RETRIEVAL OF HUMAN FACIAL IMAGES BASED ON VISUAL CONTENT AND SEMANTIC DESCRIPTION

AHMED ABDU ALI ALATTAB

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2013

RETRIEVAL OF HUMAN FACIAL IMAGES BASED ON VISUAL CONTENT AND SEMANTIC DESCRIPTION

AHMED ABDU ALI ALATTAB

THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2013

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Ahmed Abdu Ali Alattab (I.C/Passport No: 03569777)

Registration / Matric No: WHA060008

Name of Degree: Doctor of Philosophy

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):
Retrieval of Human Facial Images Based on Visual Content and Semantic Description

Field of Study: Artificial Intelligence \ Image Processing

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date

Subscribed and solemnly declared before,

Witness's Signature Date

Name:
Designation:

ABSTRACT

The significant increase in the huge collections of digital images and videos that need to be managed has led to the requirement for efficient methods for the archival and retrieval of these images. Facial images have gained its importance amongst these digital images due to its use in various aspects of life such as, in airports, law enforcement applications, security systems and automated surveillance applications. The basic content-based image retrieval (CBIR) system used for the general task of image retrieval is not effective with facial images, especially when the query is in some form of user descriptions. The current CBIR is based on low-level features such as color, texture, shape, and eigenfaces thus it cannot capture the semantic aspects of a facial image. Humans by nature tend to use semantic descriptions (high-level feature) when describing what they are looking for, and they normally encounter difficulties when using descriptions based on low-level features. This is because human beings normally perceive facial images and compare their similarities using high-level features such as gender, race, age and the rating scale of the facial traits and thus cannot relate these high-level semantic concepts directly to low-level features. In this research, we propose a semantic content-based facial image retrieval technique (SCBFIR) that incorporates multiple visual features with semantic features to increase the accuracy of the facial image retrieval and to reduce the semantic gap between the high-level query requirements and the low-level facial features of the human facial image. Semantic features were selected and weighted based on a case study, with the participation of one hundred respondents. Visual features and semantic features were extracted by different methods, so they have variant weights. A new method was proposed through a radial basis function network for both, measuring the distance between the query vectors and the database vectors of the different features for similarities finding, and for ranking and combining the similarities. A probabilistic approach was used to improve the differences observed based on humans' perception and the viewpoint that may appear during image annotation and/or query process. A prototype system of human facial image retrieval was subsequently built to test the retrieval performance. The system was trained and tested on two databases; the first database being the 'ORL Database of Faces' from AT&T Laboratories, Cambridge, while the second database consists of local facial images database of one hundred and fifty participants from the University of Malaya (UM), Kuala Lumpur,and some of their friends and families outside the UM. The results of the experiments show that, as compared to the content-based facial image retrieval technique, the proposed methods of SCBFIR achieve the best performance based on the number of semantic features used. The content-based facial image retrieval technique achieves 80.60% and 89.51% accuracy, while the SCBFIR achieves 97.85 % and 99.39% accuracy for the first and second database respectively within the top 10 retrieved facial images.

ABSTRAK

Peningkatan ketara dalam pengutipan-pengutipan imej bergidit dan video memerlukan keadah-keadah pengurusan yang cekap bagi tujuan arkib dan pembacaan semula. Imej-imej muka adalah yang terpenting antara imej-imej berdigit kerana terdapat banyak aplikasinya dalam kehidupan manusia seperti di sistem lapangan terbang, penguatkuasaan undang undang, sistem-sistem keselamatan dan sistem pengawasan automatik.*Content Based Retrieval System (CBIR)* yang asas dan digunakan bagi tujuan pembacaan imej semula secara am adalah tidak berkesan dengan imej-imej muka khususnya apabila querynya adalah dalam bentuk penggambaran pengguna.CBIR yang tersediaada adalah berasaskan pada ciri-ciri tahap rendah seperti warna, tekstur dan bentuk, maka it tidak boleh menangkap aspek-aspek semantic dalam imej-imej muka.Dalam penggambaran apa-apa yang di perlukannya, manusia secara semulajadi menggunakan ciri-ciri semantik (ciri-ciri tahap tinggi) dan biasanya menghadapi kesusahan-kesusahan apabila membuat penggambaran dengan ciri-ciri tahap rendah. Ini adalah kerana manusia biasanya mengertikan imej-imej muka dan membandingkan persamaan-persamaan antaranya dengan penggunaan ciri-ciri tahap tinggi seperti jantina, bangsa, usia dan skala penilaian sifat sifat muka. Dengan yang demikian, manuia tidak dapat menghubungkan konsep-konsep semantik tahap tinggi ini secara langsung dengan ciri-ciri tahap rendah.Dalam penyelidikan ini kita telah mencadangkan sesuatu *Semantic Content Based Facial Image Retrieval System* (SCBFIR) yang dapat menggabungkan pelbagai ciri-ciri visual bersama ciri-ciri semantik bagi tujuan meningkatkan lagi kejituan pembacaan semula imej muka dan mengurangkan perbezaan semantik di antara keperluan-keperluan query tahap tinggi dan ciri-ciri tahap rendah yang terdapat dalam imej muka manusia.Ciri-ciri semantik telah dipilih dan diberikan nilai berat yang berasaskan pada sesuatu kajian kes yang melibatkan penyertaan 100 peserta.Ciri-ciri visual dan semantik dapat diekstrakkan melalui keadah-keadah yang berbeza supaya mereka dapat nilai-nilai berat yang berlainan.Suatu rangkaian

neuro telah dicadangkan untuk (i) pengukuran jarak antara vektor-vektor query dan vektor-vektor pangkalan data bagi pelbagai ciri-ciri yang berlainan bagi tujuan kajian persamaan-persamaan dan (ii) pemberian nilai berat dan penggabungan persamaanya.

Sesuatu keadah kebarangkalian dapat digunakan untuk mengurangkan perbezaan-perbezaan yang dilihati dari segi persepsi manusia dan padangan-padangan yang dihasilkan semasa anotasi imej dan/atau proses query.Dengan ini, sesuatu sistem prototaip bagi pembacaan semula imej muka manusia dapat dibangunkan bagi tujuan menguji prestasi pembacaan semula.Sistem ini dapat dilatih dan diuji sekali pada dua pangkalan data, iaitu pangkalan data imej-imej muka manusia ORL daripada makmal AT&T, Cambridge dan pangkalan data imej-imej muka manusia tempatan yang merangkumi 150 perserta daripada staf Universiti Malaya, Kuala Lumpur dan juga kawan-kawan dan ahli keluarga mereka di luar universiti.Keputusan-keputusan eksperimen menunjukkan bahawa keadah SCBFIR telah mencapai prestasi yang paling baik berasaskan pada bilangan ciri-ciri semantik yang digunakan berbanding dengan teknik yang tipikal bagi pengenalan dan pembacaan semula imej-imej muka. Teknik yang tipikal bagi pembacaan semula imej-imej muka dapat mencapai kejituan sebanyak 80.60% dan 89.51% masing-masing bagi pangkalan data ORL dan pangkalan data tempatan untuk 10 imej muka teratas yang dibaca semula. Walaubagaimana pun teknik SCBFIR mencapai kejituan sebanyak 97.85% dan 93.39 % bagi kedua-dua pengakalan data masing masing.

ACKNOWLEDGEMENTS

It gives me great pleasure to express my gratitude, appreciation, and sincere thanks to my supervisor, Associate Prof. Datin Dr. Sameem Binti Abdul Kareem, for her support and assistance in pertinent activities of my research. Her guidance, wise academic advices, encouragement of independent thinking, valuable suggestions, and never-ending pile of ideas, has made the completion of my research possible.

I would like to thank all the people who have participated in part of the research and have allowed me to take photos of them for the purpose of the research documentation.

I would also like to express my gratitude and appreciation to all staff of the Faculty of Computer Science and Information Technology, University of Malaya for their support, kind assistance, and collaboration to facilitate the needs relevant for my research.

On a personal level, I would like to express my deepest gratitude to my relatives and friends for all their encouragement I have received during my research works.

Finally, and most importantly, I would like to thank my family members for their everlasting love, patience, and support, I required dearly while studying as well as conducting research works.

TABLE OF CONTENT

CHAPTER THREE: SEMANTIC BASED -IMAGE RETRIEVAL

CHAPTER FOUR: FACIAL FEATURES EXTRACTION AND CLASSIFICATION

CHAPTER FIVE: RESEARCH DESIGN AND METHODOLOGY

CHAPTER SIX: EXPERIMENTAL RESULTS AND DISCUSSION

LIST OF FIGURES

LIST OF TABLES

Table A.9: Lip thickness representation 240

LIST OF ABBREVIATIONS

AdaBoost	Adaptive Boosting
BP	Back-Propagation
CBFIR	Content-Based Facial Image Retrieval
CBIR	Content-Based Image Retrieval
ED	Euclidean Distance
FERET	Face Recognition Technique.
GD	Gradient Descent
HSI	(Hue, Saturation, Intensity)
HSV	(Hue, Saturation, Value)
IDM	Image Database Model
IRM	Image Retrieval Model
LDA	Linear Discriminant Analysis
LSI	Latent Semantic Indexing
MSE	Mean Squared Error
PCA	Principal Components Analysis
QBE	Query By Example
QBIC	Query By Image Content
RBF	Radial Basis Functions
RBFN	Radial Basis Function Network
RF	Relevance Feedback
RGB	(Red, Green, Blue)
SCBFIR	Semantic Content-Based Facial Image Retrieval
SSE	Sum Squared Errors
ST	Semantic Template

CHAPTER 1
INTRODUCTION

1.1 Introduction

Due to the availability of image capturing devices such as digital cameras and image scanners, there has been a significant increase in the huge collections of digital images and videos lately from various domains, including fashion, crime prevention, publishing, medicine, architecture, etc. These collections of digital images need to be managed resulting in the requirement for efficient methods for the archival and retrieval of these images. The search for solutions for image retrieval problems is becoming an active area for research and development.

Facial images have gained importance among digital images because of its use in various aspects of life, such as in airports, law enforcement applications, security systems, and automated surveillance applications.

The face is the most significant component of the human body that are normally used by humans to recognize each other; thus, facial images are the most common biometric characteristics used for human verification and identification (Jain, Hong, & Pankanti, 2000). Numerous works are emerging for various purposes of face identification, verification, and retrieval used for different applications of facial images.

A face retrieval problem is concerned with retrieving facial images that are relevant to user requests from a collection of images. The retrieval is based on the visual contents and/or on the information associated with this facial image.

Content-based facial image retrieval (CBFIR) is a computer based vision technique that is applied to the problem of facial image retrieval, especially when searching for digital images of faces in a comprehensive database with similar features, and making the exact retrieval of the target face is difficult or almost impossible through traditional techniques such as content-based image retrieval (CBIR) and face recognition technique (FERET). Although the main purpose of a face recognition system is to find the facial images of the same person for identification or verification task, a face retrieval system is also required to figure out facial images that look similar to the query face (Datta, Joshi, Li, & Wang, 2008).

The basic image retrieval system mostly use visual features, such as color, texture, and shape features. These features are usually referred to as low-level features. Low-level features are extracted automatically using image processing methods to represent the raw content of the image. Image retrieval based on color usually yields images with similar colors, whereas image retrieval based on shape yields images that clearly have the same shape, and so on (Datta, et al., 2008; Lew, Sebe, Djeraba, & Jain, 2006). From the discussion above, such systems used for the general purpose of image retrieval using low-level features are not effective for facial images, especially when the user query is a verbal description, since the semantic aspects of a face are not captured with these features, while humans in their nature tend to use the semantic descriptions (high-level features) in order to find what they are looking for, and they encounter difficulties in using the language of low-level features, for instance, color and texture. This is because human beings normally perceive facial images and compare their similarities using high-level features such as gender, race, age, and the rating scale of the facial traits, and thus cannot relate these high-level semantic concepts directly to the low-level features. Traditional systems use visual features and are usually based on a query by example strategies for navigating through the image database. If an example image is not available, such systems are not likely to perform the task of facial images retrieval efficiently. Generally, facial images differ from other images because facial images are complex, multidimensional, and similar in overall configuration.

There have been many discussions on image visual features and the ability of the human for face recognition. Jain et al. (Jain, et al., 2000) indicated that it is questionable whether the visual features of the face itself, without any contextual information, is a sufficient basis for recognizing a person from a large number of identities with an extremely high level of confidence. This is confirmed by Sinha et al. (Sinha, Balas, Ostrovsky, & Russell, 2007) who suggested that humans are good at recognizing because they process the input facial features holistically. Image contains much information that can be perceived easily by human vision, but is still difficult to extract automatically. The human ability to recognize faces and distinguish individuals is effective at distance and under different illumination and weather conditions.

Human beings are much better than computers at making use of high-level semantic information from facial images. A complete facial image understanding consists of interpreting face image objects and their relationship (Datta, et al., 2008; Lew, et al., 2006).

Although, “a picture is worth a thousand words”, one of the best methods used to represent high-level concepts in a computer system is the text-based description. Different ways have been used to incorporate textual information into image retrieval. Up until now, neither of these two types of features has individually been satisfactory in retrieving facial images namely text-based description and visual features. There is still a huge gap that needs to be filled in the area of these researches.

The proposed work in this research is a semantic-content based facial image retrieval (SCBFIR) model that incorporate multiple visual features with semantic description. The aim is to increase the accuracy of the facial image retrieval and to reduce the semantic gap between the high-level query requirement and low-level facial features of the human facial image, enabling the model to meet human natural tendencies and needs in the description and retrieval of face images.

Visual features represent the raw content of the human facial image, while the semantic features are obtained by textual annotation. Semantic features were selected and weighted based on a case study, with the participation of one hundred respondents. Visual features and semantic features are extracted by different methods, so they have variant weights. There is therefore, a need for distance measurements between the vectors of these features in order to measure the degree of similarity of each semantic or visual feature. Some features may be considered more important than others, so features weighting is used to distinguish the importance of the various features. A Neural network is proposed for both, measuring the distance between the query vectors and the database vectors of the different features for similarities finding, and for weighting and combining the similarities. A probabilistic approach is used to improve the differences observed based on humans perception and the viewpoint that may appear during image annotation and / or query process.

1.2 Research Importance

Images and videos have dramatically entered our lives excessively throughout the last decades. They are indeed likely to play an increasingly important role in our live; this is because of the advances in digital imaging technologies and devices. The steady growth on the number of digital images generated and an explosion in the amount of digital images available has led to an increase on data storage capacity. The difficulties of locating a desired image in a large and varied collection are now the current main problem in this field. In order to search in such a large and varied images' collection, there is a growing need for efficient storage and retrieval techniques.

Image retrieval systems are developed in order to search the target image more easily, speedily, and at a lower cost of retrieval. In content-based image retrieval techniques, the visual features particularly color, texture, and shape are extracted as uncorrelated characteristics based on pixel values, and aggregated information derived digitally from larger segments of the image. The techniques in such system uses the representation of these features that reflect a global description of images to calculate the similarity and matching between images without considering the physical extension of objects on their primitive features and do not take into account the image contents. This leads to the failure to consider the implicit semantics of an image. As such, the CBIR approach is still far from enabling semantic-based access, in other words, the inability of automatic understanding. This is one of the limitations facing the current CBIR system, humans compare and measure the similarities between images, and the semantic content found therein, whereas a computer-based system uses low-level features and image semantics is not intrinsically expressed in image pixels. Humans are interested in the content of images at the semantic level, e.g., humans, looking at a facial image; will consider the features of the face parts (and their correlation) and other description such as gender, age, etc. They will expect to retrieve the target facial images from a database, while a computer-based system would "look for" images with certain features such as color, textures and shape. The mismatch between human expectations and the system performance gives rise to the difference between the humans' frameworks for interpreting the semantics description of the query image and the aforesaid low-level features abstraction from the visual content- leading to the semantic gap.

Suitable ways of describing image content is by text (concept) because humans understand and expressed things in keywords more easily. Expressing characters using keywords symbols are more effective compared to specifying exactly using visual features. The symbolic features are conceptual, and they are easy to manipulate.

Therefore, users create their queries with a higher semantic level while an image-processing algorithm extracts visual data at a non-semantic level. Therefore, it is very important to bridge these two levels together and support the mapping of low-level visual features to the high-level semantic concepts. Thus, we need to deal with two types of data, visual and textual information. Metadata that is extracted from visual content and text caption should be integrated to facilitate the semantic based image retrieval system.

A promising idea is to represent images as 'words' analogous to text retrieval solutions. Using text caption enhances the image classification and interpretation process. The matching query way reflects human similarity judgments, understanding users' needs, and information-seeking behavior. In this research, a combination of textual information of the human facial image description with visual features information has been proposed to improve image search results.

Content based image retrieval systems has gained interest among research scientists for efficient image searching, browsing and retrieval methods that are required for various domains and applications , these are (da Silva Torres & Falcão, 2006; Liu, Zhang, Lu, & Ma, 2007):

- Journalism and Advertising
- Education and Training
- Biodiversity Information Systems
- Travel and Tourism
- Crime prevention including Fingerprint Recognition and Face Recognition
- Home Entertainment
- Fashion, Architecture and Engineering
- Surveillance System
- Historic and Art Research
- Digital libraries
- Medical Diagnosis
- Web Searching

The corresponding application area of the proposed facial image retrieval system is in law enforcement agencies. This is to assist witnesses to use their verbal descriptions of suspects to retrieve the facial image of the suspect from the police records of the past suspects' facial image. The facial image is probably one of the most important tools in criminal investigation where identification is often the hardest part of a police investigation. Law enforcement agencies usually keep large archives of visual evidence; one of them is the past suspects' facial images that are known as mugshots or booking photographs. In the context of law enforcement, the mugshot registers a photographic record of the arrested individual for victim and investigator identification. Whenever a crime occurs, they can compare the description of the suspect in the crime from eyewitnesses, who can provide the similarity to the records in their archives. Hard-copy image formats were the initial support for crime prevention with maintenance, storage room, difficulty of search and retrieval contributing for its secondary role. Digital images, the soft-copy format, are the current alternative. The police have a computerized facial image system; containing a huge database of images. On the other side, the witnesses always have a mental image of the suspect alone. The description of the suspect whom they give is generally verbal in nature. Since the database is large, manually inspecting every image is impractical, how will they find a particular face in this very large database? In addition, to sketch a suspect's face as described verbally by a witness would entail time not only for sketching but also for matching the sketch with available facial images. However, in the proposed system, the verbal description provided by the witness will be matched with the semantic descriptions of faces in the database. The former description can be the input into the system while the latter process gives the output comprising a list of ranked facial images of past suspects in the database. The proposed system has provided methods, which narrow down the number of images to be searched in the database for matching with the queried image.

Using the system in law enforcement applications for searching through a database of criminals is an application example, while there are various other areas where the need for efficiently retrieving the facial image is required , and the proposed research can be used. Another one of them is for personal use, such as, searching through a personal collection of facial images and the other is commercial applications, for instance, searching through the web.

1.3 Problem Statement

Despite the various efforts to improve the image retrieval system during the past years, the current systems still suffer from many problems that can degrade their performance and keep them from achieving users' expectation. Two main issues of these problems that are associated with the current system and keep floating on the surface. The first issue is the semantic gap and the second issue is the subjectivity of human perception.

1.3.1 Semantic Gap Problem

Human facial image features can be obtained from the whole image, or from segmented parts of the image. Image features include visual content that is so called low-level features automatically extracted using computer vision techniques, and semantic content that is so called high-level features. Semantic content is described either directly through descriptions and textual annotations or by complex inference procedures based on visual content (Long, Zhang, & Feng, 2003).

Most of the current systems are specialized for image based matching and retrieval based on the low-level features such as color, texture and shape. Such systems used for the general purpose of image retrieval using low-level features are not effective for facial images, especially when the user query is a verbal description. The systems do not capture the semantic aspects of a face. However, humans naturally tend to use verbal descriptions of the semantic features (high-level features) to find what they are looking for, and they encounter difficulties in using the language of low-level features and cannot relate these high-level semantic concepts directly to low-level features. There is no direct link between the high-level concepts and the low-level features.

Humans use keywords to characterize the face. These keywords are assigned corresponding to the representative visual features, visual impression, and inspired impressive features of the facial image. However, how the low-level features of the facial images correlate are significant to be used as features for comparing between the images in the features space in order to find the similarity, this similarity does not correlate with the similarity perceive by the humans between these facial images.

We summarize the previous discussion in the following:

- The basic content-based image retrieval (CBIR) system used for the general task of image retrieval is not effective for facial images.
- Most of researches and systems that were done in the field of content-based facial image retrieval (CBFIR) were based on low-level features such as color, texture, eigenfaces, etc. thus, it cannot capture the semantic aspects of a facial image.
- The low-level features cannot describe the high-level semantic concepts in the user's mind and high-level semantic concepts cannot directly relate to low-level features.
- Humans by nature tend to use semantic descriptions when describing what they are looking for, such as gender, race, age and the rating scale of the facial trait.

They perceive facial images and compare their similarities using high-level features.

- Similarity amongst the semantic features is not equal to the similarities in low-level features.

The above discussion described the limitation in current systems which created a gap called semantic gap that is defined as follow : “the lack of coincidence between the information that one can extract from the visual features and the interpretation that the same features have for a user in a given situation“. In other words, “It expresses the disagreement between the low-level features that can be extracted from the images and the descriptions that are meaningful for the users” (Datta, et al., 2008; Smeulders, Worring, Santini, Gupta, & Jain, 2000).

That was the first issue in the facial image retrieval. Figure 1.1 shows the issue of the semantic gap problem.

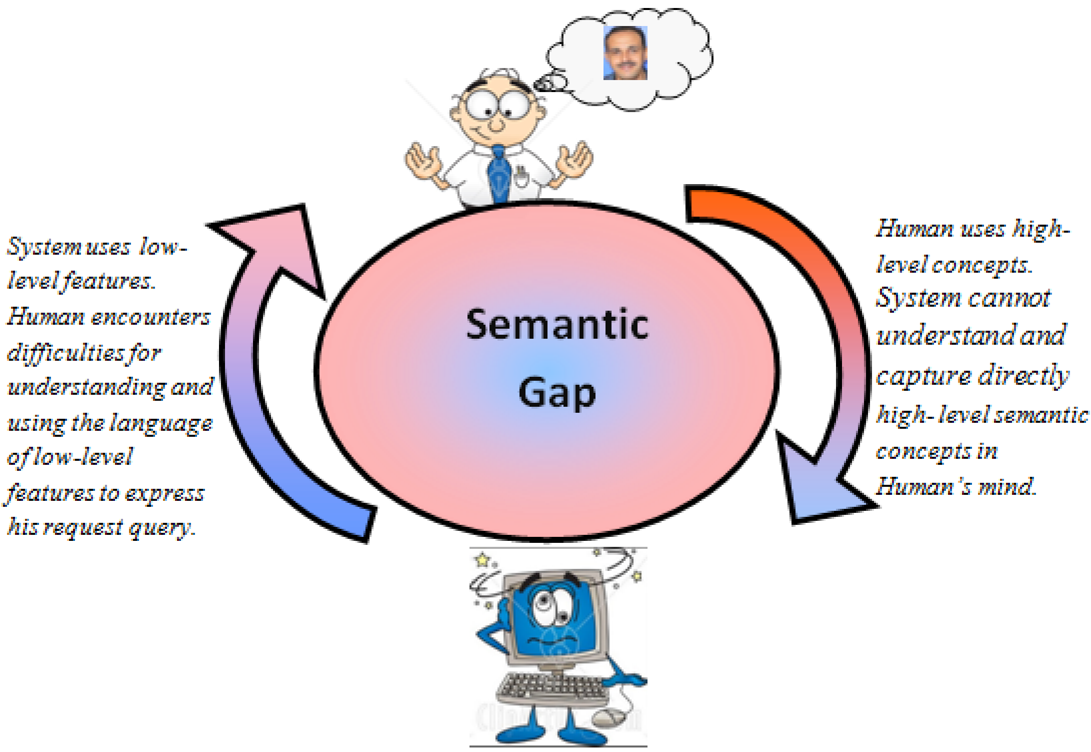

Figure 1.1**:** Illustration of the semantic gap problem, human uses high-level concepts, while systems use low-level features.

1.3.2 Subjectivity of Human Perception

Semantic attributes play a very important role in facial image recognition and retrieval because human facial image include a variety of these semantic attributes that are used for recognizing faces and characterizing them. However, the problem appears during the description of these features. Human perception and viewpoint are considered as subjective aspect, which means that different people may perceive the same facial images differently and give them different rating scales of description. As an example, one person may describe a facial image as having a short beard and flat nose while another person or the same person under a different situation may perceive the same facial image features differently. The issue of the subjectivity of human perception in facial image retrieval is depicted in Figure 1.2.

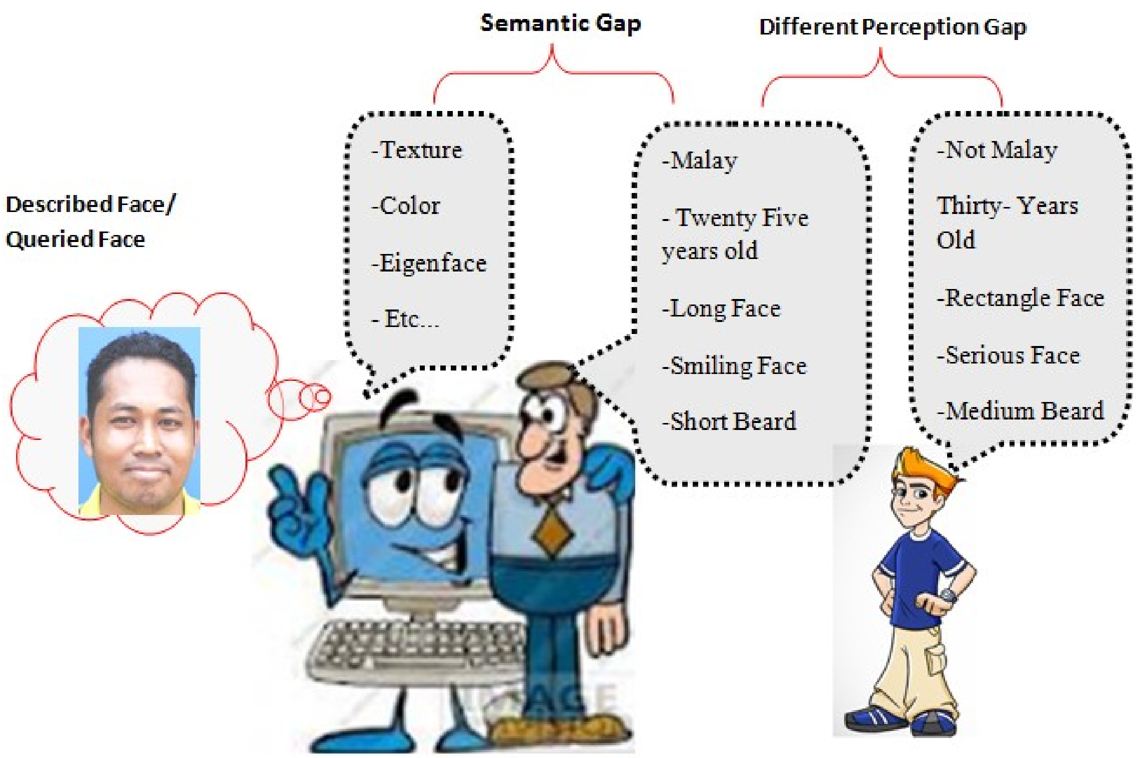

Figure 1.2: Illustration of human perception subjectivity in facial image retrieval.

