Image-Based Search Engine
For Art Exhibition Gallery
(ImBa SEA Exhibition/Gallery)

By

Mohamad Nasrie bin Ajiji

Dissertation submitted in partial fulfillment of
the requirements for the
Bachelor of Technology (Hons)
(Information and Communication Technology)

SEPT 2012

UniversitiTeknologi PETRONAS
Bandar Seri Iskandar,
31750 Tronoh
Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

Image-Based Search Engine For Art Exhibition Gallery

By

Mohamad Nasrie bin Ajiji

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

Approved by,

_____

(Dr Vivian Yong Suet Peng)

UNIVERSITI TEKNOLOGI PETRONAS
TRONOH, PERAK
SEPT 2012

CERTIFICATION OF ORIGINALITY

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 reference and acknowledgements, and that the original work contained herein has not been undertaken or done by unspecified sources or persons.

_____

(MOHAMAD NASRIE BIN AJIJI)

ABSTRACT

The image-based search engine for art exhibition/gallery (ImBa SEA) is like any search engine but will provide a better functionality. The ImBa SEA is specially design for art gallery where lies thousands of arts displayed on the exhibition. With the help of the ImBa SEA, the user can upload image that they have snapped into the system to retrieve the information about the art itself. Some features are very difficult to describe with text, some special textures and complex shapes cannot be clearly represented by alphanumeric inputs. The arts need to be digitalized and stored in the gallery's database. Based on the existing way of accessing information for art gallery, instead of directly using the image as a 'keyword' to retrieve information, each art have its own name. The name is stored in the art gallery database, with the relevant information. However, this method takes some time because some arts have the similar name or a long name. This decreases the accuracy of the search engine to retrieve information of the art wish to be accessed. With ImBa SEA, the user directly uses the image to retrieve the information on the given art. Each art will have its own unique id to ease the process of retrieving information. The image is uploaded into the search engine, and then it will process the image to look for similarity of the image with the images stored in the database. Once the similar image is found in the database, the unique id of the art is used to retrieve the information of the image. Basically,ImBa SEA compares the two images pixel by pixel to look for similarities. This technique will surely provide better efficiency and accuracy on information retrieval. This paper will discuss everything from the problem faced, and scope concentrated to achieve the objectives of the project which are discussed in Chapter 1. The system or technologies related or similar with the ImBa SEA are discussed in Chapter 2: Literature Review. Also, the research methodology of the project is discussed later in Chapter 3 together with the development methodology. The basic structure is also discussed in this paper to give a clear view on how the system works. The future plans on the ImBa SEA are discussed at the end of this paper.

ACKNOWLEDGEMENT

First and foremost, I would like to take this opportunity to thank God for giving me the opportunity to finish this project. I would like to express his greatest gratitude and appreciation to project supervisor, Dr Vivian Yong Suet Peng, who had continuously given me constructive comments, advices, and suggestions which have guided the project towards its successful final outcome.

Next, I would like to acknowledge Universiti Teknologi PETRONAS (UTP) especially the committee of Final Year Project of Computer Information Sciences (CIS) department for excellent organization and management of this course.

Last but not least I would also like to express the participant of survey questions, and not to forget my parents for their endless support throughout the project. They have given me strength, motivation and determination in all my undertaking.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES

ABBREVIATIONS AND NOMENCLATURES

ImBA SEA	Image-Based Search Engine for Art Exhibition/Gallery
PHP	Hypertext Pre-processor
SQL	Structured Query Language
CBIR	Content-based image retrieval
CLEF	Cross-Language Evaluation Forum
IRMA	Image Retrieval in Medical Applications

CHAPTER 1
INTRODUCTION

1.1 Background of Study

In the twentieth century, images have grown in numbers, availability and its importance. Nowadays, we can relate images in diverse fields such advertising, design, art, education, entertainment, medicine and many more. However, this vast growth of visual content creation has not been matched with the technologies existed to cater efficient image analysis and retrieval.

People currently looks for image using search engine which grabs information based on the user's input in textual form. A problem with the current method is that, sometimes, the user did not know what the right keywords that they should enter into the search engine to get the desired result. These have been the case for art gallery. The people who come to the art gallery are interested to know more information on the art, however, because of its complexity of the arts to describe the art using text descriptions, the information retrieve may be wrong. Some features are very difficult to describe with text, some special textures and complex shapes cannot be clearly represented.

This is why this paper proposed a slightly different type of search engine for the art gallery where the user can upload the image into the search engine and the search engine will return with the information related to the image or any other related images that are similar to the image uploaded. Each art are digitalized into the gallery's database with its own unique id. This unique id is then used to retrieve the information of the image uploaded.

1.2 Problem Statement

The existing search engines which are widely used today work when the user input some textual keywords to look for images. With the rapid growth of digitalizing visual content, technological advancement in the efficient visual analysis and retrieval is a need. However, the problem is some image is associated with keywords that are not related to the image at all which will brings error to the searched result. With the existing technology, it searched for the image using the keywords, hence, there's a probability that the result of the image may not necessarily relates to the keywords that have been submitted by the user. Some people simply named their image filename as anything that they desire giving false retrieval when it comes to alphanumeric search. Apart from that, the image metadata search (description of image) is not practical when it involves large database. Metadata requires the user to manually describe the images that is in the database [12]. So, there is a need for a more efficient image analysis and retrieval.

1.3 Objectives of study

The objectives of this project are to:

- Research on the methodologies that have been applied to achieve the image retrieval system
- Propose an algorithm for image similarity matching and information retrieval.
- Develop image-based search engine for art exhibition/gallery. The system will have the following functions:
    - Retrieve information based on user input in textual form like the current search engine
    - Retrieve information of the image based on user input based on what the user's image upload
    - Add, edit and delete function for admin for easy update of the system's database.

1.4 Scope and Limitation Of Study

Based on the project title, Image-Based Search Engine for Art Exhibition/Gallery (ImBa SEA), it is clear that this system will be in the form of a web-based project which will be the main focus for this project. Apart from that, the scope of this project is studying the aspects of image search and web-based application. The proposed project is to seek ways to further improve the existing search engine for art gallery, instead of just entering some keywords to search for information, the search engine will also provide the functionality to upload image to retrieve information from the search engine. There are certain limitations on this image-based search engine where it can only be used for a certain specific art gallery where the arts are digitalized into their database for easy information retrieval. Apart from that, as of now, the image being compare can only be compare if the image uploaded is not been skewed to the right of left. This means that the image must be at the right position in order to compare the image with the one in the database. The technologies focuses on this project are CBIR and also pattern recognition that are for the image retrieval part of the ImBa SEA.

1.5 Significance of Project

The success of this project could be a stepping stone towards a new way for image retrieval. The current way which is the metadata-based search system requires the human to personally describe the image in the database. Thus a system that can filter images based on their content would provide better indexing and return more accurate results [13]. This project is highly related to CBIR. The potential of using CBIR includes architectural and engineering design, art collections, crime prevention, intellectual property, medical diagnosis, military, photograph archives and many more. There is so much benefits that can be achieved through CBIR. In this project, the focus would be how CBIR can help in achieving more efficient retrieval on art gallery collections.

CHAPTER 2
LITERATURE REVIEW

2.1 Content-based Image Retrieval for Scientific Literature Access

This journal was written by T. M Deserno, S. Antani, and L. Rodney Long from the Department of Medical Informatics, Aachen University of Technology, Germany. The objective of this article is to estimate the benefits of using content-based image retrieval (CBIR) on article figures to augment traditional access to articles [3].

In this article it compares the traditional ways of accessing the articles by alphanumerical search on title, author or abstract and may disregard numerous figures. Also the team had conducted experiments where they are making a quantitative estimate by projecting from data from the Cross-Language Evaluation Forum (Image-CLEF) campaigns, and qualitatively validate it through experiments using the Image Retrieval in Medical Applications (IRMA) project. The result of these experiments shows the predicted accuracy up to 97.08% of article retrieval. They had conducted the experiments due to the increasing numbers of electronic articles been published and due to its limited access of searching for the right articles. Some problem is that the figures in the articles are not easily described by words which make it difficult for alphanumeric search like the traditional ways.

To conclude this article, the researchers are convince that CBIR have high potential impact in medical literature search and retrieval due its high accuracy of retrieving information.

2.2 Design and evaluation of algorithms for image retrieval by spatial similarity

This journal discuss about image retrieval using spatial similarity. It was written by Venkat N. Gudivada of Ohio University and Vijay V. Raghavan of University of Southwest Louisiana. Spatial similarity can be seen as a subset of similarity and all the entities being compared to each other have spatial components [4].

In image database applications, similarity-based retrieval of images is an important task. Users tend to request retrieving those images in the database based on the similarity of the query image. In this paper, they propose an algorithm to compute the spatial similarity between two symbolic images. A natural image or the original image cannot be stored in database systems as it is for retrieval. Therefore, the natural image is converted into symbolic image as a logical representation of the original image. The proposed algorithm can deal with translation, scale, and rotational variances in images. The characteristics of the proposed algorithm are compared with those of the previously available algorithms using experimental set of images. The comparison verified that the algorithm provides a rank ordering of images that constantly matches with the expert's expected rank ordering and also in a more efficient manner.

In conclusion, the algorithm proposed may open a new dimension to image retrieval based on the similarity pattern of the images. Maybe the alteration of the algorithm will innovate new ideas on pattern recognition on image and will surely benefit in the future for image based application technology.

2.3 Partial Similarity of Objects

This Journal is written by Alexander M. Bronstein, Michael M. Bronstein, Alfred M. Bruckstein and Ron Kimmel for the International Journal of Computer Vision. This journal discuss about how partial similarity are used to retrieve similar image. It is related to the project because it have the same concept where the system using a partial similarity to retrieve image as it compare the image uploaded by the user with the existing image in the database. After comparing it will retrieve the required information based on the image in the database. Similarity is one of the most important abstract concepts in human perception of the world. In computer vision, numerous applications deal with comparing objects observed in a scene with some a priori known patterns. Often, it happens that while two objects are not similar, they have large similar parts, that is, they are partially similar. In this journal, they present a novel approach to quantify partial similarity using the notion of Pareto optimality [5].

The main purpose of this paper, stated briefly, is to provide a quantitative interpretation to what is meant by “similar” and “large”, and derive a consistent relation between these terms. It allows us to formulate a computationally tractable problem of finding the largest most similar parts. In this approach, they use the formalism of Pareto optimality and multicriterion optimization. While well-known in fields like information theory and economics, these tools have been explored to a lesser extent in the computer vision and pattern recognition community.

In conclusion, this paper used the Pareto optimality to differentiate two objects that are very similar but different in nature. For example, a centaur and a horse, the lower part of these two may looks very similar but in nature, it is two different things. By applying Pareto optimality, it increases the partial similarity between the two objects.

2.4 The QBIC project: Querying images by content using colour, texture and shape

This paper is written by W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman and D. Petkovic. The past few years have seen many advanced techniques evolving in Content-Based Image Retrieval. Applications like art gallery, medicine, entertainment, education, manufacturing, and many more make use of massive numbers of visual data in the form of images. This envisions the need for fast and effective retrieval mechanisms in an efficient manner. A major approach directed towards achieving this goal is the use of low level visual features of the image data to segment, index and retrieve relevant images from the image database [6]. Recent CBIR systems based on features like colour, shape, texture, spatial layout, and object motion. It is shown that colour is the most dominant and distinguishing one in almost all applications.

Most CBIR systems use low-level visual features for representation and retrieval of images. Generally such methods suffer from the problems of high-dimensionality leading to more computational time and inefficient indexing and retrieval performance. This paper focuses on a low-dimensional colour and shape based indexing technique for achieving efficient and effective retrieval performance. They propose a combined index using colour and shape features. A new shape similarity measure is proposed which is shown to be more effective. Images are indexed by dominant colour regions and similar images form an image cluster stored in a hash structure. Each region within an image is further indexed by a region-based shape index. The shape index is invariant to translation, rotation and scaling. A Java based query engine supporting query-by-example is built to retrieve images by colour and shape. The retrieval performance is studied and compared with a region-based shape indexing scheme.

In conclusion, by using the low-level visual features for representation, the efficiency and the effectiveness of the accuracy on retrieval can be increase. This is because the computational time taken to process the image will be reduce, hence faster information retrieval which result in better performance for the system.

2.5 Content-Based Image Retrieval Systems

This paper is written by Venkat N. Gudivada from Ohio University and Vijay V. Raghavan from University of Southwestern Louisiana. There is a lot of possible usage of this CBIR technology. In this paper, it discuss about the possibility of CBIR technology of changing the way we use the creation and distribution of digital visual content. Some application areas in which CBIR is a principal activity are numerous and diverse are [11]:

- Art galleries and museum management
- Architectural and engineering design
- Interior design
- Remote sensing and management of earth resources
- Geographic information systems
- Scientific database management
- Weather forecasting
- Retailing
- Fabric and fashion design
- Trademark and copyright database management
- Law enforcement and criminal investigation
- Picture archiving and communication systems

In this paper also discuss the previous approaches to content-based retrieval. There are two approaches. First approach, the image contents are being modelled as a set of attributes which are extracted manually and managed within the conventional database management systems. To retrieve the information, queries are used. This approach is known as attribute-based retrieval. This first approach is not really using the content of the image itself to retrieve the information. It is more like metadata search like the one we are using right now like Google, Bing, Yahoo and so on. The second approach depends on an integrated feature-extraction or object-recognition subsystem to overcome the limitations of attribute-based retrieval. The second approach have a subsystem which carries the task of feature-extraction and object-recognition task whenever the image is

inserted into the database. However, this approach is said to be highly expensive and quite difficult to carry out.

