grouping process since we assume that each attribute has the same weight.
- Weakness of arithmetic mean is not robust to outliers. Very far data from the centroid may pull the centroid away from the real one.
- The result is circular cluster shape because based on distance.

One way to overcome those weaknesses is to use K-mean clustering only if there are available many data. To overcome outliers problem, the analysts can use median instead of mean.

Some people pointed out that K means clustering cannot be used for other type of data rather than quantitative data. The key to use other type of dissimilarity is in the distance matrix.

2.4 Hierarchical Clustering

The Hierarchical Clustering are two main methods. First method is *agglomerative approach*, where we start from the bottom where all the objects are and going up (bottom *up approach)* through merging of objects. We begin with each individual objects and merge the two closest objects. The process is iterated until all objects are aggregated into a single group. Second method is *divisive approach (top down approach)*, where we start with assumption that all objects are group into a single group and then we split the group into two recursively until each group consists of a single object. One possible way to perform divisive approach is to first form a minimum spanning tree (e.g using Kruskal algorithm) and then recursively (or iteratively) split the tree by the largest distance. In this simple tutorial will only show the example of agglomerative approach.

Step by step algorithm of agglomerative approach to compute hierarchical clustering is as follow

1. Convert object features to distance matrix.
2. Set each object as a cluster (thus if we have 6 objects, we will have 6 clusters in the beginning)
3. Iterate until number of cluster is 1
    1.) Merge two closest clusters
    2.) Update distance matrix

The flow chart of agglomerative hierarchical clustering algorithm is given below.

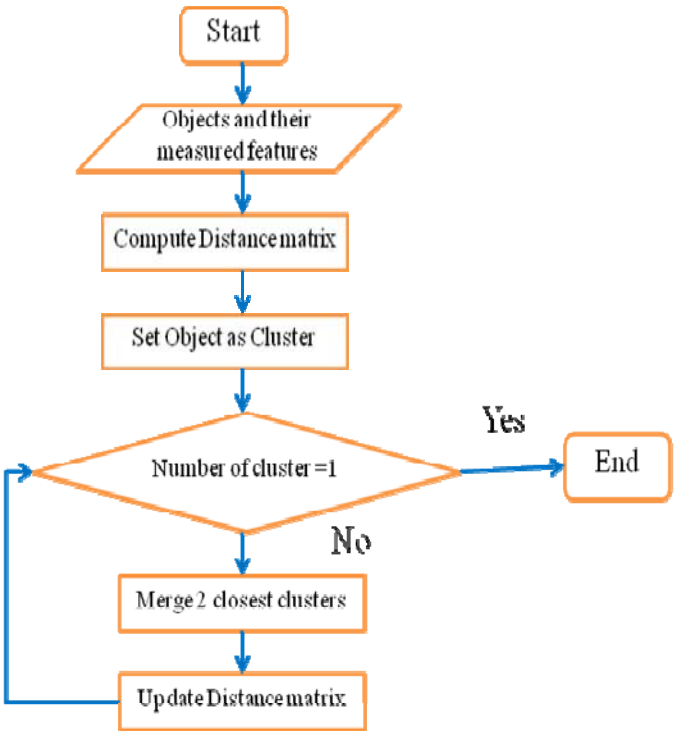


Figure 2.14 Hierarchical Clustering process

The rule of hierarchical clustering lie on how objects should be grouped into clusters. Given a distance matrix, linkages between objects can be computed through a criterion to compute distance between groups. Most common & basic criteria are Single Linkage, Complete Linkage, Average Group and Centroid distance

- Single Linkage: minimum distance criterion

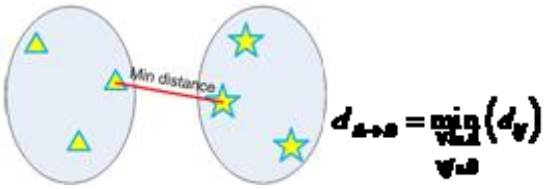
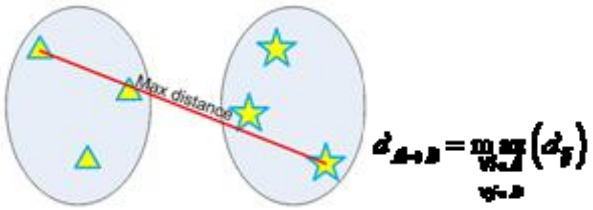
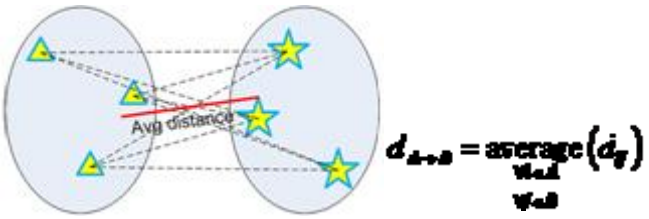


Figure 2.15 Hierarchical Clustering criterion

- Complete Linkage: maximum distance criterion



- Average Group: average distance criterion



- Centroid distance criterion

Figure 2.15 Hierarchical Clustering criterion (Cont.)

Source: (“Hierarchical Clustering Tutorial, 2009”)

CHAPTER III
MATERIALS AND METHODS

3.1 Steps of Research Methodology

The first step to study is survey and gathers research about on the taxonomy of fish species that are important to the economy of the country. To study other research institutions or experts, which are either in from of book, document, article and teaching materials online to guide of method selected for use in the research.

3.1.1 Cluster method

Cluster is automatically algorithm, which call is "Unsupervised". The clustering algorithm will automatic calculate the distance between data from a nearing data. The data is similar will be get into same group and the difference will be get into difference group. The information contained in the same group is very similar and the data in different groups will have a different.

This study will compare the clustering analysis 2 method is Hierarchical clustering analysis and K-Mean clustering analysis in the same data

3.1.2 Collect fish data

In this research were to study the fish shape with similar shape in the Order or Family. The researcher will choose in commercial fish of Thailand. Through an online database of FisBase ("FishBase-World Wide Web electronic publication," 2012), which an online database of fish is reliability. The data collected in the Order, Family, Name and image storage in database that will select only left side fish image and most obvious image.

Figure 3.1 represents data for commercial fish in Thailand; they were stored in database including image data. The example of record data as below:

Sample Data Family, Fbname, Use elsewhere and commercial fish picture

Order	Family	Species	Occurrence	FishBase name	Name	Use	Use elsewhere
Tetraodontiformes	Balistidae	*Abalistes stellaris*	native	Starry triggerfish		commercial	commercial
Perciformes	Pomacentridae	*Abudefduf vaigiensis*	native	Indo-Pacific sergeant	ตะกรับเขียวเหลือง	subsistence fisheries	subsistence fisheries
Perciformes	Sparidae	*Acanthopagrus berda*	native	Goldsilk seabream	Pla e-klud	commercial	commercial
Perciformes	Sparidae	*Acanthopagrus latus*	native	Yellowfin seabream	Pla e-klud	minor commercial	commercial
Perciformes	Acanthuridae	*Acanthurus nigricans*	native	Whitecheek surgeonfish		subsistence fisheries	subsistence fisheries
Perciformes	Acanthuridae	*Acanthurus triostegus*	native	Convict surgeonfish		commercial	commercial
Perciformes	Serranidae	*Aethaloperca rogaa*	native	Redmouth grouper	Pla karang leudnok	minor commercial	minor commercial
Rajiformes	Myliobatidae	*Aetobatus narinari*	native	Spotted eagle ray	Krabane nog	minor commercial	minor commercial
Rajiformes	Myliobatidae	*Aetomylaeus maculatus*	native	Mottled eagle ray		commercial	minor commercial
Rajiformes	Myliobatidae	*Aetomylaeus nichofii*	native	Banded eagle ray	Krabane nog	commercial	minor commercial
Albuliformes	Albulidae	*Albula vulpes*	native	Bonefish		minor commercial	minor commercial
Cypriniformes	Cyprinidae	*Albulichthys albuloides*	native		Paled barb	commercial	minor commercial
Perciformes	Carangidae	*Alectis ciliaris*	native	African pompano		minor commercial	minor commercial
Perciformes	Carangidae	*Alectis indica*	native	Indian threadfish		commercial	commercial
Perciformes	Carangidae	*Alepes djedaba*	native	Shrimp scad		commercial	commercial
Perciformes	Carangidae	*Alepes melanoptera*	native	Blackfin scad		minor commercial	minor commercial
Perciformes	Carangidae	*Alepes vari*	native	Herring scad		minor commercial	minor commercial
Lamniformes	Alopiidae	*Alopias vulpinus*	native	Thresher		commercial	commercial
Perciformes	Ambassidae	*Ambassis kopsii*	native	Freckled hawkfish		commercial	commercial
Clupeiformes	Clupeidae	*Amblygaster leiogaster*	native	Smoothbelly sardinella		subsistence fisheries	subsistence fisheries
Clupeiformes	Clupeidae	*Amblygaster sirm*	native	Spotted sardinella		commercial	commercial
Perciformes	Anabantidae	*Anabas testudineus*	native	Climbing perch	Pla mor Thai	commercial	commercial
Anguilliformes	Anguillidae	*Anguilla japonica*	not established	Japanese eel	Pla lai yee pun	highly commercial	highly commercial
Clupeiformes	Clupeidae	*Anodontostoma chacunda*	native	Chacunda gizzard shad	ตะเพียนน้ำเค็ม	commercial	commercial
Clupeiformes	Clupeidae	*Anodontostoma thailandiae*	native	Thai gizzard shad		subsistence fisheries	subsistence fisheries
Pristiformes	Pristidae	*Anoxypristis cuspidata*	native	Pointed sawfish	Knifetooth sawfish	commercial	commercial
Perciformes	Serranidae	*Anyperodon leucogrammicus*	native	Slender grouper	Pla karang-laisen	minor commercial	minor commercial
Perciformes	Lutjanidae	*Aphareus furca*	native	Small toothed jobfish		commercial	commercial
Perciformes	Lutjanidae	*Aphareus rutilans*	native	Rusty jobfish		commercial	commercial
Perciformes	Lutjanidae	*Aprion virescens*	native	Green jobfish		highly commercial	highly commercial
Perciformes	Sparidae	*Argyrops spinifer*	native	King soldier bream		commercial	commercial
Siluriformes	Ariidae	*Arius maculatus*	native	Spotted catfish		commercial	commercial