1.3.3 Research Objectives

Based on the above discussion and facts of the facial image retrieval problems, the main goal of our research is to develop a semantic-content based facial image retrieval (SCBFIR) technique; towards reducing the semantic gap problem, and enabling the facial image retrieval system to meet human natural tendencies and needs in the description and retrieval of facial images. The more detailed objectives of our research include the following:

- To develop a model that links the high-level query requirement and the low-level facial features of the human facial image.
- To retrieve facial image based on the high-level query requirement and the low level facial features efficiently and accurately.
- To compare the performance of face retrieval technique (based on low-level features) with the developed model.
- To improve the differences observed based on humans 'perception and the viewpoint that may appear during image annotation and/or query processing

Our research aims to investigate the methods that can improve the performance of the content-based facial image retrieval (CBFIR) technique. The research aims also to address the issue of combining the heterogeneous attributes of visual features and semantic features using efficient and accurate method for improving the performance accuracy of the semantic facial image retrieval and enabling the user to specify his/her query through the query by example together with the natural language descriptions.

1.4 Significance of Study

This research has proposed a new approach for facial image retrieval from a large image database. Most existing image retrieval systems are based on low-level image features for facial image retrieval without considering associated image semantic features. Given that a facial image carries a wealth of information, user expectation is therefore not met if a face is described using only single image features. Therefore, a combination of image features should be considered.

The proposed research aims to reduce the gap between content based and semantic based image retrieval systems and improve the retrieval performance. It has focused on using high-level semantic information and low-level information together to enhance the image retrieval. Its performance is evaluated to indicate the degree of improvements made. This research has indeed generated an improved method of facial image retrieval. The results of the study are contributed to the identification of new method.

1.5 Thesis Organization

The thesis is organized in seven chapters as follows:

- **Chapter One** (Introduction)

  In this chapter, we presented an introduction to our research issues, research motivation, importance, and application of the research, the problem statement, the aims, and objectives, and the outline of the research approach.

- **Chapter Two** (Content-Based Image Retrieval)

  In this chapter, we present a review of previous literature and studies relevant to the field of content based image retrieval. The chapter gives an overview of the CBIR systems and their various components. A related works in content-based facial image retrieval systems are discussed.

- **Chapter Three** (Semantic Based -Image Retrieval)

  In this chapter, we present a review of previous literature and studies relevant to the field of semantic image retrieval. A related works in semantic image retrieval and facial image retrieval systems are discussed.

- **Chapter Four** (Facial Features Extraction and Classification)

  In this chapter, we simplify and explain the details of facial features extraction and classification techniques used in this research. The techniques and approaches chosen and applied were based on the literature review in order to achieve our research objectives.

- **Chapter Five** (Research Design and Methodology)

  In this chapter, we describe and explain the research methodology used, including research design, proposed methods, procedures adopted, and data and the method of its collection.

- **Chapter Six** (Experimental Results and Discussion)

  In this chapter, we present the research results in the form of text, figures, and tables. We present a discussion and analysis of the research results. The results finding based on the comparison of previous studies are also presented.

- **Chapter Seven** (Conclusion and Futures Works)

  In this chapter, the findings are summarized and their implications discussed. The section also includes suggestion for future works.

CHAPTER 2
CONTENT-BASED IMAGE RETRIEVAL

2.1 Introduction

Among the key tasks of computer science, is the management of digital information. In the initial stages of development, when most of the data comprised of text and numbers, storage and searching were well administered by relational databases. However, because of the rapid growth of multimedia technology and an increase in image and video accumulations, the need for workable and efficient image retrieval techniques and the management of visual data has resulted in substantial research efforts in providing the needed tools.

There has been noteworthy advancement in both system development and theoretical research. However, many challenging research problems persist, which continues to attract researchers from multiple disciplines.

Image retrieval is an extension of the conventional information retrieval. Image retrieval techniques are in some ways extrapolated from established information retrieval methods, and are designed to manage the enormous amount of more versatile visual data (Lew, Sebe, Djeraba, & Jain, 2006).

Traditional information retrieval was founded mainly on text, and methods of textual information retrieval have been introduced to image retrieval in diversified ways, an example of this is traditional indexing for image retrieval which is text-based (Jörgensen, 1998). Increased interest in developing image-based solutions have arisen due to the insufficiency of text-based access to images. However, “a picture is worth a thousand words” and thus, image contents are much more impactful as compared with text, and the quantity of visual data is already extensive and still rapidly growing. Image retrieval is based on the availability of a representation scheme of image content , In the hope of dealing with these particular characteristics of visual data, content-based image retrieval methods have been introduced (Gimel'Farb & Jain, 1996; Yoshitaka & Ichikawa, 1999). Probably the most rapidly maturing application of similarity searching is content-based image retrieval, because of the limitations underlying the metadata-based systems, as well as the extensive range of potential applications of efficient image retrieval. Content-based methods try to overcome the drawbacks of text-based retrieval systems, by harnessing the advantages of the visual content of images. The evaluation of visual similarity is a natural process for people. This makes image search ideal for evaluating content-based retrieval performance.

Content-based image retrieval approaches use low-level visual features that are directly related to the perceptual facets of the contents of the image. The majority of these features are simple to extract and representing the similarity measures of these attributes using their statistical properties is convenient (Grosky, 2011).

In the content-based image retrieval technique, the images are indexed as a set of attributes. When queried, the information is extracted from previously calculated image attributes, instead of retrieving by requesting information directly from the images. A variety of content-based image retrieval techniques have been introduced in the past few years, and there are many researchers have been carried out in retrieval based content, which has been employed in many applications such as in internet searches, medical diagnosis, and trademark images. Content-based image retrieval technique is still very active research area with investigations on different image features and different features extraction methods for image retrieval (Datta, Joshi, Li, & Wang, 2008) .

The efforts in the techniques of image retrieval focused on a ‘query by example image’ paradigm. The user’s query cannot be a basic description of the requested image content such as ‘find images containing a particular human facial image’. An example image or a sketch of a face is submitted to the search engine instead.

Content-based retrieval is not dependent on mapping the content in its entirety. The description has to fit the retrieval methods, which are based on similarity. The key problem stems from trying to interpret what people perceive as similar.

Image retrieval methods should offer support for user queries in an effective and viable way, just as traditional information retrieval supports textual retrieval. However, because of the dynamic and variable characteristics of the contents of an image, costly computing and advanced methodologies are needed to process the image; visualize data, and measure similarities.

The content-based search is not yet well developed for easy use by the public. In some situations, they do not satisfy user expectations, although the search results are acceptably promising in other cases (Datta, et al., 2008).

Several factors should be considered in the image retrieval (Datta, et al., 2008). These factors include:

- a clear idea of what the user wants,
- where would the user prefer to search,
- in what form does the user post her query,
- how would the user want the results presented, and
- what is the nature of the inquirer's input/interaction.

Users may need access to images based on abstract concepts and symbolic imagery, or may select access to images through vague features such as texture, color, or shape. Currently perceiving how individuals relate to visual information is inadequate, and the technology to access these images has intensified exponentially. Low-level visual features often fail to depict the high-level semantic concepts in the user’s mind. Image semantics cannot be modeled by describing low-level features with sophisticated algorithms. In general, the relatively more important meanings of objects and scenes in images that are perceived by humans are represented by high-level concepts, though they are not expressed directly from the visual contents. The users' preferences, viewpoint, and subjectivity determine these conceptual aspects (Zhao & Grosky 2002). It is not an easy task for the machine to mine and manage meaningful semantics features automatically from the submitted image and to use them to make the retrieval step more intelligent and user-friendly. On the other hand, text descriptors usually represent high-level conceptual information conveniently.

Hence, there is a pressing need to bridge the gap between the low-level features and high-level concepts, and integrating them from a different perspective. Consequently, some of the image retrieval research communities focus their attention on the semantic problem that is related to content-based image retrieval, and its effect on the retrieval process. Hence, the research focus has been shifted from the abstract content-based image retrieval into reducing the 'semantic gap' between the visual features that are represented in the machine and the richness of human semantics. Figure 2.1 depicts the hierarchical progress in image retrieval techniques. The semantic gap refers to the inconsistency between the limited descriptive power of low-level image features and the richness of users' semantics (Liu, Zhang, Lu, & Ma, 2007).

Humans contrive texts and their interpretation. Pictures on the other hand are a projection in a real scene. While images are a combination of pixels which have no importance by itself, texts are made up of characters, and each character has a meaning. While character arrangement is predictable, pixel combinations are not. A machine can easily interpret text semantics; however, it cannot easily understand image semantics, as image semantics are not intrinsically expressed in image pixels.

Due to the lack of any integrated structure for image representation and retrieval, some methods may perform better than others under varying query conditions. Text based retrieval systems supply natural query interaction, but as they do not use any image data, may provide noisy results. Image based systems frequently give similarity based results when a sample image is used to retrieve similar looking images. It is intelligible that the image retrieval techniques should consist of an integration of both low-level visual features covering the more detailed perceptual characteristics, and high-level semantic features implicit in the broader conceptual aspects of visual data. Therefore, to facilitate efficient image data management, these schemes, and retrieval techniques need to somehow be merged and adjusted (Zhao & Grosky 2002).

The main areas of the works related to this research are:

1. textual information as related to image retrieval,
2. image content as related to image retrieval, and
3. semantic image retrieval.

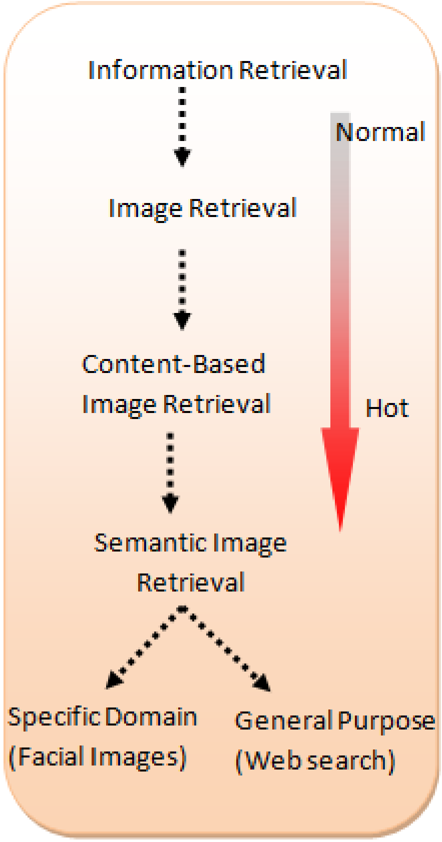

Figure 2.1: The hierarchical progress
in image retrieval techniques.

2.2 Image Retrieval Model

Before we review the existing image retrieval models and techniques, it is important to look at the general image retrieval model. Image retrieval model (IRM) covers the specification of the following: an image database model (IDM), and a query specification language for expressing user queries, and a matching or retrieval algorithm for retrieving relevant images from an image database to answer user queries (Gudivada, Raghavan, & Vanapipat, 1994; Stanchev, 1999). The image retrieval model is unique because of its broad coverage of image features. In the image retrieval model, the general image description model representation is used for searching the images where the model is based on similarity retrieval.

Let the vectors $(X_1, X_2, \ldots X_n)$ represent the features of the database of a set of images. Each image in the database has the description $X\ (x_1, x_2, \ldots x_m)$. If we suppose the query is submitted through the general image data model in an image description $Q\ (q_1, q_2, \ldots, q_m)$, then each database image, X_i is compared with the query image Q using an appropriate comparison technique, such as distance function for numerical value. The similarity value (SV) between Q and X_i is defined as $SV=$ distance (Q, X_i). The similarity can be calculated in different ways according to the Q and X_i content. That content can be symbolic, numerical or linguistic values, histograms, pictures or spatial representation characters (Deselaers, Keysers, & Ney, 2008).

2.3 Image Database Model

An image database model is used to realize the general method of image knowledge representation and, usually helps to better manage the image retrieval task and introduce the nomenclature that is related to image attributes. It determines the view(s) of the image data, and is a means of depicting entities of interest in images, their geometric characteristics and attributes values, and associations with objects within images. It is a form of data abstraction used to depict the conceptual data representation and an assemblage of concepts that may be employed to describe a database's structure. The database structure comprises of types of data, relationships and restraints that relate to the data, and can also contain a list of operations for database retrieval (Gudivada, et al., 1994; Stanchev, 1999). Generally, each of these schemes forms a symbolic image for each rendered physical image, and to reduce the search space, symbolic images are then used together with the index structures as proxies for image comparisons. Once a measure of similarity is ascertained, the actual matching images are retrieved from the database.

Because of the absence of any integrated framework for the image representation, storage, and retrieval, these symbolic representation schemes have greatly improved image management (Tao & Grosky, 1998) .

Various schemes for data modeling and image representation have been suggested. Figure 2.2 shows the schema of the image database model. The model constitutes the taxonomy founded on the systematization of existing approaches (Stanchev, 1999). It includes:

- Language approach, in which language phraseologies are used for physical and appropriate image content descriptions.
- Object-oriented approach, in which the image and the image objects are handled as objects containing relevant functions to calculate its functions. In terms of the object-oriented approach the image itself together with its semantic descriptions is processed as an object. The image is presented in two layouts (classes) - logical and physical.

Logical attributes refer to the attributes used in describing the properties of an image, regarded as either an integral entity or a collection of component objects. Logical attributes evidence the characteristics of an image and its constituent objects at different levels of abstraction.

The terminology that is associated with image attributes are categorized by three broad categories:

- meta attributes,
- semantic attributes, and
- content based attributes.

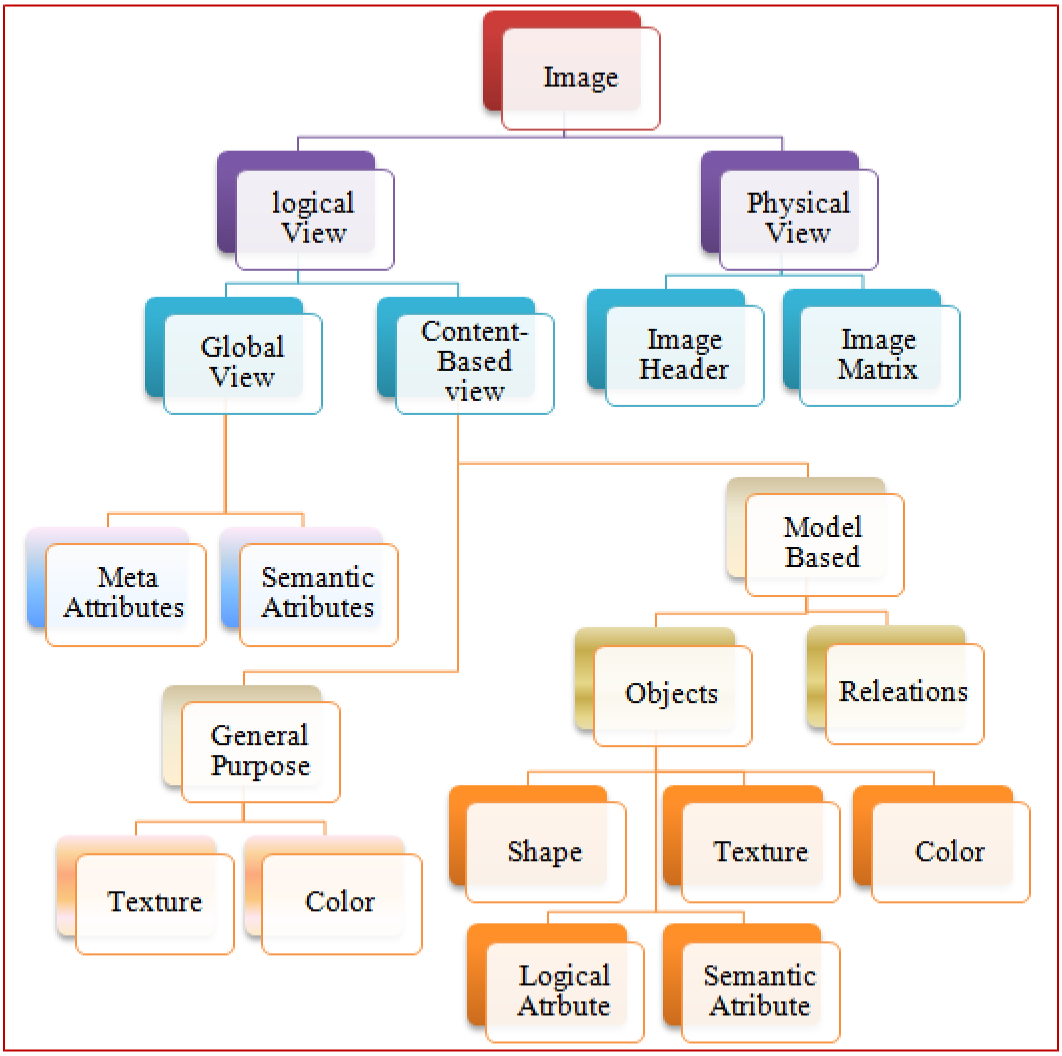

Figure 2.2: Schema of the image database model.

Meta attributes refer to those characteristics of an image that are derived externally and do not depend on the image content. Image acquisition date and identification number are some of the attributes that may be included. Image meta attributes refer to meta attributes that relate to the whole image, and the meta attributes that apply to an images' constituent objects are termed image-object meta attributes. High-level domain concepts, which the images manifest, are described using semantic attributes. Content based attributes, include the general-purpose attributes, such as the texture and color of the image, or model based attributes, such as objects and objects relations (Gudivada, et al., 1994).

2.4 Information Retrieval

Information retrieval encompasses the area of study concerned with searching for documents, for the information in documents, and for the document metadata. Information retrieval traditionally refers to retrieving documents containing text from a single source. It was developed to include information retrieval in the form of images, audio, and video from various sources. In automatic information retrieval, users submit their query to a system to search for relevant information such as from the internet. The information will be extracted and retrieved from the data store based on their relevance to that query. Text-based information representation is the standard recognized method for information retrieval. This method is known as text-based or keyword-based information retrieval. The first few automated information retrieval systems were commenced in the 1950s and 1960s (Deselaers, Weyand, Keysers, Macherey, & Ney, 2006).

2.5 Image Retrieval

Image retrieval is the task of browsing, searching, and retrieving images from a large database of digital images. Image searching is a specialized data search method used to locate digital images. The search can be through a digital image metadata search or digital image visual search where two forms of information are related to the digital image:

- The metadata, giving information about the image.
- The visual features, containing information intrinsic to the image.

Metadata comprises keywords or text while the visual features are derived through computational processes and based on the raw data's pixel values. Computational processes may comprise image processing and computational geometric routines performed on the digital image (Gupta & Jain, 1997). A user has to provide his query terms such as keyword or image example to search for images, and the system will deliver images "similar" to the query. The first automated image retrieval system was developed at Massachusetts Institute of Technology (MIT), in 1980s,by Banireddy Prasaad, AmarGupta, Hoo-min Toong, and Stuart Madnick (Prasad, Gupta, Toong, & Madnick, 1987).

2.5.1 Text-Based Image Retrieval

Long before images could be digitized, access to image collections was provided by librarians, curators, and archivists through text descriptors or classification codes. These indexing schemes reflected the one-off characteristics of a specific collection or clientele.

As defined earlier, text based information retrieval technique focuses on text documents, and is the science of searching for information within documents, or for documents themselves. However, text-based image retrieval (specific information,) is a technique used to retrieve the digital image from a digital image database, based on text or a keyword associated with the image. The text-based image retrieval approach is a well-established tradition within the field of information retrieval. It dates back to the 1970s (Liu, et al., 2007; Long , Zhang , & Feng 2003). In such technique, the images are annotated with text descriptors. The annotations or text-descriptions are rich in keying out semantic content of images. The text based image retrieval system uses the techniques of the conventional document retrieval, for instance, a user presents his inquiry as a keyword or a number of keywords. The query is compared with each text description during the retrieval process, and the images whose text descriptions match the query are retrieved (Salton, 1989).

Long before the advent of the web, text based retrieval has been used to organize keyword retrieval of images. Many techniques have been developed for text-based information retrieval, and they were very successful in indexing and querying web sites. Some of the initial image retrieval systems, including the commercially successful Yahoo image search and Google image search use text analysis to retrieve images. These systems employ text-based methods to retrieve the image, without considering a single pixel. Comprehensive surveys of early text-based image retrieval methods are presented in (Chang & Hsu, 1992; Rasmussen, 1997; Tamura & Yokoya, 1984).

Text-based indexing has lots of strengths including its capability to represent both specific and general instantiations of an image at differing complexity levels. Images can be arranged by topical or semantic structures with text descriptions. Based on the standard boolean queries, this offers easy navigation and browsing.

The most advantage of the text-based image retrieval technique is its ability to capture complex semantics contents contained in the image, such as human emotions - ‘smiling, sadness or angry’, things descriptions ‘big and small ‘and events/actions like ‘dance or bray’. In addition, text-based image retrieval is reliable and quick. However, annotation inaccuracy resulting from the subjectivity of human perception is the main drawbacks. If the descriptions of some features are omitted, or represented by unlikely terms or are different from the standard query terms, the retrieval performance would consequently be poor. The difficulty to describe some visual properties such as certain textures and shapes introduces limitations. This leads to the search for new methods to overcome these limitations, and stimulated interest in content-based image retrieval techniques for retrieving images using visual features. Generally, it is recognized that text-based image retrieval systems, are so far successful and satisfactory for the user.

2.5.2 Content-Based Image Retrieval

A query based visual information method is referred to as content-based image retrieval (CBIR). It is the application of computer vision to the image retrieval problem. Content-based image retrieval emerged as an alternative to the automated text-based image retrieval systems, although text-based image retrieval remains popular in the image retrieval system. Images are greatly rich in information; some information is translated to text description, while other information is captured by their visual representation. As compared to the text-based image retrieval techniques, CBIR techniques operate on a completely different principle, the images are retrieved from a collection by comparing features automatically extracted from the images themselves.

The term 'content-based image retrieval' was first coined and used by T. Kato in 1992 (Kato, 1992) to describe his experiments pertaining to automatic retrieval of images by color and shape, from a database. He was curious about how information on shape and color could be used to query a database of images. His term has since been used more broadly to describe any system, which uses information extracted from the content of the image to search for matches (Vasylenko, 2010). Features extraction techniques extract the visual features of the image to achieve this type of image retrieval, and use it for indexing and subsequent retrieval purposes. The signatures for each image is generated by the database management system while the visual features are extracted, providing a means of comparing visual features between images. An example of this technique is the query by example (QBE) method, which uses an example query image as a seed image to find other images, then applying the signature similarity for comparison (Xiangyu & James, 2003).

Query techniques can be used as criteria for classifying the image retrieval. If the query is represented by using a sample image the retrieval system is called a query by example system, a 'query by text' system uses keywords. Figure 2.3 shows the difference between text-based image retrieval and content–based image retrieval, based on query techniques and retrieval results.

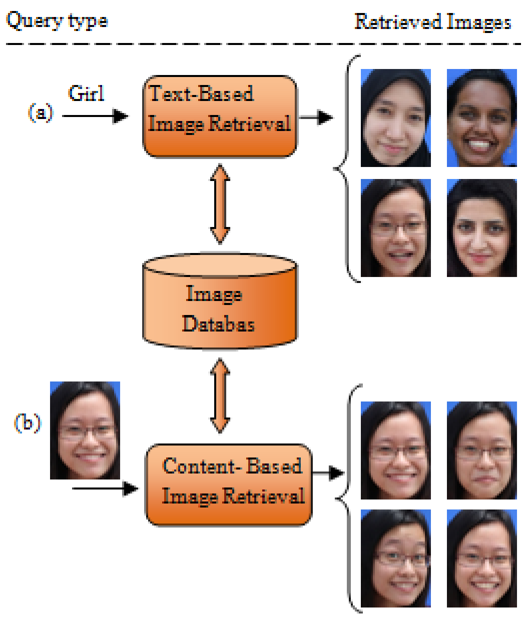

Figure 2.3: Image retrieval classification based on the query types,
(a) : text-based image retrieval technique, use keyword for query,
(b) : content-based image retrieval technique, use image for query.

Content-based image retrieval can be classified depending on the domain of the application into two types. The first type is the general-purpose applications. A query image is used to match with an arbitrary collection of images, such as in web searches. The goal is to retrieve images with similar objects to the query. As an example, a query image with a tree will find all images with trees. The second type is a domain specific application. In this type, the query image is used to match to a collection of images of a particular type. Such as, in facial images applications, fingerprints, X-ray images of a specific organ, and images of skin lesions.

2.5.3 Review of Content Based Image Retrieval

During the past several years, many different image representations have been developed. Several content-based image retrieval techniques have been suggested based on classifying images by their content using the low-level features. The most common low-level image features are texture, shape, color, and spatial layout. Since these low-level features are not enough by itself to represent image contents on the object level, researchers have concentrated on integrating different features, or different feature representations.

Some of the earlier commercial products and academic retrieval systems developed during the last decade are the CBIR systems in use by IBM's Query By Image Content (QBIC) described by Flickner et al, (Flickner et al., 1995), and VIRAGE system (Gupta & Jain, 1997) in commercial domains. In the academic domain are the MIT Photobook system (Pentland, Picard, & Sclaroff, 1996), and the WebSEEK system (Smith & Chang, 1997b) among others.

There has been a measurable increase in research publications on the techniques of user query and interaction, visual information extraction, organization, indexing and database management. Comprehensive reviews and surveys of these techniques during this period are presented in (Datta , Li , & Wang 2005; Rui, Huang, & Chang, 1999; Veltkamp & Tanase, 2000).

2.5.3.1 Low Level Features Based Image Retrieval

Low-level features are those features that can be automatically obtained from the images themselves, and permit us to examine the image's inner workings. Many image retrieval systems have evolved for general or specific image retrieval purposes, based on low-level features. The more expressive visual features are color, texture, and shape.

For image retrieval applications considerable work has gone into designing efficient descriptors for these features (Rui, et al., 1999).

2.5.3.1.1 Color Features

Color is the most widely used visual content feature representation in image retrieval systems. An important contribution is the use of color histograms that characterizes the color distribution in an image. The color histogram identifies the proportion of each pixel's color in an image, simply and in a computationally effective manner. Among the earliest application of color histograms was that by Swain and Ballard (Swain & Ballard, 1991). A high proportion of current CBIR systems now use the variants of this technique (Bagdi, Patil, & Dharaskar, 2013; Eakins & Graham, 1999).

2.5.3.1.2 Texture Features

Texture relates to the visual patterns with homogeneity properties that do not arise from a single color or intensity. Texture can offer additional information on the spatial arrangements and patterns of a varying intensity available in an image. It is an essential element in human vision and has been found to offer cues on the scene depth and surface orientation (Tsai & Hung, 2008).