The point of this paper is to get the idea on some important features needed to carry out the ImBa SEA project. Regardless of which approach is used, the generic query classes which facilitate CBIR through retrieving by:

- Colour
- Texture
- Sketch
- Shape
- Volume
- Spatial constraints
- Browsing
- Pixel comparison

The best way to retrieve image by comparing it with another image would be by texture and colour. These however need a high level knowledge of understanding in computer vision. In general, by comparing the images by its texture, the probability of getting it by similarity is higher, and by adding colour, it can confirms the similarity of the two images being compared. Pixel by pixel comparison is a good way to compare two images to test whether it have the similarity. Like for example, to compare two images whether it is similar, after comparing it by pixel, get the percentage of how many it does similar, a higher percentage would mean that the two images are very similar.

CHAPTER 3
METHODOLOGY

3.1 Introduction

In this chapter, it will explain the methodology use to carry out the research and also the development section later on. In 3.2, the research methodology will be explained in order to describe on the approach on how the research are carry out in order to solve the problems that are discussed earlier in chapter 1. This research methodology will give a clear view on how the objectives are set to solve the problems. In 3.3, the development methodology is explained in term of how the system is going to be develop later. This will give clear view on how the developments of the system are going to be carry out. In 3.4, the tools used to develop the system are discussed. This chapter also includes the key milestones, and the Gantt chart for the project.

3.2 Research Methodology

In the research methodology, there are six stages altogether which are Project Planning, Data Gathering and Analysis, Research on any Existing Similar Systems, Drafting the Main Components of the System, Develop System Basic Architecture, and Sketching the Interface of the System.

3.2.1 1st Stage: Project Planning

This is the stage where the problem related to the project is identified and the significance of the study is determined. The objective and also scope of study are outlined and the feasibility of the project work is ensured to be within the time frame given. The solution to the problem statement is studied and the types of system to be develop and tools used for developing the system are also identified through literature reviews. The objectives of the project must be able to solve the problem identified.

3.2.2 2nd Stage: Data Gathering and Analysis

Series of studies had been performed to gain further knowledge on information retrieval using image. Apart from that, the knowledge to retrieve similar image based on the image uploaded are needed in order to make this project work. A better understanding on how information retrieval from the database is needed so a lot of readings had been done. People have the difficulty to retrieve information when it comes to image. Image sometimes are indescribable where using alphanumeric input does not guarantee accuracy on the information retrieved. Finding the right algorithm to match the similarity of the image uploaded with the image in the database is the crucial part of the project. Using the wrong algorithm may lead to undesirable output on the information retrieval end.

3.2.3 3rd Stage: Research on any Existing Similar Systems

Next is the study performed to check whether there are any similar system already existed. The main objective of doing research on similar existing system is to know how it works, what concept is being applied in the system, what is being computed by the system and how the system helps solving the problem. By doing extensive research on similar existing systems, a better understanding on the system architecture can be comprehend nicely. Apart from that, this research opens up opportunities to enhance the existing system to make it better or simply avoiding the problems encountered by the existing systems.

3.2.4 Drafting the Main Components of the System

After the research on similar existing systems, the next step needed is to identify what will be the main component that made up the system to be developed. As the intention of this system is create a search engine based on image uploaded for the art gallery, a few main functionalities need to be identified so that the development of the system later on will be in a systematic manner. The main components of this system are:

- Uploading image architecture
- Image processing based on the similarity
- Information retrieval based on the unique id on the image

3.2.5 Develop System Basic Architecture

Once knowing how the similar existing system works, the next phase to develop the architecture on how the system will works. This will give the clear picture and understanding on how the system will operate and to avoid developing a system that does not solving the problem it intended to solve. **Figure 3.1** below is the basic system architecture of how these systems will works.

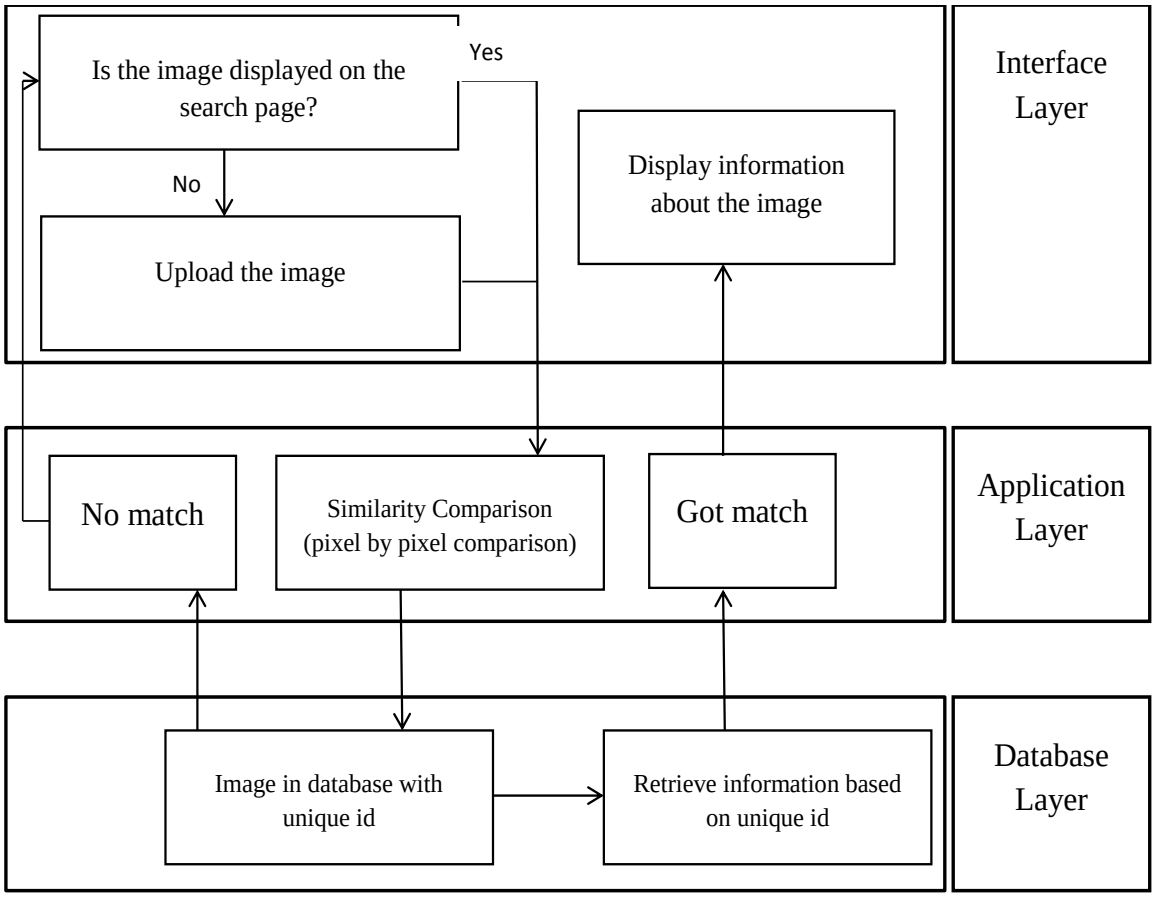


Figure 3.1: System basic architecture

3.2.6 Sketching the Interface of the System

Once all the functions performed by the system are identified, the last step is to design the interface of the system. Developing the interface of the system will make it easier for the system to be developed during the development phase.