Figure 3.1 Data for commercial fish in Thailand

minor commercial	minor commercial	

Figure 3.2 Data group minor commercial fish example

highly commercial	highly commercial	

Figure 3.3 Data group highly commercial fish example

commercial	commercial	

Figure 3.4 Data group commercial fish example

subsistence fisheries	subsistence fisheries	

Figure 3.5 Data group subsistence fisheries fish example

3.1.3 Studies using MATLAB

This research is study and analyze image. The researcher need to study the use Matlab function about detect and trace border to create fish shape and cluster analyze is following by:

- Image processing function is studying on the edge function of the image (trace border) to create fish shape from the original image. To find the center of the fish shape and distance between the center point of the fish shape and border shape of fish.
- Clustering Function is studying the function of clustering processing both 2 method is Hierarchical clustering analysis and K-Mean clustering analysis, which used to analyze group shape data.

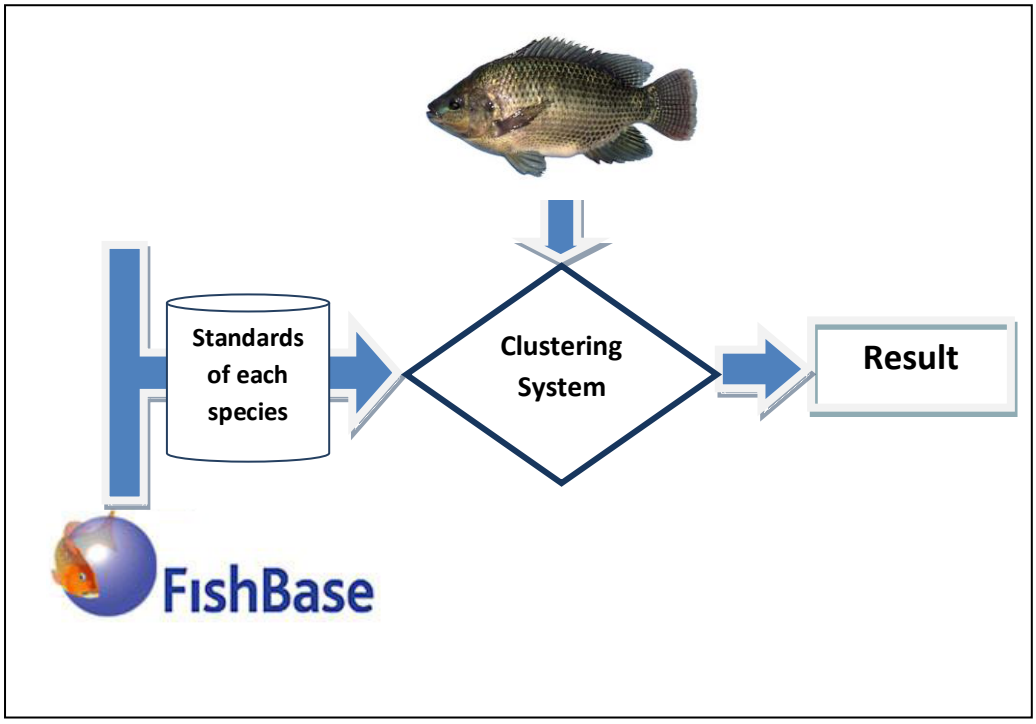


Figure 3.6 System Architecture

3.1.4. Step of research

1. Collect image in data stored and used to image processing function. The program has output is distance between the center point of the fish shape and border shape of fish. The process is as follows:

- The program will convert the image from color to black and white (image as binary) for used to edge function and bwtraceboundary function, which the program will trace the edge at start point and drag it around until back to the beginning is the end. The program is return value of each row and column of points to plot fish shape. After that it will calculate to find the midpoint of the shape.
- The program is find distance between the centers points along the border with the collected them to excel. After that researcher used to data for cluster analyze grouping similar data.

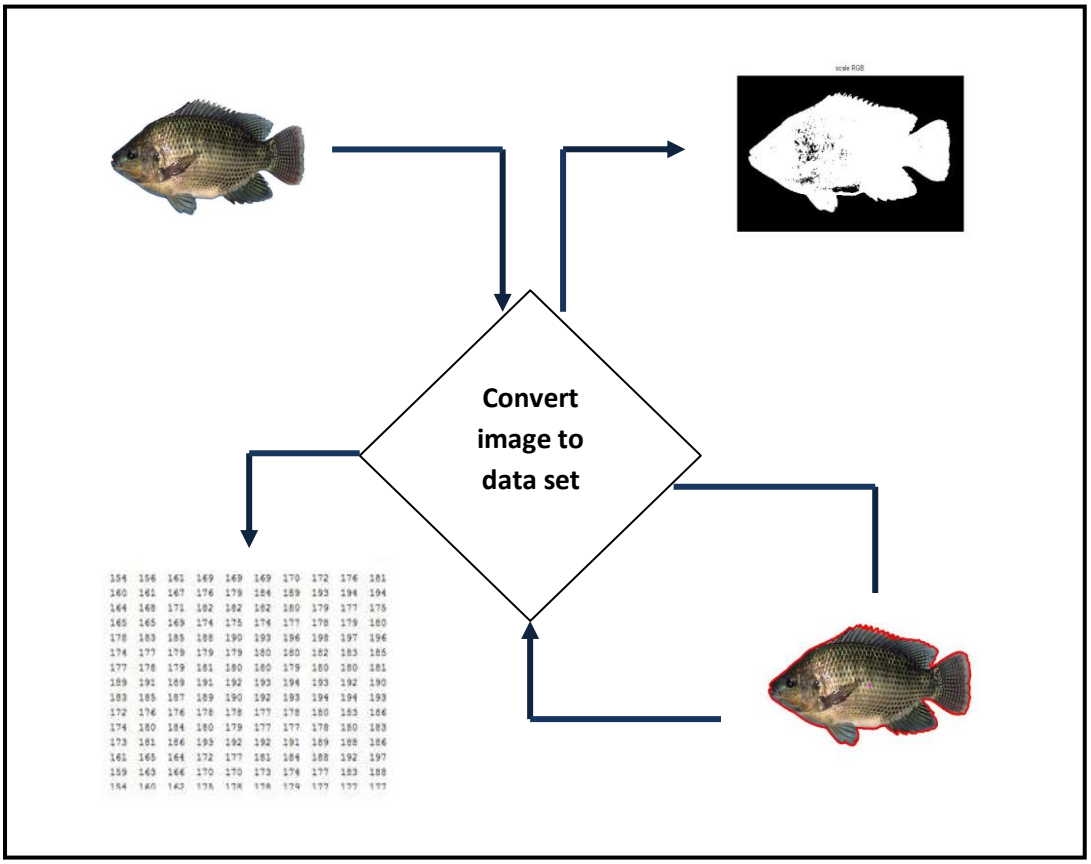


Figure 3.7 System convert fish picture to dataset

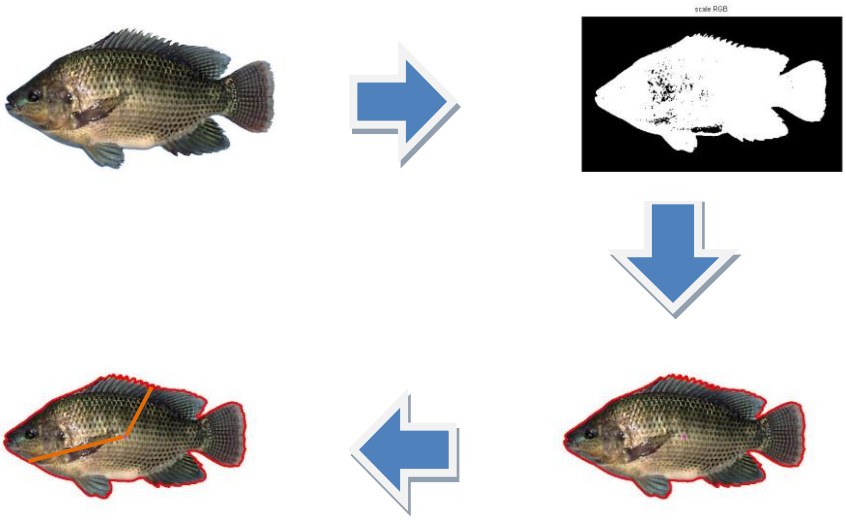


Figure 3.8 System to trace border and calculate distance

2. The distance between the midpoints of the line shape analysis to group the 2 method is follow:

- Hierarchical clustering analysis is agglomerative approach, where researcher starts from the bottom where all the objects are and going up (bottom up approach) through merging of objects. Researcher begin with each individual objects and merge the two closest objects. The process is iterated until all objects are aggregated into a single group. (“Teknomo, Kardi. (2009) Hierarchical Clustering Tutorial”)