A variety of techniques has been used to measure texture similarity, based on the texture analysis approach that can be divided into statistical and structural approach. Statistical approach characterizes texture using the statistical properties of the gray levels of the pixels forming the image. These compute the relational brightness of the chosen pixels' pairs of each image. Following this method, it is possible to calculate the image texture properties such as, the degree of contrast, coarseness, directionality, and regularity. Structural techniques characterize texture by texels composition (texture pixels). These texels are arranged regularly on a surface based on some specific arrangement rules.

One of the early works found was by Manjunath and Ma (Manjunath & Ma, 1996) who focused on using texture information for browsing and retrieving the textured regions in images, based on the similarity to automatically-derived code words exemplifying key texture classes within the collection. Gabor wavelet features was used for texture analysis.

2.5.3.1.3 Shape Features

Shape, in this context, does not indicate the shape of the whole image but a specific region of interest. Shape features can represent the spatial information not represented by texture or color. It contains all the geometrical information of an object in the image, which remains unchanged even if the object's orientation or location is changed. Shape information is one of the most difficult features to extract for describing the object(s) of an image, since there is no unified mathematical definitions for shape similarity. Unlike color and texture features, shape features are normally described after segmentation of images into regions or objects. The shape representations can be divided into two categories:

- boundary-based (or edge detection) and
- region-based.

The Fourier descriptor and moment invariants are the most accepted representation for these two categories. Representation should be unvarying to basic transformations such as rotation, scale etc.

Several works used the low-level features for image matching and retrieval. One of the earliest systems is a query by image content system (QBIC) (Flickner, et al., 1995) designed to work on general image databases. QBIC permits the user to use a color wheel in order to select a color to paint a query. The result is an image object that can be directly compared against the database images, which represents the images based on the average color, color distribution, mathematical representation of texture coarseness, shape, and contrast. Another work is that by (Jacobs, Finkelstein, & Salesin, 1995) who used multi-resolution wavelet decompositions of the query and database images. In this system, as in QBIC, the user can paint a rough sketch of the image query in the query image interface. While the query image is being created, the database displays the most similar images, and with every change to the query image, the displayed images are updated. It is difficult and time consuming to construct an accurate specific image from scratch with a painting tool, because the system does not offer an interface for specific images such as facial image to be painted skillfully. Photobook system (Pentland, et al., 1996) provides methods to search for several types of image databases including faces. The image data is compressed into a relatively small set of perceptually significant coefficients that represent the face features, from which a lost version of the original image can be created. The disadvantage of this system is that this system does not capture the specifications of the face given by the users. Another disadvantage of this system is that following the hill-climbing search algorithm strategy. When a user is faced with the local maxima problem that is associated with this strategy, he will be stuck with the same set of images without making any further progress. In PicToSeek system (Gevers & Smeulders 1997) the invariant color image features are specifically extracted from the images. Using the image analysis methods, the collected images are automatically classified into an assortment of image styles and types: JFIF–GIF, gray–color, photograph–synthetic, size, date of creation, and color depth. For the same group of (Gevers & Smeulders, 2000) color and shape invariants the feature set were combined for discriminatory object retrieval from the database consisting of images taken from the multicolored man-made objects. A similar work is WebSeek (Smith & Chang, 1997a). In this system, Smith & Chang semi-automatically classified images into taxonomy of categories, with related text and filename cues. Images within a category or over the entire catalog with similar color contents can then be found by applying a color histogram-based similarity matching. In the Blobworld system (Carson, Belongie, Greenspan, & Malik, 2002), pixels are clustered by their color and texture properties. These clusters supposedly represent the image content using the color distributions, the mean value and the standard deviation, to distinguish similar images from the extensive database.

2.5.3.1.4 Objects and Spatial Relationships

Color, texture, and shape features are used as low-level features for image content representation and retrieval. Besides applying these features, objects and the spatial relationships among objects in an image which are also low-level features are used to represent the image content. The relationships can be to the left or right of the object, inside the object, and above object. Some image retrieval systems compute image similarity using the properties of individual image regions. Region-based visual signatures have been a growing trend in the last decade. Together with advances in image segmentation, improved methods have surfaced. Datta et al. (Datta, et al., 2008) believes the shift towards local descriptors was sparked by "a realization that the image domain is too deep for global features to reduce the semantic gap”. such works that applied this concept was the works by (Carson, Thomas, Belongie, Hellerstein, & Malik, 1999), who represented the image by the number of the image parts, which corresponds to different objects in the image. The features that are used include color, texture, location, and the region's shape. The description of the objects that the image contains can be used in query processing. In the work by (Stricker & Dimai, 1996) , the image is defined by a number of overlapping fuzzy regions. Each region was indexed by three moments of color distribution extracted from the same region. To retrieve images, a measurement function is defined to find the similarity of two color feature vectors.

The Simplicity system by (Wang, Li, & Wiederhold, 2001) define images by groups; graphs vs. photograph and textured vs. non-textured. With such method, a segmentation model is used to define the images firstly by regions. Regions ideally correspond to different objects and then these regions are used for retrieval. With this strategy of grouping and extraction, some semantically adaptive search methods are attempted. Another work is by (Vu, Hua, & Tavanapong, 2003) who introduced an image retrieval system based on regions of interest. Each region contains relevant objects regarding the submitted image. One of the drawbacks of image similarity measure based on image objects is the position dependence. By using a fixed image segmentation strategy, the image objects cannot be rotated within an image. Moreover, each image object may appear differently, depending on the viewpoint, occlusion, and deformation. However, it is more meaningful to represent the spatial distribution of color information based on image objects or regions. With region-to-region similarity as a ground of the comparison, the user has to pick a restricted number of regions from the given image in order to begin a query processing. Consequently, it is often not easy for users to decide which regions he has to use for a particular query. As discussed in (Wang, et al., 2001) due to the uncontrolled nature of visual contents in an image, extracting image objects automatically and precisely is still beyond the reach of the state-of-the-art in computer vision.

2.6 Content Based Facial Image Recognition and Retrieval

Content-based facial image retrieval (CBFIR) is a computer based vision technique that is applied to the problem of facial image retrieval, especially when searching for digital images of faces in a comprehensive database with similar features, and making the exact retrieval of the target face is difficult or almost impossible through traditional techniques such as content-based image retrieval (CBIR) and face recognition technique (FERET). This is because digital images of faces are unique and different from other digital images.

The face is the most significant component of the human body that people use to recognize each other. Consequently, facial images are probably the most common biometric characteristic used by humans to make personal verification or identification, typically based on the location and shape of facial attributes, and their spatial relationships. It is easier for human to identify ethnicity; gender and age of a person from a face. Thus, facial images are high in demand in airports and other public places for automated surveillance applications.

For decades, facial image applications have posed a problem for computer vision, biometrics, and pattern recognition. However, apart from their use as a hard biometric and instead of uniquely identifying a person by his or her face, researchers are now using the “soft” traits of face modality to group people. Face retrieval is one of the more interesting applications that are based on faces as soft biometric.

Usually, there is a need to query a given facial image from a large database to decide its identity for (i) security reasons; facial image retrieval is also concerned with, (ii) human computer interaction applications, (iii) law enforcement applications.

Basically, the fundamentals of content-based facial image retrieval are based on the fundamentals of the CBIR technique and face recognition technique (FERET). Usually the facial image will be retrieved from the database based on the geometric or statistical features of these images. Face recognition systems use query by example to solve identification and verification problems. The recognition processing typically begins with an example of a digital facial image that is submitted to the system to be verified or identified by comparing it to facial images of known individuals in the database. The essential differences between the face recognition and face retrieval is that while a face recognition system's purpose is to recognize the facial images of the same person, because the purpose is to do an identification task, a face retrieval system must retrieve facial images that look very similar to the query face. Another main difference is that user cannot always provide a digital facial image to be used as the query image.

2.6.1 Facial Image Recognition

A lot of works in computer recognition (not retrieval,) of the face have been done, such as by (Alfalou, Brosseau, Katz, & Alam, 2012; AF Alsamman & Alam, 2002; A Alsamman & Alam, 2005; Fromherz, Stucki, & Bichsel, 1997; Tolba, El-Baz, & El-Harby, 2005; Zhao, Chellappa, Phillips, & Rosenfeld, 2003). Two basic methods were applied for face recognition tasks. The first method was information theory-based recognition, where a computational model that best describes a face, is used to extract the most relevant information contained in that face. Eigenfaces approach (Turk & Pentland, 1991) is one, which uses a small set of characteristic pictures to trace the difference between the facial images. Using this method, information that clearly describes a face is extracted from the whole face image. Different algorithms have been developed, two of which have been well investigated in the face recognition literature. These are the Principal Components Analysis (PCA) and Linear Discriminant Analysis (LDA).

Feature based recognition is another technique used for face recognition. Deformable templates and active contour models with excessive geometry and extensive mathematics are applied to extract the feature vectors of the basic parts of a face, such as the nose, eyes, mouth, and chin. Together with their relationships to each other, the information is gathered from parts of a face and then transformed into a feature vector. The example of this method is discussed in (Yuille, Hallinan, & Cohen, 1992), who made a big contribution to adapting deformable templates for contour extraction of face images (Agarwal, Jain, Kumar, & Agrawal, 2010; Atalay 1996). However, such approaches are complex. It is tough to apply to multiple views, and it has often been regarded as quite flimsy, needing a good initial guess to guide them (Turk & Pentland, 1991).

2.6.2 Facial Image Retrieval

Because of the limited inter-class variation in the face database, researchers face a significant challenge in automatic similarity retrieval from a face database. Human faces are structurally the same, with only minor variations between different individuals.

An effective facial retrieval needs a strong features extraction method that is able to attain satisfactory retrieval performance in a larger face database through rigid similarity measures on low-level features. However, many factors can degrade facial image retrieval performance. Intrinsic factors like facial expressions, makeup styles, and aging vary facial appearances, as does extrinsic factors such as illumination and pose variations, and partial occlusions. Without any descriptive information, the geometry of the face itself is insufficient for confidently retrieving a facial image from a large number of identities. These factors further complicate the facial retrieval task, making it the most challenging problem in image retrieval.

In the traditional facial image searching systems (manual search system), users descriptions are usually used for searching and finding faces. Such systems were used by law enforcement agencies employing sketch artists and Identikits (Laughery &

Fowler, 1980). An early attempt to automate such systems was by (Johnston & Caldwell, 1991; Penry, 1974), who developed Compusketch system, a computerized version of the Photofit system, which is used to create composite facial photographs. However, users may have specific details of the semantic description like race sex, or age, and the matching process for the actual retrieval does not consider the semantic descriptions of the face, only the entire facial image.

The FacePrints system (Caldwell & Johnston, 1997) provides an interface for the user to use a composite of facial parts for the face query. Each face is represented by six facial parts, together with a set of position coordinates for each part. A genetic algorithm is used to define mating and mutation operators. Thirty randomly generated composites faces would be displayed, one at a time, according to its similarity to the query image. The user rates each generated image and a new generation of faces is produced based on this rating. This process is continued until the required face is found. Johnston and Caldwell contended that this method is more effective than systems such as Compusketch, as it uses a recognition-based strategy rather than an individual feature recall strategy, and relates better to the way people usually remember faces. They contended that the genetic code for a system developed by FacePrints "may offer a convenient way of searching a database of known criminals to identify those that most closely match a generated composite" (Caldwell & Johnston, 1997). One potential limitation in the FacePrints representation is that a single "bit" mutation could generate a face where one of its parts is the only difference from the original face, but that part may be totally different from the original one. A more gradual change or alteration in the individual's facial features may be caused by another representation, and this might have a telling impact on the performance of the search procedure. Another possible problem with the FacePrints representation is that two perceptually similar faces may appear representationally quite different if they should happen to be composed of different parts that are somehow similar in appearance. Brunelli & Mich in (Brunelli & Mich, 1996) applied PCA to facial features, such as the hair, eyes, nose, and mouth, using eigenfeatures (Turk & Pentland, 1991). Using the interface, the user can slide to select the desired feature's coefficients. The system continuously responds to these selections by updating the reconstructed image. The database then displays faces that are similar to the reconstructed image. The disadvantage of this system is that the image features extracted by the PCA computation do not always correspond to those features that people understand intuitively; this may make the system more difficult to use.

In (Pcyuen, Feng, & Dai, 1998), they combined the wavelet transform with the principal component analysis. Wavelet transform is used for image analysis, while PCA is used for finding the features. EvoFIT system developed by (Frowd, Hancock, & Carson, 2004) evolves the required face from user feedbacks on faces present in the database. EvoFIT starts by creating a set of faces with random facial shapes and facial textures. A user selects the shapes and textures that most resemble a query. These selections then serve as the "parents" of the next population. The components of the selected faces are combined to produce another generation. The limitation of such works is that the retrieval process depends on image matching, not on semantics features. The problem lies in not just how clearly we describe, but also in how the system will interpret and understand this description.

A learning framework to automatic annotation of photographs in a family photograph management system was developed in (Zhang, Chen, Li, & Zhang, 2003). Latinic semantic index was applied in the work by (Ito & Koshimizu, 2004), where some face parts sizes and lengths were employed as a face description vector. These could be the size of the pupil of an eye, the length between two eyes, the length between the pupils of two eyes, and the width and height of a face. In the works by (Fang & Geman, 2005), an interactive system was proposed as a series of visual queries and answers between a user and the system. The system displays a set of images from the database, and the user provides feedback to the system. The purpose is to retrieve the target image in user's mind from the image database. However, the disadvantage of this type of method is the difference between mental matching and feature-based matching, where the system is a content-based image retrieval technique and the user feedback is the image example for the system. On the other hand, deciding which image to display at each iteration is a challenge in mental picture retrieval.

(Deselaers, Rybach, Dreuw, Keysers, & Ney, 2005), introduced a framework to retrieve general images based on depicted faces. However, the aim of this work is only to retrieve images of groups of people with the same face draught as in the query image based on the low-level features.

In the work by (Gao & Qi 2005), the representation of structural information was used to indicated the connectivity of the edge points of the face objects characteristics, and the viewing direction to improve the face identity description for similarity matching.

The work by (Le, Satoh, & Houle, 2007) used the relevant set correlation (RSC) clustering model to organize similar faces into clusters, and then display only the representative faces of the clusters asked in the user query. The kd-tree index structure is used in (Vikram, Chidananda Gowda, Guru, & Urs, 2008) to store face descriptors that are based on the landmarks of the face. In the work by (Bau-Cheng, Chu-Song, & Hui-Huang, 2008), a set of Haar-like features that is a set of rectangular features was extracted, and integrated with supervised manifold learning, to retrieve facial images from large databases. This was an interactive process designed to incrementally obtain knowledge about the target from the responses of the user to a series of multiple-choice questions. Daidi and Irek (Daidi & Irek, 2008) introduced a framework for the unification of statistical and structural information for pattern retrieval based on local feature sets. The relationship between structural and statistical features of pattern description is examined, and a unified framework was proposed. Local feature descriptors in the form of parameterized feature vectors were constructed from the coefficients of the quantized block transforms. Feature vectors statistics describe local feature highlighted by histograms, which were treated as vectors. This method is work on the general images; it was not exclusive to faces. Vijaya et al. (Vijaya Lata, Tungathurthi, Rao, Govardhan, & Reddy, 2009) used the eigenfaces features for developing their face recognition system. The system detects pictures of faces captured by a digital camera, and then identify by comparing with a training image dataset, based on the extracted features. Shih & Liu in (Shih & Liu, 2005) used the principle component analyses algorithms for face retrieval in varying configurations of different color models. Kam-art et al. (Kam-Art, Raicharoen, & Khera, 2009) suggested the feature extraction method for face recognition. The face image and its components initially are converted to grayscale images. The features are then extracted from the grayscale image. The edges of a face image and its corresponding face components are detected by using the canny algorithm. Anew descriptor was introduced in (Thang, Rasheed, Lee, Lee, & Kim, 2011) where the constrained independent component analysis (CICA) method was used.

The limitations of the above works are their weakness to deal with semantic feature of the facial image, however they deal only with low-level features, such as structural information and the connectivity of the edge points of the face objects characteristics (Gao & Qi 2005), the landmarks of the face (Le, et al., 2007), Haar-like features (Bau-Cheng, et al., 2008), statistical and structural information of the local feature sets of the face (Daidi & Irek, 2008), PCA (Vijaya Lata, Tungathurthi, Rao, Govardhan, & Reddy, 2009)(Shih & Liu, 2005), and the edges of a face image and its corresponding face components (Kam-Art, et al., 2009). The work principles of the above systems are based on image based matching and retrieval technique. The retrieval objective in most of these approaches is simply to match images and display top images.

2.7 Neural Networks and Image Retrieval

A neural network was used with several image retrieval works for classification and retrieval purposes, such as in (Fournier , Cord , & Philipp-Foliguet 2001). The training back-propagation (BP) neural network is used to obtain the initial retrieval result. The user labeled the related image from the retrieved result. The neural network then, adjusted the network weight according to the user's feedback. The relevance feedback algorithm's goal was to minimize the difference in the error between the expected output and the actual output. Similar works can be found in (Han , Huang , Lok , & Lyu, 2005) who, firstly select a typical image from the storehouse's to use as the training set for the network . Then, based on the BP network’s output and the differential value between the user submission's query image and the images in the storehouse, the number of images will be retrieved and displayed for the user. The user selects the related image from the retrieved result to train the BP network. The process then revises the network’s weight. In the work by (Park, Lee, & Kim, 2004) a neural network was used for automatic image classification, based on its content objects. Park et al. built a classifier model based on a neural network that uses the learning pattern of the texture feature to reflect the shape of the object. A comparison is carried out based on the objects extracted with and without the background. Li et al. (Li , Shi , & Luo, 2007) suggested a neural network approach to model texture perception, and to express the fuzzy texture semantic feature, using linguistic expression based image description (LEBID) framework. They established a semantic-based image retrieval system using texture image archives. Each texture description was defined with an explicit language.

For face retrieval, Navarrete and Ruiz-Del-Solar (Navarrete & Ruiz-del-Solar, 2002) organized facial images in a tree structured self-organizing map (TS-SOM). Projections of the principal component analysis (PCA) were used to form the map for features representing the facial image in the image space. Each facial image represents a cluster in the whole image space. The user selects facial images that are considered similar to his query, the image that have neighbor positions in the map with query image are subsequently retrieved. The process is iterated until the requested face image is found. Actually, the user may be trapped in a loop as PCA-representation together with the similarity measure used in the off-line TS-SOM training means there were no on-line training when use on line images. The query and retrieval used off-line training. In addition, the search in a TS-SOM is very complex, when the database is extensive. However, this work did not consider the users conceptual query directly.

2.8 Challenges of Content Based Image Retrieval

Results of content-based image retrieval are moderate, despite the recent progress in both the features selection techniques and matching and retrieval techniques after years of research efforts. This is mainly the result of the semantic gap problem between the low-level image features and high-level semantic contents of images. The lack of semantic interpretation is the major drawback of the current content-based image retrieval systems. Its level of success is devalued when it is implemented in practical image retrieval applications (Lew, et al., 2006).

To simply distinguish between images automatically, visual features alone are not enough. Representing an image by simple features usually leads to the loss of information, so different pictures may map onto the same set of features.

There might be two images, for example, one of a blue sky and gray desert, and the other of a blue sea and gray beach, as shown in Figure 2.4.(a). With color, texture and other attributes, they might appear similar, but are completely different semantically. Another example, based on the idea from (Hove, 2004) is depicted in Figure 2.4.(b). One image is of a banana, and the other of a dolphin. Both images are without color and have a distinct curved shape. From a syntactic viewpoint, these two images share similar features, although they do not appear similar to the human eye. Hence implementing the query by example (QBE) system, may return images that are considered matching based only on the extracted features. However, the results can be somewhat random if we do not consider semantic information (Hartvedt 2007).

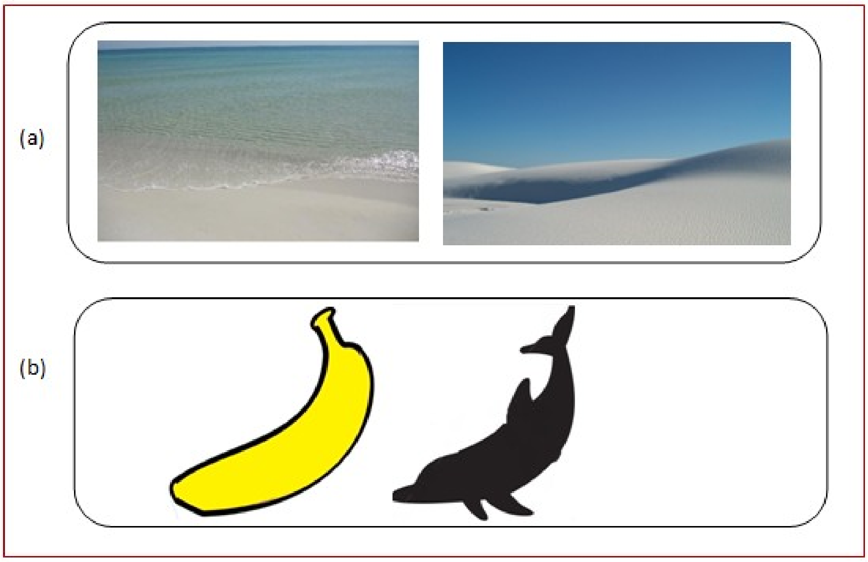

Figure 2.4: (a) Images color similarity vs. different of images content.
(b) Images shape similarity vs. different of images content.

We will show through our experiments on CBFIR systems that using visual features that is represented by low-level features often misses describing the meaningful of the similarity between the images in the users' mind. This is because; there is a deficiency in the connection between pixel statistics expressed by low-level features, and the interpretation of images by human observers. Current systems face the challenge of overcoming this problem in order to match the capabilities of the human visual system. It is expected that image retrieval systems should offer maximum support in removing the gap between the low-level visual features and the depth of human semantics.

2.9 The Needs of Current Systems

From the above review, discussion, and limitations of the previous works, we summarize some of the research needs of current systems, based on Torres and Falcão (da Silva Torres & Falcão, 2006) :

- Developing formalisms to depict image content descriptions and associated services. This formalism can lead in the design and implementation of new applications based on image content.
- Realizing the users' needs and information-seeking behavior. These need of a match query and stored images in a way that indicates human similarity.
- Addressing the semantic gap presented in images that is not available in current techniques, and textual descriptions and tools that can automatically extract semantic features from the images.
- New data fusion algorithms have to be formulated to combine information of different varied formats. Text mining techniques might be integrated with visual-based descriptions.
- Research on new user interfaces, based on image content for annotating, browsing, and searching is needed.
- Methods of matching query and images in the database need to be developed to simulate human similarity judgments.

2.10 Summary

We have reviewed the important points, and some of the existing strategies of content-based image retrieval and neural network with image retrieval, and we have reviewed the major works of content-based facial image retrieval. We have also discussed some of the disadvantages, challenges, and needs of the current systems. Most of the proposed works were based on:

- A feature vector derived from images in the database.
- Database feature vectors ordered as a database index.
- A user-submit his query through an example image, sketch or from an image montage.
- Query image features vector extracted and matched against the feature vectors in the database index.

However, the essential differences between the various works lie in the low-level features used, and in the algorithms that are applied for feature vectors comparison.

As explained above, every work has its limitations as long as the semantic gap exists with the current general content based and the specific domain content based facial image retrieval systems. A more detailed discussion is given in section 3.5.

CHAPTER 3
SEMANTIC-BASED IMAGE RETRIEVAL

3.1 Introduction

The existence of retrieval systems that can understand human high-level requests have become necessary because of the growing demands of computer users and the availability of digital image databases.

Semantic features are useful in delineating high-level features, which appear in images or can be estimated and measured semantically. These features are essentially in supporting image retrieval systems. Using semantic features with image retrieval is important to eliminate the misinterpretations arising when the present retrieval systems try to identify the basic objects and their relationships in the image.

Various semantic levels lie between the human comprehension of image contents and the raw image representation. This includes extracting descriptors, identifying and labelling objects, and objects semantics relationships. In the works of John Eakins et al., (John Eakins & Graham, 1999) and Liu et al., (Liu, Zhang, Lu, & Ma, 2007) the problem image query-processing characteristics were discussed in the form of levels to highlight the correlation between semantic features and the nature of queries submitted by the users. This discussion can be summarized as follows:

- Level one: The query is formulated based on basic primary features such as texture, color, shape or the spatial location of image elements. For example, 'retrieve pictures that look similar to this', or 'find image containing a red spot in the top right corner'. In such types of query and retrieval, features are obtained from the images themselves without the necessity of an extraneous knowledge base. Queries of these types are termed 'query by example'.

- Level two: Here, a query is formulated based on an object's identity within the picture. Extraction of features from the picture is based on primitive features and external knowledge. Some logical inference to the identified object in the image is needed. An example of such query may be, ‘find a picture of a double floor villa’, or ‘find a picture of Barack Obama’, or ‘find a picture of the Petronas Twin Towers’. In such types of query and retrieval, some prior understanding of the image is necessary.
- Level three: Here, the query is formulated based on abstract attributes, which requires some kind of high-level reasoning based on the objects or the depicted scenes of the images. Such queries may include the retrieval of named events or types of activity: for example, ‘find a picture of the Dance Festival’ or pictures with emotional or religious import like ‘find a picture of a cheerful crowd’. These type of queries are hard to answer automatically, as subjective judgments and complex reasoning are needed to relate the abstract image concepts and the image content (J. Eakins, 2001; John Eakins & Graham, 1999).

Levels 2 and 3 are together referred to as semantic image retrieval. Most current systems based on image content are within the first level, where the image's semantic data is not used during retrieval (Wang & Ma 2005).

This categorization of query types reflect the denotation of high level semantic features and is also useful for exemplifying the limitations of the present image retrieval techniques. So far, the best method for representing the expression of high-level semantic attributes is by using text, because humans interpret images and measure their similarity using high-level concepts, which can be easily represented by the concepts of keywords or text description. Data stored with text are much easier for human interpretation. Text consists of words that are well-defined concepts, making human communication, and understanding possible. While words may be equivocal, they are usually easily defined by contents. With text, users are free to compose queries using varied words. Each byte is a numerical code for a character in the text files. Hence, strings of bytes correspond to words that, in turn, convey semantic meaning. In pictures, however, each byte or group depicts the color at a particular location (pixel). These pixels are quite distant from components that have a semantic meaning (Pavlidis, 2008). Of course, it is undeniable that without the support of visual features, it is impossible to deduce an image's semantics, unless they are annotated. One of the most important factors for measuring the semantic similarity between images is to look at the objects in the image and try to find relationships between these objects and not just look at the image generally (Sridhar, Nascimento, & Li, 2002).

We can define image retrieval systems depending on the features that are used:

- Existing systems, which extract low-level visual features from images.
- A semantic retrieval system, where interpreting and meaning are extracted from raw images. A construction key is then drawn from these semantic items. The query is characterized using some combinations of the semantics extracted from the images, and the retrieval is realized by applying a suitable similarity measure to figure out the distance between a query and the images in the database. The images are then ranked based on their distances.