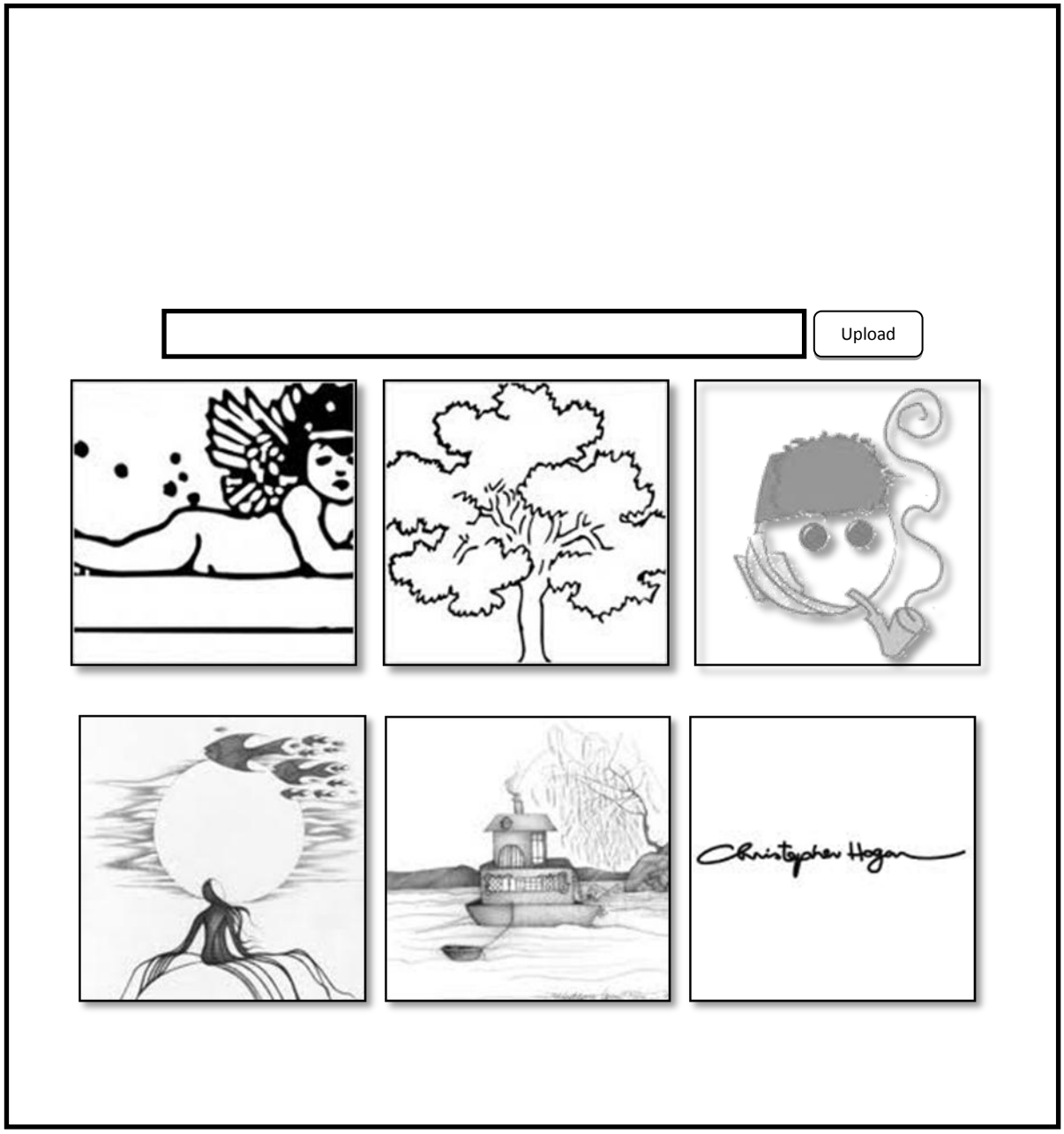


Figure 3.2: Proposed Interface for the system

3.3 Flowchart of the System

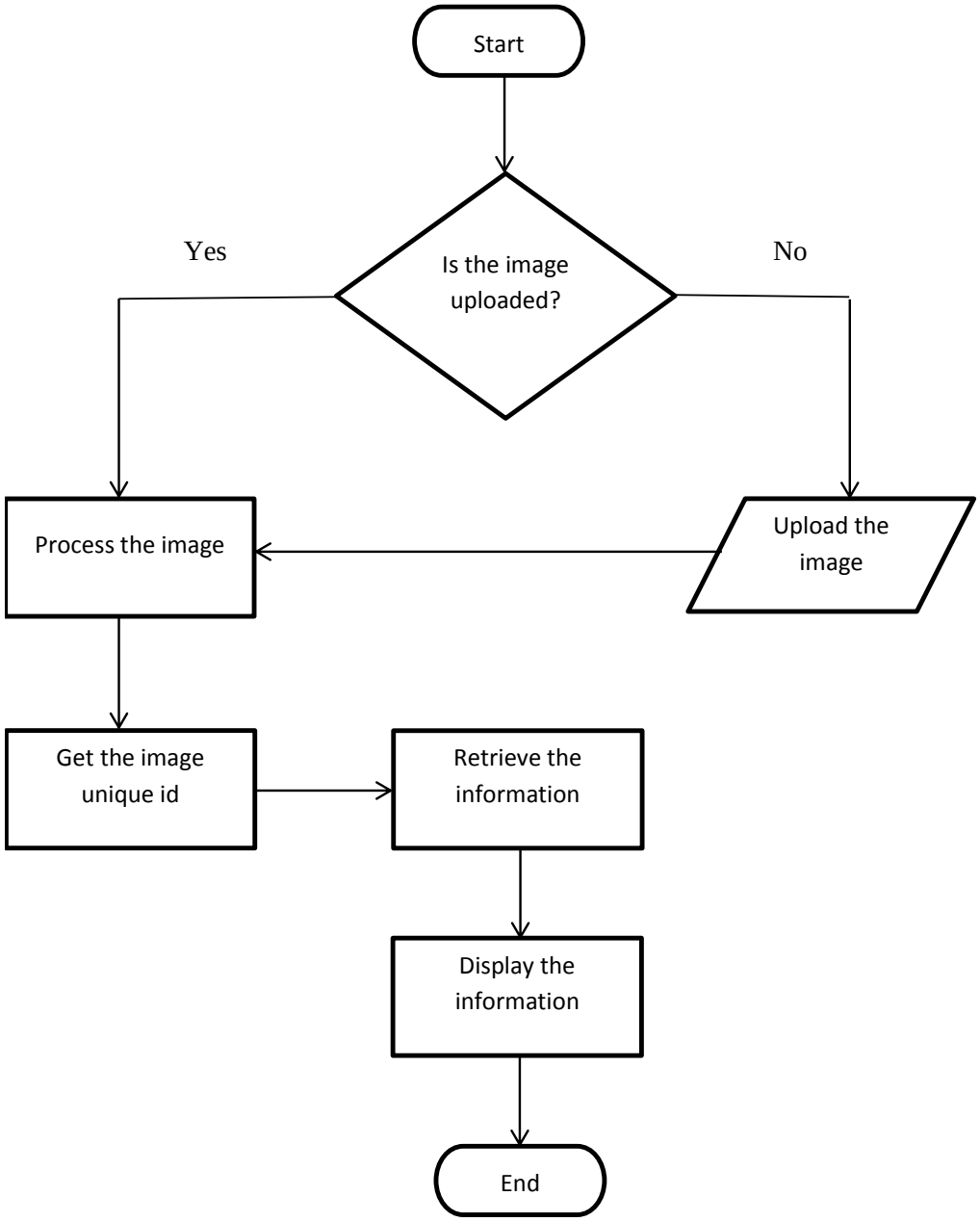


Figure 3.3: Flowchart of the basic system architecture

3.4 Development Methodology

In completing this project, the development methodology to be chosen is Incremental Prototyping. Prototyping is the process of quickly putting together a working model in order to test various aspects of a design, illustrate ideas or features & gather early user feedback. It is believed to reduce project risks and cost. Incremental Prototyping is a prototype technique where the final product is built in several prototypes. Some parts of the functionalities are produced before others in this incremental prototyping process. This enables us to view the early visibility of some key aspects. Also, some or all the most critical functions can be done early for this prototyping process. The advantage of prototyping is that the development provides the developers with insight to how the system should be structured.

The benefits of using this methodology is it allows any changes to be made during the development phase if there is needs to review and recheck at any other phase of project development. This is important as it provides flexibility throughout completing the project such as debugging process. In order to make it clearer on the process, the whole project will be divided into four main phase as illustrated in Figure 3.1:

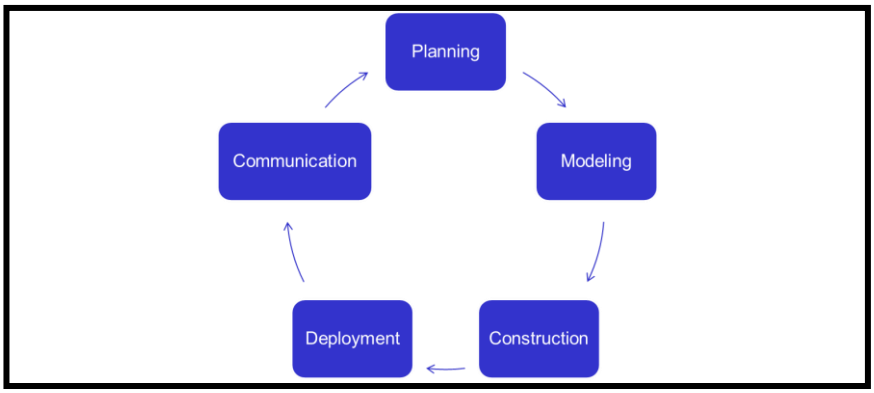


Figure 3.3: Incremental Prototyping Development Cycle

3.5 Tool

The design of the background will be created using Adobe Photoshop just to give a better look on the search engine. The programming language intended to use is PHP as the programming language because of its security to the system. The database for the search engine will primarily use SQL database. The platform for the SQL would be MySQL. XAMPP will be used as the server to test for its functionality.

3.6 Algorithm Used

3.6.1 Grayscale image algorithm

There are three known algorithm to grayscale an image which are the lightness method, the average method and the luminosity method [14]. The lightness method averages the most prominent and least prominent colours. The average method simply averages the values. The luminosity method is a more sophisticated version of the average method. It also averages the values, but it forms a weighted average to account for human perception. The weights are based on studies of the human eye which have shown that the sensitivity to red, green, and blue differs, therefore the luminance has different scaling factors for each colour channel. We are more sensitive to green than other colours, so green is weighted most heavily. R refers to red, G refers to green and B refers to blue. This algorithm is important before comparing the two images. Both images must be convert into grayscale first before comparing them so that both picture can be compare easily without the color affecting its differences. In other word, after converting it into grayscale, we can focus on the differences between images' shape, edge, texture and not the colour. Sometimes, there is a possibility that the two images compared with have two different colours. Therefore, converting it into grayscale, eliminate the comparison between its colours.

Below are the formulas of the said grayscalling method:

Lightness method	:	(max(R, G, B) + min(R, G, B)) / 2
Average method	:	(R + G + B) / 3
Luminosity method	:	0.212671 R + 0.715160 G + 0.072169 B

Equation 3.1: Formula of the grayscalling method

The method that have been chosen to grayscale the images are the luminosity method. The numbers in the luminosity method equation above refers to the weightage that are being studied by (Porle, Chekima, Wong, & Sainarayana, 2009) related to the sensitivity of colours to the human eyes.

3.6.2 Pixel-based similarity check algorithm

After the grayscale of images are done, the two images are compared pixel by pixel. The algorithm used is based on the pixel-based identity check algorithm. The idea is that, the algorithm goes through the image pixel by pixel and calculates the percentage of its similarity with the image being compared to. The algorithm splits every pixel in three sub-pixels of red, green and blue. Every colour has its own hex value, so, by getting the differences in the hex value of the two images, we then calculate the offset of each pixel then sum up the total to get the percentage of its similarities. This however got some limitation because it can only compare the images if the said images are not skewed or rotated.

3.7 Comparison Method

The concepts on how the image comparison work will be explain in detail in this section. The image uploaded will be compared with the images that are stored in the database by its pixel. Before it is compared, both image will be resize so both image will be on the same size. Then, both images are transformed into a grayscale images using the colour mapping technique. Afterwards, the images are compared using an algorithm which will compare the two images pixel by pixel. The difference between two images is calculated by finding the difference between each pixel in each image. Then, after comparing every pixel by pixel, it will calculate the percentage of its similarity. By a certain percentage, let say, by 70% and above, the images are consider to be similar. Similarity percentage of lower than 70% will be rejected. After getting the percentage of similarity by 70% and above, the images with those percentages will be displayed to the user and the user will choose the image that they intend to retrieve its information. After choosing the image, the information of that image will be retrieved from the database. Each image stored in the database has its own unique id that is linked with its relevant information. With this unique id, the information about the said image will be retrieve and displayed to the user.

3.8 Suggested Algorithm for Future Development

A few suggestions is made in order to improve the current prototype of the system. This suggestion can make the system more reliable and hopefully can increase the efficiency of the system in the future. Based on the current prototype, the two images that are being compared must both positioned at the same angle. So, for suggested improvement, if the image uploaded is not the same angle as the image of the one in the database, the image uploaded can be rotated so that both image can be compare correctly. The theory is that by using the PHP function of *imagerotate()* [15]. This function can be used to rotate the image by a certain degree to match the image in the database. In theory, we can use array function to rotate the image from 0^o to 360^o until it matches the image being compared to. Then, compare the image for similarity.

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 Data Gathering and Analysis

Based on the survey questions that have been distributed to random respondents, I have come up with a quantitative analysis to know the reactions of people about the acceptance of my project. Up until 5th August 2012, a total of 48 responses have been recorded.

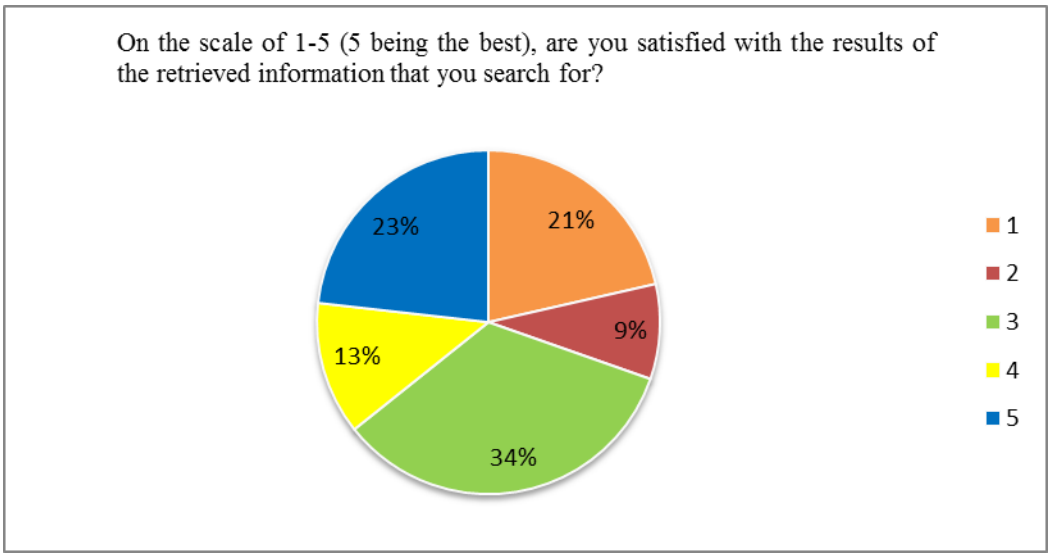


Figure 4.1: Pie Chart for Survey Question 1

Figure 4.1 shows how the people's satisfaction on the results of the retrieved information with the existing search engine that they use. It shows 40% on the scale at average satisfaction on the existing results while quite a number rate the results retrieved are at the worst at 25%. This means, that there is a room for improvement for the current method for the search engine.

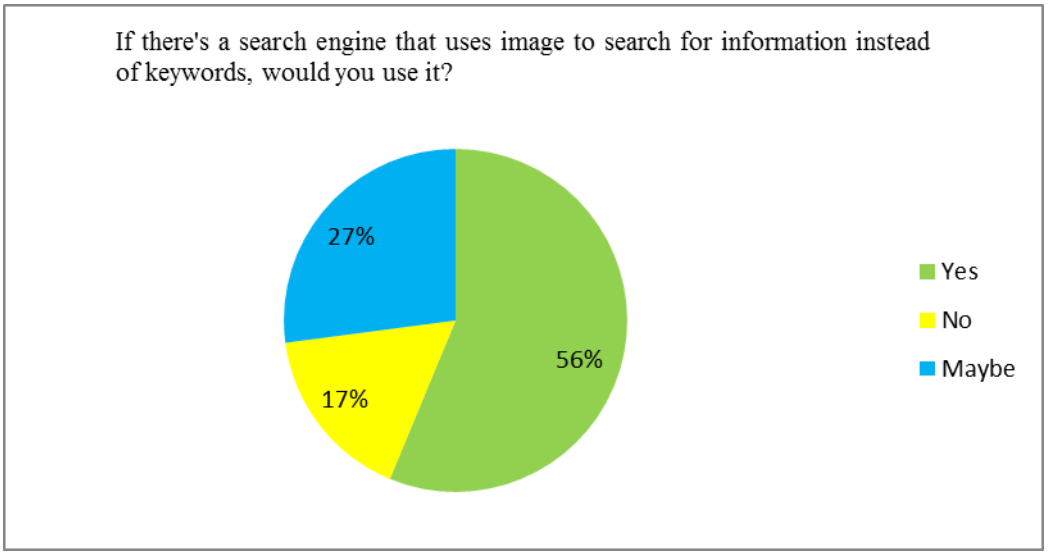


Figure 4.2: Pie Chart for Survey Question 2

Figure 4.2 shows the reactions of the respondents if there is a new way of searching for information are introduced which is using image. Out of 48 respondents, 27 respondents thinks that having a new way of searching information would be great and they are willing to use it. This shows interest on the proposed new way of searching for information. However, quite a number of them are still unsure of the possibility of using image to search for information.