- K-Mean clustering analysis is algorithm was developed by J. MacQueen(1967) and then by J.A. Hartigan and M.A. Wong around 1975. Simply speaking k-means clustering is an algorithm to classify or to group your objects based on attributes/features into K number of group. K is positive integer number. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid. Thus the purpose of K-mean clustering is to classify the data (“Teknomo, Kardi. (2006) Hierarchical Clustering Tutorial”). The researcher will use K = 5 because fish shape is determined by side view in this follow:

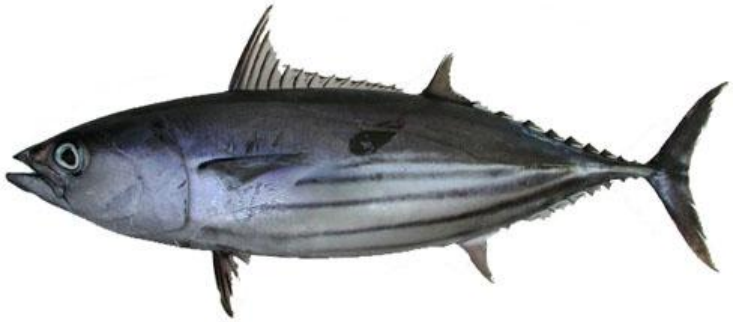


Figure 3.9 Fusiform shape

Figure 3.10 Angguiliform shape

Figure 3.11 Globiform shape

Figure 3.12 Arrow-shape

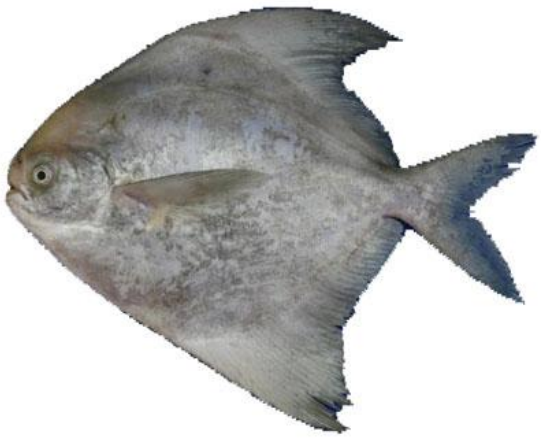


Figure 3.13 Compressed shape

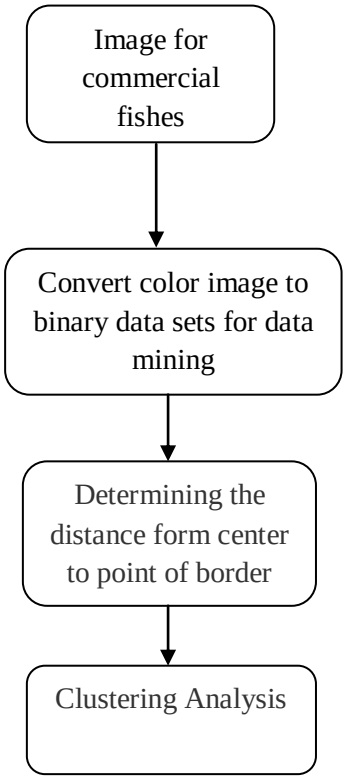


Figure 3.14 Summary process

3.2 Database design

Excel is the data platform that well done and effectively for runs application. It help to save time, reduce course for manage application, availability, scalability

3.3 Programming Technique

MATLAB (matrix laboratory) is a numerical computing environment and fourth-generation programming language. Developed by MathWorks, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, Java, and Fortran.

Although MATLAB is intended primarily for numerical computing, an optional toolbox uses the MuPAD symbolic engine, allowing access to symbolic

computing capabilities. An additional package, Simu link, adds graphical multi-domain simulation and Model-Based Design for dynamic and embedded systems.

3.4 Testing and system evaluation

3.4.1 Hierarchical clustering

The data will analysis by calculating the distance to the Euclidean distant of a single linkage criterion and average criterion, which show result in dendogram to the relevance of similar shape.

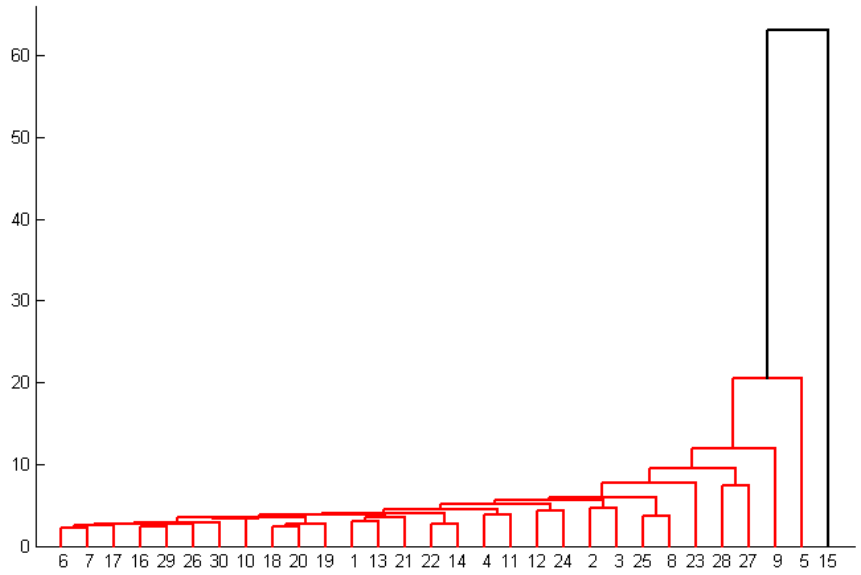


Figure 3.15 Dendogram

3.4.2 K-mean clustering

The data will analysis K=5 by calculating the distance to the Euclidean distant, which show result in table

3.5 Required tools and Time frame

3.5.1 Hardware

Personal computer

(Personal computer is used for developing application)

3.5.2 Software

Programming Language:

Matlab 7.9.0b (used for development the system)

Database Management System (DBMS):

Excel 2007

Operating System:

Windows XP professional (Windows XP will be installed on personal computer that is used for development program.)

Gantt chart:

Activities	**Time (Months)**	
	1	**2**
1. Literature Review		
2. System Analysis		
3. System Design		
4. System Development		
5. Testing		
6. Conclude the results		
7. Final document		

Figure 3.16 Gantt chart

CHAPTER IV
EXPERIMENT AND RESULTS

This chapter presents the image processing result and the clustering analysis result Researcher will show the collect data analyzed. This chapter is divided into two parts. The first part presents image processing result from Matlab function. And, the second part presents the clustering analysis result for 2 method are the hierarchical cluster analysis and the k-mean cluster analysis.

4.1 The Image Processing Results

The first develop source code in Matlab program for converting images to dataset. The color image cannot change to dataset in one step but requires a multi-step process. The information system must be run as follows:

1) The program can use images stored correctly follow by user input path directory image source. Matlab program use imread function for read image from graphics file to unit8 format.

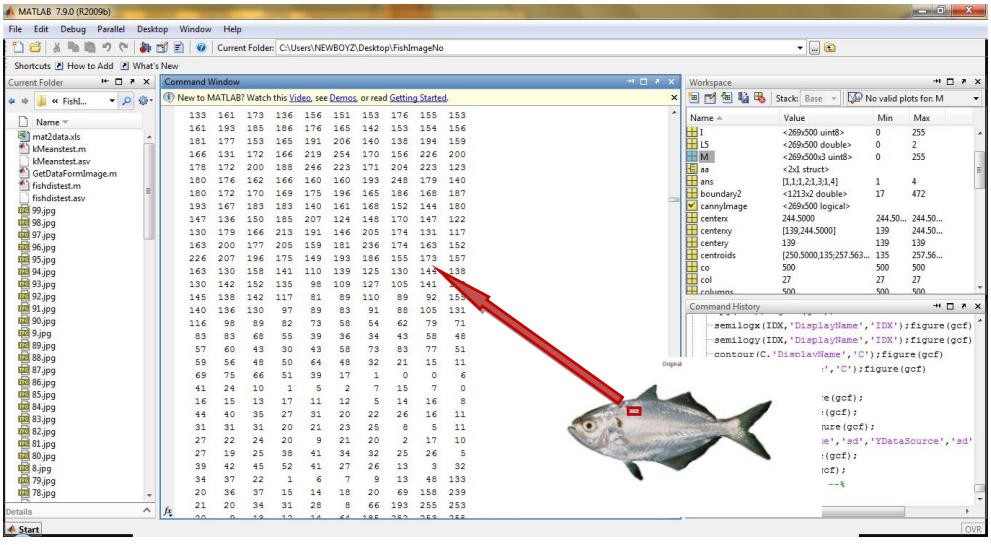


Figure 4.1 Result imread function

2) The program use rgb2bw function for convert the true color image from RGB to black and white image. After that remove noise from image by imfill function on screen and lets you define the region to fill by selecting point.