3.2 Semantic Attributes Extraction

The method of semantic features extraction may differ from one user to another. There is no unified model to capture the semantic attributes that differ from domain to domain and from user to user within the domain. Semantic attributes associated with the whole image are termed image semantic attributes, while those that concern the constituent image objects are termed image-object semantic attributes (Gudivada, Raghavan, & Vanapipat, 1994) .

There have been several attempts to derive some high-level semantic features automatically using supervised or unsupervised machine learning techniques. The goal of supervised learning for image semantics is to predict an outcome's representation value (semantic concepts for example,) based on a set of image features or input images. The state-of-the-art techniques in semantic-based image retrieval can be sorted by various points of view. For example, authors in (Smeulders, Worring, Santini, Gupta, & Jain, 2000) categorize search by association, aimed search and category search, each of which calls for dissimilarity.

Some other classification may consider the application domain or techniques used to extract high-level semantics. Liu et al. in (Liu, et al., 2007) proposed 5 categories to derive high-level semantics, based on the following techniques:

- Defining high-level concepts, using object ontology. In the work by Mezaris et al. (Mezaris, Kompatsiaris, & Strintzis, 2003), who presented a typical example of such ontology-based methods. Here each region of an image is described by its vertical and horizontal position, its size and shape, and its average color in the color space.
- Associating low-level features with query concepts using machine learning tools.
- Continuous learning of users' intention by introducing relevance feedback (RF) into retrieval loop. The system try to merge the user's continuous feedback towards learning more about the user's query and adjusting the parameter, semantic, feature, or classification spaces to show the relevant and irrelevant instances. A typical scenario for relevance feedback in CBIR is as below (Zhou & Huang, 2003) :
  (i) Through a query-by-example, sketch, etc., the system provides initial retrieval results.

(ii) User judges to what degree the results are relevant (positive examples) / irrelevant (negative examples) to the query.

(iii) Machine learning algorithm is applied to learn from the user' feedback.

(iv) Step 2 and 3 are repeated until the results are satisfactory.

- Generating semantic template (ST) to aid high-level image retrieval. Semantic template is a representation between low-level visual features and high-level concepts. The representative feature of a concept is determined from a collection of sample image.
- Integration of the visual content of images and the textual information. To support the semantic retrieval effort, it is important to fuse the evidences from these two techniques.

Textual information with images described can be associated in two ways - annotation and categorization. Keywords or explicit text descriptions are associated with an image in annotation, whereas in categorization, each image is assigned to one of several predefined categories (Chen & Wang, 2004). Categories can be more generalized in to two category classification, such as indoor/outdoor (Luo & Savakis, 2001) or city/ landscape in (Vailaya, Figueiredo, Jain, & Zhang, 2001) to particular classification such as fashion, the Asian people ,and fishes in (Chen & Wang, 2004). To further process the image, categorization provides an initial step toward image understanding. For example, in (Wang, Li, & Wiederhold, 2001), a categorization is made into graph/photograph and textured/non-textured categories in a pre-processing step.

3.2.1 Image Annotation

Bridging the semantic gap for image retrieval is not an easy task. The nearest solution to the well-known image retrieval problem maybe image annotation. Using image annotation for querying image databases with text have been tried by several researchers. Satisfactory progress has been achieved by anticipating that users can manage imperfect retrieval results and fulfil images retrieval, with the probability of incorporating a particular concept of interest.

The information from the images, which are directly related to its visual content, are content-dependent metadata, this is related to low-level features and content-descriptive metadata, and is the information which characterize the relationships between image entities and real-world entities or events, emotions and meaning associated with visual signs and scenes.

The images in the database can be classified into different categories by mapping their metadata descriptors to interpretation, which acts as high-level semantics. As an example, 'Sky' can be mapped to the region of 'light blue' (color), 'uniform' (texture), and 'upper' (spatial location) (Liu, et al., 2007).

3.2.1.1 Metadata Specification

One or more of the following approaches can be used to specify content-descriptive metadata for the images: keyword annotation, free text annotation and, ontology-based annotation.

3.2.1.1.1 Keyword and Free Text Annotation

In keywords annotation, the image is annotated by linking a list of keywords with it. There are two options for selecting the keywords:

(i) The annotator, as needed, can use arbitrary keywords.

(ii) The annotator limits keywords to a pre-defined list.

Two levels of specificity can be used to associate this information:

(i) A list of keywords related to the entire image, specifying what is in the image.

(ii) Image segmentation together with keywords associated to each of the region, with keywords describing the whole image. Figure 3.1 shows the image annotation levels.

Figure 3.1: Image annotation levels: (a) entire image annotation, (b) segmentation's region annotation.

In free text annotation, the user can use any combination of words or sentences, such as highlights or underlining, comments, footnotes, tags, and links. This method is easy to implement, but more difficult to use the annotation later for image retrieval.

3.2.1.1.2 Ontology Based Annotation

In ontology-based annotation approach, the ontology acts as a specification of a conceptualization. It essentially includes concepts together with their rules and relationships. The taxonomy is produced by appending a hierarchical structure to a keywords collection (Hanbury, 2008).

Keywords can be assigned to the visual features using different techniques. The techniques aim to identify the correlation between high-level semantics and low-level visual features. Three techniques of image annotation are used; manual image annotation, Semi-automatic image annotation and automatic image annotation.

3.2.1.1.3 Manual Image Annotation

Manual image annotation is the familiar way to describe an image. When the images are loaded or browsed, users have to include some descriptive keywords. Applications that provide storage for annotations, such as disk space or a database are required. It is the most accurate annotation method, since keywords are based on how humans interpret the image's semantic contents. However it needs more effort and time (Barnard & Shirahatti, 2003; Stamou, 2006).

3.2.1.1.4 Semi-automatic Image Annotation

Semi-automatic image annotation is less accurate, compared to the manual annotation. In semi-automatic image annotation, the user provides an initial query at the beginning. The system parses the human’s query, and extracts semantic information in order to carrying out the annotation. Visual information is taken from the raw image contents. These contents are then mapped with semantically rich keywords. Machine learning together with the user’s feedback help to make use of previously annotated images (Pagare & Shinde, 2012). The annotation quality improves after correction, an example of this technique is in (Wenyin et al., 2001). The user has to provide feedback while examining the retrieval results. This method has three main parts: The query interface, the image browser, and the relevance feedback interface. When a user submits a query, the search results rank the relevance of the images against query. According to the ranked list order, images are displayed on the image browser for user viewing. After browsing, the user can provide feedback through the relevance feedback interface. The system returns the refined retrieval results based on the user's feedback and presents the results in the browser. This method is particularly suited to a dynamic database system, in which new images are constantly being introduced.

3.2.1.1.5 Automatic Image Annotation

Automatic image annotation is the best in terms of effort and time but is a less accurate annotation method. In automated image annotation, the system generates a set of keywords that help to describe the scene represented in the image. One such method is in the works of (Jeon, Lavrenko, & Manmatha, 2003). A training image set is used to annotate the images automatically. A vocabulary of blobs describes regions of the image. By using the image training set with annotated keywords, the probability of obtaining a label for the blobs in the image is predicted. The image can be seen as a collection of blobs. For each, there is a probability distribution known as the relevance model of the image. This relevance model can be accepted as a container holding all possible blobs that exist in the image, and containing the keywords that exist in the image. With the help of a training set of images with annotated labels, the possibility of producing a tag specifying the blobs in an image can be guessed (Pagare & Shinde, 2012). Automatic image annotation can be a global feature based image annotation, or block based image annotation technique. The global annotation utilizes the properties of global image features such as global color and texture distributions. Torralba and Oliva (Torralba & Oliva, 2003) use global features for predicting the presence of objects or classifying natural images into semantic categories. To identify real-world objects within the image, the block based image annotation uses an automatic segmentation step prior to the actual learning stage.

This annotation scheme use image segmentation algorithms to zone images into their constituent pseudo-objects. Statistical models of their co-occurrence with annotation words are then found. The success of this approach would hinge to a large degree on the accuracy of the image segmentation algorithms. Depalov et al. (Depalov, Pappas, Li, & Gandhi, 2006) assigned semantic labels to image segments. Their proposed approach depended on a spatially adaptive, perceptually based, color-texture subdivision scheme. To classify the segments into semantic categories they used linear discriminate analysis techniques.

Several studies pointed out to the poor performance of the current automatic image annotation techniques compared to the manual annotation.

Enser et al. (Enser, Sandom, & Lewis, 2005) highlight two limitations of the automated image annotation, as compared to the conventional manual annotation by human. One of these limitations is that the keywords in the annotation vocabulary have to relate to visible entities within the image. However, users frequently submit search requests addressing the significance of depicted objects or scenes. The so-called visibility limitation tries to describe how automated image tagging algorithms typically depend on linking visible image features to words successfully. It is very difficult for automated algorithms to capture content and contextual information from images that do not have any associated image features. A prime example of content that would be hard to automatically extract from images would be a CBIR query, “find a picture of the first public engagement of Prince Charles”. In the second limitation the author moves on to state an additional limitation in the form of generic object limitation, which questions the use of very generic label for the images such as “sun”, “grass” and “tiger”. They share the generic property of visual stimuli, which needs a minimally interpretive response from the viewer. In the generic nature of keywords in annotation vocabularies, the authors argue that they “appear to have the common property of visual stimuli which needs a minimally-interpretive response from the viewer” (Moran, 2009).

Enser et al. mentioned many studies that show most users tend to use queries that refer to objects by proper name, which normally have limited visual stimuli association in images. Enser et al. cited studies indicating that search requests for images with features uniquely described by proper names are very common, and where such visual prominence does not serve a useful role. Enser et al. wraps up his work by saying that any needed textual annotations will always have to be assigned to images manually. Regardless of the progress in image analysis, specifying textual annotations will have to be done manually (Moran, 2009). Analyzing and understanding images automatically is still extremely difficult. The image representation has to be very particular to semantically discriminate between similar objects. Moreover, any representation must be constant to various confusing factors contained in the images as shown in Figure 3.2,

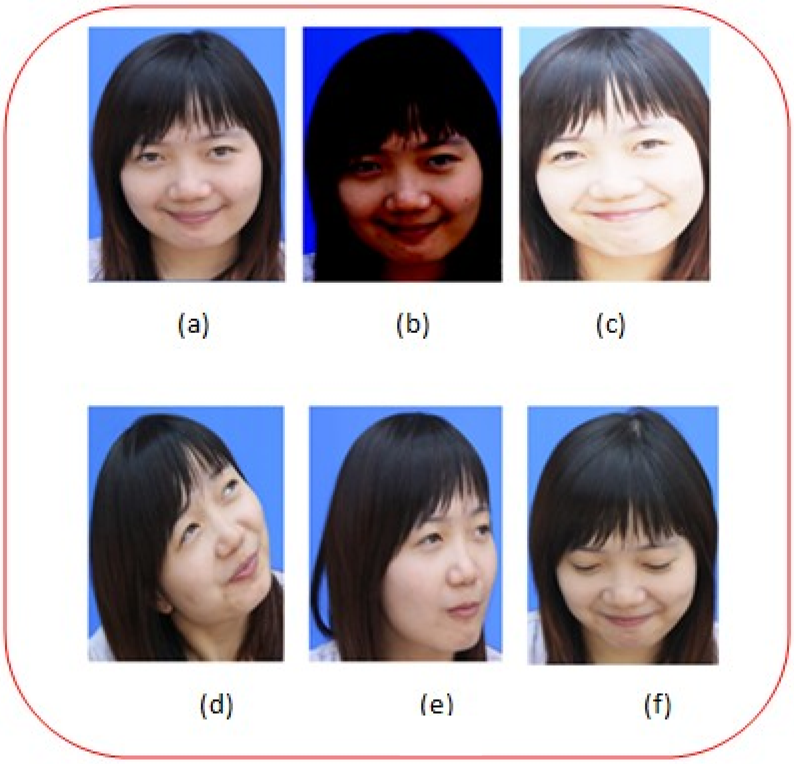

Figure 3.2 : Various confusing factors contained in facial images of the same person: (a), (b) and (c) contain various illumination and scale (zooming), and (d), (e) and (f) contain different angles and orientation.

such as illumination, occlusions, scale, deformation, different angles and orientation, and viewpoint variations. Such factors can make the same face look very dissimilar for the machine.

3.3 Text-Based and Content-Based Image Retrieval

There has been much works and progress in both content-based image retrieval for research applications and text-based search on the web. However, there has been limited work to combine these fields to provide a large-scale, content-based image retrieval approach, especially in the facial image retrieval domain. Both text and content-based techniques have their own advantages and limitations. Neither of these is sufficient for retrieving or managing visual data in an effective way. Regrettably, keywords may not precisely describe the image content. The image itself must offer cues about its content. In some instances, it is difficult to characterize certain important real world concepts, entities, and attributes using text only.

There have been some attempts to merge images information with text for a range of tasks including search, automatic labelling of images with keywords, image clustering, and labelling regions within images. Early attempts were made to integrate text and color as in (Smith & Chang, 1996), using text with color histograms and user relevance feedback for sorting information into a predetermined taxonomy, for browsing and searching images. In the work by (La Cascia , Sethi , & Sclaroff 1998) a system that integrates textual and visual statistics in a single index vector for retrieval image based on content was suggested. With latent semantic indexing (LSI) based on the document's text, textual statistics are represented in a vector form. Color histograms are used to capture visual statistics in the vector form. Low-dimensional vectors that can be matched against user queries represent text documents. iFind system in (Hongjiang , Wenyin, & Hu, 2000) incorporate several low-level MPEG-7 (Manjunath, Salembier, &

Sikora, 2002) visual feature types with keywords to retrieve images. A different approach in the utilization of textual and visual information is employed in (Srihari, 1995). They proposed a method to index magazine pictures. The image caption is used as an indication to recognize human faces in an accompanying newspaper photograph.

Wang & Ma (Wang & Ma 2005) used the color moments and discrete cosine transform coefficients as the system input, and the semantic labels predefined by the system experts as its output. Neural networks mapped the low-level feature vectors to their corresponding semantic labels that are predefined earlier. During retrieval, the neural network weights are updated based on the user's relevance feedback. In the testing stage, all features vectors of the images in the database were inputted to the network and the output of the network was their semantic vectors.

Xiang & Huang (Xiang & Huang, 2000) automatically created a thesaurus of semantic collection. The intention is to use texture, color, and structure features, with text annotation. A concept similarity matrix is generated and the update is based on user browsing and feedback. A Hopfield network, is used to classify these concept based on its semantic relevance.

Grouping concepts based on semantics similarity (for example car and motorcycle,) is not practical for face retrieval purposes. As concepts grouping does not bridge the semantic gap in facial image retrieval. For instance, user looking for specific and corporate features, as an example "Glasses" or not - or " big lip", "thick lip", "race", "gender" etc. where , each of the attributes of the face itself is a class.

Gao et al. (Gao et al., 2005) used a tripartite graph to model the relations between visual features, and the texts surrounding each image. Representing different categories within a collection of images from the Photography Museums and galleries of the Yahoo! Directory.

In (Li, Shi, & Luo, 2007) The fuzzy color semantics of the image is extracted and described based on the human color perception model. Linguistic variables are used to depict the image color semantics for the model such as ‘mostly red’.

Latent semantic indexing technique (LSI) has, for a long time, been used for textual information retrieval in several works (Dumais, 2004). Latent semantic indexing technique was introduced to overcome the fundamental problem that plagues existing textual based documents retrieval techniques. Users want to retrieve documents based on term, while individual terms do not provide a reliable conceptual meaning of a document. The concept of the term can be expressed and represented in many ways. Therefore, the literal terms of a user query may not be in harmony with those of a relevant document. In addition, most words are used in different contexts and have multiple meanings. Hence, a user's query concept may match the concept in documents that are of no interest to the user (Rong & Grosky, 2002).

With image retrieval, the Latent semantic indexing technique is used to analyze text that looks close to a given image. An image feature vector is then divided into two parts, one part for the visual features and the other for the textual information transformed by applying the Latent semantic indexing technique.

Cascia et al. (La Cascia, Sethi, & Sclaroff, 1998) uses the latent semantic indexing technique of the text with visual statistics, to compute a representative vector for a content-based search of the web image. Textual statistics are represented in vectors, depending on the text in the HTML document to be integrated with the visual statistics of color and the orientation histograms.

In a similar work by (Rong & Grosky, 2002), there was an attempt to transform low-level features to a high-level of meaning. Firstly, a global feature vector was developed consisting of textual feature and visual feature. Then, the Latent semantic indexing technique is applied on this global feature vector.

3.4 Semantic Based Facial Image Retrieval

Semantic face retrieval refers to the retrieval of facial images based on the semantics features of the facial images themselves. In facial image application, domain users prefer to express the query with some keywords. Such keywords correspond to the symbolic features of the face parts, visual impression and inspired characters, etc. such as the description of the person's nose, face shape, race, age, etc. These semantic features inherently encode geometric relationships (scaling, rotation, translation, and shearing) among facial components (Hsu & Jain, 2002) and are used for recognizing faces and characterizing them.

As we discussed earlier in section 2.6.2, most current facial retrieval methods depend on query-by-example, starting from a digital facial image as the query image. The goal is to search the faces in the database in which their visual features are similar to the query image, however , most of the time there is no actual image to be provided to the system as the query image, only a mental image, that is represented by information on attributes, subjective impressions and opinions about the target face. Another example of a current facial retrieval technique that is also based on the query-by-example, beginning with a sketch or a synthetic digital facial image of the target face based on the description of the face. This is usually what the police do when they obtain the description of a suspect from a witness. An image match of the created image with those in the database is then performed. In such systems, the process is time consuming, because it needs time to sketch or to accomplish the synthesis of the query image to start the retrieval task and in the final stage the process is an image match between the formed image and images in the database based on the visual features represented by the low-level features. This method is far from the most natural way, in which people describe faces and measure the similarity through the semantic facial features. Such methods can take advantage of features description if all are described, but cannot take advantage of the semantic features matching process method. This is because of the weakness of the visual similarity approach of the system, which is not the same as the similarity criteria adopted by humans for comparing images.

Semantic interpretation of facial images requires an accurate interpretation in order to become usable in a general context. Humans are good at recognizing and interpreting facial images because of their holistic processing of visual input. For example, the measurement of whether a face is beautiful or ugly, there is no fixed features or single measurement that can be modelled, extracted automatically, and employed to be used by the system. However, human beings have a natural ability to discern and capture characteristics that are present, absent or not visible in the image itself.

The semantic descriptions, provided by humans are protected to picture quality and other effects that reduce the efficiency of the face image retrieval and they are used to enrich the retrieval process. If a user was to search using the term 'beard' for example, the system will return image of the a human face with the attribute “beard” directly. Therefore, when comparing the traditional facial retrieval to the semantic retrieval systems, the latter has the advantage of a higher level of abstraction, and easier query specification through the natural language.

Keywords are assigned based on the visual features of individual face images. Moreover, there is the case that expressions of characters, which are symbolized by keywords, are more effective than the exact specification using visual features, in intuitive. The visual features are represented by means of the size and lengths of some face parts. When the symbolic features are compared with the visual features, the symbolic features are conceptual, and they are easy to understand and to manipulate.

3.4.1 Related Works

The existing systems query and searching strategy for facial image retrieval do not directly address the human verbal description of the face. The systems developed do not have the advantage of the semantic features description for retrieving the faces according to their semantic contents. Some early works have been attempted to employ some advantages of these semantic features. for face recognition system, geometric feature and elastic graph matching-based face recognition methods have been applied in the work by (Brunelli & Poggio, 1993). In this application, facial features such as eyes and mouth were detected and the properties and geometric relationship between these features were used to describe the faces. Wiskott et al. used elastic bunch graph (Wiskott , Fellous , Kuiger , & von der Malsburg 1997), to represent the faces. The objective of this representation is to allow the system to determine the presence or absence of some features in the face. For example, if the person is using glasses or otherwise, or whether the person has a beard or not. The task of glasses detection, whether the face includes any eye-glasses or not, has been worked upon in similar works by Jiang et al. and Wu et al. (Jiang, Binkert, Achermann, & Bunke, 2000; Wu, Ai, & Liu, 2004). However, in all these works, glasses detection has been the only task that was addressed.

Hsu & Jain in (Hsu & Jain, 2002) suggested a 3D generic face model to drive a semantic face graph. The semantic face graph is used to provide a depiction of the face and its facial components for face recognition purposes. Each node of the graph model is a representation for a facial component (e.g., eyes, mouth). The match between faces is based on these components. The application of this work was for managing consumer photographs. In the works by (Zhou , Yuan , & Sadka 2008), the extraction of facial semantic features was integrated with tensor subspace analysis for the task of face recognition. The semantic features consist of the eyes and mouth, and the part defined by the centres of the three components. The limitation of this work is the using of limited semantic features of the face (three features,) and the retrieval process is still dependent on image matching, not on semantics features.

A probabilistic approach was proposed for face retrieval in (Sridharan, Nayak, Chikkerur, & Govindaraju, 2005). Hybrid Markov Chain sampling model was applied to perform the localization of the facial features. The proposed method tries to avoid pruning images from the data based on faulty user descriptions. While the limitation in this method is that if the features do not exactly match, it will be ignored. The system avoids images pruning from the database, however, the mismatched features is ignored.

In our research, we have proposed a method for facial image retrieval to enable the use of a user's descriptions of a face to retrieve the desired images from a large database of facial images. The objective is to identify semantic facial features for more accurate retrieval. Facial images are annotated with semantic terms, enabling a user to specify his or her queries through natural language descriptions. The overriding aim is to match the verbal queries of a user to the corresponding representation values of semantics features of the face.

The probabilistic approach is incorporated to address the problems associated with image pruning from the database. In addition, the problems associated with mismatched features are also addressed. Our proposed method of probabilistic approach is an improvement to the weakness of the method in the work by (Sridharan, et al., 2005). An illustration of the weakness of the previous work and our proposed method will be discussed in section 5.9.

3.5 Discussion

Most early efforts in image retrieval problem focused on solving this problem completely within a query and retrieval based on image content (Datta, Joshi, Li, & Wang, 2008; Liu, et al., 2007).

Here the user's query is no longer a simple instruction of desired image content. Instead, example images or sketches of the query are presented to a search engine with the intention of retrieving similar-looking images. Vectors of low-level features are typically generated to represent images or sketches. The similarity between images is reasoned as some inverse function of a metric distance between their corresponding vectors. Such as of these works is the system QBIC by Flickner et al. The disadvantages of this system are that it is prepared to search for general tasks. It does not have features specifically used for searching an image database containing only faces, and it fails to consider the semantic aspects of a face.

Systems like Photobook (Pentland, Picard, & Sclaroff, 1996) have the advantage that some of the image's semantic aspects is preserved through PCA features because this features are based on the compression of the images that is statistical in nature. At the same time, there is no way to capture the specifications of the face as given by a human.

A mechanism where query images are submitted as color sketches had been successfully applied in image retrieval (Flickner et al., 1995; Jacobs, Finkelstein, & Salesin, 1995), however, the user is severely limited in having to expand his or her needed information through this querying mechanism. It is a difficult and time-consuming task to construct an accurate face from scratch with a painting tool. The systems do not provide an interface for creating accurate images of faces.

Other systems deal with the problem of applying the user's specifications directly for face retrieval by composing the face using user feedback (Johnston & Caldwell, 1991; Penry, 1974) ; however, users may have specific details of the semantic description like race, sex, or age. Furthermore, the matching process for the actual retrieval does not equal the semantic descriptions, only the entire facial image.

In a similar work, the synthesizing process involves choosing similar faces and combining them (Caldwell & Johnston, 1997; Frowd, Hancock, & Carson, 2004).
The retrieval process however is still dependent on image matching, not on semantics features. The problem lies in not just how clearly we describe, but also in how the system will interpret and understand this description. While this type of work may solve the query problem of the query image, it does not work out the difficulty of facial semantic features matching and retrieval. Furthermore, the drawback of the draw and synthesize method is that some problems do arise from lighting, pose, and feature size differences. For instance, minimal smoothing the edges is not always enough to harmonize the differences when a feature from a very dark complexion is superimposed on a very light one.
The drawing or synthesis of a facial image requires a set of complete tools, excellent skills, and the proper selection of many components of the desired image. The final process is a kind of matching between the low-level features of the images in the database, and those of the drawn face. The computed accuracy of similarity and the effectiveness of the retrieval process reside heavily on the accuracy of the created face.

Genetic algorithm operation or crossover and mutation are also used to synthesize faces (Caldwell & Johnston, 1997). The method allows users to select precise facial features for creating the face that manifests some of the semantic aspects of the target face; this is considered as an advantage in such work. The disadvantages are, firstly, if the genetic algorithm is stuck in local maxima during the retrieval process, the user will be stuck on looping with the same set of images. Secondly, the matching process for the actual retrieval does not relate to the semantic description but only the entire facial image, and the retrieval process still involves image matching.

Content-based image retrieval methods based on statistical computation of visual features that process raw images without regards to contents are useful for visual similarity retrieval only. They are semantically poor, express only partial statistical relationships, fulfil little public expectations, and fail to capture similarities that can easily be inferred by humans; a consequence that is now normally termed as the semantic gap.

To obtain the high-level features, which is desired with image retrieval, regional information is insufficient. In addition, automatic partitioning is time consuming and not always reliable. In designing CBIR systems, object extraction can be ignored for some applications. This is because the CBIR system's objective is to retrieve some semantically applicable images from the databases instead of recognizing objects from images.

Some current image semantic retrieval systems (general domain purpose) are based on classifying an image into one of the many predefined categories such as ‘indoor’ or ‘ outdoor’ images (Chen & Wang, 2004; Luo & Savakis, 2001; Vailaya, et al., 2001). Evidently, the description ability for such method is limited since the predefined categories are limited. Additionally, the subjectivity and fuzziness in human image understanding are ignored, since it focuses on the objective statistics of some images features. However, the image semantics should be defined with a more complete and extensive linguistic expression set. Other similar works are where the concepts were grouped based on its similarity in semantics: for example, car and motorcycle share the same semantic meaning, however, this method is not practical for face semantic features and does not bridge the gap in facial retrieval.

Latent semantic indexing has been proposed by some to reduce the semantic gap (Dumais, 2004; Ito & Koshimizu, 2004; La Cascia , et al., 1998; Rong & Grosky, 2002). Many of them address web document retrieval. The applied scheme attempted to explore the correlation between semantic and visual features. However, it did not provide explicit semantics description using the natural language.

Machine learning approaches with ontology techniques are utilized in some works (Hanbury, 2008; Mezaris, et al., 2003; Pagare & Shinde, 2012; Smeulders, et al., 2000) to define high-level concepts by using semantic templates or a dictionary to interpret low-level features.

Little guarantee that the automatically semantic annotations are optimal for the retrieval is provided by these approaches. The approach usually needs restrictive independence assumptions on the relationship between the visual components and the text. Understanding image automatically is a challenging mission and much effort is required to achieve satisfactory results.

A relevance feedback technique was used in some works for the semantic problem (Fang & Geman, 2005; Pagare & Shinde, 2012; Wenyin, et al., 2001; Zhou & Huang, 2003). Relevance feedback works on the low-level features and based on the user's feedback to refine weights given to the features. However, once the retrieval based on the low-level features fails, appropriate user's feedback will not be offered. Relevance feedback does not provide semantic retrieval functionality for users.