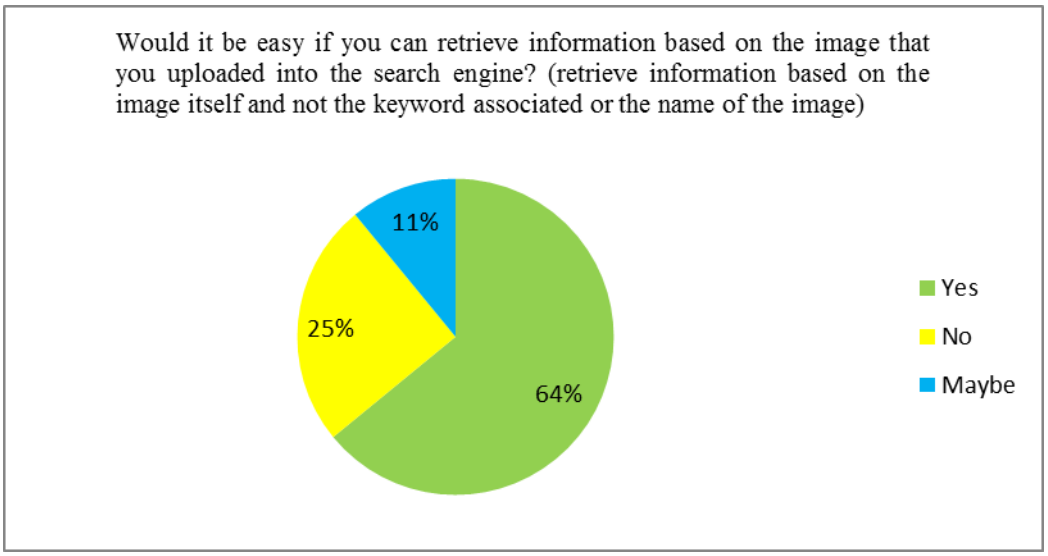


Figure 4.3: Pie Chart for Survey Question 3

Figure 4.3 shows a majority of respondents thinking that it would be easy for them to search for information using image as an improvement to the current search engine. This shows that this technology may assist the user on information retrieval at a more accurate rate.

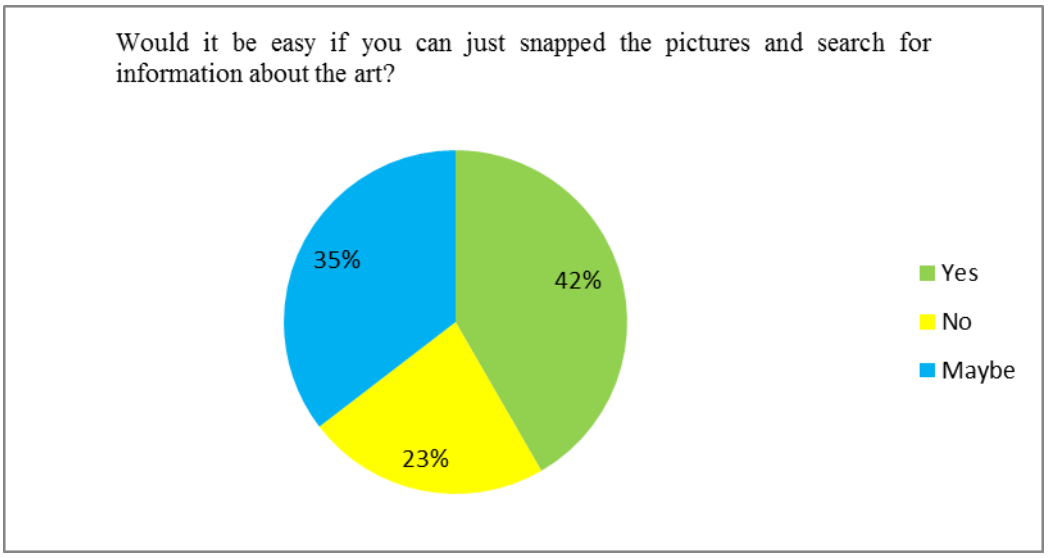


Figure 4.4: Pie Chart for Survey Question 4

Based on Figure 4.4, the respondents are ask whether it is applicable if it is used to search for information on arts that have been snapped. 42% agreed that it would be easy to search for the information with the arts that have been snapped while another 35% are still unsure whether it will bring them any good with the picture that they have snapped. It means that, they are still quite a numbers of respondents that have a feeling that using image is not really an improvement on information retrieval.

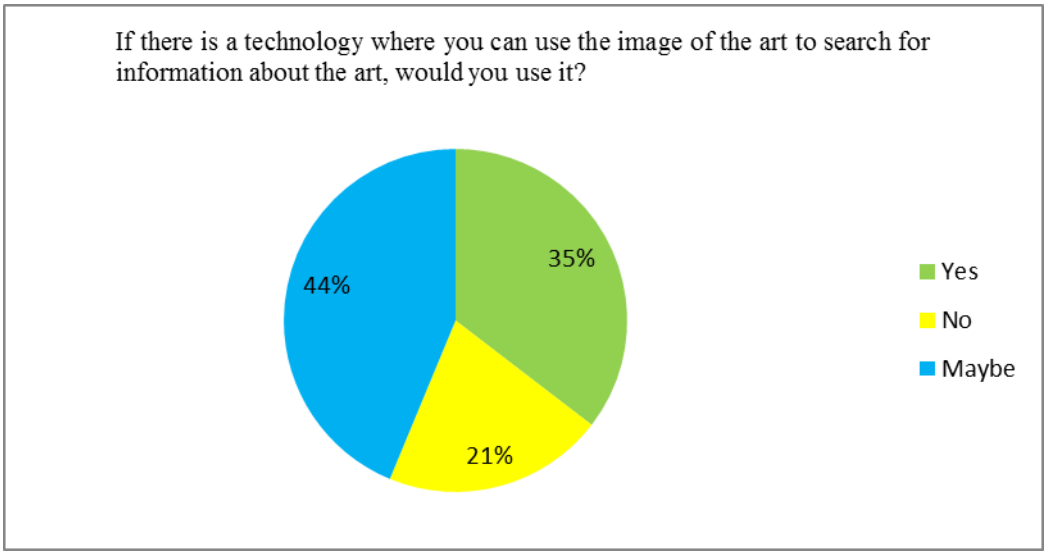


Figure 4.5: Pie Chart for Survey Question 5

Based on the results on Figure 4.5, majority are unsure whether they are going to use this technology if such technology exists. Maybe it is because they have not seen the technology yet to experience it by themselves. Some people prefer to try it first before saying that they will use the technology. Quite a numbers of them are willing to use the technology is it exist, which means that they are people who actually want to use this technology. Hence, the existence of this technology may improve the ability to retrieve information.

4.2 Experiment I

An experiment to test the code had been carried out. The objective of this experiment is to test whether the code can be used to find the similarity of the image that the users upload with the image that is stored in the database. All of the images are compared with the original. In this experiment, the constant variable would be the original image. The images that are compared with the original image are said to be the responding variable or the variable that are changed in order to achieve the objective of the experiment. The image used in the experiment can be referred in Appendix B. Three sets of experiments had been carried out to determine which colormap will determine the best choice to use for the system. The tables below are the result of the experiments.

Table 4.1: Experiment I Results

Table 4.1.1: Using colormap Green>Red>Blue (0.212671 * R + 0.715160 * G + 0.072169 * B)

Responding variable	Image of different dimension	Edited image	Image of different file type	Different image
Percentage of similarity	92.5%	85.75%	99.5%	43%

Table 4.1.2: Using colormap Red>Green>Blue (0.412453 * R + 0.357580 * G +0.180423 * B)

Responding variable	Image of different dimension	Edited image	Image of different file type	Different image
Percentage of similarity	93.75%	86.5%	99.5%	44%

Table 4.1.3: Using colormap Blue>Green>Red (0.019334 * R + 0.119193 * G + 0.0.950227 * B)

Responding variable	Image of different dimension	Edited image	Image of different file type	Different image
Percentage of similarity	90%	80.75%	96.75%	43.25%

Based on the experiments, the best colormap to use is the colormap where the Green weighted value is more than Red and Red weighted value is more than Blue. This is because, the human eyes are more sensitive to green than other colours, so green is weighted most heavily. Also, based on the results, the colormap Green>Red>Blue retrieve the most accurate results.

4.3 Experiment II

Another experiment is carried out to test the code with three different algorithms to use. The objective of this experiment is to determine which of the three algorithms is the best method to use in the comparison mechanism of the system. All of the images are compared with the original. In this experiment, the constant variable would be the original image. The images that are compared with the original image are said to be the responding variable or the variable that are changed in order to achieve the objective of the experiment. The experiment will be carry out three times. The first would be using

the lightness method. The second would be using the average method and third method would be using the luminosity method. The image used in the experiment can be referred in Appendix B.

Table 4.2: Experiment II Results

Responding variable		Image of different dimension	Edited image	Image of different file type	Different image
Percentage of similarity	Lightness	93.75%	85.35%	98%	46%
	Average	94.5%	85.75%	99.5%	46.25%
	Luminosity	92.5%	85.75%	99.5%	43%

Based on the results that are displayed on Table 4.2, we can assume that the best method would be using Luminosity method. This is because, based on the percentage of different image compared to the original image, both Lightness and Average method calculate the similarity by 46% and 46.25% respectively. This is quite a large value that may change the accuracy of the system on comparing two images that are totally different. Therefore, it can be concluded that the most suitable method to use is using the Luminosity.

4.4 Screenshot of the system

There are two parts of the system, one for users and another one for admin. Below are the user interfaces of the system.

Figure 4.11: Main page for User

Figure 4.12: Image being uploaded by user

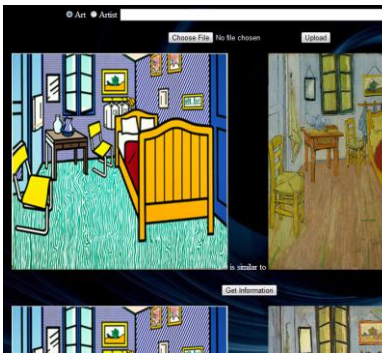


Figure 4.13: Comparison result

Figure 4.14: Art information retrieved

Below are the screenshot for the admin interface. The larger versions of the screenshots of all the system's interface are attached in Appendix C.