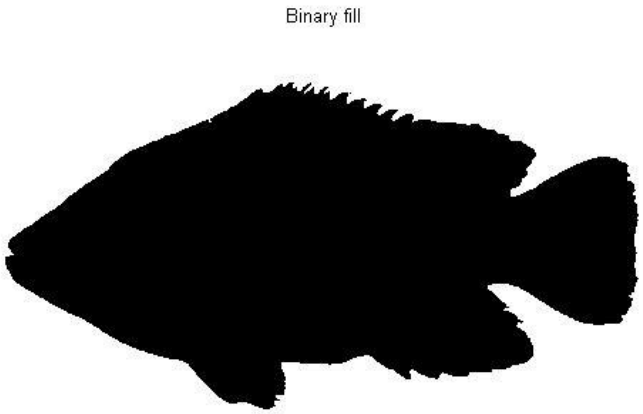


Figure 4.2 Result imfill function

3) The program use bwperim function returns a binary image containing only the perimeter pixels of objects in the input image for edge detect border.

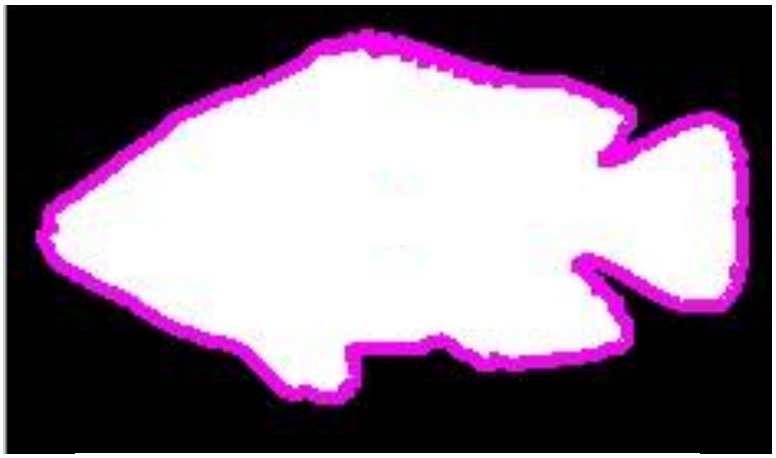


Figure 4.3 Result bwperim function

4) The program use bwtraceboundary function for traces the outline of an object in binary image get two – element vector specifying the row and column coordinates of the point on the object boundary where you want the tracing to begin.

```
[ro,co]=size(BW5);
plotxy = 0;
for y1 = 1:co;
  for x1 = 1:ro;
    ppp = BW5(x1,y1);
    if (ppp == 0)
       plotxy = 1;
           break;
    end
  end
 if plotxy == 1
    break;
 end
end
x1 = x1-1;
boundary2 = bwtraceboundary(BW5,[x1, y1],'N');
```

Figure 4.4 Function bwtraceboundary result

5) The program will finds centroid of shape by centroid function in Matlab.

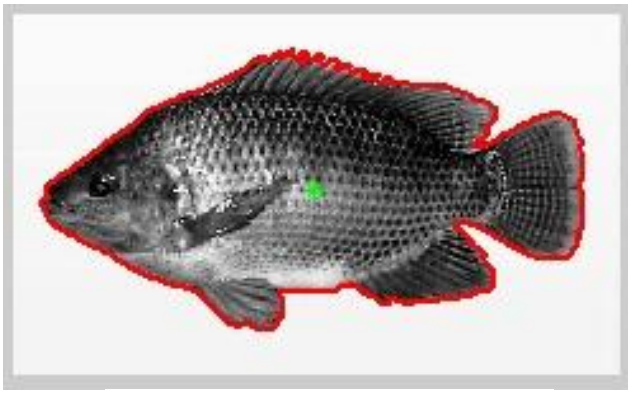


Figure 4.5 Function centroid

6) The program will find the distance between centroids and points on shape by distance function.

```
distpoint2 = distance('gc',centerxy(1,1:2),boundary2);
disstat2 = distpoint2(1:1,1);
distpointscale2 = distpoint2*(100/disstat2);
```

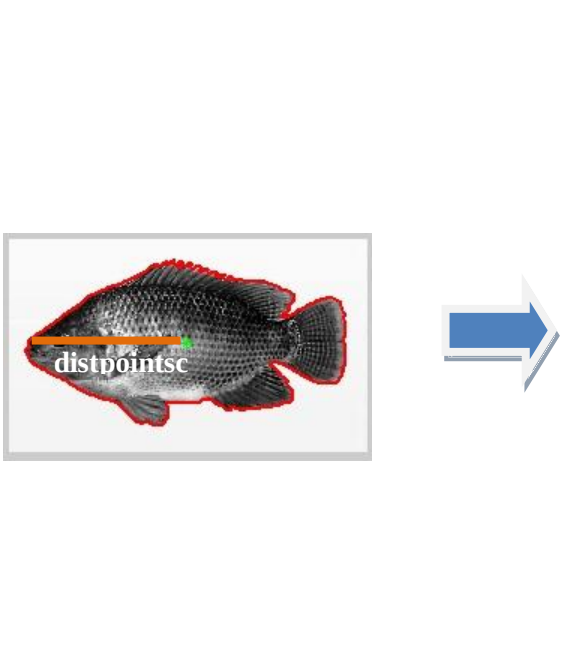


Figure 4.6 Dispointscale convert to data

The program run code for 100 example fish find data distance of fish shape. After program complete get dataset.

Name	X1	X2	X3	X4	X5	X6	X7	X8
Pla Krai	100	129.5909	148.0813	162.2262	171.9198	181.0805	185.6742	193.6629
Pla Salad	100	113.1078	106.3605	96.71003	104.3047	97.6909	84.07657	90.80574
Pla Kraho	100	106.132	117.935	112.8318	120.3245	129.8478	121.6258	112.5597
Pla Yeesok	100	121.9597	132.6234	146.2474	143.3361	137.0908	123.7697	130.9195
Pla Tapienkhaw	100	111.5087	123.0498	134.376	142.0526	146.8014	144.2921	137.2383
Pla Kra mang	100	140.7828	165.9602	176.9592	192.0808	202.2918	189.9244	196.4572
Pla Nai	100	95.84989	97.89668	93.77613	91.92105	86.22679	80.57323	73.95129
Pla Chao	100	102.5592	113.3415	122.6786	125.947	121.3051	114.1602	104.3697
Pla Leng	100	105.8744	115.7501	121.645	119.0888	113.6808	105.3528	96.45963
Pla Yeesoktad	100	115.2685	139.121	138.6571	118.4442	116.8254	132.7531	126.9322
Pla Khao	100	59.79126	15.3032	24.55039	76.09611	122.2588	135.728	129.0665
Pla Deng	100	62.82695	38.99617	25.87249	10.20736	9.965302	23.28715	38.5626
Pla Kot	100	99.67115	95.39802	90.02548	90.25051	92.67363	91.78989	91.93976
Pla Buk	100	135.8272	168.2276	200.8024	219.9737	240.2433	240.9293	242.0365
Pla Thepho	100	112.0264	124.6449	140.2089	154.1307	165.447	169.8702	173.3107

Figure 4.7 Dispointscale result

4.2 The Cluster Analysis results

4.2.1 Hierarchical Clustering Method

The program will compute hierarchical clustering using hierarchical function in Matlab. The hierarchical clustering using agglomerative technique and validate the clustering using Cophenetic Correlation Coefficient.

The rule of hierarchical clustering lie on how objects should be grouped into clusters. Given a distance matrix, linkages between objects can be computed through a criterion to compute distance between groups. Most common and basic criteria are Single Linkage, Average Group and Complete Linkage

The program run in Single Linkage is show dendogram in figure 4.8. The single linkage using criterion is minimum distance criterion for grouped dataset. Dendogram show fish shape data are short distance such as fish type number 95, 96, 88, 98, 56, 43, 37, 45 are not different. For fish type number 100 is morphologically distinct from other because distance than others, but there are also similar to those for fish type number 71, 72, 35. The minimum distance criterion have results more similar than another criterion by using to grouping data have similar back to top of the hierarchy. The image of fish type number in dendogram can see in appendix.