Most of the efforts on CBIR and semantic retrieval have been on the general domain images (Datta, et al., 2008; Heesch, 2008; Lew, Sebe, Djeraba, & Jain, 2006; Liu, et al., 2007; Veltkamp & Tanase, 2000). Very few of such works have been applied in the specific domain of facial retrieval. The works carried out on facial image retrieval have not direct addressed the semantic facial image retrieval problem. Some of the suggested facial retrieval systems deal only with the face model, and user's feedback. These methods are based on image based matching and retrieval technique. The retrieval objective in most of these approaches is simply to match images and display top images.

Generally, the main disadvantages in most retrieval methods are that they do not directly express the faces' semantic features in the database. They do not capture the face's semantic aspects, especially when the query is some kind of user description. However, a general description of face is semantic (verbal) in nature. This can be realized through simple verbal descriptions by a person; these descriptions narrow down the candidate faces efficiently and speed up the query retrieval process measurably.
A satisfactory method for quantifying human vision, more explicitly in the context of understanding and explaining images is needed.

The lack of an efficient approach to content-based image retrieval on the one hand, and the presence of an efficient technique for text retrieval (though image retrieval based on only text is not accurate in itself) together provide the motive for combining content and context information to reduce the semantic gap and improve the image retrieval. The combination of the content and context information creates a semantic space of image and words. The textual features indicate the external description and interpretation of the facial image from the people's viewpoint, while visual features relate to the image's internal attributes.

CHAPTER 4
FACIAL FEATURES EXTRACTION AND CLASSIFICATION

4.1 Introduction

Raw image datasets are not useful in most computer vision tasks. This is due to not only the high dimensionality of the raw image making it difficult to utilize the whole image but also the redundant information in these raw images. Therefore, it is pertinent to extract a good representation of the significant information contained in the raw image for analysis, in the application concerned.

A feature is a data derivative from the image content; it can be defined as a function of measurements to specify a quantifiable property of an object. It quantifies some significant characteristic of the object.

In a broad sense, image content may include visual content (so-called low-level features) and semantic content (so-called high-level features). Semantic content is described either directly by textual annotation or by complex inference procedures based on visual content (Long , Zhang , & Feng 2003). Visual content can be classified as general or domain-specific:

- General visual content: Application-independent features such as color, texture, shape, and spatial relationship. A visual content descriptor feature can be divided into pixel-level features, global features or local features. A global descriptor feature is calculated over the whole image or regular sub-area of an image, whereas a local descriptor feature uses the visual features of regions or objects to describe the image content. Local visual descriptors are obtained by dividing the image into parts of equal size and shape first. Pixel-level features include those calculated at the pixel level, e.g. color or location.

- Domain specific visual content: Application-dependent features such as those that are extracted from human faces and fingerprints, and may involve domain knowledge. These features are often a synthesis of low-level features for a specific domain (Rui, Huang, & Chang, 1999).

A wide variety of features has been proposed for image retrieval in the global or specific domain visual content. The fundamental issue is feature selection in the design of a content-based image retrieval system. The following issues should be considered in the feature selection process:

- The features possess sufficient image information and there should be no requirements for domain-specific knowledge for their extraction.
- Computation of the features should be easy to facilitate large image collection and rapid retrieval.
- There should be a good relationship between features and the human perceptual characteristics, to ease users decisions on the suitability of the retrieved images (T. Deselaers, Keysers, & Ney, 2008; Rui, et al., 1999).

Other considerations for features selection include general data reduction to limit dataset storage and increase algorithm speed, feature set reduction, thus saving resources in the next round of data collection or utilization (Guyon & Elisseeff, 2006). Features selected for accurate image retrieval should conform with semantics, robustness to noise level and invariance to background. They should also be robust to scale and environmental changes.

In this research, before implementing the features extraction algorithms, we proposed applying some of the important processes that may play a vital role in the accuracy of the extracted features of the image. These processes include pre-processing methods, face detection methods, and image segmentation methods.

The background of the techniques used in this research will be discussed in the following sections, while our proposed methods will be discussed in chapter five.

4.2 Image Pre-processing

The goal of pre-processing is to enhance image quality and consequentially, improving the image retrieval performance. This is because raw images are usually noisy, particularly camera noise gained when the image is taken. This noise degrades the capabilities used in the feature extraction module. Different methods of image pre-processing may be implemented in a face retrieval system based on the need:

- Normalization of the image size. This is implemented to change the image size to default image.
- Image enhancement or image-processing filters. These operations include noise reduction, smoothing, or sharpening. Median filtering can "clean" the noise in an image while keeping the original information of the image. High-pass filtering is useful to emphasize some details of a facial image such as contours. As a result, some important facial features are more obvious for the feature extraction module, which can radically improve the facial retrieval systems performance.
- Background removal. Important features of the facial image are concentrated in the primary information of the face itself, so the information that is extracted from the background would be considered noisy within the raw original

information. Other pre-processing operations can be implemented, such as illumination normalization and face rotational normalizations.

4.3 Face Detection

Face detection processing is used to detect and determine any existence of faces from the image in selecting regions of interest using an appropriate feature extraction algorithm. A collection of the image of human faces in different positions, scales, orientations, poses, and lighting conditions are located and identified in the face detection module. This is a challenge that has confronted the researchers because human faces are highly non-rigid and vary greatly in size, shape, color, and texture. In addition, the obvious changes in facial appearances are attributed to varying facial expression and lighting conditions.

It has also been observed that camera limitations and pose variations in real life surveillance and biometrics would result in more dispersed and complicated distribution of the human faces in the feature space compared to that of frontal faces. This situation would further worsen the problems confronting robust face detection.

As face detection techniques have been researched for years, it is observed that most of the face detection methods have focused on detecting frontal faces with good lighting conditions. Yang et al. (Yang, et al., 2002) have classified single image face detection methods into four types: knowledge-based methods, feature invariant approaches, template matching methods and appearance-based methods.

The above listed methods can employ color segmentation, pattern matching, statistical analysis and complex transforms, to achieve classification with minimum error. Although classification accuracy varies from method to method, higher accuracies have been observed in techniques, which have adopted dynamic models or classification rules derived from machine learning processes.

Of the four categories, the appearance-based approach is preferred as it relies on statistical and machine learning techniques to characterize face and non-face images. Learning-based face detection techniques are the most successful in terms of detection accuracy and speed. One of the most popular of these methods, which will be employed in the current work, is the Viola and Jones method.

In this research, the Viola-Jones face detection algorithm **(**P. Viola & M.J. Jones, 2004**)** was adopted in combination with the skin color face detection method (Pai, Ruan, Shie, & Liu, 2006)(Hsu, Abdel-Mottaleb, & Jain, 2002). Application of Viola-Jones algorithm was based on the trained classifier from (Bradski, 2000).

4.3.1 Viola-Jones Face Detection Method

Paul Viola and Michael Jones (Viola & Jones, 2004) have proposed the frontal-view face detection framework, which works real-time and yields high detection rates. This technique relies on a set of a simple rectangular feature (so- called Haar-like features) to detect the face. These features are reminiscent of the Haar basis functions, which have been used by Papageorgiou et al. (Papageorgiou, Oren, & Poggio, 1998). Paul Viola and Michael Jones have also introduced the concept of an integral image as a new image representation, which allows for a fast and efficient feature computation. Rectangle features can be computed very rapidly using an intermediate representation for the image, called the integral image. By utilizing the integral image representation, the simple features can be rapidly computed in linear time. In the learning stage, the AdaBoost algorithm (Freund & Schapire, 1995) is employed to select a reduced number of critical and important features from a huge library of potential features, which are used to create very efficient and simple classifiers.

The number of Haar-like features is far larger than the number of pixels within an image sub-window. Therefore, to speed up the classification process, the learning process excludes the majority of features through the selection of a small set of critical features.

The classifiers are arranged in a cascade architecture, which can achieve increased detection performance, while simultaneously reduce computation time and successively discard background regions, by focusing more on regions, which have passed previous filters thus increasing the chance for potential facial regions. Many of the negative sub-windows are rejected, while detecting almost all positive instances.

Four kinds of rectangle features are used with varying numbers of sub-rectangles - two two-rectangle features, one three-rectangle feature and one four-rectangle feature, as shown in Figure 4.1. For a given feature, the sum of pixels in the white rectangles is subtracted from the sum of pixels in the black rectangles (Viola & Jones, 2004).

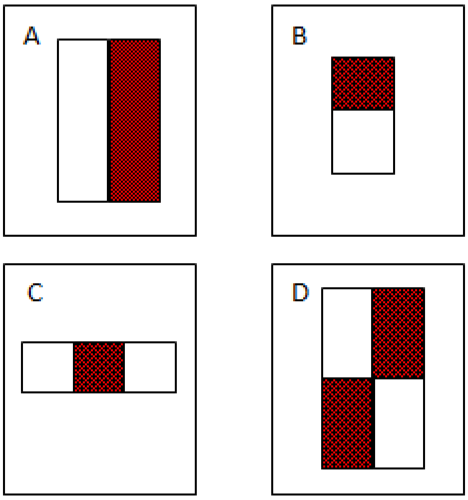

Figure 4.1: Example rectangle features shown relative to the enclosing detection window. The sums of the pixels, which lie within the white rectangles, are subtracted from the sum of pixels in the grey rectangles. Two-rectangle features are shown in (A) and (B), figure (C) shows a three-rectangle feature, and (D) a four-rectangle feature.

4.3.1.1 Integral Image

Rapid computation of rectangle features can be achieved using an intermediate representation for the image, the integral image. As shown in the equation below the integral image at location (*x,y)* is the sum of the pixels above and to the left of (*x,y*) inclusive:

$$ii(x,y) = \sum_{m=0}^{x} \sum_{n=0}^{y} i(m,n). \tag{4.1}$$

Where $i(x,y)$ represents the original image and $ii(x,y)$ is the integral image as appeared in Figure 4.2. The integral image can be efficiently computed using the following pair of recurrences:

$$(x,y) = s(x,y-1) + i(x,y)\ . \tag{4.2}$$
$$ii(x,y) = ii(x-1,y) + s(x,y)\ . \tag{4.3}$$

Where $s(x,y)$ is the cumulative row sum, $(x,-1) = 0$, and $ii(-1,y) = 0$ is the integral image, which is computed using a one pass over the original image.

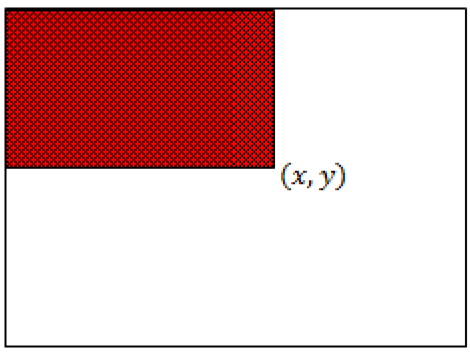

Figure 4.2 The value of the integral image at point (*x,y*) is the sum of all the pixels above and to the left.

As shown in Figure 4.3, any rectangular sum is computed in four array references using the integral image. The value of a two-rectangle feature is computed as the difference between the sum of pixels within two rectangular regions, having the same size and shape and is adjacent horizontally or vertically. The value of a three-rectangle feature is computed as the sum of pixels within two outside rectangles subtracted from the sum of pixels in a center rectangle. Finally, the value of a four-rectangle feature is computed as the difference of sums of pixels between diagonal pairs of rectangles (Viola & Jones, 2004).

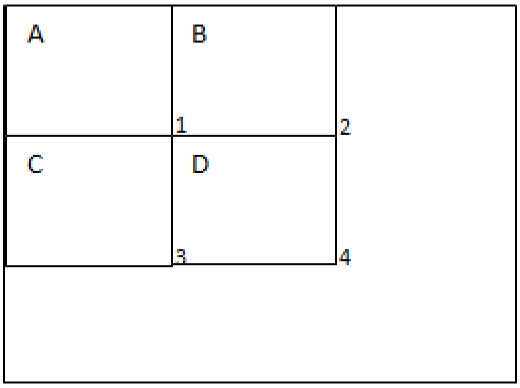

Figure 4.3 : The sum of the pixels within rectangle D can be computed with four array references. The value of the integral images at location 1is the sum of the pixels in rectangle A. The value at location 2 is A + B, a t location 3 is A + C, and at location 4 is A + B + C + D. The sum within D can be computed as 4 + 1 − (2 + 3).

4.3.1.2 AdaBoost

In this technique of face detection, within any image sub-window the total number of Haar-like features is generated based on the integral image method, as the four rectangular feature numbers is very large. To speed up the classification process, the learning process must exclude the majority of available features, and instead focus on a small set of critical features. Boosting is a method of improving the effectiveness of predictors. It relies on the existence of weak learners. A weak learner is a “rough and moderately inaccurate” predictor, but one that can predict better than chance. Boosting shows the strength of weak learners in combination (Meir & Rätsch, 2003; Schapire, 2003).

The AdaBoost algorithm (Adaptive Boosting) was introduced in 1995 (Freund & Schapire, 1995) as an algorithm for solving classification problems. It is used to boost the classification performance of a simple learning algorithm by combining weak classification functions to form a stronger classifier. It has been successfully employed in the selection of a reduced number of critical features. A small number of important features are used to create very efficient classifiers, which in turn trains an over-completed feature set to obtain a reduced set of critical features used for classifying scanned image sub-windows as faces or non-faces.

As shown in the equation below the weak classifier $h_j(x)$ consists of a feature f_j, a threshold θ_j and a parity p_j, indicating the direction of the inequality sign (Lai, Marculescu, Savvides, & Chen, 2008):

$$h_j(x) = \begin{cases} 1, & if\ p_j f_j(x) < p_j \theta_j \\ 0, & otherwise \end{cases} \quad . \tag{4.4}$$

The final strong classifier is shown in Eq. (4.5). Given a test sub-window *x*, the strong classifier would classify *x* as a face if the output is one.

$$h(x) = \begin{cases} 1, & if\ \sum_{t=1}^{T} a_t h_t(x) > \theta \\ 0, & otherwise \end{cases} \quad . \tag{4.5}$$

Where $h_t(x)$ is the weak classifier, and a_t is the coefficient for h_t.

4.3.1.3 The Cascade of Classifiers

A reduced set of features was not enough to reduce the vast amounts of computation in a detector task. To reduce the degree of computation, increase the speed of the detection process, and increase the detection performance, Viola & Jones describe the degenerative tree, where the classifiers are arranged in cascade architecture as shown in Figure 4.4. In the cascade architecture, a series of classifiers are applied to every sub-window. Negative sub-windows will be rejected and positive sub-windows will be detected, in the beginning stages by the initial classifier, with fewer features and less computational time. Subsequent layers eliminate any additional negatives but require additional computation. The cascade classifiers in the final stages then evaluate only the sub-windows that have passed the simple classifiers.

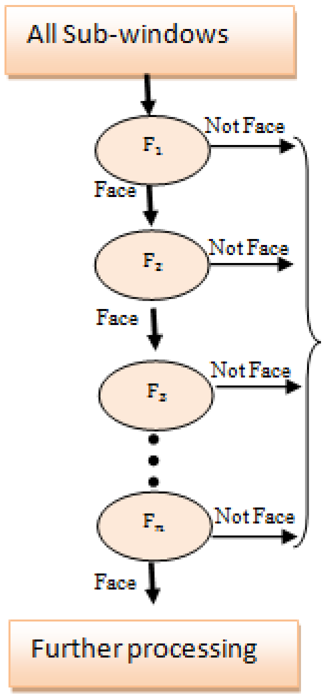

Figure 4.4: Schematic depiction of a detection cascade.

After several stages of processing, many of the negative sub-windows are rejected while detecting almost all positive instances. The background region will be eliminated, while the focus will be more on those regions in the face-like region.

AdaBoost algorithm can be summarized by the following steps (Gao , Sang , & Tang 2010)**:**

1- Let the set of training pairs be $(x_1, y_1), \dots, (x_n, y_{n)}$ where $y_i = \{0,1\}$ for negative and positive samples respectively.

2- Initialize the weight $W_{1,i} = \frac{1}{2m}\ for\ y_i = 0\,, and\ W_{1,i} = \frac{1}{2l}\ for\ y_i = 1$ where the *m* and *l* are the number of negatives and positives training sets respectively.

3- For t =1,…T :

3.1- Normalize the weights

$$W_{t,i} \leftarrow W_{t,i} / \sum_{j=1}^{n} W_{t,j}\ , \tag{4.6}$$

w_t is a probability distribution.

3.2- For each feature, j, train a classifier h_j which is restricted to using a single feature. The error is evaluated with respect to

$$W_t\ , \varepsilon_j = \sum_i w_{t,i} \left| h_j(x_i) - y_i \right| . \tag{4.7}$$

3.3- Choose the classifier, h_t , with the lowest error ε_t.

3.4 - Update the weights:

$$W_{t+1,i} = W_{t,i}\ \beta_t^{1-e_i} \quad , \tag{4.8}$$

where $e_i = 0$ if example x_i is classified correctly,

$e_i = 1$ otherwise , and $\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$.

4- The final strong classifier is:

$$H(x) = \begin{cases} 1\,, \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0, \qquad\qquad otherwise \end{cases} , \tag{4.9}$$

where $\alpha_t = log \frac{1}{\beta_t}$.

4.3.2 Detection Based On Skin Color

The second method of face detection is based on color in combination with the feature-based detections. In this method of face detection, lighting compensation was used to improve the performance of the color-based system as well as to minimize the computation complexity of the feature-based scheme (Hsu, et al., 2002; Pai, Ruan, Shie, & Liu, 2006; Vezhnevets, Sazonov, & Andreeva, 2003).

Pai et al. (Pai, et al., 2006) reported that, the method has been proven effective on facial variations such as dark/bright vision, closed eyes, open mouth, half-profile face, and pseudo-faces with complex backgrounds and cartoon/human face discrimination. The algorithm steps applied to implement this method of face detection are explained in the following sections based on Pai et al. (Pai, et al., 2006).

4.3.2.1 Skin Color Detection

This method requires the application of the color-based technique - YC_bC_r in color space, to separate skin regions from non-skin regions. The follow up extraction of the facial features is done based on the human eyes, mouth, and the height to width ratio of the face. As the luminance of every image differs, the resultant images would naturally have different colour distribution. Therefore, the lighting compensation has rested on luminance to modulate the range of skin colour distribution. Firstly, the average luminance Y_{aveg} of an input image is computed as given in the equation below:

$$Y_{aveg} = \sum Y_{i,j} \ , \qquad (4.10)$$

where $Y_{i,j} = 0.3R + 0.6G + 0.1B$, $Y_{i,j}$ is normalized to the range (0,255), and i and j are the index of a pixel. Based on the Y_{aveg}, the compensated image $C_{i,j}$ is determined based on the equations:

$$R'_{ij} = (R_{ij})^T \ , \tag{4.11}$$
$$G'_{ij} = (G_{ij})^T \ , \tag{4.12}$$
$$C_{ij} = \{R'_{ij}, G'_{ij}, B_{ij}\} \ , \tag{4.13}$$

where $$T = \begin{cases} 1.4\ , \ Y_{aveg} < 64 \\ 0.6\ , \ Y_{aveg} > 192 \\ 1, otherwise \end{cases} . \tag{4.14}$$

The chrominance C_r is defined as follow:

$$C_r = 0.5R' - 0.419G' - 0.081B \ . \tag{4.15}$$

R and *G* are compensated to reduce the computation. Human skin is defined by a binary matrix:

$$S_{ij} = \begin{cases} 0\ , t_1 < C_r < t_2 \\ 1, otherwise \end{cases} . \tag{4.16}$$

Where t_1, t_2 are two thresholds experimentally defined as 1.5 and 0.8 respectively , "0" is the white point, and *1* is black point.

A low pass filter of 5 × 5 is used to remove high frequency noise. Firstly, S_{ij} is segmented into 5 × 5 blocks, where the number of white points of individual blocks is then computed. Next, every point of a 5 × 5 block is set to the white point for cases where the number of white points is greater than half the number of the total points. However, if the number of black points exceeds half the number of total points, the 5 × 5 block concerned is then modified to form a complete black block.

After the 5 × 5 low pass filter, there exist several skin-color regions, which could be candidate blocks for further human face analysis as defined in S_{ij}. To demarcate these regions for determining the skin colour blocks, four rectangular vertices are registered and stored. These are the leftmost, rightmost, uppermost, and downmost points. A rectangular region is then created through these four points thus forming a skin colour candidate block for the detection of facial features (Hsu, et al., 2002; Pai, et al., 2006).

4.3.2.2 Height to Width Ratio Detection

After candidate blocks localization, several regions, which could be the human face, are obtained. Then, the features-height to width ratio, mouth, and eyes are detected sequentially for every candidate block. Any of these three detections can eliminate the candidate blocks, thus the low computation module is given a higher priority for processing. The height to width ratio is a very fast and simple detection process. The size of the candidate block is assumed to the h x w .The height to width ratio *(h: w)* is defined to be out of range t_1 and t_2, candidate block is rejected as a face .The two thresholds t_1 and t_2 is experimentally defined as 1.5 and 0.8 respectively (Pai, et al., 2006).

4.3.2.3 Mouth Detection

For mouth detection ,a formula proposed by Araki, Shimada & Shirai (Araki, Shimada, & Shirai, 2002) is used to define the value of θ, using a vertical based histogram and some thresholds for locating the mouth and eyes pixels from the face block. The mouth region in the face block is then detected using a more complex detection algorithm:

1- Determine the height to width ratio for the candidate face.

Use θ to find the mouth pixels. The θ value is calculated for all pixels within a candidate block as defined by the equation below.

$$\theta = cos^{-1}(\frac{0.5(2R'-G'-B)}{\sqrt{(R'-G')^2+(R'-B)(G'-B)}}) \ . \tag{4.17}$$

2- The pixel will be determined to be part of the mouth by a binary matrix M :

$$M_{pq} = \begin{cases} 0, \ \theta < 90 \\ 1 \ , otherwise \end{cases} , \tag{4.18}$$

where “0” indicates that the pixel is mouth.

After finding the mouth pixels, the vertical based histogram is used to determine whether the mouth is in this block. The number of mouth pixels having the same y-coordinates is calculated.

3- Use w_h to store the values of the different y-coordinates. The maximum value of w_h is denoted by w_{max}, and the y-coordinate of w_{max}, represented by h_m . Thus, if w_{max} is less than the threshold *ths*, experimentally defined as (1/6) of the block width *w*, the block will be rejected.

4.3.2.4 Eyes Detection

After the mouth detection stage, the *y*-coordinate (h_m) of the mouth is defined. The y-coordinate of the eyes should be smaller than the y-coordinate of the mouth. Therefore height of the eye region must be less than h_m . This information allows detecting human eyes within smaller regions. These regions are defined through y-coordinate values between zero to (h_m-w_{may}). Due to the deeper lineaments around the human eye region, the existence of human eye pixels through an appropriate luminance could be detected, which is supposed to be slightly darker than the average skin-color. The pixels around human eyes are defined by $E_{\hat{h}w}$:

$$E_{\hat{h}w} = \begin{cases} 0, th_1 < Y < th_2 \\ 1, otherwise \end{cases} , \tag{4.19}$$

where $\hat{h} = h_m - w_{max}$, and the two thresholds th_1 and th_2 is experimentally defined as 65 and 80 respectively. It is assumed that the candidate block has human eye pixels if there exist α values greater than the threshold β. The α and β values were determined by $\alpha = 0.5w_{max}$ and $\beta = w_{max}$. The blocks, which pass through three feature detections, height to width ratio, mouth detection, and eyes are considered as human face (Pai, Ruan, Shie, & Liu, 2006).

4.4 Image Segmentation

The process of partitioning a digital image into various segments has been referred to as 'Image Segmentation' in the field of computer vision. This process simplifies and/or changes the image representation for the purpose of more rapid and accurate analysis. This operation is extremely relevant in many applications of digital image processing and computer vision since it is the initial step of low-level image analysis, processing, and information extraction. The objective of this operation is to cluster pixels into salient image segments, decomposing the image into parts useful for the application concerned.

Image segmentation is a multiple objective operation, involving processes such as pattern representation, feature selection, feature extraction, pattern recognition, image compression, and image editing. The quality of the segmentation depends on the input digital image (Thomas Deselaers, Rybach, Dreuw, Keysers, & Ney, 2005; Gupta, Saxena, Singh, Dhami, & Singh, 2012; Thilagamani, 2011).

In image retrieval, either a local or a global visual content descriptor is employed. The global descriptor describes the visual features of the whole image, whereas a local descriptor focuses only on the visual features of regions or objects. To utilize the local visual descriptor, the prerequisite is to divide an image into parts. The simplest way of image segmentation is to stack the image into tiles of equal size and shape using a digital partition. This does not generate perceptually meaningful regions but represent the global features of the image at a finer resolution. A more advanced method is to divide the image into homogenous regions based on criterion defined in respective region segmentation algorithms (Long, Zhang, & Feng, 2003).

Some image retrieval systems retrieve the image based on objects, affecting therefore only part of the database. In this case, image segmentation is typically used to locate objects and boundaries (lines, curves, etc.). A more complex image segmentation procedure entails a complete object segmentation to obtain semantically meaningful objects (like ball, car, and horse with a general-purpose system). Currently, it is doubtful where automatic object segmentation can be successful in broad domains of general images. As an image normally contains more than one objects, the challenge confronting researches is to segment the image based on object features to extract meaningful objects (Long, Zhang, & Feng, 2003)(Thomas Deselaers, et al., 2005).

Facial image segmentation is applied in some face detection systems to help locate a face in a given large image, since most face classification techniques work only with face images. Therefore, face segmentation has to correctly extract only the face portion of a given large image. The technique is carried out based on a skin color segmentation algorithm that classifies skin-colors and non-skin-colors (Aiping, Lian, Yaobin, & Ning, 2010; Lakshmi & PatilKulakarni, 2010).

Facial image segmentation based on template matching are employed in some previous works for the extraction of facial features such as eye corners and centers, mouth corners and center, and nose corners etc. to be used for further processing.

4.5 Visual Contents of the Image

Visual contents are pertinent in content-based image retrieval to facilitate fast and efficient retrieval of similar images from the image databases. Retrieving images by their content, as opposed to external features, is becoming more universally accepted. What is fundamentally important is that content-based image retrieval rests on the technique employed for comparing images.

Visual contents of the images in the database are extracted and described in multi-dimensional feature vectors. These extracted feature vectors will then form the feature database. There is no single feature, which could sufficiently perfectly represent the whole content of an image. The combination of two or more features best represents image content. In our current research, we employed two types of features to represent facial image content:

1- General visual content features, represented by the color features of the facial image.
2- Domain-specific visual content features, represented by eigenfaces features of the facial image.

4.5.1 Domain-Specific Visual Content

Human faces are quite complex and multidimensional. Changes in facial identity and variation among images of the same face do occur. In the field of computer vision, dealing with facial image has been regarded as the most complex and challenging issues. This is due to problems arising from the following factors as reported in (Yang , Kriegman , & Ahuja 2002) :

1. Pose: Varying face images because of different camera-face positions (frontal, 45 degree, profile, upside down), and partial or whole occlusions of facial features such as an eye or the nose.
2. Structural component: Presence or absence of features such as beards, mustaches, and glasses of varying shape, colour and size.
3. Facial expression: Different facial expressions would result in different facial appearances.
4. Occlusion: Partial occlusion of faces by objects including faces of other people particularly in-group photographing.
5. Image orientation: Variation of face images resulting from rotation about the camera's optical axis.