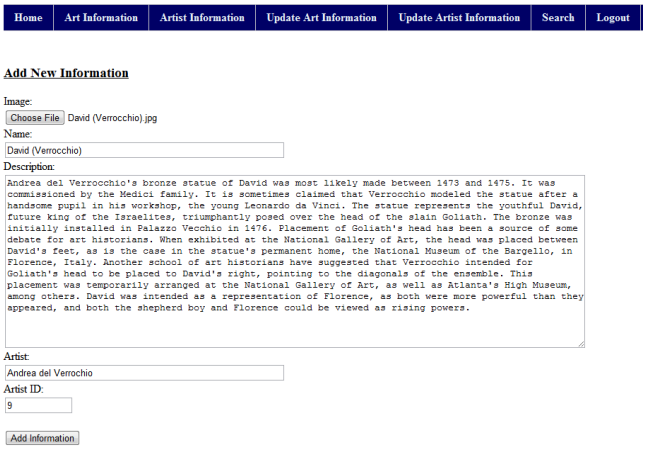


Figure 4.15: Add new information for Art

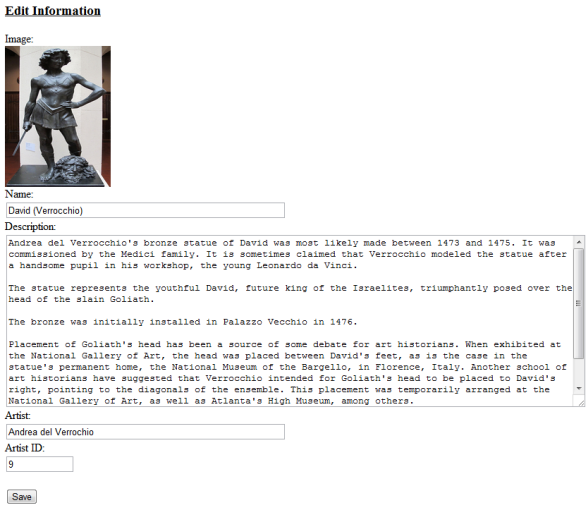


Figure 4.16: Edit information for Art

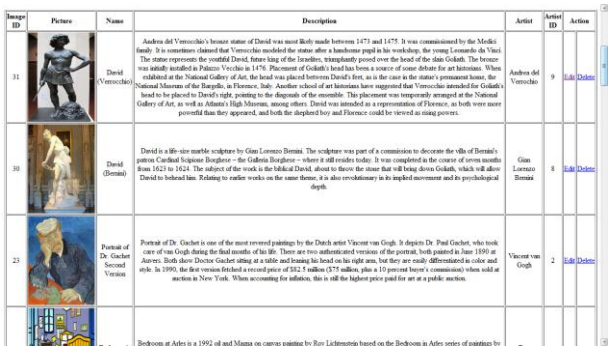


Figure 4.17: View information for Art

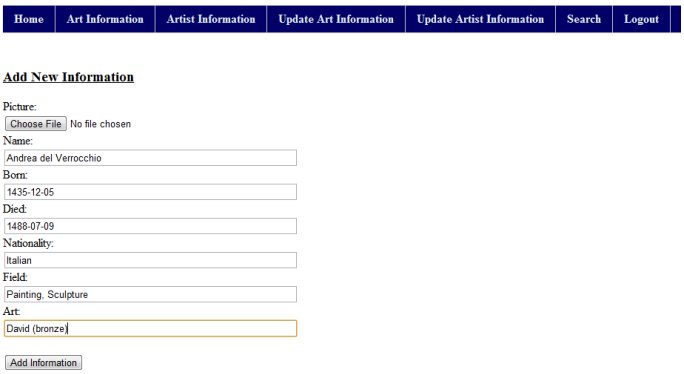


Figure 4.18: Add new information for Artist 37

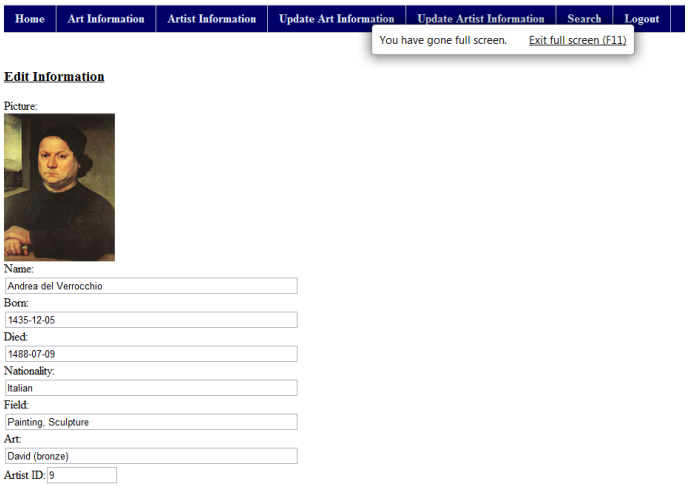


Figure 4.19: Edit information for Artist 38

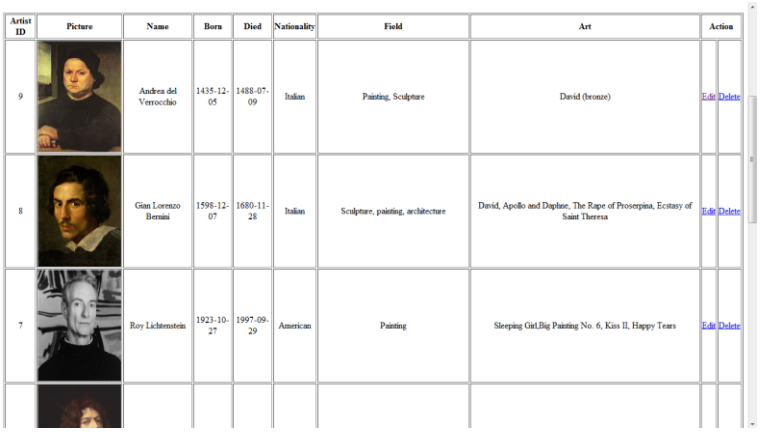


Figure 4.20: View information for Artist 38

4.5 Description of the screenshot

4.5.1 Description of user interface

In Figure 4.11 shows the main page of the system seen by the users. The user can use the traditional way of retrieving information by using alphanumeric input in the search text box. If the user intend to retrieve information about the artist, the user can tick the "Artist" radio button and proceed by entering some keyword into the search text box and hit the "Search" button. If the users want to upload their own image, the user can click the "Choose File" button to upload their image. Then, just press the "Upload" button after choosing their desired image.

Next, in Figure 4.12 shows the image had already been uploaded. If the image is correct, the user can proceed to compare the image with the one stored in the database. There will be background process where the image are converted into grayscale for easier comparison with the one stored in the database. Then, if the uploaded image is similar by 70% or more with the one in the database, the image with 70% or more similar will be retrieved.

Proceed to Figure 4.13, the system retrieved the similar image with the one in the database. Note that the image on the left is the uploaded image by the user and the image on the right would be the image retrieve from the system's database. If the user found the image that they are looking for, they can press the "Get Information" button to retrieve the information.

Finally in Figure 4.14, the information are retrieve from the database. The idea on how the system work is that, the information is retrieve based on the unique id associated with the image. Based on that unique id, the system will pass the unique id to another page to retrieve the information.

4.5.2 Description of admin interface

In the admin page, there are two main functions. One for updating the art information and another one would be to update the artist information. These two main functions have the similar sub-functions. Each function has another three sub-functions which are Add, Edit and View functions. The use of this admin interface is so that the database can be updated easily. Figure 4.15 and 4.18 are the sub-functions for Add. This function is particularly for the admin to add new information to the server directly from the webpage without going to the server and add it from there. Figure 4.16 and 4.19 are the sub-functions for Edit. The admin can edit the existing art or artist information from these sub-functions. The last sub-functions are View which is used to view the information on arts and artist in a tabulated format. The View sub-functions can be referred to Figure 4.17 and 4.20.

4.6 Process Flow of the Comparison Part of the System

Basically there are three process flows for the comparison part of the system which are Resizing, Grayscalling and Image differencing. Each process will be explained in further detail below.

4.6.1 Resizing process

Resizing process is where the two images will be resized into the same height and width. For example, if the image uploaded is at a resolution of 800x600 and the image in the database are set fixed at 1600x1200, the image uploaded will be resize into the same resolution as the image in the database which are 1600x1200. This is to ensure that every pixel can be compared easily. The method used to resize and resampled the image is *imagecopyresampled().* This method will take a portion of one image and compare it with another image to check for its difference in height and width. This process will be in a loop process where it will resampled and resize the whole image.

4.6.2 Grayscalling process

This process will make the two images into a grayscale picture. This process is important as it eliminates the comparison of the colour it contains in the two images. Comparing the two images with grayscale value will be much easier than to compare it when it is coloured. This process will be carry out using the *colormap().* This method will be using the grayscalling algorithm that had been explained in section 3.5.1.

4.6.3 Image differencing process

This is where the comparison will take part. Basically the idea is to convert the grayscaled images into hex value. Then, compare the two images' hex value to calculate the differences. These differences will determine whether the two images are similar or not.

4.7 Discussion

The image compared with the percentage of higher than 70% are considered to be similar. Based on the experiment, three out of four images are accepted as similar to the original image. One of the images is only similar by 43%. Hence, it is rejected as it not very similar with the original image. Based on this experiment, it is safe to say that it reached it objective to test the code used to find the similarity of two images by its pixel. Improvement of the code will possibly achieve the objective of this project. The next step will be utilizing the percentage in order to rank the images by its similarity. After ranking them, the user will choose the image intended to look for its information. Based on my interpretation of the results, I can say that using image as a 'keyword' for information retrieval can increase the accuracy. Apart from that, supported by the literature review, this technology of using image to retrieve information can increase accuracy and it can be applicable in various fields such as medicine, arts, education, entertainment and many more.

4.8 Challenges Faced at Development Stage

One of the big challenges faced during development is the problem with the system to retrieve the images of the uploaded file and forwarding it to the comparison page. This however came to a solution where to make a variable to read the source and forward the variable to the comparison page. Another challenge is to pick the right value for the similarity to work. As of right now, the system will only consider the image uploaded and the image stored in database at 70% similarity. Anything lower than that will be consider as not similar and will not be shown as a result in the system. According to the suggested algorithm in section 3.7, that can be used to make the comparison even more efficient, there is a possibility that the system will take a very long time to run the comparison algorithm. The proposed algorithms are still in development stage. There are still a lot of areas that can be improved such as the method of comparing the two images, the time it takes to run the algorithm and so many more.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

The field of Content-Based Image Retrieval (CBIR) has become one of the most important field in image processing. The traditional way or textual or metadata image retrieval have its own limitation where an image can be poorly described by the user hence making it unreliable when it is described. Similar image can be described differently by different people. Some filename is not even related to the image making retrieval of the image undependable. Extensive research on CBIR to enable information retrieval using image more reliable is a necessity nowadays to cater the advancement of technology in virtual data. This project achieved its objectives to come up with an image-based search engine for art gallery.

5.2 Recommendations

In this section, there are a lot of future plans or suggestions to improve the project. The limitations of this project have been set up where the target of users for this system is focus mainly on the art gallery visitors who wish to know more about the art that are exhibited. Apart from the focus scope on art, this system has the limitation where it can only compare two to find its similarity only by pixel to pixel comparison. In other words, it would not work if the image compared to with the one in the database is skewed. So, for suggestions, another algorithm can be implementing to find the correct way to compare if the uploaded image is skewed.

For future plan, the search engine can be access directly using mobile phones where the user can easily snapped the pictures and directly search for the information using their mobile phones. An extension for the application is a good example to further improve this system. For example like TinEye, they have Google Chrome's extension where the user can directly search for the similar image just by using right click directly from the image that they wish to search.

Apart from that, in the future, hopefully, this type of search engine will be implemented widely more than just a search engine for art gallery. This may be stepping stone to create a new search engine by adding image as an "input" instead of just using the alphanumeric input which we used today.

REFERENCES

[1] Arnold W.M. Smeulders, M. W. (2000, December). Content-Based Image Retrieval at the End of the Early Years. *TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 22(12).

[2] Henning Müller, N. M. (2004). A Review Of Content-Based Image Retrieval Systems In Medical Applications—Clinical Benefits And Future Directions. *International Journal of Medical Informatics*, 1—23.

[3] T. M. Deserno, S. A. (2009, April). Content-based Image Retrieval for Scientific Literature Access. *Methods Inf Med*.

[4] Venkat N. Gudivada, V. V. (1995, April). Design And Evaluation Of Algorithms For Image Retrieval By Spatial Similarity. *ACM Transactions on Information Systems (TOIS)*, 13(2), 115-144.

[5] Alexander M. B. et al., (2009, August). Partial Similarity of Objects. *International Journal of Computer Vision*, 163-183

[6] Niblack W. et al., (1993) The QBIC project: Querying images by content using color, texture and shape. *In Storage and Retrieval for Image and Video Databases (SPIE)*, 173-187

[7] Anick, P.J. (1994)Adapting a full-text information retrieval system to the computer troubleshooting domain, *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 349-358,

[8] S. K. Chang, E. Junger and G. Tortora, Spatial reasoning, image indexing and retrieval using symbolic projects.

[9] S. Zhu and A.L. Yuille, "Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 9, pp. 884-900

[10] J.P. Eakins, M.E. Graham, (2000, September) Content-based image retrieval, *Tech. Rep. JTAP-039, JISC Technology Application Program, Newcastle upon Tyne*

[11] Venkat N. Gudivada , Vijay V. Raghavan, Content-Based Image Retrieval Systems, *Computer*, v.28 n.9, p.18-22, September 1995 [doi>10.1109/2.410145]

[12] *Metadata definition.* (n.d.) Retrieved from The Tech Terms Computer Dictionary:http://www.techterms.com/definition/metadata

[13] *Content-based Image Retrieval-Wikipedia,the free encyclopedia*.(n.d.). Retrieved from Wikipedia, the free encyclopedia:http://en.wikipedia.org/wiki/Content-based_image_retrieval

[14] Cook,J.D. (2009,August 24). *Three algorithms for converting color to grayscale - The Endeavour*. Retrieved from The Endeavour — The blog of John D. Cook:http://www.johndcook.com/blog/2009/08/24/algorithms-convert-color-grayscale/

[15] *Scaling and rotating: imagecopyresized(), imagecopyresampled(), and imagerotate() | Practical PHP Programming | TuxRadar Linux*.(2011,February 11).Retrieved from TuxRadar Linux:http://www.tuxradar.com/practicalphp/11/2/12

[16] Porle,R.R., Chekima,A.,Wong,F.,& Sainarayana, G.(2009).Performance of Histogram-Based Skin colour Segmentation for Arms Detection in Human Motion Analysis Application. *World Academy of Science, Engineering and Technology.*

[17] *YUV - Wikipedia, the free encyclopedia*.(n.d.). Retrieved from Wikipedia, the free encyclopedia:http://en.wikipedia.org/wiki/YUV#Formulas

[18] Kaplan, A., & Rivlin, E.(2004,August). "Robust Feature matching across widely separate Color Images, "*Proceedings of the 7th International Conference on Pattern Recognition*, Vol. 2, pp. 136-139

[19] Tseng,V.S., Lee, C.J., & Su., J.H. (2005,August 21). "Classify By Representative or Associations (CBROA):A Hybrid Approach for Image Classification", MDM/KDD 2005 Chicago, pp. 61-69 Chicago, Illinois, USA

[20] Abu Bakar, M.H. (2007,January).*Content Based Image Retrieval.* Final Year Project, Universiti Teknologi PETRONAS, Computer & Information Sciences, Tronoh

APPENDICES

Appendix A: Survey questions

IMAGE-BASED SEARCH ENGINE FOR ART EXHIBITION/GALLERY

(IMBA SEA)

A search engine for art exhibition/gallery that uses image instead of keywords to retrieve relevant information about the art. The user just need to upload the pictures of the art to the system, and the system will retrieve the information based on the image itself.