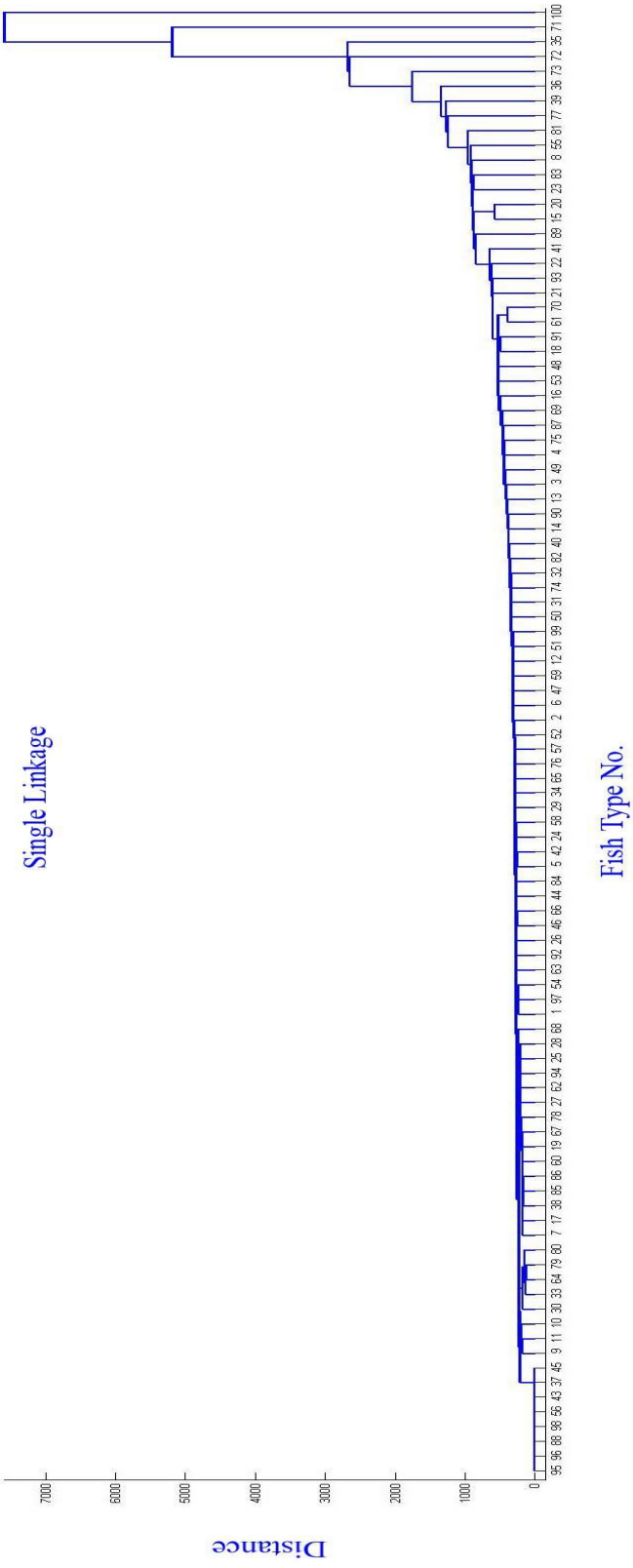


Figure 4.8 Dendogram Single Linkage of 100 species

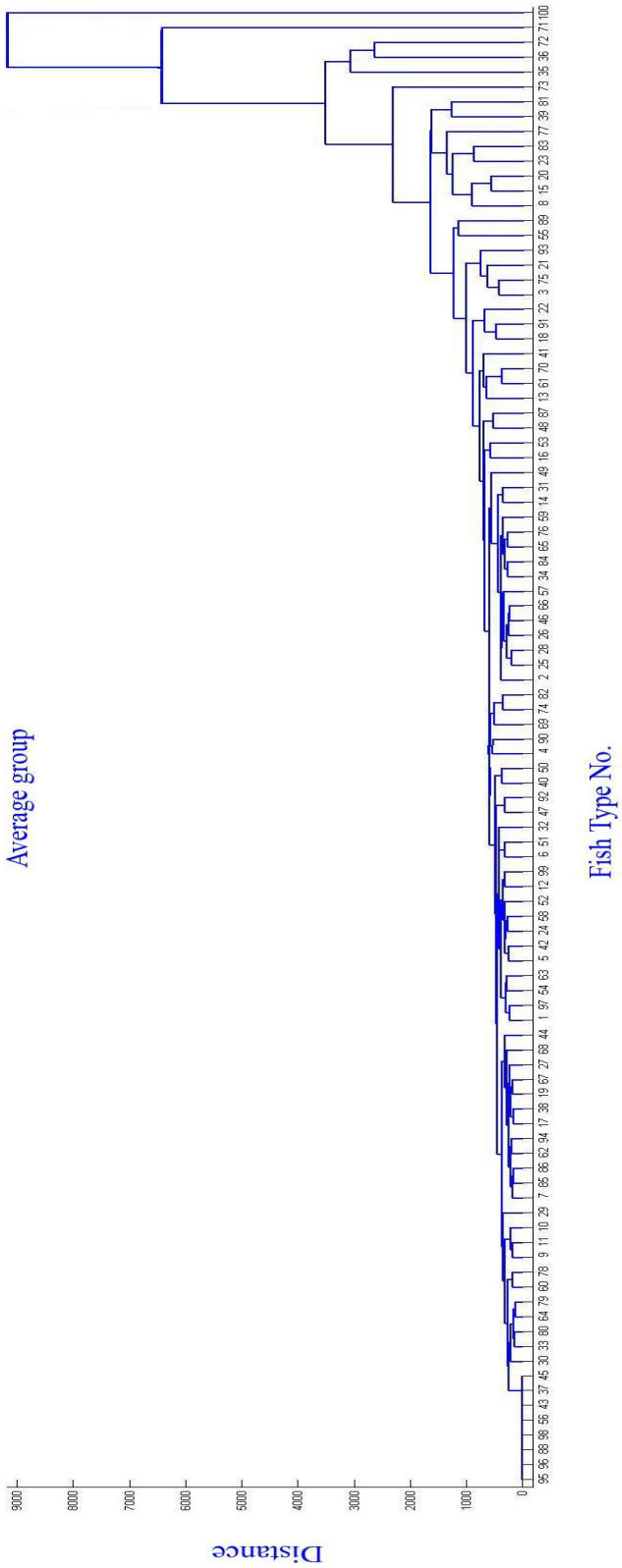


Figure 4.9 Dendogram Average Group of 100 species

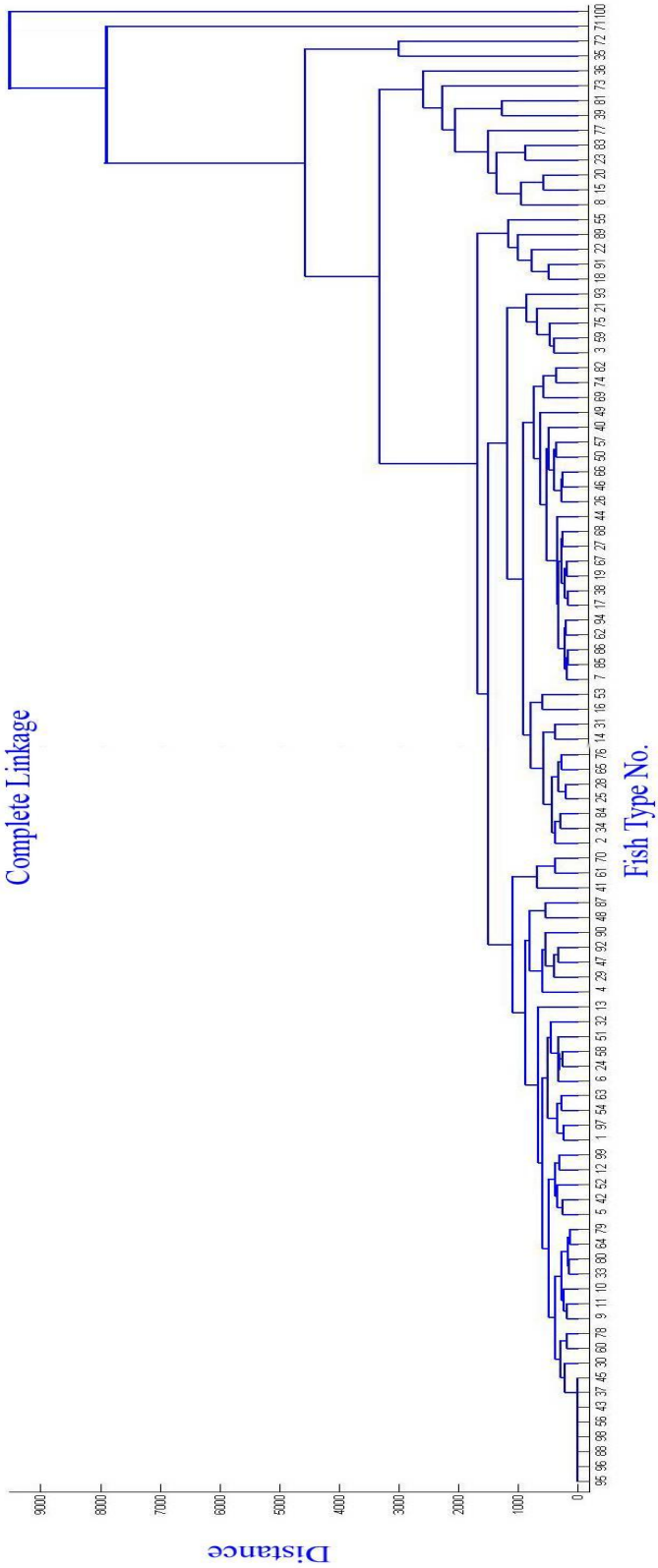


Figure 4.10 Dendogram Complete Linkage of 100 species

The program run in Average Group is show dendogram in figure 4.9. The Average Group using criterion is average distance criterion for grouped dataset. Dendogram show fish shape data are short distance such as fish type number 95, 96, 88, 98, 56, 43, 37, 45 are not different. For fish type number 100 is morphologically distinct from other because distance than others, but there are also similar to those for fish type number 71, 72 same The Single Linkage. The average distance criterion results have pairing data for grouping data in the middle of dendogram by the result using to grouping data have similar back to top of the hierarchy. The image of fish type number in dendogram can see in appendix.