6. Imaging conditions: Face appearance is affected by environmental factors such lighting (spectra, source distribution, and intensity) and camera characteristics (sensor response, lenses).

Therefore, analyzing the facial image is a very high-level computer vision task, where many vision techniques are involved.

Extracting relevant features from facial images is the initial step in human face identification. Research in this field has primarily intended to generate sufficient reasonable familiarities of human faces to facilitate the correct face identification by users. Several researchers in recent years have indicated that certain facial characteristics have been utilized by users to identify faces. Numerous face recognition methods have been suggested (Zhao, Chellappa, Phillips, & Rosenfeld, 2003). Two basic techniques are frequently used for feature extraction.

- The first technique is information-theory based face recognition, or finding a computational model that best describes a face by extracting the most relevant information it contains. Application of an algorithm called the Principal Components Analysis (PCA) to a database of standardized faces (called eigenfaces) can derive the information that best describes a face from a given image. A small set of characteristics is used to describe the variation between faces.
- The other technique, feature-based recognition, uses deformable templates and active contour models with complex geometry mathematics. This method employs a different algorithm to extract mathematical descriptions of basic facial components - eyes, nose, mouth, and chin - as well as their relationships to each other. This information is gathered and converted into a feature vector. Such method is used by (Yuille, Hallinan, & Cohen, 1992). They played a great role in adapting deformable templates to contour the extraction of face images

(Agarwal, Jain, Kumar, & Agrawal, 2010; Atalay 1996). This technique requires detailed geometrical data, extensive computation and is highly complex. It does not deal well with multiple views and has often proven to be fragile, requiring a good "initial guess" as a guide (Turk & Pentland, 1991).

4.5.1.1 Eigenfaces Features

Eigenfaces are features, which characterize global variation among face images. They are essentially a set of eigenvectors used in computer based facial recognition, ,where the input signals of the faces are grouped into classes based on both facial characteristic features (eyes, nose, mouth) and relative distances among these features. The features are extracted from the face images using a mathematical tool, namely, the Principal Component Analysis (PCA). The PCA transforms each original training set image into corresponding eigenfaces.

Each eigenfaces represents certain features of the face, and is provided with a certain weight, which specifies the extent of the specific feature occurring in the original image. Eigenfaces reduce the computation and space complexity. As each facial image is represented by a limited number of dimensions. The "best" eigenfaces is given the largest eigenvalues and eigenfaces that have low eigenvalues are omitted. The high valued eigenfaces will form the "face space" of all the images. The eigenfaces approach has been regarded as the first working facial recognition technology, and it has become one of the top commercial face recognition products (Vijaya Lata, Tungathurthi, Rao, Govardhan, & Reddy, 2009). Relevant face information and their variations are extracted from the principal components of the distribution of faces, or the eigenvectors of the covariance matrix of the set of face images. The use of principal components to represent human faces was originated by Sirovich and Kirby (Sirovich & Kirby, 1987) and used by Turk and Pentland (Turk & Pentland, 1991) for face detection and recognition.

4.5.1.1.1 Principal Component Analysis

Principal component analysis (PCA), a mathematical procedure, transforms orthogonally a set of correlated variables into a new set of unrelated variables, called the principal components. The number of principal components generated is less than or equal to the number of original variables. This statistical tool has been used in many applications including image compression and pattern recognition of high dimensionality image data sets. Among others, it has been employed in facial recognition as it is easy to describe and understand mathematically (Asadi, Rao, & Saikrishna, 2010). In face recognition, PCA has been used to compute eigenvectors of a covariance matrix, transforming the original high dimensionality data sets into a lower-dimension feature space, defined by eigenvectors with large eigenvalues. The advantages of PCA are summarized as follows:

- PCA used in face recognition is based on the information theory approach, where the relevant information in a face image is extracted and encoded efficiently. Recognition is performed on a face database that consists of similarly encoded models.
- PCA is most efficient in data dimensionality reduction, in terms of data compression. This has enabled high dimensionality image data sets, to be represented by lower dimensionality data sets, reducing the complexity of grouping the images.
- There is no data redundancy, as the principal components are orthogonal (uncorrelated). With PCA, the complexity of grouping the images is reduced (Asadi, et al., 2010).
- The trained images are not stored as raw images, rather as weights, determined by transforming individual trained image sets into the corresponding sets of eigenfaces.

4.5.1.1.2 Extraction of Eigenfaces

To extract the eigenfaces by principle component analysis, the following steps are applied based on Turk and Pentland (Turk & Pentland, 1991):

1. Prepare the data

   In the beginning, the faces constituting the training set should be prepared for processing.

2. Convert images to vectors

   Convert each image $I_i(x, y)$ into a vector T_i, and represent the whole matrix (RxT) where R is the number of training image.

$$I_i = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{bmatrix}_{NxN} \xrightarrow{Concatenation} T_i = \begin{bmatrix} a_{11} \\ \vdots \\ a_{1N} \\ \vdots \\ a_{2N} \\ \vdots \\ a_{NN} \end{bmatrix}_{N^2x1}, \quad (4.20)$$

   Let this set of face images be $T_1, T_2, \dots T_M$, N is the image size and M is the size of the training set.

3. Calculate the mean

   The average face of the set is defined by:

$$\Psi = \frac{1}{M} \sum T_M. \quad (4.21)$$

4. Subtract the mean

   Subtract the mean face from each original face vector T_i and the result stored in the variable Φ_i where

$$\Phi_i = T_i - \Psi\ . \quad (4.22)$$

   The purpose of subtracting the mean image from each image vector is to be left with only the distinguishing features from each face and "removing" any common information.

5. Calculate the covariance matrix

The covariance matrix has simply been made by putting one modified image vector obtained in one column each. The covariance matrix c is calculated according to

$$c = \frac{1}{M}\sum_{n-1}^{M} \Phi_n \Phi_n^T \; . \tag{4.23}$$

$$c = \frac{1}{M}\sum_{n-1}^{M} \Phi_n \Phi_n^T \quad = A \, . A^T . \tag{4.24}$$

T is the transpose of the matrix. The matrix $A = \Phi_1 \quad \Phi_2 \quad ... \; \Phi_M$,

$C \;\; is \; N^2 x N^2 and \;\; A \; is \;\; N^2 x M$.

6. Calculate the eigenvectors and eigenvalues of the covariance matrix

In this step it is necessary to find the eigenvectors v of matrix C, however, $C \; is \; a \; N^2 x N^2$ matrix, which means it will produce N^2 eigenvectors of N^2 dimensional and this is a huge number .

Consider the eigenvectors v of $A \, . A^T$ such that

$$A^T A v_i = u_i v_i \; . \tag{4.25}$$

Multiplying both sides by A ,we have

$$A A^T A v_i = u_i A v_i \; . \tag{4.26}$$

The Av_i are the eigenvectors of $C = AA^T$.

From these analysis, they construct an MxM matrix, $L = AA^T$, $L_{mn} = \Phi_m^T \Phi_n$ and find M eigenvectors, v_i of L. These vectors determine linear combinations of the *M* training set face images to form the eigenfaces u_i,

$$u_i = \sum_{k-1}^{M} v_{lk} \Phi_i \, , \; l = 1, \dots, M, \tag{4.27}$$

Where l is an MxM matrix, v are *M* eigenvectors of l and u are eigenfaces.

With this analysis, the calculations are greatly reduced, from the order of the number of pixels in the images (N^2) to the order of the number of images in the training set (M). In practice, the training set of face images will be relatively small (M <=N^2), and the calculations become quite manageable.

4.5.1.1.3 Classification A new Projected Face

The eigenfaces components of the new face image is found by the operation

$$W_k = u_k^t(T - \Psi),\ k= 1,...M', \tag{4.28}$$

where each normalized training image is represented as a vector.

$$\Omega_i = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \end{bmatrix},\ i = 1,2, \dots M.,\mathrm{k} = 1, \dots M', \tag{4.29}$$

where Ω_i is the projected face and w is the contribution of a single eigenface. Such vectors must be calculated for every image in the training set and stored as a template. The high dimensional space with all the eigenfaces is called the face space (feature space). In addition, each image is actually a linear combination of the eigenfaces. Face images lie in a low dimensional space. Facial images of the same person are close together to one class whereas facial image of different people are further away, Figure 4.5 and Figure 4.6 adapted from the work of (Golland, 2005; Turk & Pentland, 1991) illustrate the faces space and classification of the new face in the face space.

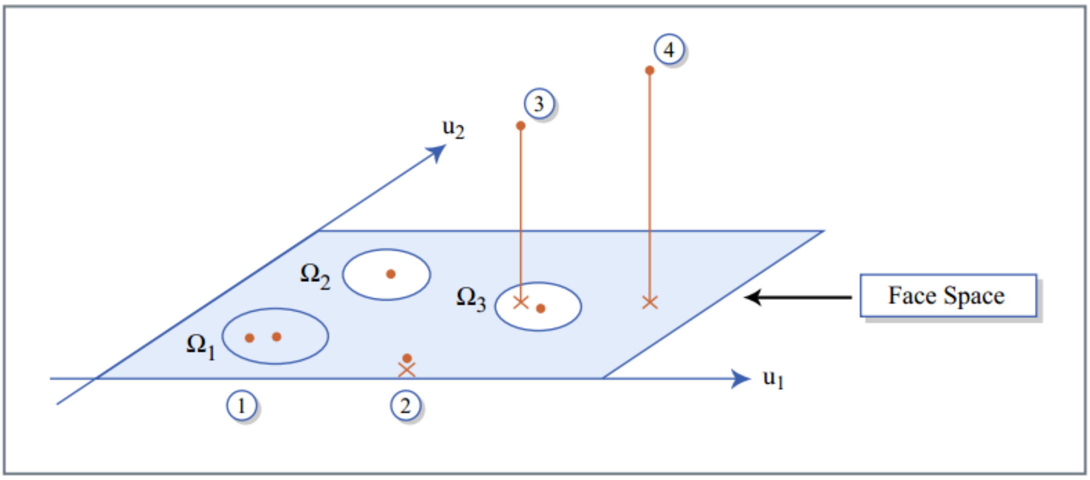

Figure 4.5: Face space illustration.

The new projected image can be classified as follows:

The new image is a face if

$$\|T - \Psi - \sum w_k u_k\| < \theta_\varepsilon \ . \tag{4.30}$$

Otherwise, it is not a face .The new face belongs to class n if

$$\|\Omega - \Omega_n\| < \theta_\delta \ . \tag{4.31}$$

Otherwise, it is a new face belonging to a new class.

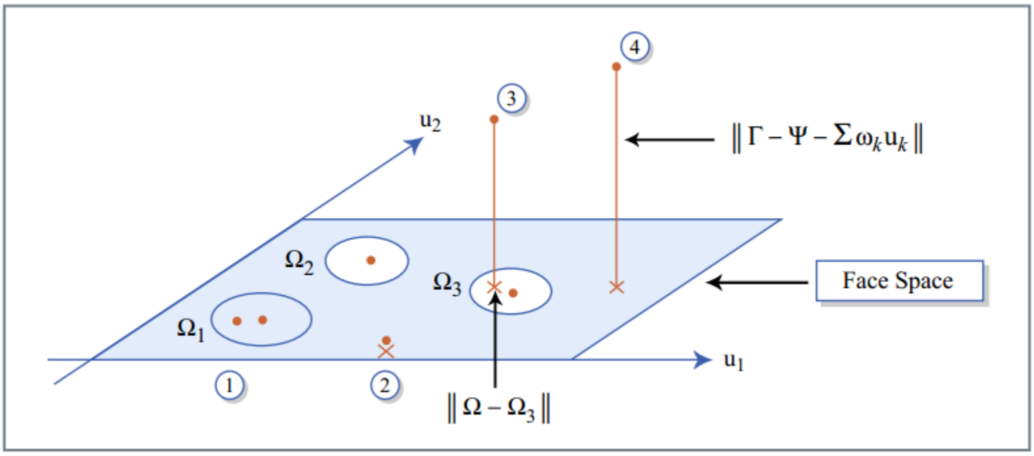

Figure 4.6: Classification of the new face in the face space.

If the residual is too high, it is not a face and if the projection face is close to one class it belongs to this class. Otherwise, it is a new face. θ_ε *and* θ_δ of equation 4.30 and 4.31 are chosen thresholds.

Using the feature vectors and the eigenfaces, an image in the face space can be reconstructed as follows.

$$\mathrm{T}' = \Psi + \Phi_\mathrm{f} \ , \tag{4.32}$$

where $$\Phi_f = \sum_{i=l}^{M} w_i u_i \ . \tag{4.33}$$

Eigenfaces with a contribution of *w* to the average of the training set images can be used to rebuild the required face image in the face space.

From the previous steps of eigenfaces extraction, the produced eigenfaces are equal to the produced eigenvalues that equals the training set. It is not practical to use all training vector weights in the eigenfaces vectors, because more eigenfaces require more time during the retrieval process, especially when the database is large. There is a tradeoff between the time needed and retrieval accuracy (Turk & Pentland, 1991).

4.5.2 General Visual Content - Color Feature

Colour is one important feature that has enabled recognition of images by humans. Its hue and intensity is dependent on the reflection of light to the eye and the processing of information received in the brain. Colour has assisted us to differentiate among objects and places according to the time of the day. With computer vision, colour is the most intuitive information that can be extracted for comparison of image characteristics, which has been widely used as a visual feature in image retrieval. This is justifiable because color is a powerful descriptor for image objects identification, and humans can discern thousands of shades and intensities of color, compared to about two dozen shades of gray.

An important criterion is that the color system is independent of the imaging devices used, especially when different imaging devices such as scanners, cameras, and camera-recorders (e.g. images on Internet) are used to record image sets. Another prerequisite is that the color system should exhibit perceptual uniformity, meaning that numerical distances within the color space should conform as close as possible to human perceptual differences. This is important when images to be retrieved are required to be visually similar (Kaur & Banga 2011).

Several algorithms have been developed since the late 1980s to extract color information from images. One basic form of color retrieval involves specifying targets of color values that can be searched from the image data sets. This basic method is even confronted with operational challenges due to the different manners in which computers and humans 'perceive' colors. Computers perceive all visible colors with a combination of color components. Thus, images perceived by a computer as having a large component of red may not necessarily appear “reddish” to the human eye (Chakravarti & Meng, 2009).

Color features can be represented by numerous descriptors. The commonly used color descriptors are color moments, color histograms, color coherence vectors, and color correlograms. A previous study (Kodituwakku & Selvarajah, 2004) carried out an experimental comparison of these different color descriptors for content-based image retrieval. The results indicated that the color histogram had performed well compared to other descriptors. This is an efficient representation of color content and it is fairly insensitive to variations caused by camera rotation or zooming (Smeulders, Worring, Santini, Gupta, & Jain, 2000).

Color descriptors of images are both global and local. Both techniques are proven useful for the retrieval of images and are suitable for different query types.

A global descriptor is recommended for a sample image query. For example, in the current research, if the user is interested in finding a photo of a desired person, then providing one sample image of the person would allow other images to be found. Global color descriptors are suitable in this case because the user does not require information of positions of colored regions in the images. However, if the user requires locating image colored regions, the global color descriptor does not provide the means to do so. A localized or regional color caters for partial or sub-image matching between images. For example, if the user is interested in finding images with a red spot in the upper right corner, then a regional descriptor allows this query to be answered.

An operational system is needed for the automated extraction and efficient representation of color in both local and global descriptors. A localized or regional color descriptor generally requires more effective extraction and representation as it deals with local regions (Smith & Chang 1996).

Three colour coordinates are required to determine color position within the color space. This concept is described and illustrated below.

4.5.2.1 Color Space

The color of an image is represented through color models. The color model describes the color information of the image. The model enables a user to specify, create, and visualize color. Humans define a color by its brightness, hue, and colorfulness.

A computer describes a color through the amount of red, green and blue phosphor emissions required to match a color (Ford & Roberts, 1998; Tkalcic & Tasic, 2003).

The purpose of a color model is to facilitate the specification of colors to a common standard. Several color representations are currently in use for color image processing. However, the most popular and commonly used ones include RGB (red, green, blue), HSV (hue, saturation, value) and HSI (hue, saturation, intensity) also known as HSL (hue, saturation, lightness/luminance).

4.5.2.1.1 RGB Color Space

The humans perceive color as a combination of primary colors -Red, Green, Blue, which form a color space. Additive colors are also obtained by varying the combinations of these primary colors. The colour guns of red, green, and blue are combined to create color composites in the computer monitor. It is not perceptually uniform, meaning this colour composite variation is not always perceived as the same color variation in the human brain. Practically speaking, this means that the measure of the variation perceived by a human is different from the computer based mathematical distance.

The RGB colour space is defined as a unit cube with red, green, and blue axes as illustrated in Figure. 4.7. Thus, a vector with three co-ordinates of RGB represents the colour in the colour space, for instance black is represented by the RGB coordinates of 0,0,0 and white color is represented by 255,255,255. Other color spaces operate in a similar fashion but with a different perception.

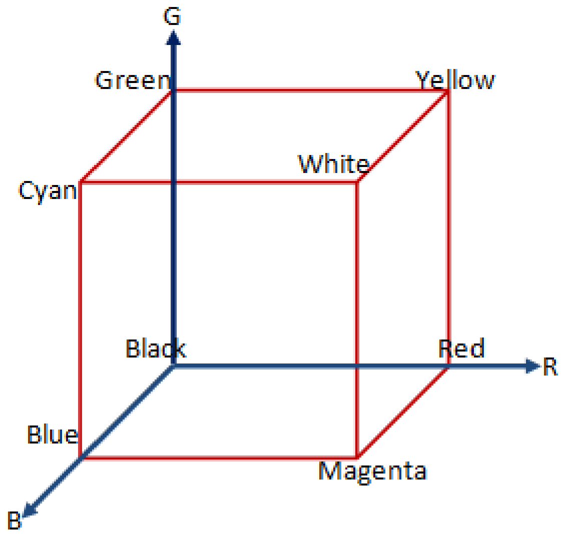

Figure 4.7: RGB color space coordinates.

4.5.2.1.2 HSV Color Space

The HSV (Hue, Saturation, and Value) color space is a simple transform from the RGB color space, in which all the existing image formats are represented. The HSV color space is a popular choice for manipulating color. It was developed to provide an intuitive representation of color and to approximate the way in which humans perceive and manipulate color, but are perceptually not uniform. Figure 4.8 shows the HSV color space coordinate system (Smeulders, et al., 2000).

RGB to HSV is a nonlinear, but reversible transformation. *H*, the hue, represents the chromatic component in this model and it is the definition of a color by the combination of the primary colors. It specifies one color family from another, as red from yellow, green, blue, or purple. Saturation or *S* refers to how little the color is mixed with white light and the *V*, the value, refers to how little the color is mixed with black.

Saturation refers to the predominance of a particular hue in a color. The hue (color) is invariant to the illumination and camera direction, and thus suitable for object recognition.

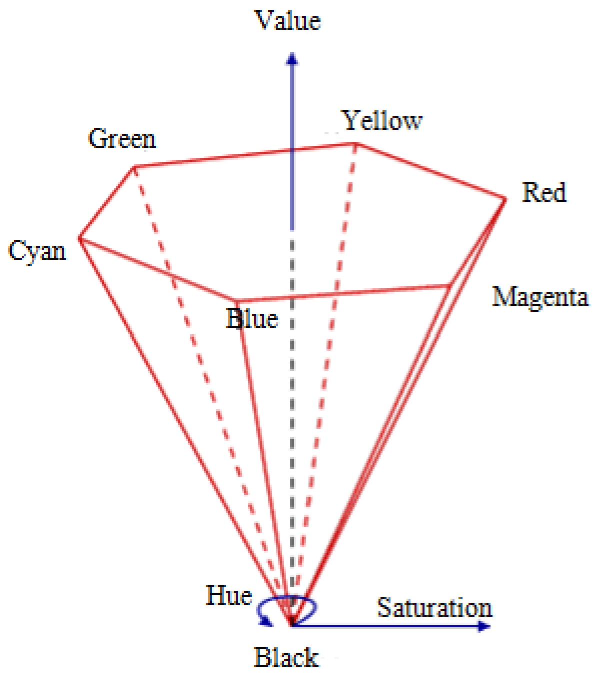

Figure 4.8: HSV color space coordinates.

The HSV color space model is derived from the RGB space cube (Choras, 2007; Tkalcic & Tasic, 2003), where the hue is given by :

$$H = cos^{-1}\left(\frac{1/2[(R-G)+(R-B)]}{[(R-G)^2+(R-B)(G-B)]^{1/2}}\right), \tag{4.34}$$

the saturation S is given by :

$$S = 1 - \left(\frac{3}{R+G+B}\right)\min(R, G, B), \tag{4.35}$$

and the value V is defined as the largest component of a color .

$$V = \max(R,G,B). \tag{4.36}$$

4.5.2.1.3 HIS Color Space

The HIS (Hue, Intensity, and Saturation) color model's hue and saturation are derived from the RGB space cube (Tkalcic & Tasic, 2003) in similar way to the HSV derivation. While the simplest definition of intensity is just the average of the three components, so the intensity, I is defined by:

$$I= (R, G, B)/3. \tag{4.37}$$

4.5.2.2 Color Histogram

Color histogram represents the intuitive information that can be extracted from images. It is the most effective and direct way in distinguishing visually, colour features available in an image. Images characterized by the color histogram features have many advantages, which are listed below (Chen, Gagaudakis, & Rosin, 2000; Choras, 2007; Hu, 1962; Swain & Ballard, 1991):

- Robustness- The image color histogram is invariant to image rotation at its view axis, as well as changes in small steps when rotated or scaled. It is also insensitive to changes to image and histogram resolution and occlusion.
- Effectiveness- Relevance between the query image and the extracted matching images remains high.
- Implementation simplicity- The colour histogram construction process is straightforward, which includes image scanning, colour assignment based on histogram resolution and histogram building using color indices.
- Low storage requirements- The color histogram storage size is significantly smaller than the image itself, assuming color quantization.

However, each feature descriptor has its drawbacks. For this descriptor, its feature vector dimension is quite large. For example, the number of bins in a typical color histogram ranges from tens to a few hundreds. The high dimensionality of the feature vectors would result in high computational cost in distance calculation for similarity retrieval, as well as to the search inefficiency.

One method proposed in previous works to overcome these problems is the color moments descriptor. The color moments descriptor proposed by (Stricker & Orengo, 1995) has a compact representation, which included average, variance, and the third-order moment of the colors in the image. Ma and Zhang (Ma & Zhang, 1998) showed that the color moment descriptor has performed slightly worser than a high-dimension color histogram. One drawback observed is that the average of the colors is different from any of the original colors. This means that it is difficult to recover the actual colors in the image. However, the color histogram is also quite compact, and requires only a small number of colors to characterize the color information in an image region (Deng, Manjunath, Kenney, Moore, & Shin, 2001).

4.5.2.2.1 Color Histogram Quantization

The difficulty with histogram-based retrieval as has been mentioned before is the high dimensionality of color histograms. For a true color image, the number of colors is 256 x 256 x 256, that is, $2^{24} = 16{,}777{,}216$ colors. A huge amount of time will be needed to compute and compare the bins of one histogram with others. In order to reduce the computation without a significant reduction in image quality, a representative color is extracted, to represent the image, thereby reducing the storage space and enhancing speed.

A color quantization technique is a process that reduces the number of distinct colors used in an image. Which means that some pre-specified colors is present in the image and each color is mapped to some of these pre-specified colors. The intention of color quantization is that the new image should be as visually similar as possible to the original image.

A standard quantization scheme divides each axis of the image color space into a certain number of fractions. If the axes are divided into *r, g,* and *b* parts, the used colors number that will represent an image will be: *n*= *r. g. b* (Chakravarti & Meng, 2009).

4.5.2.2.2 Histogram Generation

Regardless of which color space is used, the histogram of color images is generated by counting the number of the pixels that correspond to the specific color in the uniform quantization color. A color histogram H. for an image is defined as a vector $H = h_1$,$h_{2,}. . . h_j . . . , h_M$, where j represents a color in H , h_j is the number of pixels in color j , and M is the number of bins in H. A Color histogram refers to the probability mass function (PMF) of the image intensities and can be defined by:

$$H_{A,B,C}(a,b,c) = N\ x\ P(A - a, B = b, C = c). \tag{4.38}$$

Where A, B and C represent the three-color channels and N is the number of pixels in the image (Smith & Chang, 1996).

In order to compare images of different sizes, the color histogram values are normalized by dividing the number of pixels in each histogram bin by the number of pixel values used in the comparison as given in the equation below.

$$h(i) = \frac{h(j)}{N}, j \in [0, \dots, M]. \tag{4.39}$$

Figure 4.9 depicts a color histogram as a bar graph, where each bar represents a particular color of the color space being used. The bars in a color histogram are referred to as bins and they represent the x-axis, meaning the number of bins will depend on the number of colors occurring in an image. The y-axis denotes the number of pixels there are in each bin.

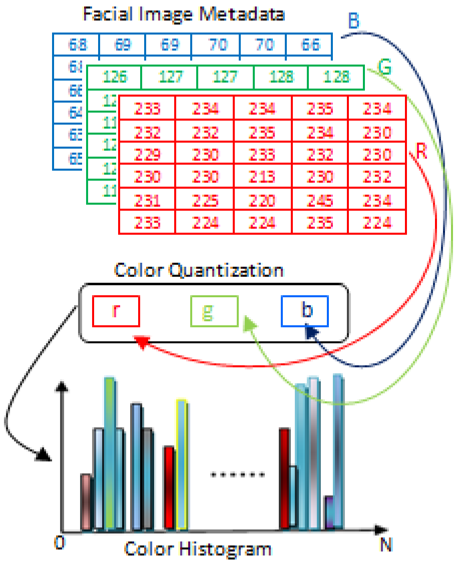

Figure 4.9: Illustration of the color histogram generation.

4.6 Image Semantic Contents

An image attribute is any kind of feature or component that can be represented by an information processing system. Semantic contents attributes are those used to describe high-level concepts that appear in images. The semantic attributes of the image generally can be in different types such as (i) perceptual attributes that are directly related to a visual stimulus (e.g., color, shape, texture, body parts, motion, visual component), (ii) interpretive attributes requiring both interpretation of perceptual cues and a general level of knowledge or inference from that knowledge (e.g., the artist of a painting, relationship, activity, event, similarity) and (iii) reactive attributes describing a personal response or emotion (e.g., the attractiveness of a face, personal reaction). Using these semantic concepts, we can extract the hidden attributes in the image as well as exploit the semantic relations between the images through the image semantic space and the relationships among these attributes themselves. Furthermore, it also allows the measurement of their semantic similarity (Jörgensen, 1998).