* Required

1. On the scale of 1-5 (5 being the best), are you satisfied with the results of the retrieved information that you search for? *

   The search engine like Google, Yahoo, Bing, or just any search engine that you use.

   - 1
   - 2
   - 3
   - 4
   - 5

2. If there's a search engine that uses image to search for information instead of keywords, would you use it? *

   The search engine like Google, Yahoo, or Bing, uses keywords to search for information, how about using image as a keywords?

   - Yes
   - No
   - Maybe

3. Would it be easy if you can retrieve information based on the image that you uploaded into the search engine? (retrieve information based on the image itself and not the keyword associated or the name of the image) *

   The traditional way of searching for image information using the current search engine is using the keywords associated with the image, (i.e, the name of the image itself). But, what if the name of the image is not related to the image itself, or it is simply hard to describe the image.

   - Yes
   - No
   - Maybe

4. Would it be easy if you can just snapped the pictures and search for information about the art? *

   Imagine you're in an art gallery and you are fascinated by the arts and want to know more, but no enough information on the title of the art.

   - Yes
   - No
   - Maybe

5. If there is a technology where you can use the image of the art to search for information about the art, would you use it? *

   Art Exhibition/Gallery usually gives the name or the artist for the arts displayed, but not enough information about the art itself.

   - Yes
   - No
   - Maybe

Appendix B: Images used for Experiment I

Figure 4.6: JPG Original (810x1080)

Figure 4.7: JPG Different Dimension (405x540)

Figure 4.8: JPG Edited (405x540)

Figure 4.9: PNG Different File type (405x540)

Figure 4.10: JPG Different (405x540)

Appendix C: Screenshot of the system

User Interface:

Figure 4.11: Main page for User

Figure 4.12: Image being uploaded by user

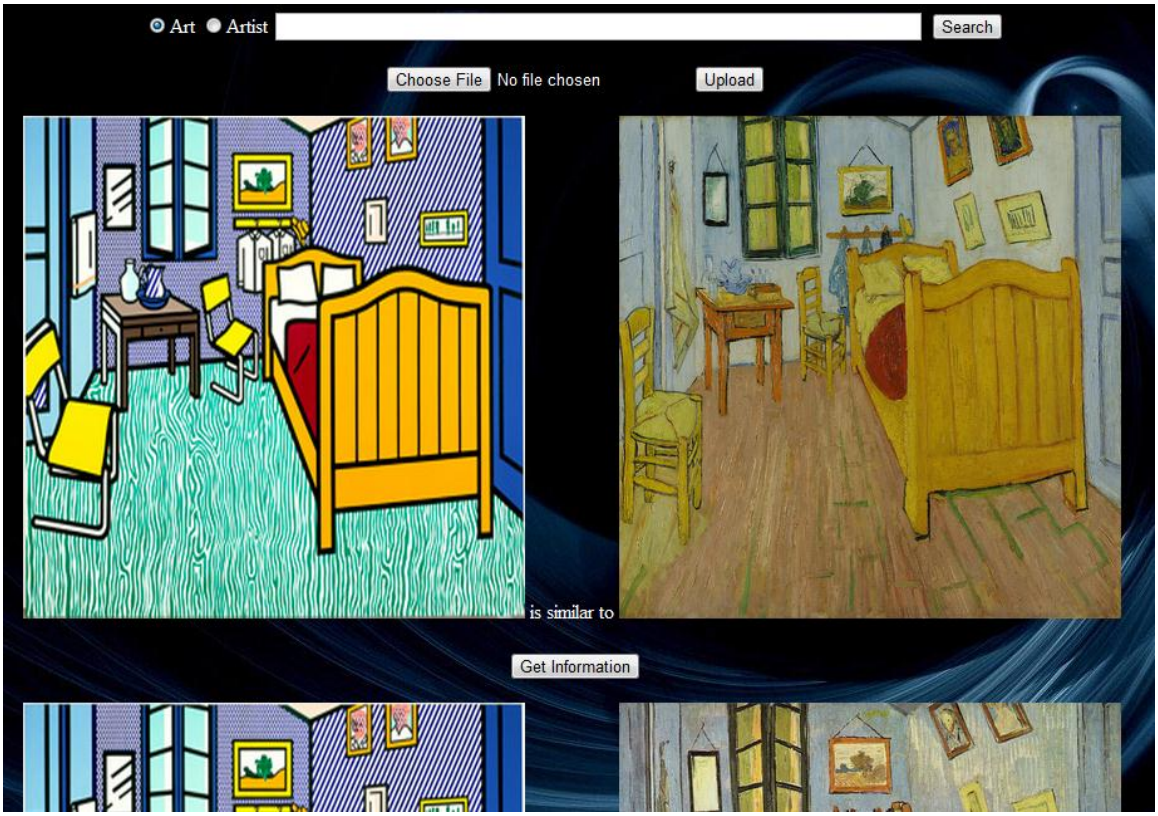


Figure 4.13: Comparison Results

Figure 4.14: Art information retrieved

Admin Interface:

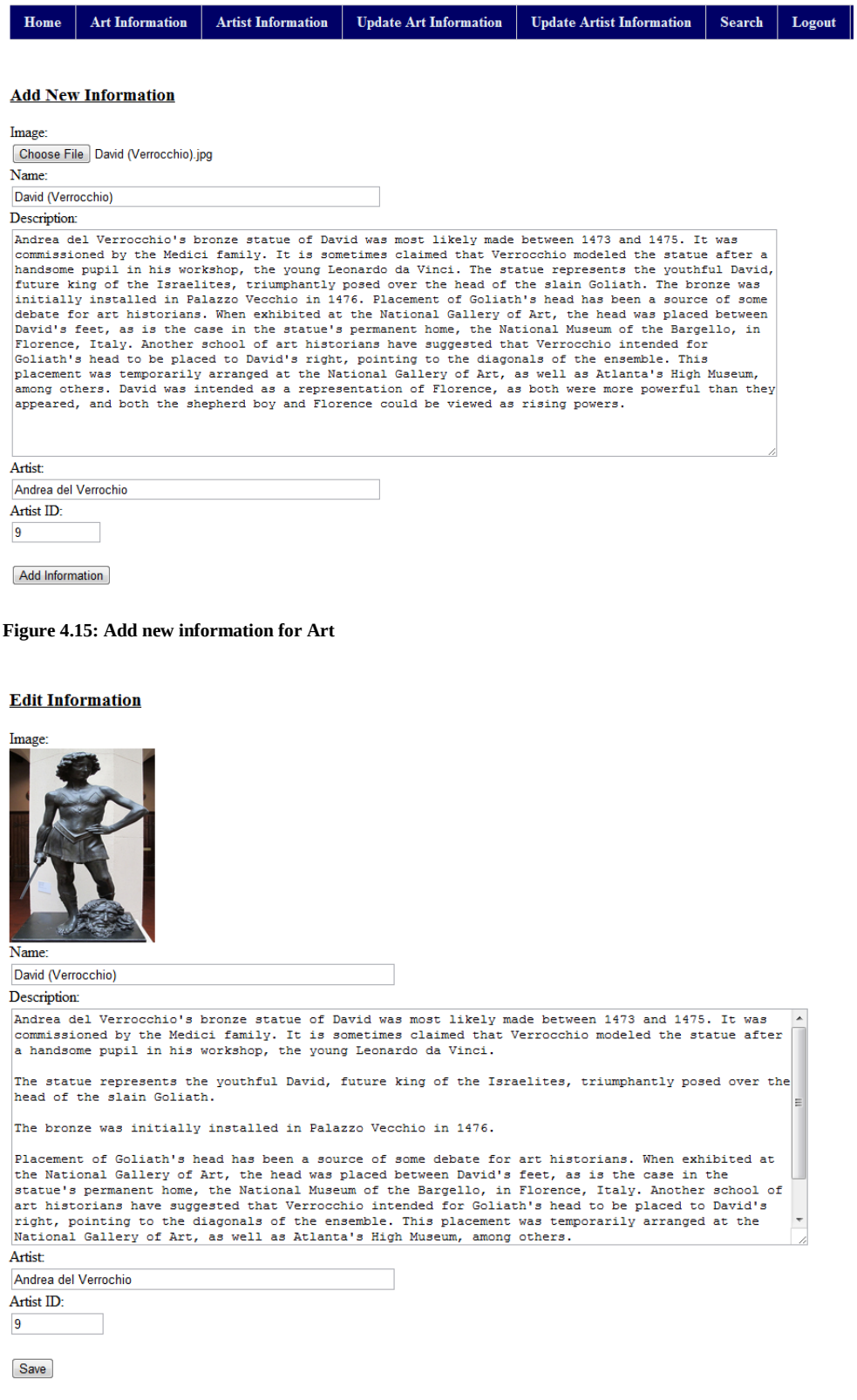


Figure 4.15: Add new information for Art

Figure 4.16: Edit information for Art

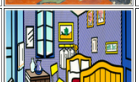
Image ID	Picture	Name	Description	Artist	Artist ID	Action	
31		David (Verrocchio)	Andrea del Verrocchio's bronze statue of David was most likely made between 1473 and 1475. It was commissioned by the Medici family. It is sometimes claimed that Verrocchio modeled the statue after a handsome pupil in his workshop, the young Leonardo da Vinci. The statue represents the youthful David, future king of the Israelites, triumphantly posed over the head of the slain Goliath. The bronze was initially installed in Palazzo Vecchio in 1476. Placement of Goliath's head has been a source of some debate for art historians. When exhibited at the National Gallery of Art, the head was placed between David's feet, as is the case in the statue's permanent home, the National Museum of the Bargello, in Florence, Italy. Another school of art historians have suggested that Verrocchio intended for Goliath's head to be placed to David's right, pointing to the diagonals of the ensemble. This placement was temporarily arranged at the National Gallery of Art, as well as Atlanta's High Museum, among others. David was intended as a representation of Florence, as both were more powerful than they appeared, and both the shepherd boy and Florence could be viewed as rising powers.	Andrea del Verrochio	9	Edit	Delete
30		David (Bernini)	David is a life-size marble sculpture by Gian Lorenzo Bernini. The sculpture was part of a commission to decorate the villa of Bernini's patron Cardinal Scipione Borghese – the Galleria Borghese – where it still resides today. It was completed in the course of seven months from 1623 to 1624. The subject of the work is the biblical David, about to throw the stone that will bring down Goliath, which will allow David to behead him. Relating to earlier works on the same theme, it is also revolutionary in its implied movement and its psychological depth.	Gian Lorenzo Bernini	8	Edit	Delete
23		Portrait of Dr. Gachet Second Version	Portrait of Dr. Gachet is one of the most revered paintings by the Dutch artist Vincent van Gogh. It depicts Dr. Paul Gachet, who took care of van Gogh during the final months of his life. There are two authenticated versions of the portrait, both painted in June 1890 at Auvers. Both show Doctor Gachet sitting at a table and leaning his head on his right arm, but they are easily differentiated in color and style. In 1990, the first version fetched a record price of $82.5 million ($75 million, plus a 10 percent buyer's commission) when sold at auction in New York. When accounting for inflation, this is still the highest price paid for art at a public auction.	Vincent van Gogh	2	Edit	Delete
			Bedroom at Arles is a 1992 oil and Magna on canvas painting by Roy Lichtenstein based on the Bedroom in Arles series of paintings by				

Figure 4.17: View information for Art

Figure 4.18: Add new information for Artist

Edit Information

Picture:

Name:
Andrea del Verrocchio
Born:
1435-12-05
Died:
1488-07-09
Nationality:
Italian
Field:
Painting, Sculpture
Art:
David (bronze)
Artist ID: 9

Figure 4.19:Edit information for Artist

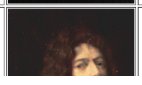
Artist ID	Picture	Name	Born	Died	Nationality	Field	Art	Action	
9		Andrea del Verrocchio	1435-12-05	1488-07-09	Italian	Painting, Sculpture	David (bronze)	Edit	Delete
8		Gian Lorenzo Bernini	1598-12-07	1680-11-28	Italian	Sculpture, painting, architecture	David, Apollo and Daphne, The Rape of Proserpina, Ecstasy of Saint Theresa	Edit	Delete
7		Roy Lichtenstein	1923-10-27	1997-09-29	American	Painting	Sleeping Girl,Big Painting No. 6, Kiss II, Happy Tears	Edit	Delete
									

Figure 4.20: View information for Artist

Image-Based Search Engine For Art Exhibition Gallery

(ImBa SEA Exhibition/Gallery)

Yong Suet Peng
Computer and Information Sciences Department
Universiti Teknologi PETRONAS
Bandar Seri Iskandar, 31750, Tronoh, Perak, Malaysia
yongsuetpeng@petronas.com.my

Mohamad Nasrie bin Ajiji
Computer and Information Sciences Department
Universiti Teknologi PETRONAS
Bandar Seri Iskandar, 31750, Tronoh, Perak, Malaysia
naszyuki@gmail.com

ABSTRACT

The image-based search engine for art exhibition/gallery (ImBa SEA) is like any search engine but will provide a better functionality. The ImBa SEA is specially design for art gallery where lies thousands of arts displayed on the exhibition. With the help of the ImBa SEA, the user can upload image that they have snapped into the system to retrieve the information about the art itself. Some features are very difficult to describe with text, some special textures and complex shapes cannot be clearly represented by alphanumeric inputs. The arts need to be digitalized and stored in the gallery's database. Based on the existing way of accessing information for art gallery, instead of directly using the image as a 'keyword' to retrieve information, each art have its own name. The name is stored in the art gallery database, with the relevant information. However, this method takes some time because some arts have the similar name or a long name. This decreases the accuracy of the search engine to retrieve information of the art wish to be accessed. With ImBa SEA, the user directly uses the image to retrieve the information on the given art. Each art will have its own unique id to ease the process of retrieving information. The image is uploaded into the search engine, and then it will process the image to look for similarity of the image with the images stored in the database. Once the similar image is found in the database, the unique id of the art is used to retrieve the information of the image. Basically, ImBa SEA compares the two images pixel by pixel to look for similarities. This technique will surely provide better efficiency and accuracy on information retrieval. This paper will discuss everything from the problem faced, and scope concentrated to achieve the objectives of the project which are discussed in Chapter 1. The system or technologies related or similar with the ImBa SEA are discussed in Chapter 2: Literature Review. Also, the research methodology of the project is discussed later in Chapter 3 together with the development methodology. The basic structure is also discussed in this paper to give a clear view on how the system works. The future plans on the ImBa SEA are discussed at the end of this paper.