The program run in Complete Linkage is show dendogram in figure 4.10. The Complete Linkage using criterion is maximum distance criterion for grouped dataset. Dendogram show fish shape data are short distance such as fish type number 95, 96, 88, 98, 56, 43, 37, 45 are not different. For fish type number 100 is morphologically distinct from other because distance than others, but there are also similar to those for fish type number 71, 72, 35 same The Single Linkage. The maximum distance criterion can grouping data more than another criterion by the result using to grouping data have similar back to top of the hierarchy. The image of fish type number in dendogram can see in appendix.

4.2.2 K-Mean Clustering Method

The program runs result grouped by user design. In paper researcher determine number of cluster K and assume the centroid or center of these clusters. Researcher can take any random objects as the initial centroids or the first K objects in sequence can also serve as the initial centroids.

This research use K = 5 cause research want the program group fish shape is analyze result show in figure 4.11. The result is show number of group, which the program grouping data have more similar is in same group but data have difference is in other group.

Table 4.1 K-Mean result

No.	Family	Species	Thai Name	Group
1	Muraenesocidae	Muraenesox cinereus	Pla Yodjak	5
2	Synodontidae	Saurida micropectoralis	Pla Pakkom	2
3	Clupeidae	Clupeoides lile	Pla Katakkhao	3
4	Chirocentridae	Chirocentrus dorab	Pla Dab lao	5
5	Cyprinidae	Catlocarpio siamensis	Pla Kraho	3
6	Cyprinidae	Probarbus jullieni	Pla Yeesok	3
7	Cyprinidae	Puntius gonionotus	Pla Tapienkhaw	3
8	Cyprinidae	Puntioplites proctozysron	Pla Kra mang	3
9	Cyprinidae	Cyprinus carpio	Pla Nai	3
10	Cyprinidae	Ctenopharyngodon idella	Pla Chao	2
11	Cyprinidae	Hypophthalmichthys molitrix	Pla Leng	2
12	Cyprinidae	Labeo rohita	Pla Yeesoktad	3
13	Chanidae	Chanos chanos	Pla Dok mai	3
14	Mugilidae	Mugil vaigiensis	Pla Krabok	2
15	Notopteridae	Notopterrus chitala	Pla Krai	4
16	Notopteridae	Notopterus notopterus	Pla Salad	4
17	Channidae	Channa (Ophicephalus) striatus	Pla Chon	2
18	Channidae	Chana (Ophicephalus) micropeltes	Pla Chado	2
19	Cichlidae	Oreochromis nilooticus	Pla Nil	3
20	Anabantidae	Anabas testudinus	Pla Mor	3
21	Eleotridae	Oxyeleotris marmoratus	Pla Buzai	2
22	Sphyraenidae	Sphyraena jello	Pla Sakdam	5
23	Polynemidae	Eleutheronema tetradactylum	Pla Kurao	5
24	Lates perches	Lates calcarifer	Pla Kapongkeo	3
25	Serranidae	Epinephelus areolatus	Pla Karanghangtud	3
26	Serranidae	Plectropomus maculatus	Pla Kudsalak	3
27	Sillaginidae	Sillago sihama	Pla Chonsaikaeo	2
28	Lactariidae	Lactarius lactarius	Pla Bikanun	3
29	Carangidae	Alectis indica	Pla Shomngam	3

Table 4.1 K-Mean result (Cont.)

No.	Family	Species	Thai Name	Group
30	Carangidae	Caranx ignobilis	Pla Sikunphueak	3
31	Carangidae	Caranx boops	Pla Sikunthong	3
32	Carangidae	Caranx leptolepis	Pla Sikunkanglieng	3
33	Carangidae	Caraangoides armatus	Pla Takong	3
34	Carangidae	Seriolina nigrofasciata	Pla Sumlee	3
35	Carangidae	Scomberoides lysan	Pla Chaliap	3
36	Carangidae	Scomberoides tol	Pla Sisiad	3
37	Carangidae	Megalaspis cordyla	Pla Kakai	3
38	Carangidae	Decapterus russelli	Pla Thu	3
39	Rachycentridae	Rachycentron canadus	Pla Chontaley	2
40	Coryphaenidae	Coryphaena hippurus	Pla Etomon	2
41	Lutjanide	Asrion virescens	Pla Kapongkeaw	2
42	Lutjanide	Lutjanus lunulatus	Pla Kapongkanglieng	3
43	Lutjanide	Lutjanus vitta	Pla Kanglueng	3
44	Lutjanide	Lutjanus sebae	Pla Kapongdaengthabdeng	3
45	Lutjanide	Lutjianus malabaicus	Pla Daeng	3
46	Lutjanide	Lutjanus lutjanus	Pla Kapong	3
47	Lutjanide	Lutjanus russellii	Pla Kangpan	3
48	Lutjanide	Nemipterus hexodon	Pla Unggolee	3
49	Lutjanide	Nemipterus japonicus	Pla Saidaenghangyao	3
50	Lutjanide	Nemipterus peronii	Pla Saidaengkadong	3
51	Leiognathidae	Secutor insidiator	Pla Panpangmoo	3
52	Leiognathidae	Eubleekeria splendens	Pla Paenkrasouy	3
53	Leiognathidae	Equulites elongatus	Pla Paen	3
54	Pomacentridae	Abudefduf vaigiensis	Pla Trakrup	3
55	Pomatomidae	Pomatomus saltatrix	Pla Numngen	5
56	Sciaenidae	Otolithes ruber	Pla Juadteinkeiw	5
57	Sciaenidae	Pennahia anea	Pla Juadkrebtao	2
58	Sciaenidae	Nibea soldado	Pla Juad	2
59	Sciaenidae	Boesemania microlepis	Pla Hangkew	2

Table 4.1 K-Mean result (Cont.)

No.	Family	Species	Thai Name	Group
60	Lethrinidae	Lethrinus nebulosus	Pla Moosee	3
61	Mullidae	Upeneus moluccensis	Pla Paithong	2
62	Mullidae	Upeneus sundaicus	Pla Pailay	2
63	Mullidae	Parupeneus heptacanthus	Pla Paithongleing	2
64	Siganidae	Siganus guttatus	Pla Salidhinjudleing	3
65	Siganidae	Siganus punctatus	Pla Baikanoonleang	3
66	Siganidae	Siganus canaliculatus	Pla Salidhinjudkaow	3
67	Siganidae	Siganus javus	Pla Salidtab	3
68	Trichiuridae	Trichiurus haumela	Pla Dapngen	5
69	Scombridae	Rastrelliger brachysoma	Pla Thumong	3
70	Scombridae	Rastrelliger kanagurta	Pla Lung	3
71	Scombridae	Scomberomorus commerson	Pla Inseebung	2
72	Scombridae	Scomberomorus guttatus	Pla Inseekhaotok	2
73	Scombridae	Scomberomorus lineolatus	Pla Thangdinsor	2
74	Scombridae	Auxis rochei rochei	Pla Thukak	2
75	Scombridae	Euthynnus affinis	Pla Oklab	3
76	Scombridae	Katsuwonus pelamis	Pla Otab	3
77	Istiophoridae	Istiophorus platypterus	Pla Katongtangkluy	5
78	Stromateidae	Pampus argenteus	Pla Jalamedkaow	3
79	Stromateidae	Pampus chinensis	Pla Jalamedtao	3
80	Stromateidae	Parastromateus niger	Pla Jalameddum	3
81	Xiphiidae	Xiphias gladius	Pla Katongtangdab	5
82	Scombridae	Auxis thazard	Pla Okaow	2
83	Scombridae	Thunnus tonggol	Pla Odum	2
84	Psettodidae	Psettodes erumei	Pla Sigdeaoy	3
85	Cynoglossidae	Cynoglossus bilineatus	Pla Yodmoung	4
86	Cynoglossidae	Cynoglossus puncticeps	Pla Changchun	4
87	Cynoglossidae	Cynoglossus lingua	Pla Linma	4

Table 4.1 K-Mean result (Cont.)