4.6.1 Human Face Semantic Feature Types

Semantic features of the human face are expressed by verbal descriptions. Each description consists of keywords and specification of sizes and lengths of face parts. There are five types of semantic features:

1. Demographic features such as age, race, and gender.
2. Impressions implied from a face image, using descriptive keywords for character or personalities, such as “serious” and “happy”.
3. Skin color of a face and face parts such as fair and dark complexions, tanned face, and blue eyes.
4. Features of a face image – face parts, such as a flat nose, and large eyes; related to size and lengths of face parts, which are informal (natural) components of any human face, arranged in a somewhat similar configuration and are what has made individual human faces unique lies in the subtle variations in the form and configuration of the features.
5. Description of other components and accessories of a face such as hairstyle, moles, a pair of glasses, and earrings. Basically, these are artifacts and extra components on faces (Ito & Koshimizu, 2004; Jörgensen, 1998).

4.6.2 Facial Image Annotation

Image annotation, also known as image tagging or linguistic indexing, is the process of assigning metadata in the form of keywords to a digital image. The annotation process enables a user to retrieve, index, organize and understand large collections of image data sets. The goal of image annotation is to assign a few relevant text keywords to a given image, reflecting its visual content.

This technique has been used in image retrieval systems to organize and locate images of interest from a database. It can be regarded as a type of image classification of a very large number of classes. The challenge to researcher now is to assign a richer, more relevant set of keywords to further exploit the fast indexing and retrieval architecture of image search engines to enhance image search performance (Sumathi, 2011).

It is apparent that image retrieval techniques like the content-based image retrieval approach, complement deficiencies of previous information retrieval techniques. In this technique of search and retrieval, the actual contents of the image are analyzed using image analysis techniques. Unfortunately, this method of retrieval requires a user to submit a query as image content. Furthermore, certain image features occurring in sampled images may override the focused concept of the user concerned. Generally, current image search solutions have failed to effectively utilize image content for image search, leading often to search results of limited applicability. On the other hand, image annotation allows the user to more naturally specify queries, which is an advantage over the content-based retrieval technique (Makadia, Pavlovic, & Kumar, 2010).

Image annotation continues to be an important research issue in the information retrieval communities. Ongoing researches in this area have indeed led to several interesting techniques. The current technique of image annotation is manual operated, whereas ongoing researches have focused on the development of automatic image annotation. This has been a difficult task for two main reasons:

First is the well-known pixel-to-predicate, or semantic gap problem, which points to the fact that it is hard to extract semantically meaningful entities using just low-level image features, unambiguous recognition of thousands of objects or classes reliably is currently an unsolved problem.

The second difficulty arises due to the lack of correspondence between the keywords and image features in the training data. For each image, one has access to keywords assigned to the entire image and it is not known which features of the image correspond to these keywords. This makes the direct learning of classifiers by assuming each keyword to be a separate class, difficult (Makadia, et al., 2010)(Sumathi, 2011).

To resolve these problems at the human level, it is important that the lack of scene understanding be addressed first. Currently, identifying objects, events, and activities in a scene is still being researched into with limited success. In the absence of such scene information, most of the image annotation methods have opted to model the joint distribution of keywords and images to further understand the association of keywords and low-level image features. These probabilistic model based methods could only infer the correlations or joint probabilities between images and annotations. The classification-based methods have treated keywords (concepts) as classes and input images annotated by trained classifiers based on classification results. However, these state-of-the-art techniques would require elaborate modelling and training (Makadia, et al., 2010).

The challenge facing researchers in automatic annotations is to develop suitable models to assign visual terms to an image to successfully describe it. To-date these state-of-the-art image annotation methods have performed unsatisfactory. By and large, the new annotations algorithms that have been developed have poorly performed specifically in the context of image retrieval (Sumathi, 2011).

4.7 Image Classification and Similarity Measure

The key component for image retrieval is the similarity measure. A common approach to compute a similarity metric among the patterns to be classified is by using the distance-based method. The distance metric, which defines the neighbours of a query point, is fundamental in the accuracy of classification and retrieval. As observed in previous works, the classification problem is treated with a traditional distance metric learning algorithm. By and large, the Euclidean distance has been widely used as a similarity measure. Puzicha et al. in (Puzicha, Buhmann, Rubner, & Tomasi, 2001) compared nine image dissimilarity measures empirically and showed that no measure has achieved the best overall performance. In this context, the selection of different measures will depend on the sample distributions. Yang and Jin (Yang & Jin, 2006) conducted a comprehensive survey of distance matrices. In spite of many successful works on distance matrices, it was found that these algorithms could not easily solve the problem of integrating varied features and finding the distance similarity among the vectors of these features, to generate a unique value for similarity ranking.

This traditional distance metric has logically been chosen in situations where it is fair to assume that all features are equally scaled and equally relevant. However, in most cases the data distribution is such that distance analysis along some specific directions in features space is more informative than along other directions.

For applications, where different algorithms and techniques extract the feature attributes, the above methods would be inefficient. This is also applicable to semantic features, extracted by different methods, resulting in variable weights. There are also other situations, where some features are considered more important than others, or some features would reflect negative effects on other features if they are not combined in a suitable way. In addition, there is also a multiclass situation, where an image contains multiple features of varying classes or where similar images belong to different feature classes. The features information and discriminative directions should be given due considerations, when ranking these images.

4.7.1 Euclidean Distance

Euclidean distance (ED) is used to calculate the similarity distance between two vectors for image retrieval system as shown in Figure 4.10. Differences are calculated by comparing each pair of values from the two vectors. These differences are squared and summed together. The square root of this value is taken as the following (Long , et al., 2003):

$$FacesSimilarityDistance\,(Q\,,D\ ) = (\textstyle\sum_{i=1}^{n}(Q_{\ i} - D_{\ i})^2)^{\frac{1}{2}}), \tag{4.40}$$

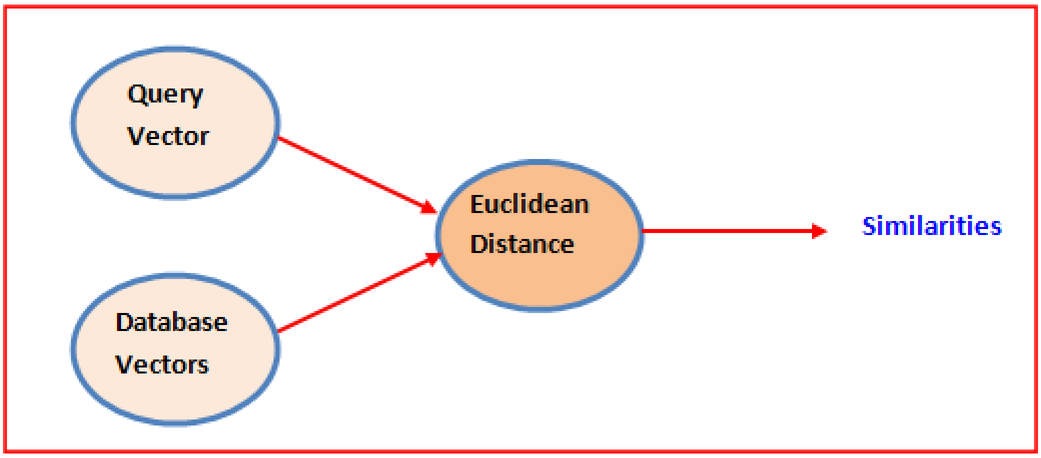

Figure 4.10: Euclidean distance for similarity measurement.

4.7.2 Radial Basis Function network

A Radial Basis Function network (RBFN) is an artificial neural network that uses radial basis functions as activation functions. Its theory is based on the function approximation theory. Radial basis functions (RBF) were first introduced by Powell to solve the real multivariate interpolation problem (Powel, 1981,; Renals, 1989). RBF networks were first used by Broomhead and Lowe (Lowe & Broomhead, 1988). Other major contributions to the theory, design, and applications of RBF networks can be found in papers by Darken, Poggio, and Girosi (Moody & Darken, 1989; Poggio & Girosi, 1990) (Sahin, 1997; Salnn, 1998).

RBF networks have been proven superior over other neural networks approaches (Sahin, 1997; Salnn, 1998; Yousef & Hindi, 2005), because of the following reasons:

- RBF networks are capable of approximating nonlinear mappings effectively.
- The training time of the RBF networks is significantly lower compared to that of other neural network approaches as the input layer and the output layer of an RBF network are trained separately and sequentially.
- RBF networks are quite successful for identifying regions of sample data not in any class because it uses a non-monotonic transfer function based on the Gaussian density function.
- The technique forms a strong link among fields such as function approximation, regularization, and pattern recognition. Therefore, it is an excellent candidate for pattern applications and many researchers have been successful in employing the RBF to hasten the learning process, normally required for the multi-layer feed forward neural networks (Haddadnia, Ahmadi, & Faez, 2002). It has gained wide acceptance in the pattern recognition and signal processing areas and has been employed mostly to address classification problems (Yousef & Hindi,

2005) confronting several applications which included face recognition ,speech recognition, medical diagnosis, and digital mapping (Simon, 2002).

4.7.2.1 Structures of Radial Basis Function Network

A Radial Basis Function Network has the architecture of a traditional three-layer feed forward neural network. Its design in its most basic form consists of three separate layers with respective feed forward architecture as appeared in Figure 4.11. The first layer is the input layer - a set of units of dimension k of the input feature vector x. The k-dimensional inputs $(x_1, \dots, x_k)$ broadcast the inputs to the second layer, which is a hidden layer of a set of units equal to the number of the training vector in the input feature vector. The output layer is the third layer, composed of nodes to respond to the activation patterns applied to the input layer through a summation of the output layer from the hidden layer. The weights $(w_{11}, \dots, w_{ij})$ transmit outputs of the hidden layer to the output layer.

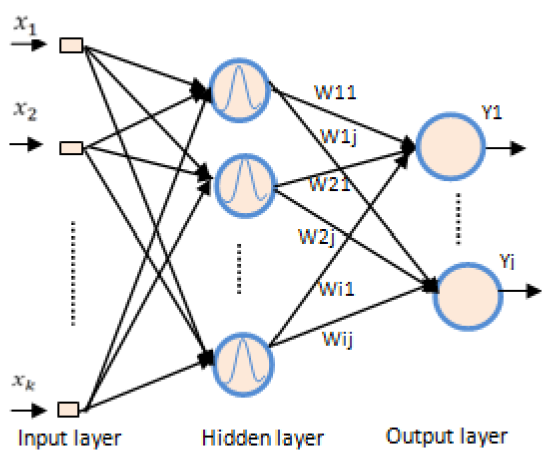

Figure 4.11: Radial Basis Function Network structure.

The mapping from the input space to the hidden-unit space is nonlinear with nonlinear activation function whereas the mapping from the hidden space to the output space is linear with a linear activation function. The distance between the input vector and a prototype vector (center vector) determines the activation function of the hidden units (Kurban & Beşdok, 2009; Simon, 2002). The activation function of the RBFN is expressed as follows:

$$h(x)_i = h_i\left(\frac{\| X\text{-}C_i \|}{\sigma_i}\right) \qquad \mathrm{i} = 1,2,\ldots.\mathrm{r}\ . \tag{4.41}$$

Where, *x* is an *n*-dimensional input feature vector, *c* is a n-dimensional vector called the center of the RBF unit, σ is the width of RBF unit and *r* is the number of the RBF units. This is typically derived from a Gaussian function with center *c* and width σ as follows:

$$h(\|X - C_i\|) = exp\left(\frac{(\mathrm{X} - C_i)^2}{\sigma_i}\right). \tag{4.42}$$

The parameter σ represents the standard deviation for the Gaussian function.

Let the input vector, the dimensional, and the scalar output be $X = (\, x_1, x,, \ldots\ldots, x_n)$, $n\ and\ F(x)$ respectively, then the input and output relationship is expressed as

$$F(x) = b_j + \sum_{i=1}^{I} w_{ij} h(\|x - c_i\|), \tag{4.43}$$

Where, w_{ij} is the RBF weight that connects the hidden neuron *i* to the output node *j*, C_i is the RBF center of the neuron *i*, and *I* is the number of neurons in the hidden layer. The norm is typically the Euclidean distance between the input x and one of the center c, and *h* is the Radial Basis Function, a Gaussian function. The function *F* is the RBF network, since it is expressed as a linear combination of RBF, *b* is the bias of the *i j-th* output. The bias is omitted in this network to reduce network complexity as shown in the equation below:

$$F(x) = \sum_{i=1}^{I} w_{ij} h(\|x - c_i\|). \tag{4.44}$$

We can write that as:

$$Y = \begin{bmatrix} w_{11} & w_{12} & \cdot & w_{1j} \\ w_{21} & w_{22} & \cdot & w_{1j} \\ \cdot & \cdot & \cdot & \cdot \\ w_{i1} & w_{i2} & . & w_{ij} \end{bmatrix} . \begin{bmatrix} h(\|x - c_1\|) \\ h(\|x - c_2\|) \\ \cdot \\ h(\|x - c_i\|) \end{bmatrix} \quad and \ \ Y = W.H, \tag{4.45}$$

where the weight matrix is represented as W, h matrix is represented as H and Y is the actual output. The RBF weight parameter W is determined by solving the following linear equation:

$$(H \ + \ \lambda I)W = D, \tag{4.46}$$

where $D = (d_1, d_2, \dots, d_M)$ is the desired output, and $\lambda \geq 0$ is adjusted value (Neruda & Vidnerová, 2009; Sahin, 1997; Salnn, 1998; Simon, 2002; Yousef & Hindi, 2005).

4.7.2.2 Radial Basis Function Network Training

In RBF network, many parameters need to be chosen to adapt the network for a particular task. The first parameter is the number of neurons in the hidden layer, which is very important in neural networks. Using more neurons than needed will cause an over learned network, which in turn will increase the complexity and eventually affects the generalizing capability of the network. Therefore, accurate determination of the number of neurons in the hidden layer is important. The second parameter is the center vectors c, which influences the performance of the radial basis function network in the hidden layer. Therefore, finding the optimal locations of the centers is significant. The third parameter that has to be chosen is a suitable activation function. As each neuron has an activation function, it is necessary to choose the suitable activation function for the network. From the literature, the most preferred activation function using the RBFN is the Gaussian function, which has a spread parameter that controls the behaviour of the function (Kurban & Beşdok, 2009) .

Finally and most importantly is determining the weights vector parameter, w between the hidden layer and the output layer. This parameter is very critical in the training process within the RBFS environment.

The supervised learning of neural networks is often used to address function approximation problem. Given the data set x, we are looking for the function that approximates the unknown function $f(x)$. Therefore, the goal is to approximate a given function using a set of examples – training set.

To measure the quality of the network an appropriate error function is used such as sum of squared errors (SSE), where the network is learned by minimizing the error between the desired and computed unit values. The network computes the following function:

$$f: R^n \rightarrow R^m \ , \ T = \{(x(t), d(t); t = 1, .., k\}, \tag{4.47}$$

where T is a training set — a set of examples of network inputs $x(t) \in R^n$ and desired outputs $d(t) \in R^m$. For every training example we can compute the actual network output $f(x(t))$ and error $e_j(t)$ of each of the output units: $e_j(t) = d_j(t) - f_j(t)$. The instantaneous error $\varepsilon(t)$ of the whole network and the overall error E are computed (Neruda & Vidnerová, 2009; Simon, 2002) by:

$$\varepsilon(t) = \frac{1}{2} \sum_{j=1}^{J} e_j^2(t), \tag{4.48}$$

$$E = \sum_{t=1}^{k} \varepsilon(t) \ . \tag{4.49}$$

Therefore, the goal is to minimize the over-all error function:

$$E = \frac{1}{2} \sum_{t=1}^{k} \sum_{j=1}^{J} (d_{tj} - f_{tj})^2. \tag{4.50}$$

There exists many algorithms for RBF network learning of which two most significant ones are three step learning method and gradient learning method (Neruda & Vidnerová, 2009). Among others, the gradient descent (GD) training of RBF networks has proven to be much more effective (Simon, 2002). This is a first-order derivative based optimization algorithm used for finding a local minimum of a function, from which the aim is to determine a set of weights that minimizes the error (Kurban & Beşdok, 2009) .

Training the RBFN by using the GD method requires several iterations - use a set of inputs, compute the output and then adjust the weights based on the errors of the first iteration. This process can be implemented in two different ways: batch mode and incremental mode. In batch mode, all training vectors are applied to the network before the weights are updated. In incremental mode, after each training, the vector is applied to the network, the gradient is computed and the weights updated.

The RBFN algorithm based on (Karayiannis, 1999; Kurban & Beşdok, 2009; Simon, 2002) are described as following:

1) Select $m, \eta \; and \; \epsilon$, initialize w_{ij} with zero values; randomly initialize the center vectors $c_i, 1 \leq i \leq I; \; set \; h_{0,k} = 1, \forall k$.
2) Compute the initial response:

$$h_{i,k} = (g(\|x_k - c_i\|^2)^{\frac{1}{1-m}} \quad , \quad \forall k, i, \tag{4.51}$$

$$h_k = \left[h_{0,k}, h_{1,k}, \ldots h_{I,k}\right]^T, \forall k, \tag{4.52}$$

$$y_{j,k} = w_j^T h_k \quad \forall k, j. \tag{4.53}$$

3) Compute

$$E = \frac{1}{2} \sum_{k=1}^{M} \sum_{j=1}^{n} (d_{j,k} - y_{j,k})^2. \tag{4.54}$$

4) Set $E_{old} = E$.
5) Update the adjustable parameters

$$\varepsilon_{j,k}^{o} = d_{j,k} - y_{j,k} \; \forall k, j \, , \tag{4.55}$$

$$w_{ij} \leftarrow w_{ij} + \eta \sum_{k=1}^{M} \varepsilon_{i,k}^{o} h_k, \forall i, \quad (4.56)$$

$$\varepsilon_{i,k}^{h} = \frac{2}{m-1} g'(\|x_k - c_i\|^2)\left(h_{i,k,}\right)^m \sum_{j=1}^{n} \varepsilon_{j,k}^{o} w_{ij} \forall j, k., \quad (4.57)$$

$$c_i \leftarrow c_i + \eta \sum_{k=1}^{M} \varepsilon_{i,k}^{h} (x_k - c_i), \forall i. \quad (4.58)$$

6) Compute the current response:

$$h_{i,k} = (g(\|x_k - c_i\|^2)^{\frac{1}{1-m}} \ , \ \forall k, i, \quad (4.59)$$

$$h_k = \left[h_{0,k}, h_{1,k}, \dots h_{I,k}\right]^T, \forall k, \quad (4.60)$$

$$y_{j,k} = w_j^T h_k \ \ \forall k, j. \quad (4.61)$$

7) Compute

$$E = \frac{1}{2} \sum_{k=1}^{M} \sum_{j=1}^{n} (d_{j,k} - y_{j,k})^2. \quad (4.62)$$

8) If: $(E\_old - E)/E\_old > \epsilon$ then go to step 4.

Where *m* is equal to 3, *I* is the number of neurons in the hidden layer, $i \in \{1,2,\dots,I\}$, *n* is the number of the neuron in output layer, $j \in \{1,2,\dots,n\}$, w_{ij} is the weight of the i^{th} neuron and j^{th} output, g is the radial basis function, *x* is the input data vector, c_i is the center vector of i^{th} neuron, y_j is the actual output of the output neuron j^{th}, d_j is the desired output of the output neuron j^{th}, and M the number of training vector and $k \in \{1,2,\dots,M\}$.

4.8 Summary

In this chapter, we described and explained the details of facial features extraction and classification techniques used in this research. We discussed the following:

- Image pre-processing method that is applied for normalization of the image size, image enhancement, and image background removal.
- Face detection technique and some of their limitations. Two method of face detection were discussed intensively.
    - Viola-Jones face detection algorithm that is based on machine learning techniques to characterize face and non-face images
    - Skin color face detection method that is based on the skin color detection and some of face features including the face height to width ratio and mouth and eyes location.
- Image segmentation technique and some of its applications in facial image.
- Facial image features and their extraction method. Facial features include :
    - Visual content
    - Semantic content

      Semantic content is expressed by the description of the human face while visual content include the general and domain specific features that were represented by the color histogram and eigenfaces features in this research. Eigenfaces and color histogram features were discussed intensively while the semantic features types were described.
- Three-color space models were described, including the RGB (red, green, blue), HSV (hue, saturation, value) and HSI (hue, saturation, intensity).

- A similarity measure technique and a machine classification technique have been described in this chapter. Euclidean distance was discussed as a similarity distance metric while the RBFN was discussed as a classification machine learning technique.

CHAPTER 5

RESEARCH DESIGN AND METHODOLOGY

5.1 Introduction

In our current research, we focused on the semantic-content-based facial image retrieval (SCBFIR). SCBFIR involves the retrieval of facial images based on their visual content including the color and eigenfaces features, and the semantic content including the semantic features such as race, age, gender, and face shape. SCBFIR based on the combination of CBIR and FERET techniques, and the facial image semantic features description and annotation method; using the proposed methods of image segmentation, probabilistic approach, and neural network approach.

In the following sections, we explain and discuss the approaches and techniques that are used to achieve the objectives of this research.

5.2 General Principles of Content Based Image Retrieval

A typical content-based image retrieval system consists of four principle units as shown in Figure 5.1. These are the query-processing unit, features extraction unit, similarity calculation unit and the storage unit. The query-processing unit enables a user to specify a query through a query pattern, translate the query into an internal form, and visualize the retrieved similar images. The features extraction unit extracts a feature vector from each image in the image database. Finally, the storage unit stores the generated feature space. Essentially, the feature space of the queried image is compared with those available in the feature database, one by one, before the images with the smallest feature distance are retrieved. The compared images are then ranked in decreasing order of similarity with the queried image. The user is also requested to provide an example through the query-processing unit.

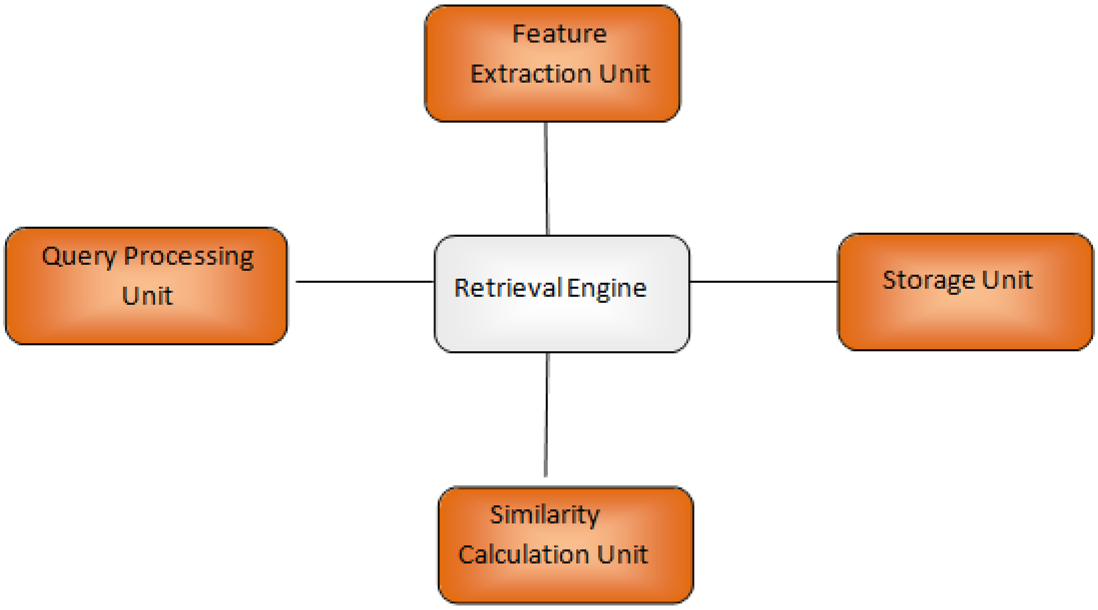

Figure 5.1: The principal components of content-based image retrieval.

5.3 Proposed Framework for Semantic-Facial Image Retrieval

The proposed framework include the following models: data collection, preprocessing, face detection, face segmentation, feature extraction and image annotation, neural network development, and probabilistic approach. Figure 5.2 shows the block diagram of our research design and methodology.

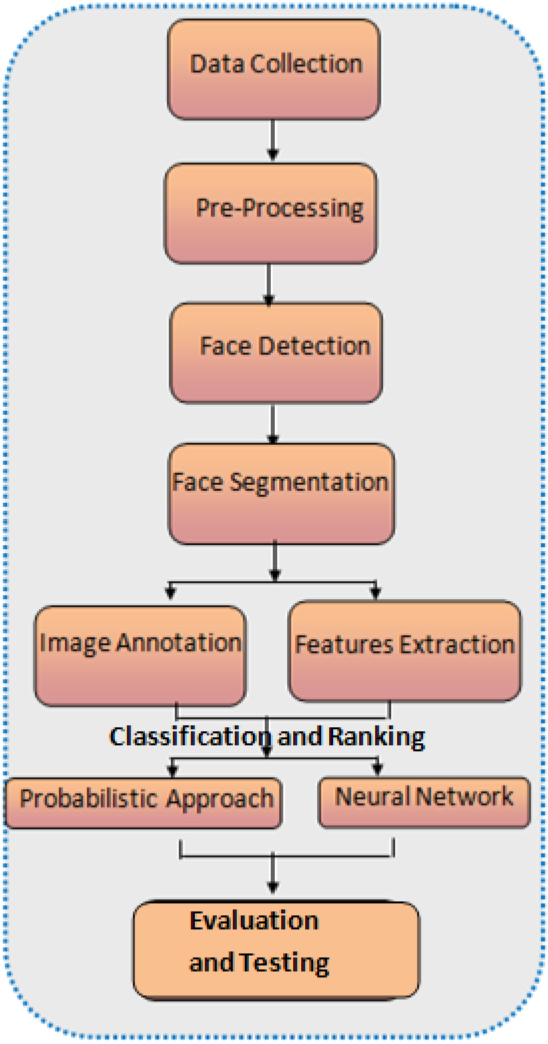

Figure 5.2: Methodology of the current research.

5.4 Data Collection

The main purpose here is to collect and prepare suitable facial image datasets for system training and testing in order to evaluate the performance of the proposed methods. Two databases are used in this research for training and testing the proposed approaches. They are:

- Local facial images database - This consists of 1500 local facial images of 150 participants from the University of Malaya (UM), Kuala Lumpur and their friends and families of different races, gender, ages, and skin colour, etc. Ten different images were taken for each participant, showing different facial expressions such as open/closed eyes and smiling/not smiling; and carrying different facial features such as having spectacles/no spectacles and bearded/not bearded. All images were prepared with a blue background with the image number against a red background. The image subjects are all in frontal positions with some orientation (upright, rotated) tolerance. Figure 5.3 shows some sample image from the local database.

Figure 5.3: Local database sample (150 participants).

- The second database is the Olivetti Research Laboratory (ORL) Database of Faces available at the AT&T Laboratories Cambridge website. It is well known, publicly available and has been used as a standard database in many face recognition systems. The database contains 400 greyscale images of 40 participants. These images were taken at different times under different lighting conditions with the image subjects having different facial expressions and facial features. The background of all the images is dark and homogeneous while the image subjects are positioned upright and frontal with some side movement tolerance. The database was developed between April 1992 and April 1994 to be used for a face recognition project by the project team. Figure 5.4 shows some sample images from the ORL database. The complete listing of the concepts frequency of the semantic features of each database is provided in Appendix B.