I. INTRODUCTION

In the twentieth century, images have grown in numbers, availability and its importance. Nowadays, we can relate images in diverse fields such advertising, design, art, education, entertainment, medicine and many more. However, this vast growth of visual content creation has not been matched with the technologies existed to cater efficient image analysis and retrieval.

People currently looks for image using search engine which grabs information based on the user's input in textual form. A problem with the current method is that, sometimes, the user did not know what the right keywords that they should enter into the search engine to get the desired result. These have been the case for art gallery. The people who come to the art gallery are interested to know more information on the art, however, because of its complexity of the arts to describe the art using text descriptions, the information retrieve may be wrong. Some features are very difficult to describe with text, some special textures and complex shapes cannot be clearly represented.

This is why this paper proposed a slightly different type of search engine for the art gallery where the user can upload the image into the search engine and the search engine will return with the information related to the image or any other related images that are similar to the image uploaded. Each art are digitalized into the gallery's database with its own unique id. This unique id is then used to retrieve the information of the image uploaded.

A. *Problem Statement*

The existing search engines which are widely used today work when the user input some textual keywords to look for images. With the rapid growth of digitalizing visual content, technological advancement in the efficient visual analysis and retrieval is a need. However, the problem is some image is associated with keywords that are not related to the image at all which will brings error to the searched result. With the existing technology, it searched for the image using the keywords, hence, there's a probability that the result of the image may not necessarily relates to the keywords that have been submitted by the user. Some people simply named their image filename as anything that they desire giving false retrieval when it comes to alphanumeric search. Apart from that, the image metadata search (description of image) is not practical when it involves large database. Metadata requires the user to manually describe the images that is in the database [12]. So, there is a need for a more efficient image analysis and retrieval.

B. *Objective*

The objectives of this project are to:

- Research on the methodologies that have been applied to achieve the image retrieval system
- Propose an algorithm for image similarity matching and information retrieval.
- Develop image-based search engine for art exhibition/gallery. The system will have the following functions:
    - Retrieve information based on user input in textual form like the current search engine
    - Retrieve information of the image based on user input based on what the user's image upload
    - Add, edit and delete function for admin for easy update of the system's database.

C. *Scope and Limitation of Study*

Based on the project title, Image-Based Search Engine for Art Exhibition/Gallery (ImBa SEA), it is clear that this system will be in the form of a web-based project which will be the main focus for this project. Apart from that, the scope of this project is studying the aspects of image search and web-based application. The proposed project is to seek ways to further improve the existing search engine for art gallery, instead of just entering some keywords to search for information, the search engine will also provide the functionality to upload image to retrieve information from the search engine. There are certain limitations on this image-based search engine where it can only be used for a certain specific art gallery where the arts are digitalized into their database for easy information retrieval. Apart from that, as of now, the image being compare can only be compare if the image uploaded is not been skewed to the right of left. This means that the image must be at the right position in order to compare the image with the one in the database. The technologies focuses on this project are CBIR and also pattern recognition that are for the image retrieval part of the ImBa SEA.

D. *Significance of Project*

The success of this project could be a stepping stone towards a new way for image retrieval. The current way which is the metadata-based search system requires the human to personally describe the image in the database. Thus a system that can filter images based on their content would provide better indexing and return more accurate results [13]. This project is highly related to CBIR. The potential of using CBIR includes architectural and engineering design, art collections, crime prevention, intellectual property, medical diagnosis, military, photograph archives and many more. There is so much benefits that can be achieved through CBIR. In this project, the focus would be how CBIR can help in achieving more efficient retrieval on art gallery collections.

II. LITERATURE REVIEW

A. *Content-Based Image Retrieval for Scientific Literature Access*

This journal was written by T. M Deserno, S. Antani, and L. Rodney Long from the Department of Medical Informatics, Aachen University of Technology, Germany. The objective of this article is to estimate the benefits of using content-based image retrieval (CBIR) on article figures to augment traditional access to articles [3].

In this article it compares the traditional ways of accessing the articles by alphanumerical search on title, author or abstract and may disregard numerous figures. Also the team had conducted experiments where they are making a quantitative estimate by projecting from data from the Cross-Language Evaluation Forum (Image-CLEF) campaigns, and qualitatively validate it through experiments using the Image Retrieval in Medical Applications (IRMA) project. The result of these experiments shows the predicted accuracy up to 97.08% of article retrieval. They had conducted the experiments due to the increasing numbers of electronic articles been published and due to its limited access of searching for the right articles. Some problem is that the figures in the articles are not easily described by words which make it difficult for alphanumeric search like the traditional ways.

To conclude this article, the researchers are convince that CBIR have high potential impact in medical

literature search and retrieval due its high accuracy of retrieving information.

B. *Design and Evaluation of Algorithms for Image Retrieval by spatial similarity*

This journal discuss about image retrieval using spatial similarity. It was written by Venkat N. Gudivada of Ohio University and Vijay V. Raghavan of University of Southwest Louisiana. Spatial similarity can be seen as a subset of similarity and all the entities being compared to each other have spatial components [4].

In image database applications, similarity-based retrieval of images is an important task. Users tend to request retrieving those images in the database based on the similarity of the query image. In this paper, they propose an algorithm to compute the spatial similarity between two symbolic images. A natural image or the original image cannot be stored in database systems as it is for retrieval. Therefore, the natural image is converted into symbolic image as a logical representation of the original image. The proposed algorithm can deal with translation, scale, and rotational variances in images. The characteristics of the proposed algorithm are compared with those of the previously available algorithms using experimental set of images. The comparison verified that the algorithm provides a rank ordering of images that constantly matches with the expert's expected rank ordering and also in a more efficient manner.

In conclusion, the algorithm proposed may open a new dimension to image retrieval based on the similarity pattern of the images. Maybe the alteration of the algorithm will innovate new ideas on pattern recognition on image and will surely benefit in the future for image based application technology.

C. *Content-Based Image Retrieval Systems*

This paper is written by Venkat N. Gudivada from Ohio University and Vijay V. Raghavan from University of Southwestern Louisiana. There is a lot of possible usage of this CBIR technology. In this paper, it discuss about the possibility of CBIR technology of changing the way we use the creation and distribution of digital visual content. Some application areas in which CBIR is a principal activity are numerous and diverse are [11]:

- Art galleries and museum management
- Architectural and engineering design
- Interior design
- Remote sensing and management of earth resources
- Geographic information systems
- Scientific database management
- Weather forecasting
- Retailing
- Fabric and fashion design
- Trademark and copyright database management
- Law enforcement and criminal investigation
- Picture archiving and communication systems

In this paper also discuss the previous approaches to content-based retrieval. There are two approaches. First approach, the image contents are being modelled as a set of attributes which are extracted manually and managed within the conventional database management systems. To retrieve the information, queries are used. This approach is known as attribute-based retrieval. This first approach is not really using the content of the image itself to retrieve the information. It is more like metadata search like the one we are using right now like Google, Bing, Yahoo and so on. The second approach depends on an integrated feature-extraction or object-recognition subsystem to overcome the limitations of attribute-based retrieval. The second approach have a subsystem which carries the task of feature-extraction and object-recognition task whenever the image is inserted into the database. However, this approach is said to be highly expensive and quite difficult to carry out.

The point of this paper is to get the idea on some important features needed to carry out the ImBa SEA project. Regardless of which approach is used, the generic query classes which facilitate CBIR through retrieving by:

- Colour
- Texture
- Sketch
- Shape
- Volume
- Spatial constraints
- Browsing
- Pixel comparison

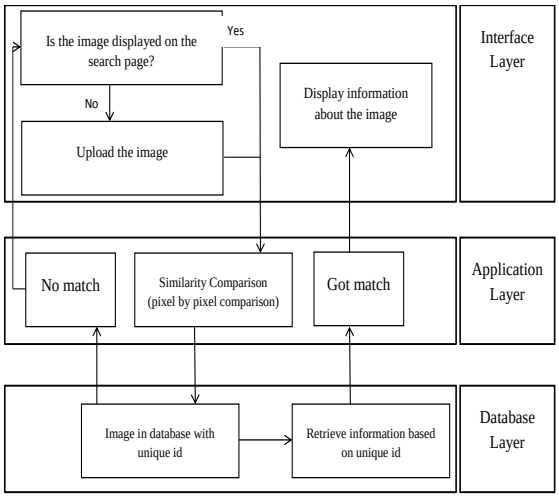
The best way to retrieve image by comparing it with another image would be by texture and colour. These however need a high level knowledge of understanding in computer vision. In general, by comparing the images by its texture, the probability of getting it by similarity is higher, and by adding colour, it can confirms the similarity of the two images being compared. Pixel by pixel comparison is a good way to compare two images to test whether it have the similarity. Like for example, to compare two images whether it is similar, after comparing it by pixel, get the percentage of how many it does similar, a higher percentage would mean that the two images are very similar.

III. METHODOLOGY

A. *System Basic Architecture*

Once knowing how the similar existing system works, the next phase to develop the architecture on how the system will works. This will give the clear picture and

understanding on how the system will operate and to avoid developing a system that does not solving the problem it intended to solve. **Figure 3.1** below is the basic system architecture of how these systems will works.



B. Development Methodology

In completing this project, the development methodology to be chosen is Incremental Prototyping. Prototyping is the process of quickly putting together a working model in order to test various aspects of a design, illustrate ideas or features & gather early user feedback. It is believed to reduce project risks and cost. Incremental Prototyping is a prototype technique where the final product is built in several prototypes. Some parts of the functionalities are produced before others in this incremental prototyping process. This enables us to view the early visibility of some key aspects. Also, some or all the most critical functions can be done early for this prototyping process. The advantage of prototyping is that the development provides the developers with insight to how the system should be structured.

The benefits of using this methodology is it allows any changes to be made during the development phase if there is needs to review and recheck at any other phase of project development. This is important as it provides flexibility throughout completing the project such as debugging process

C. System Flowchart

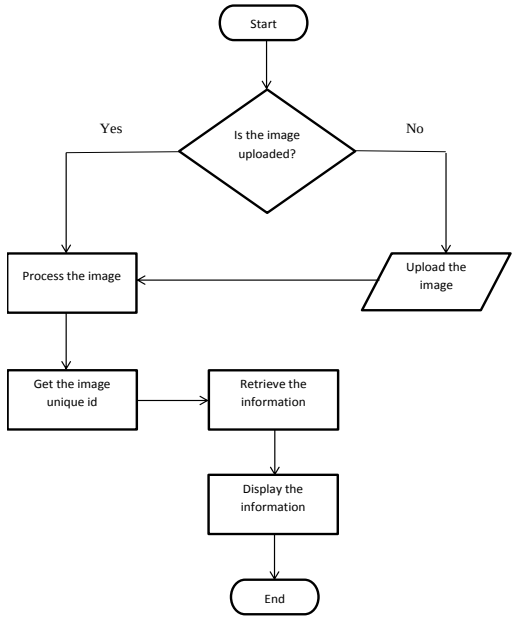


Figure 3.3: Flowchart of the basic system architecture

D. Tools Required

The design of the background will be created using Adobe Photoshop just to give a better look on the search engine. The programming language intended to use is PHP as the programming language because of its security to the system. The database for the search engine will primarily use SQL database. The platform for the SQL would be MySQL. XAMPP will be used as the server to test for its functionality.