No.	Family	Species	Thai Name	Group
88	Platycephalidae	Inegocia japonica	Pla Huaban	5
89	Platycephalidae	Platycephalus indicus	Pla Hangkway	5
90	Siluridae	Wallagonia attu	Pla Khao	4
91	Siluridae	Kryptopterus bleekeri	Pla Deng	4
92	Bagridae	Mystus nemurus	Pla Kot	4
93	Pangasiidae	Pangasianodon gigas	Pla Buk	4
94	Pangasiidae	Pangasius larnaudii	Pla Thepho	4
95	Clariidae	Clarias macrocephalus	Pla Dukuey	4
96	Clariidae	Clarias batrachus	Pla Dukdan	4
97	Ariidae	Arius thalassinus	Pla Riwkiw	2
98	Plotosidae	Plotosus lineatus	Pla Pinkaeo	4
99	Synbranchidae	Monopterus albus	Pla Lai	2
100	Molidae	Mola mola	Pla Praatid	1

CHAPTER V
CONCLUSION AND DISCUSSION

The objective of this research is to cluster the commercial fish's image for automatic grouped by shape. The research used image processing and data mining techniques (Clustering) to reach the objective.

5.1 Image processing techniques

Matlab for image processing applications require a knowledge of mathematics and knowledge about the function in the program. To get the correct value needs to be tested frequently and use a variety of experiments. The image convert to the shape data needed to determine the direction of the image of fish to be tested. In this research, the image is the left side of the fish and white background of the image and visibility of the fish.

Problem is the quality of color that is soft similar to the background, the resulting in incorrect data. Edge of the shape fish with noise pixel color resulting process does not match the image fish. System image processing that does not convert images to grayscale will not be converting to data.

5.2 Data mining techniques

The cluster techniques used in this research, are logistic hierarchical method and k-mean method.

Hierarchical method is automatic grouping for dataset and shown similar data to a tree-like diagram which is called a dendogram. The method is suitable for the numbers of dataset are not so many, because the result data is contiguous which not easy to read. In this paper researcher use dataset is similar so the result of hierarchical to both criteria are not difference, which the result can group similar into group but

some data is not true similarity like data in group. As a result are these because shape image are not difference and convert to distance having error some data.

K-Mean method is automatic grouping base on features into k number of group. The result is grouping is similar true data when k number is 5 dataset were separated by distance shape data and create new group. K-mean is appropriate for the numbers of dataset are many, because the result is nearest to true data.

5.3 Discussion

1. The convert images commercial fishes to data shape, should be collected the same and one time the conversion by develop code Matlab to cover all image to reduce working time.

2. The cluster shape have an error in convert shape processing cause image shape and image background so similar.

3. In next time should be tried to add the number of distance shape in dataset and should be diverse example for see the difference.

4. In next time should be include other data for improve system to cluster in other formats.

5.4 Summary

The system to convert images to dataset can work well if the scope is defined. The system can convert the correct shape fish and convert to as distance. The k-mean clustering method is better than hierarchical clustering method because the result of k-mean clustering method easy to read group of data. The cluster will be effective if data have different shape or include more information and should be improve by dynamic time warping.

REFERENCES

Froese, R. & Pauly D. (2012). FishBase,World Wide Web electronic. Retrieved January 18, 2012, from http://www.fishbase.org/search.php

Kantardzic, M. (2001). Data Mining Concepts, Models, Methods, and Algorithms. Wiley-InterScience, New York.

Maneesriwongul, Amnuay. (1998). Cluster Analysis in research journal. Retrieved January 18, 2012, from http://rci2010.files.wordpress.com/2010/06/cluster-analysis.doc

Numerical Example of K-Means Clustering. (2006). Retrieved January 18, 2012, from http://people.revoledu.com/kardi/tutorial/kMean/NumericalExample.htm publication. Retrieved January 18, 2012, from

Sittikraiwong, P. (2012). Education instruction online. Retrieved January 18, 2012, from http://courseware.rmutl.ac.th/courses/107/unit000.html

Sompode. (2012). Classify variable groups of Cluster Analysis by Sompode. Retrieved January 18, 2012, from http://www.saruthipong.com/port/document/299-705/299-705-10.pdf

Tan, P. N., Steinbach, M., & Kumar, V. (2007). Introduction to Data Mining (Pearson International Edition), Addison-Wesley, New York.

Teknomo, K. (2009). Hierarchical Clustering Tutorial, Retrieved January 18, 2012, from http://people.revoledu.com/kardi/tutorial/clustering

The MathWorks. (2012). Tools for image processing. Retrieved January 18, 2012, from http://www.mathworks.com/help/toolbox/images/ref/bsemyqp-1.html

APPENDIX

No.	Order	Family	Species	Common Name	Thai Name	Image
1	Anguilliformes	Muraenesocidae	*Muraenesox cinereus*	Conger	Pla Yodjak	
2	Aulopiformes	Synodontidae	*Saurida micropectoralis*	Shortfin lizardfish	Pla Pakkom	
3	Clupeiformes	Clupeidae	*Clupeoides lile*	White sardine	Pla Katakkhao	
4	Clupeiformes	Chirocentridae	*Chirocentrus dorab*	Dorabs	Pla Dab lao	
5	Cypriniformes	Cyprinidae	*Catlocarpio siamensis*	Giant carp	Pla Kraho	
6	Cypriniformes	Cyprinidae	*Probarbus jullieni*	Jullien's golden price carp	Pla Yeesok	

No.	Order	Family	Species	Common Name	Thai Name	Image
7	Cypriniformes	Cyprinidae	*Puntius gonionotus*	Barb	Pla Tapienkhaw	
8	Cypriniformes	Cyprinidae	*Puntioplites proctozysron*	Smit's barb	Pla Kra mang	
9	Cypriniformes	Cyprinidae	*Cyprinus carpio*	Common carp	Pla Nai	
10	Cypriniformes	Cyprinidae	*Ctenopharyngodon idella*	Grass carp	Pla Chao	
11	Cypriniformes	Cyprinidae	*Hypophthalmichthys molitrix*	Silver carp	Pla Leng	
12	Cypriniformes	Cyprinidae	*Labeo rohita*	Rohu	Pla Yeesoktad	

No.	Order	Family	Species	Common Name	Thai Name	Image
13	Gonorynchiformes	Chanidae	*Chanos chanos*	Milk fish	Pla Dok mai	
14	Mugiliformes	Mugilidae	*Mugil vaigiensis*	Grey mullet	Pla Krabok	
15	Osteoglossiformes	Notopteridae	*Notopterrus chitala*	Featherback	Pla Krai	
16	Osteoglossiformes	Notopteridae	*Notopterus notopterus*	Bronze featherback	Pla Salad	
17	Perciformes	Channidae	*Channa (Ophicephalus) striatus*	Striped snake head	Pla Chon	
18	Perciformes	Channidae	*Chana (Ophicephalus) micropeltes*	Giant snake head	Pla Chado	

No.	Order	Family	Species	Common Name	Thai Name	Image
19	Perciformes	Cichlidae	*Oreochromis nilooticus*	Nile tilapia	Pla Nil	
20	Perciformes	Anabantidae	*Anabas testudinus*	Climbing perch	Pla Mor	
21	Perciformes	Eleotridae	*Oxyeleotris marmoratus*	-	Pla Buzai	
22	Perciformes	Sphyraenidae	*Sphyraena jello*	Barracuda	Pla Sakdam	
23	Perciformes	Polynemidae	*Eleutheronema tetradactylum*	Fourfinger threadfin	Pla Kurao	
24	Perciformes	Lates perches	*Lates calcarifer*	Giant sea perch	Pla Kapongkeo	

No.	Order	Family	Species	Common Name	Thai Name	Image
25	Perciformes	Serranidae	*Epinephelus areolatus*	Areolated grouper	Pla Karanghangtud	
26	Perciformes	Serranidae	*Plectropomus maculatus*	Blue spot grouper	Pla Kudsalak	
27	Perciformes	Sillaginidae	*Sillago sihama*	Silver sillago	Pla Chonsaikaeo	
28	Perciformes	Lactariidae	*Lactarius lactarius*	Trvally	Pla Bikanun	
29	Perciformes	Carangidae	*Alectis indica*	Threadfinned trevally	Pla Shomngam	
30	Perciformes	Carangidae	*Caranx ignobilis*	Giant trevally	Pla Sikunphueak	

No.	Order	Family	Species	Common Name	Thai Name	Image
31	Perciformes	Carangidae	*Caranx boops*	Oxeye scad	Pla Sikunthong	
32	Perciformes	Carangidae	*Caranx leptolepis*	Yellow stripe trevally	Pla Sikunkanglieng	
33	Perciformes	Carangidae	*Caraangoides armatus*	Longfin trevally	Pla Takong	
34	Perciformes	Carangidae	*Seriolina nigrofasciata*	Black-banded trevally	Pla Sumlee	
35	Perciformes	Carangidae	*Scomberoides lysan*	Yellow greenfish	Pla Chaliap	
36	Perciformes	Carangidae	*Scomberoides tol*	Talang leatherskin	Pla Sisiad	

No.	Order	Family	Species	Common Name	Thai Name	Image
37	Perciformes	Carangidae	*Megalaspis cordyla*	Hardtail scad	Pla Kakai	
38	Perciformes	Carangidae	*Decapterus russelli*	Round scad	Pla Thu	
39	Perciformes	Rachycentridae	*Rachycentron canadus*	Black king fish	Pla Chontaley	
40	Perciformes	Coryphaenidae	*Coryphaena hippurus*	Common dolphin	Pla Etomon	
41	Perciformes	Lutjanide	*Asrion virescens*	Green jobfish	Pla Kapongkeaw	
42	Perciformes	Lutjanide	*Lutjanus lunulatus*	Bigeye snapper	Pla Kapongkanglieng	

No.	Order	Family	Species	Common Name	Thai Name	Image
43	Perciformes	Lutjanide	*Lutjanus vitta*	Brownstrip red snapper	Pla Kanglueng	
44	Perciformes	Lutjanide	*Lutjanus sebae*	Emperor red snapper	Pla Kapongdaengthabdeng	
45	Perciformes	Lutjanide	*Lutjianus malabaicus*	Malabar red snapper	Pla Daeng	
46	Perciformes	Lutjanide	*Lutjanus lutjanus*	Olivestriprd snapper	Pla Kapong	
47	Perciformes	Lutjanide	*Lutjanus russellii*	Russell's snapper	Pla Kangpan	
48	Perciformes	Lutjanide	*Nemipterus hexodon*	Ornate threadfin bream	Pla Unggolee	

No.	Order	Family	Species	Common Name	Thai Name	Image
49	Perciformes	Lutjanide	*Nemipterus japonicus*	Japanese threadfin bream	Pla Saidaenghangyao	
50	Perciformes	Lutjanide	*Nemipterus peronii*	Notched treadfin bream	Pla Saidaengkadong	
51	Perciformes	Leiognathidae	*Secutor insidiator*	Pugnose ponyfish	Pla Panpangmoo	
52	Perciformes	Leiognathidae	*Eubleekeria splendens*	Splendid ponyfish	Pla Paenkrasouy	
53	Perciformes	Leiognathidae	*Equulites elongatus*	Pony fish	Pla Paen	
54	Perciformes	Pomacentridae	*Abudefduf vaigiensis*	Indo-Pacific sergeant	Pla Trakrup	

No.	Order	Family	Species	Common Name	Thai Name	Image
55	Perciformes	Pomatomidae	*Pomatomus saltatrix*	Bluefish	Pla Numngen	
56	Perciformes	Sciaenidae	*Otolithes ruber*	Tigertooth croaker	Pla Juadteinkeiw	
57	Perciformes	Sciaenidae	*Pennahia anea*	Bigeye croaker	Pla Juadkrebtao	
58	Perciformes	Sciaenidae	*Nibea soldado*	Pawak croaker	Pla Juad	
59	Perciformes	Sciaenidae	*Boesemania microlepis*	Croaker	Pla Hangkew	
60	Perciformes	Lethrinidae	*Lethrinus nebulosus*	Pig - face bream	Pla Moosee	

No.	Order	Family	Species	Common Name	Thai Name	Image
61	Perciformes	Mullidae	*Upeneus moluccensis*	Goldband goatfish	Pla Paithong	
62	Perciformes	Mullidae	*Upeneus sundaicus*	Ochrebanded goatfish	Pla Pailay	
63	Perciformes	Mullidae	*Parupeneus heptacanthus*	Cinnabar goatfish	Pla Paithongleing	
64	Perciformes	Siganidae	*Siganus guttatus*	Goldlined spinefoot	Pla Salidhinjudleing	
65	Perciformes	Siganidae	*Siganus punctatus*	Goldspotted spinefoot	Pla Baikanoonleang	
66	Perciformes	Siganidae	*Siganus canaliculatus*	White-spotted spinefoot	Pla Salidhinjudkaow	

No.	Order	Family	Species	Common Name	Thai Name	Image
67	Perciformes	Siganidae	*Siganus javus*	Streaked spinefoot	Pla Salidtab	
68	Perciformes	Trichiuridae	*Trichiurus haumela*	Hairtail	Pla Dapngen	
69	Perciformes	Scombridae	*Rastrelliger brachysoma*	Short mackerel	Pla Thumong	
70	Perciformes	Scombridae	*Rastrelliger kanagurta*	Indian mackerel	Pla Lung	
71	Perciformes	Scombridae	*Scomberomorus commerson*	Narrow-barred Spanish mackerel	Pla Inseebung	
72	Perciformes	Scombridae	*Scomberomorus guttatus*	Indo-Pacific king mackerel	Pla Inseekhaotok	

No.	Order	Family	Species	Common Name	Thai Name	Image
73	Perciformes	Scombridae	*Scomberomorus lineolatus*	Streaked seerfish	Pla Thangdinsor	
74	Perciformes	Scombridae	*Auxis rochei rochei*	Bullet tuna	Pla Thukak	
75	Perciformes	Scombridae	*Euthynnus affinis*	Eastern little tuna	Pla Oklab	
76	Perciformes	Scombridae	*Katsuwonus pelamis*	Skipjack tuna	Pla Otab	
77	Perciformes	Istiophoridae	*Istiophorus platypterus*	Banana sailfish	Pla Katongtangkluy	
78	Perciformes	Stromateidae	*Pampus argenteus*	Silver pomfret	Pla Jalamedkaow	

No.	Order	Family	Species	Common Name	Thai Name	Image
79	Perciformes	Stromateidae	*Pampus chinensis*	Chinese silver pomfret	Pla Jalamedtao	
80	Perciformes	Stromateidae	*Parastromateus niger*	Black pomfret	Pla Jalameddum	
81	Perciformes	Xiphiidae	*Xiphias gladius*	Swordfish	Pla Katongtangdab	
82	Perciformes	Scombridae	*Auxis thazard*	Frigate mackerel	Pla Okaow	
83	Perciformes	Scombridae	*Thunnus tonggol*	Longtail tuna	Pla Odum	
84	Pleuronectiformes	Psettodidae	*Psettodes erumei*	Indian halibut	Pla Sigdeaoy	

No.	Order	Family	Species	Common Name	Thai Name	Image
85	Pleuronectiformes	Cynoglossidae	*Cynoglossus bilineatus*	Fourlined tonguesole	Pla Yodmoung	
86	Pleuronectiformes	Cynoglossidae	*Cynoglossus puncticeps*	Speckled tonguesole	Pla Changchun	
87	Pleuronectiformes	Cynoglossidae	*Cynoglossus lingua*	Long tongursole	Pla Linma	
88	Scorpaeniformes	Platycephalidae	*Inegocia japonica*	Japanese flathead	Pla Huaban	
89	Scorpaeniformes	Platycephalidae	*Platycephalus indicus*	Bartail flathead	Pla Hangkway	
90	Siluriformes	Siluridae	*Wallagonia attu*	Wallago	Pla Khao	

No.	Order	Family	Species	Common Name	Thai Name	Image
91	Siluriformes	Siluridae	*Kryptopterus bleekeri*	Sheath fish	Pla Deng	
92	Siluriformes	Bagridae	*Mystus nemurus*	Catfish	Pla Kot	
93	Siluriformes	Pangasiidae	*Pangasianodon gigas*	Makong giant catfish	Pla Buk	
94	Siluriformes	Pangasiidae	*Pangasius larnaudii*	Catfiish	Pla Thepho	
95	Siluriformes	Clariidae	*Clarias macrocephalus*	Gunther's walking catfish	Pla Dukuey	
96	Siluriformes	Clariidae	*Clarias batrachus*	batrachian walking catfish	Pla Dukdan	

No.	Order	Family	Species	Common Name	Thai Name	Image
97	Siluriformes	Ariidae	*Arius thalassinus*	Giant catfish	Pla Riwkiw	
98	Siluriformes	Plotosidae	*Plotosus lineatus*	striped sea catfish	Pla Pinkaeo	
99	Synbranchiformes	Synbranchidae	*Monopterus albus*	Swamp eel	Pla Lai	
100	Tetraodontiformes	Molidae	*Mola mola*	Ocean sunfish	Pla Praatid	

BIOGRAPHY

NAME	Miss Janket Pengpanit
DATE OF BIRTH	29 October 1983
PLACE OF BIRTH	Saraburi, Thailand
INSTITUTIONS ATTENDED	Kasetsart University, 2003-2006 Bachelor of Science (Fisheries) Mahidol University, 2007-2012 Master of Science (Technology of Information System Management)
RESEARCH GRANTS	-
HOME ADDRESS	64/91 Prayasurant 12 Prayasurant Rd. Bangchan Klongsamwa Bangkok 10510
E-MAIL ADDRESS	noon.accurate@gmail.com
EMPLOYMENT ADDRESS	Accurate Engineering Co, Ltd 544 Srinakarin Rd. Suanluang Suanluang Bangkok 10250
PUBLICATION / PRESENTATION	-