E. Algorithms

- Grayscale image algorithm
  - Lightness method
  - Average method
  - Luminosity method
- Pixel-based similarity check algorithm

IV. RESULTS AND DISCUSSION

A. Experiment I

An experiment to test the code had been carried out. The objective of this experiment is to test whether the code can be used to find the similarity of the image that the users upload with the image that is stored in the database. All of the images are compared with the original. In this experiment, the constant variable would be the original image. The images that are compared with the original image are said to be the responding variable or the variable that are changed in order to achieve the objective of the experiment. The

image used in the experiment can be referred in Appendix B. Three sets of experiments had been carried out to determine which colormap will determine the best choice to use for the system. The tables below are the result of the experiments.

Table 4.1: Experiment I Results

Table 4.1.1: Using colormap Green>Red>Blue (0.212671 * R + 0.715160 * G + 0.072169 * B)

Responding variable	Image of different dimension	Edited image	Image of different file type	Different image
Percentage of similarity	92.5%	85.75 %	99.5%	43%

Table 4.1.2: Using colormap Red>Green>Blue (0.412453 * R + 0.357580 * G +0.180423 * B)

Responding variable	Image of different dimension	Edited image	Image of different file type	Different image
Percentage of similarity	93.75%	86.5 %	99.5%	44%

Table 4.1.3: Using colormap Blue>Green>Red (0.019334 * R + 0.119193 * G + 0.0.950227 * B)

Responding variable	Image of different dimension	Edited image	Image of different file type	Different image
Percentage of similarity	90%	80.75 %	96.75%	43.25%

Based on the experiments, the best colormap to use is the colormap where the Green weighted value is more than Red and Red weighted value is more than Blue. This is because, the human eyes are more sensitive to green than other colours, so green is weighted most heavily. Also, based on the results, the colormap Green>Red>Blue retrieve the most accurate results.

B. Experiment II

Another experiment is carried out to test the code with three different algorithms to use. The objective of this experiment is to determine which of the three algorithms is the best method to use in the comparison mechanism of the system. All of the images are compared with the original. In this experiment, the constant variable would be the original image. The images that are compared with the original image are said to be the responding variable or the variable that are changed in order to achieve the objective of the experiment. The experiment will be carry out three times. The first would be using the lightness method. The second would be using the average method and third method would be using the luminosity method.

Responding variable		Image of different dimension	Edited image	Image of different file type	Different image
Percentage of similarity	Lightness	93.75 %	85.35%	98%	46%
	Average	94.5%	85.75%	99.5 %	46.25 %
	Luminosity	92.5%	85.75%	99.5 %	43%

Based on the results that are displayed on Table 4.2, we can assume that the best method would be using Luminosity method. This is because, based on the percentage of different image compared to the original image, both Lightness and Average method calculate the similarity by 46% and 46.25% respectively. This is quite a large value that may change the accuracy of the system on comparing two images that are totally different. Therefore, it can be concluded that the most suitable method to use is using the Luminosity.

C. System Screenshot

There are two parts of the system, one for users and another one for admin. Below are the user interfaces of the system.

Figure 4.11: Main page for User

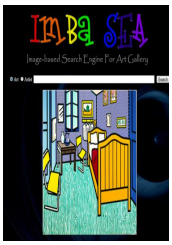


Figure 4.12: Image being uploaded by user

Figure 4.13: Comparison result

Figure 4.14: Art information retrieved

Below are the screenshot for the admin interface.

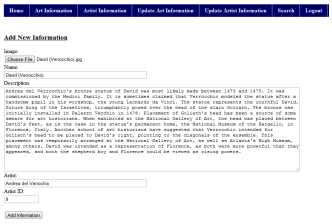


Figure 4.15: Add new information for Art

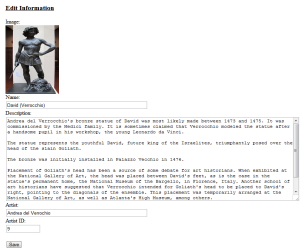


Figure 4.16: Edit information for Art

Figure 4.17: View information for Art

Figure 4.18: Add new information for Artist 37

Figure 4.19: Edit information for Artist 38

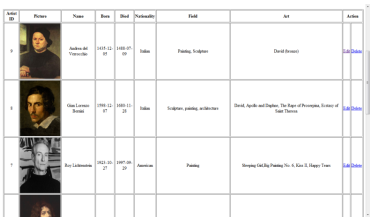


Figure 4.20: View information for Artist 38

In Figure 4.11 shows the main page of the system seen by the users. The user can use the traditional way of retrieving information by using alphanumeric input in the search text box. If the user intends to retrieve information about the artist, the user can tick the "Artist" radio button and proceed by entering some keyword into the search text box and hit the "Search" button. If the users want to upload their own image, the user can click the "Choose File" button to upload their image. Then, just press the "Upload" button after choosing their desired image.

Next, in Figure 4.12 shows the image had already been uploaded. If the image is correct, the user can proceed to compare the image with the one stored in the database. There will be background process where the image are converted into grayscale for easier comparison with the one stored in the database. Then, if the uploaded image is similar by 70% or more with the one in the database, the image with 70% or more similar will be retrieved.

Proceed to Figure 4.13, the system retrieved the similar image with the one in the database. Note that the image on the left is the uploaded image by the user and the image on the right would be the image retrieve from the system's database. If the user found the image that they are looking for, they can press the "Get Information" button to retrieve the information.

Finally in Figure 4.14, the information are retrieve from the database. The idea on how the system work is that, the information is retrieve based on the unique id associated with the image. Based on that unique id, the system will pass the unique id to another page to retrieve the information.

In the admin page, there are two main functions. One for updating the art information and another one would be to update the artist information. These two main functions have the similar sub-functions. Each function has another three sub-functions which are Add, Edit and View functions. The use of this admin interface is so that the database can be updated easily. Figure 4.15 and 4.18 are the sub-functions for Add. This function is particularly for the admin to add new information to the server directly from the webpage without going to the server and add it from there. Figure 4.16 and 4.19 are the sub-functions for Edit. The admin can edit the existing art or artist information from these sub-functions. The last sub-functions are View which is used to view the information on arts and artist in a tabulated format. The View sub-functions can be referred to Figure 4.17 and 4.20.

D. Process Flow of the Comparison Part of the System

Basically there are three process flows for the comparison part of the system which are Resizing, Grayscalling and Image differencing.

- **Resizing process**
  Resizing process is where the two images will be resized into the same height and width. For example, if the image uploaded is at a resolution of 800x600 and the image in the database are set fixed at 1600x1200, the image uploaded will be resize into the same resolution as the image in the database which are 1600x1200. This is to ensure that every pixel can be compared easily. The method used to resize and resampled the image is *imagecopyresampled().* This method will take a portion of one image and compare it with another image to check for its difference in height and width. This process will be in a loop process where it will resampled and resize the whole image.

- **Grayscalling process**
  This process will make the two images into a grayscale picture. This process is important as it eliminates the comparison of the colour it contains in the two images. Comparing the two images with grayscale value will be much easier than to compare it when it is coloured. This process will be carry out using the *colormap().*

- **Image differencing process**
  This is where the comparison will take part. Basically the idea is to convert the grayscaled images into hex value. Then, compare the two images' hex value to calculate the differences. These differences will determine whether the two images are similar or not.

V. CONCLUSION

The field of Content-Based Image Retrieval (CBIR) has become one of the most important field in image processing. The traditional way or textual or metadata image retrieval have its own limitation where an image can be poorly described by the user hence making it unreliable when it is described. Similar image can be described differently by different people. Some filename is not even related to the image making retrieval of the image undependable. Extensive research on CBIR to enable information retrieval using image more reliable is a necessity nowadays to cater the advancement of technology in virtual data. This project achieved its objectives to come up with an image-based search engine for art gallery. There are a lot of future plans or suggestions to improve the project. The limitations of this project have been set up where the target of users for this system is focus mainly on the art gallery visitors who wish to know more about the art that are exhibited. Apart from the focus scope on art, this system has the limitation where it can only compare two to find its similarity only by pixel to pixel comparison. In other words, it would not work if the image compared to with the one in the database is skewed. So, for suggestions, another algorithm can be implementing to find the correct way to compare if the uploaded image is skewed. For future plan, the search engine can be access directly using mobile phones where the user can easily snapped the pictures and directly search for the information using their mobile phones. An extension for the application is a good example to further improve this system. For example like TinEye, they have Google Chrome's extension where the user can directly search for the similar image just by using right click directly from the image that they wish to search. Apart from that, in the future, hopefully, this type of search engine will be implemented widely more than just a search engine for art gallery. This may be stepping stone to create a new search engine by adding image as an "input" instead of just using the alphanumeric input which we used today.

VI. ACKNOWLEDGEMENT

First and foremost, I would like to take this opportunity to thank God for giving me the opportunity to finish this project. I would like to express his greatest gratitude and appreciation to project supervisor, Dr Vivian Yong Suet Peng, who had continuously given me constructive comments, advices, and suggestions which have guided the project towards its successful final outcome. Next, I would like to acknowledge Universiti Teknologi PETRONAS (UTP) especially the committee of Final Year Project of Computer Information Sciences (CIS) department for excellent organization and management of this course. Last but not least I would also like to express the participant of survey questions, and not to forget my parents for their endless support throughout the project. They have given me strength, motivation and determination in all my undertaking.

REFERENCES

[1] Arnold W.M. Smeulders, M. W. (2000, December). Content-Based Image Retrieval at the End of the Early Years. *TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 22(12).

[2] Henning Müller, N. M. (2004). A Review Of Content-Based Image Retrieval Systems In Medical Applications—Clinical Benefits And Future Directions. *International Journal of Medical Informatics*, 1—23.

[3] T. M. Deserno, S. A. (2009, April). Content-based Image Retrieval for Scientific Literature Access. *Methods Inf Med.*

[4] Venkat N. Gudivada, V. V. (1995, April). Design And Evaluation Of Algorithms For Image Retrieval By Spatial Similarity. *ACM Transactions on Information Systems (TOIS)*, 13(2), 115-144.

[5] Alexander M. B. et al., (2009, August). Partial Similarity of Objects. *International Journal of Computer Vision*, 163-183

[6] Niblack W. et al., (1993) The QBIC project: Querying images by content using color, texture and shape. *In Storage and Retrieval for Image and Video Databases (SPIE)*, 173-187

[7] Anick, P.J. (1994)Adapting a full-text information retrieval system to the computer troubleshooting domain, *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 349-358,

[8] S. K. Chang, E. Junger and G. Tortora, Spatial reasoning, image indexing and retrieval using symbolic projects

[9] S. Zhu and A.L. Yuille, "Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 9, pp. 884-900

[10] J.P. Eakins, M.E. Graham, (2000, September) Content-based image retrieval, *Tech. Rep. JTAP-039, JISC Technology Application Program, Newcastle upon Tyne*

[11] Venkat N. Gudivada , Vijay V. Raghavan, Content-Based Image Retrieval Systems, *Computer*, v.28 n.9, p.18-22, September 1995 [doi>10.1109/2.410145]

[12] *Metadata definition.* (n.d.) Retrieved from The Tech Terms Computer Dictionary:http://www.techterms.com/definition/metadata

[13] *Content-based Image Retrieval-Wikipedia,the free encyclopedia*.(n.d.). Retrieved from Wikipedia, the free encyclopedia:http://en.wikipedia.org/wiki/Content-based_image_retrieval

[14] Cook,J.D. (2009,August 24). *Three algorithms for converting color to grayscale - The Endeavour*. Retrieved from The Endeavour — The blog of John D. Cook:http://www.johndcook.com/blog/2009/08/24/algorithms-convert-color-grayscale/

[15] *Scaling and rotating: imagecopyresized(), imagecopyresampled(), and imagerotate() | Practical PHP Programming | TuxRadar Linux*.(2011,February 11).Retrieved from TuxRadar Linux:http://www.tuxradar.com/practicalphp/11/2/12

[16] Porle,R.R., Chekima,A.,Wong,F.,& Sainarayana, G.(2009).Performance of Histogram-Based Skin colour Segmentation for Arms Detection in Human Motion Analysis Application. *World Academy of Science, Engineering and Technology.*

[17] *YUV - Wikipedia, the free encyclopedia*.(n.d.). Retrieved from Wikipedia, the free encyclopedia:http://en.wikipedia.org/wiki/YUV#Formulas

[18] Kaplan, A., & Rivlin, E.(2004,August). "Robust Feature matching across widely separate Color Images, "*Proceedings of the 7th International Conference on Pattern Recognition*, Vol. 2, pp. 136-139

[19] Tseng,V.S., Lee, C.J., & Su., J.H. (2005,August 21). "Classify By Representative or Associations (CBROA):A Hybrid Approach for Image Classification", MDM/KDD 2005 Chicago, pp. 61-69 Chicago, Illinois, USA

[20] Abu Bakar, M.H. (2007,January).*Content Based Image Retrieval.* Final Year Project, Universiti Teknologi PETRONAS, Computer & Information Sciences, Tronoh