Figure 5.4: ORL database sample (40 participants), grayscale images.

5.5 Pre-processing Model

In this research, image normalization and filtering are implemented. Normalization is implemented to change the image size to a default image size of 92 x 112 on which the developed system operates. The aim of resizing the image dimensions is for reducing the complexity of the computations, while in the final stage of retrieval and display; the images keep their original size. Filtering is implemented to enhance the images through noise reduction, and to emphasize some details of the facial image, as a result, some important facial features are more obvious for the feature extraction module, which can radically improve the facial retrieval systems performance. Nine different filters were experimented; the eigenfaces features showed the best performance with Prewitt filter while color histogram features showed the best performance with Unsharp filter.

After the preprocessing process, the images will undergo rough other processes of face detection, image segmentation, and low-level feature extraction. In this research, the database facial images are processed off-line.

5.6 Face Detection Model

The face detection method of Viola and Jones described in section 4.3.1 is used in this research. The aim of face detection is to determine the position and size of a face in the entire image. As shown in Figure 5.5, the algorithm detector runs on the test image using a sliding window. The sliding window is moved across the image with an increment of one pixel. A basic window operates on a 24 x 24 pixels size. The detector is applied on a gray-scale image, while the scanned window adopted is scaled by factors

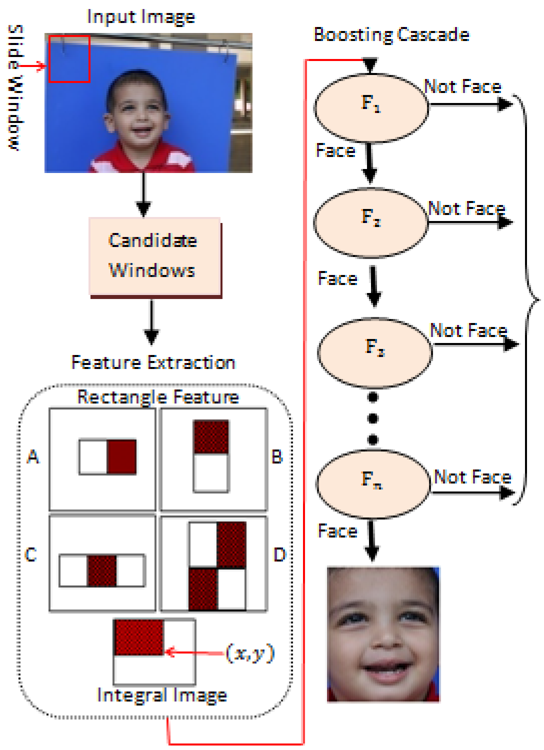

Figure 5.5: Face detection using Viola Jones method.

of 1.25 to detect faces at multiple scales, and Haar-like features are extracted from the sub-windows of a sample image. These sub-window patches of features are then treated by the classifiers. Negative sub-windows are rejected, while the positive ones detected.

The classifiers determine whether a patch is a face or non-face. In the event that it is a face, the output of the algorithm will then be a location of the detected face, otherwise the negative value will be returned.

The face detection method of Viola and Jones is fast and accurate dealing with frontal view faces as shown clearly in Figure 5.6 but it has not the same capability with the non- frontal faces. Figure 5.7(a) shows some non-frontal faces that were not detected accurately by Viola and Jones, and the results of the detection were negative values (non-faces).

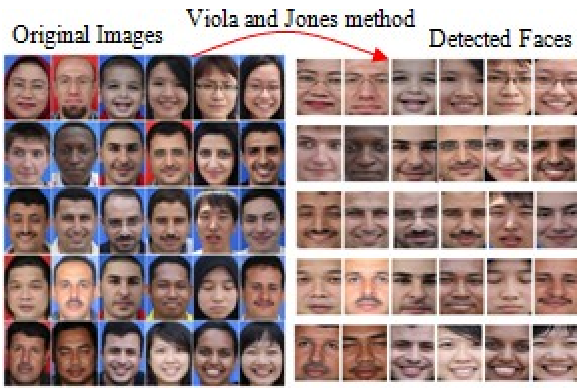

Figure 5.6: Samples of face detection results using Viola and Jones method.

Thus, to reduce the limitations of the Viola Jones method the algorithm of face detection based on skin color is combined with that of face features. This approach is described in section 4.3.2. The detector of this algorithm starts scanning the image, when the output of the Viola and Jones classifier is of negative value. The image is then subjected to treatment of the color-based technique - YC_bC_r in color space to separate skin regions from non-skin regions. The candidate blocking is localized to several regions, which could belong to the human face. The height to width ratio is used to determine whether the candidate block is a face or not. If it is a face, its location is then returned. Figure 5.7 (b) shows some results generated from the skin-color based human face detection.

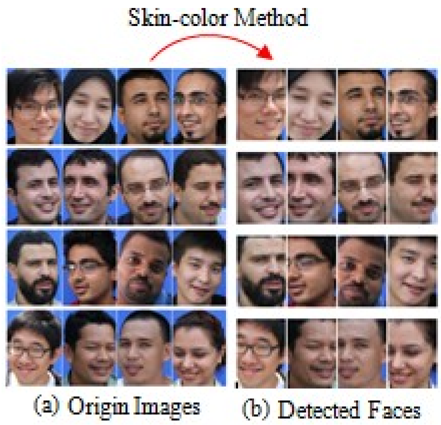

Figure 5.7: (a) Non-frontal view facial image, Faces detection results were negative values through Viola Jones method.
(b) Result of detecting the faces in (a) using skin-color based face detection.

5.7 Facial Image Segmentation Model

In this research, a new facial image segmentation technique is proposed to improve the accuracy of facial image retrieval performance.

The proposed method is based on the fact that every sub-facial image contains spatial information regarding orientation and specific scale relevant to this sub-image. A combination of the features vectors of each sub-image, independently extracted, is expected to produce more robust features vectors.

We suggested that the facial image be segmented into four partitions based on human eyes and mouth and the ratio of their respective heights to face height, based on the assumption that an image will always have at least one face. Each detected facial image is scaled to a fixed size beginning with the face detection step to optimize the candidate for face segmentation.

In order to segment the candidate's faces in the image, a template-matching technique needs to be employed. The template consists of four sub-templates as shown in Figure 5.8. The first sub-template is used for matching the upper region of the face at the eye level, the second for matching the middle region of the face between the eye level and the mouth level, the third for matching the lower region of the face starting from the mouth level and finally, the fourth sub-template is used for matching the region of the facial image center.

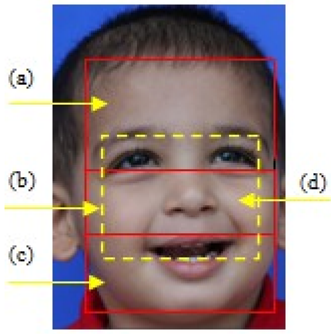

Figure 5.8: Proposed template for facial image segmentation, includes four sub–templates (a, b, c, and d).

The sub templates are scaled based on the intensity of extraction from facial images of different people. The aim of the sub-templates is to match each region of the face to be extracted independently. After the facial image and the template have been matched, the segmented regions are projected into the feature extraction algorithm to extract the features in each segment separately.

The proposed algorithm of facial image segmentation can be summarized as follows:

1. Suppose the candidate image is x.
2. The face will be detected and scaled to fixed size.
3. The four sub-template are applied on the candidate facial image, this will segment the face into four regions :
   a) The upper region of the face at the eye level.

b) The middle region of the face between eye level and mouth level.

c) The lower region of the face starting from mouth level.

d) The region of the facial image center.

4. The visual features are extracted from each region and are stored as vectors.
5. The various vectors are concatenated into one vector to represent the candidate image vector features.
6. The feature vector undergoes further processing for comparing, ranking, and retrieval.

5.8 Features Extraction Model

Retrieval and recognition systems are based on features extraction, which is the process that transfers the content of images into various content features, commonly called feature vectors, literally mapping the image from image space to the feature space as depicted in Figure 5.9. This constitutes the basis of content-based image retrieval.

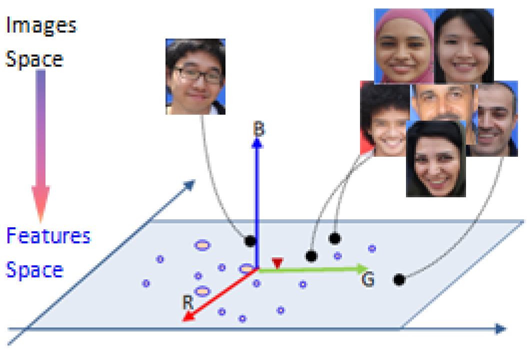

Figure 5.9: Feature extraction transfers the content of images into various content features.

Image content may include visual content called the low-level features and semantic content called the high-level features. Visual content as described in section 4.5 can be classified into general or domain-specific types. General visual contents are application-independent features, such as color. Domain-specific visual contents are application-dependent features, such as human faces. In domain-specific classification, human facial image features are extracted by two methods. The first method is the information theory concept that seeks a computational model that provides the best description of a face by extracting the most relevant information contained in that face. The other method is components-based, in which deformable templates and active contour models with excessive geometry and mathematics extract the feature vectors of the basic parts of a face. Of the two approaches, the PCA-eigenfaces method seemed appropriate for this research. Therefore, the method that was followed in this research for the purpose of facial image retrieval is the information theory concepts based recognition method. In this method, the excessive geometry and computation, time, space and processing complexity is avoided. The idea is to implement a face retrieval system, based on well-studied and well-understood features.

5.8.1 Visual Features Extraction

In this research color histogram is used as general visual content, while eigenfaces features are employed as domain specific visual content.

Eigenfaces is based on an information theory approach that decomposes facial images into a small set of characteristic feature images called eigenfaces. The idea is to find the principal component of the distribution of the set of facial images to extract information and capture the variation contained in these faces.

Based on the eigenfaces algorithm and PCA technique the facial images are transformed into a set of eigenfaces through the mean face, the eigenvectors, and eigenvalues of the training set. The weight vectors (eigenfaces vectors) are calculated for each facial image and stored in the database. Each weight vector is regarded as a point in space. Figure 5.10 shows example of eigenfaces extraction. When a user submits his or her query by example, the weight of the query image is calculated. The similarity distances between the query image weight and the database image weight are calculated, the images with the smaller distance will be displayed at the top of the list for the user.

The dimension of eigenfaces vector will base on the size of the training set. Based on the experimental results (will be discussed in section 6.3.1,) the suitable size of the eigenfaces vector used in this research for facial image retrieval is 20.

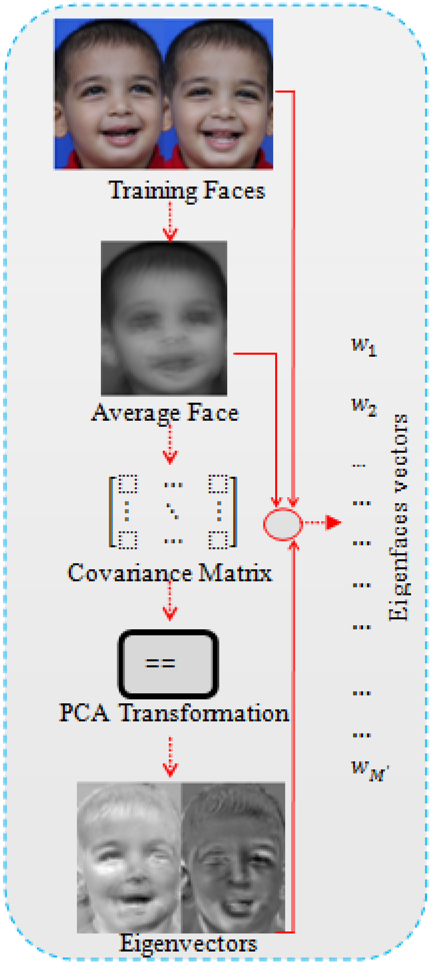

Figure 5.10: Sample of eigenfaces extraction.

The regions of human face contain unique characteristics of color distribution. In this research, color histograms are used to capture the special relations of these unique regions characteristics. A color histogram of a facial image is prepared by counting the number of pixels that correspond to a specific color in quantized color space.

Color descriptors of images are both global and local. Global descriptor enables whole images to be compared while local descriptor enables only matching between regions within an image or between images. In this research, using color for a facial image retrieval system is based on comparing the color content of the query image histogram to those of the images in the database. The query is based on the global descriptor of the facial image while the comparing is based on the local descriptor of the facial image .For each facial image, color histograms are generated to show the relative proportions of pixels within certain values. Facial images are generally represented as a series of pixel values, each corresponding to a visible color and similar images contain similar proportions of certain colors.

The color models are available for image processing, but it is important to use the appropriate color space for each application. In this research, we investigate the capability and effectiveness of the RGB, HSV and HIS models with regards to the performance and accuracy of the facial image retrieval system. The results of this study will be discussed in the next chapter.

With color histogram features, the uniform histogram is used. The RGB color space is quantized into 256 bins by using a uniform quantization, 16 for R, 4 for G, and 4 for B, in view of the fact that human eyes' response to red light is much stronger than is its response to blue and green light. Each element corresponds to one of the bins in the quantized histogram. The HSV and HIS color space is quantized into 256 bins by using uniform quantization, 16 for H, 4 for S, and 4 for I and V. The reason to assign more bits to hue than to the other components is that hue (H) carries more importance to the human visual system than the other components. One histogram bin corresponds to one color in the quantized color space. The following, as shown in Figure 5.11, is an algorithm that utilizes the color-histogram approach for facial image retrieval:

- Read and convert facial images in the database to the required color model, and extract pixel information from each image.
- A color histogram H_i is generated for each facial image i read from the database.
- Read and convert the query image to the required color model and extract the pixel information.
- Create query image histograms.
- The histograms are normalized so that its sum equals unity to remove the size of the image.
- Compute the distance by comparing the query image histograms to that of each image in the database.
- Sort images in database in order of the ascending distance of the query image and return the result.

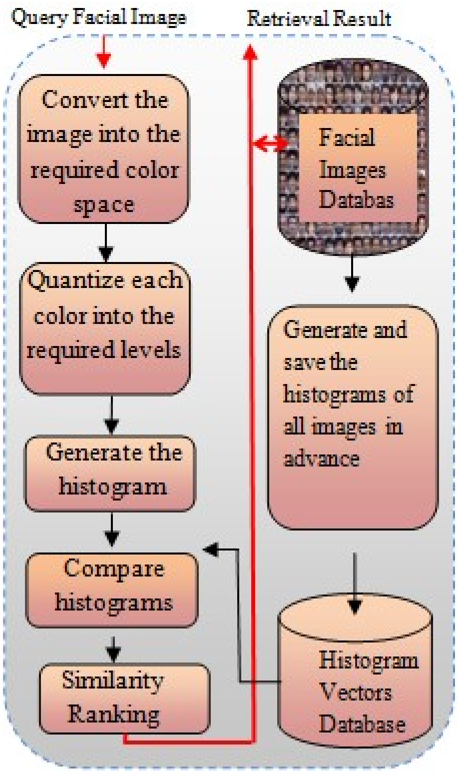

Figure 5.11: Image retrieval based on color histogram extraction.

5.8.2 Semantic Features Model

Semantic features extraction model is employed to provide direct access to specific image attributes. Image processing techniques are utilized in the content-based image retrieval to automatically extract image descriptions from the image visual contents. The processing is conducted in feature space based on the assumption that similar images are close in the feature space. However, related images are not necessarily visually similar, and may even be located in multiple disjointed semantic classes in the feature space. Depending on the user's subjective interpretation, images can be grouped into many semantic classes having varying degree of relevance. Images sharing the same semantic concept such as "Gender" and "Visual Impression", may be separated by irrelevant images in the low-level feature space, as depicted in Figure 5.12.

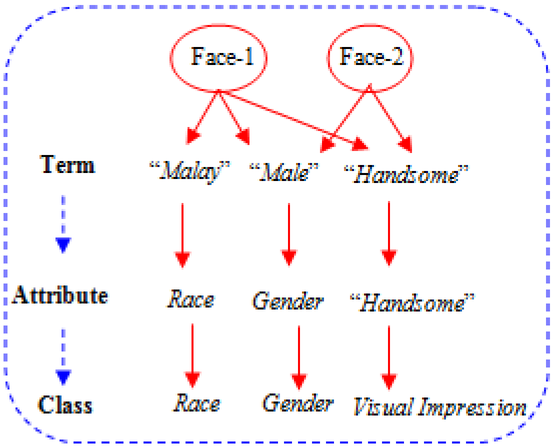

Figure 5.12: Two different images, semantically similar and visually not similar, similar image may belong to many classes.

Image access and retrieval systems need data, which can assist in the user`s selection of the relevant image attributes as well as deal with their nature and distribution based on the tasks specified. While human queries are in linguistic terms, computer based image matching is based on low-level image features that may not fulfill the query concept.

Generally, semantic attributes play a very important role in facial image recognition and retrieval because human beings measure the similarity among faces using these semantic concepts.

Human image semantic attributes can be observed from the whole human image, or from the segmented parts of the image. Human images include a variety of these semantic attributes that represent the features of the face and the visual impressions such as race, shape, size, and color of facial parts, and other features of the described face. These features are used for recognizing faces and characterizing them.

Each facial feature is described by a set of linguistic terms, where the machine should be ready to meet human natural tendencies and needs.

With our proposed method, these description features are later organized into a vector. Each description is called a face description vector and later used in the search and comparison process to define and retrieve the desired facial images in the image database.

5.8.2.1 Semantic Features Selection

In face image applications, users tend to express some semantic attributes using a rating scale. For example, a user may need to specify in his query one of the values: oval, round, long or square for an attribute that describes face shape. An application must be able to match human descriptions to retrieve facial images.

Some semantic features are considered more important than others from the aspect of the frequency of use, the simplicity of their description and the ability to distinguish between faces accurately. Most of these semantic features are difficult to extract by the system automatically, therefore they must be described by the user. In this section, we list the semantic attributes, which we used in this research and the methods that we followed to select these semantic attributes. These attributes represent the main traits for the human face. In this research, the semantic features used, were selected based on a case study of 100 volunteers of different gender, race and age. Each volunteer was required to rank 20 traits, as listed in the case study form as shown in Figure 5.13, based on their significance and ability to distinguish among individuals. Table 5.1 shows some outputs of the ranking process.

Could you assist me for ordering the following concepts of human face based on their importance and ability to distinguish between individuals?
Note: Please, write the suitable number near to the suitable concept.

No.	Suitable Order	The Concept
1		Skin Color
2		Hair color
3		Hair Length
4		Gender
5		Age
6		Race
7		Moustache size
8		Beard Size
9		Facial Marks
10		Nose shape
11		Hair Type
12		Face Shape
13		Glasses Frame
14		Eyes color
15		Eyebrows Thickness
16		Lip Thickness
17		Mouth Size
18		Ears size
19		Eyes size
20		Forehead length

Figure 5.13: Case study form.

Table 5.1: Case study output, R1: R20 are the ranks from 1 to 20.

Rank	Skin Color	Hair Color	Hair Length	Gender	Age	Race	Moustache Size	Beard Size	Facial Marks	Nose Shape
R1	3	2	1	76	8	1	0	0	0	0
R2	20	0	2	9	34	22	1	1	1	2
R3	15	7	6	5	17	36	3	1	0	3
R4	32	10	5	2	11	8	3	4	5	4
R5	3	25	2	3	5	9	3	2	8	5
R6	6	9	16	0	6	2	4	3	7	4
R7	1	12	16	0	2	1	8	6	9	3
R8	1	9	11	0	1	1	8	9	11	4
R9	4	5	16	0	1	1	11	12	1	9
R10	1	4	7	2	2	3	14	14	8	4
R11	2	4	6	0	1	2	6	11	5	4
R12	0	2	2	0	0	2	4	13	5	6
R13	2	3	0	1	1	1	6	9	4	8
R14	0	4	0	1	2	2	5	8	2	8
R15	1	0	0	0	1	1	5	3	6	5
R16	1	2	1	0	1	1	7	1	6	8
R17	1	0	3	0	1	1	5	0	3	5
R18	2	0	0	1	1	0	1	1	3	3
R19	1	1	2	0	0	1	1	0	7	4
R20	0	0	3	0	0	0	3	0	3	2
Weight	29.0997	18.3525	15.6615	83.6705	36.6927	29.6211	11.3023	11.1139	11.0468	10.6503

Rank	Hair Type	Face Shape	Glasses Shape	Eye Color	Eyebrows Thickness	Lip Thickness	Mouth Size	Ears Size	Eyes Size	Forehead Length
R1	0	0	0	1	0	0	0	0	1	1
R2	2	1	0	4	0	0	1	0	1	0
R3	1	3	0	4	1	0	1	0	0	0
R4	12	1	12	6	1	1	0	1	2	0
R5	15	4	5	9	3	3	1	1	0	0
R6	16	5	9	11	2	2	2		2	1
R7	10	8	3	9	2	4	2	1	3	3
R8	13	6	9	9	2	1	2	4	4	2
R9	8	5	10	7	4	3	6	3	3	5
R10	2	2	10	7	2	6	4	2	4	1
R11	1	7	8	5	10	5	4	5	6	1
R12	1	11	9	4	11	7	5	8	5	2
R13	4	8	7	6	12	13	15	7	5	2
R14	1	6	4	2	11	7	10	9	6	4
R15	3	6	1	3	8	5	11	1	0	6
R16	1	10	0	1	7	9	5	6	2	3
R17	1	10	6	2	3	11	9	1	1	7
R18	6	0	3	2	7	11	7	2	8	7
R19	0	3	0	0	8	4	6	16	8	8
R20	2	3	1	1	5	5	4	8	12	15
Weight	15.4504	10.5498	12.1024	15.2886	8.43862	7.9655	8.0974	5.782	7.41	5.75713

5.8.2.2 Semantic Features Ordering and Weighing

From Table 5.1, it is noted that each feature was given different rankings by different participants. For example, skin color was ranked as R_1 by 3 participants, R_2 by 20 participants, R_3 by 15 participants and so on. To compute the ranking and weight of each feature, the following proposed statistical analysis was applied:

$$w(F) = \sum_{i=1}^{n} x_i . \frac{1}{i} \quad (5.1)$$

Where, w is the weight of feature F given by the participants. This weight reflects the ranking of the feature concerned, n is the number of rank positions and x is the value that feature F received in position i. This proposed method is based on the assumption that the feature in the first position, R_1 is weighted heavier than the feature in the second position, R_2, which is similarly weighted heavier than the feature in the third position R_3, and so on.

It is assumed that the weight parameter p_i of position, i is $p_i = \frac{1}{i}$. Consequently, the weight w_i of value x in the position of i is given by the product of the value of x and p_i that is, $w_i = x.p_i$. The final weight of each feature is the sum of the weight vector. The final position and weight of each feature is shown in Figure 5.14.

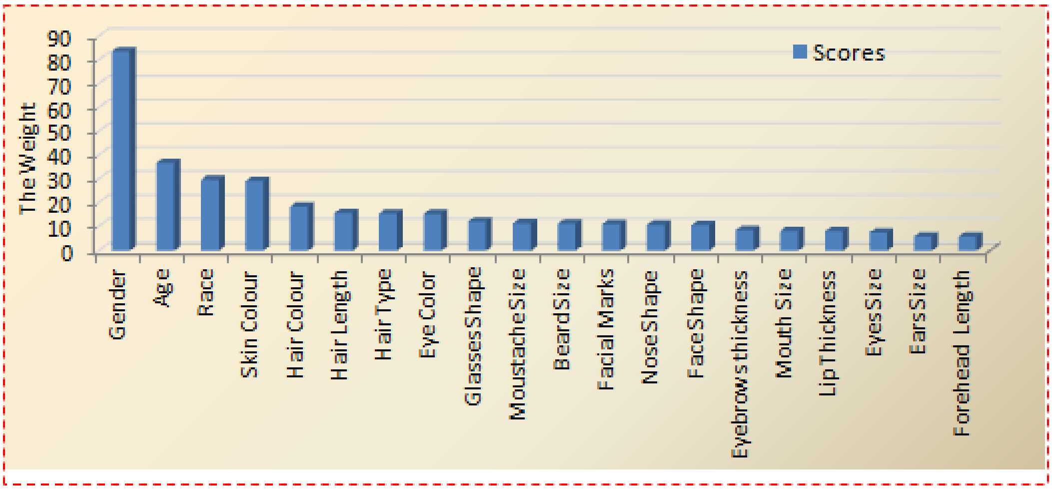

Figure 5.14: Final ranking and weighting of the face semantic features **b**ased on the case study.

Based on the results of the case study, 17 relevant features were used in this research. Table 5.2 shows the selected 17-semantic features of facial images including the hair color length, and type**.** These features involved 70 different concepts.

Table 5.2 : The semantic feature terms
with the descriptions.

Features	**Description**
Gender	Male, Female
Age	Infant, Child, Adolescent, Young Adult , Middle Adulthood, Senior
Race	Malay, Chinese, Indian, Middle Eastern, European, African
Skin Colour	Black, Brown, Tan, White
Hair Colour	Black, Brown ,Blond ,Red, Gray ,Covering Head
Hair Length	Short, Medium, Long, Bald, Covering Head
Hair Type	Curly, Wavy, Straight, Covering Head
Eye Colour	Black, Brown, Blue, Green
Glasses Shape	Oval, Circular, Square, Rectangle
Moustache Size	Short, Medium, Long
Beard Size	Short, Medium, Long
Facial Marks	Mole, Scar, Freckles
Nose Shape	Flat, Rounded, Straight, Wide, Convex , Concave
Face Shape	Oval, Round, Long, Square, Heart
Eyebrows Thickness	Normal, Bushy
Mouth Size	Small, Medium, Big
Lip Thickness	Thin, Medium, Thick

5.8.2.3 Image Annotation

Though effort and time is required, manual image annotation is considered the best approach and maintains a position ahead of other image annotation techniques in terms of simplicity, comprehensive concept and keywords, efficiency and performance. This is likely, because keywords are selected and assigned based on human determination of the semantic contents of images.

To carry out our experiments in this research, facial images are annotated based on a number of volunteers. Participant annotation was collected based on a facial image annotation forms shown in Figure 5.15, given to each participants, and some of these annotations were collected using the research prototype interface. The interface was designed to show the image and semantic terms for annotation, as shown in Figure 5.16.

Gende	Age	Race	Skin Color	Hair Color	Hair Length	Hair Type	Face Shape
Male	1-3	Malay	Black	Black	Short	Curly	Oval
Female	3-12	Chinese	Brown	Brown	Medium	Wavy	Round
	13-19	Indian	Tan	Blond	Long	Straight	Long
	20-40	Middle Eastern	White	Red	Bald	Covered Head	Square
	40-65	European		Gray	Covered Head		Heart
	65-	African		Covered Head			
		Others					

Glasses Shape	Moustache Size	Beard Size	Facial Marks	Nose Shape	Eyes Color	Eyebrows Thickness	Mouth Size	Lip Thickness
None	None	None	None	Flat	None	None	None	None
Oval	Medium	Medium	Mole	Straight	Black	Normal	Medium	Medium
Circular	Short	Short	Scar	Wide	Brown	Bushy	Small	Thick
Square	Long	Long	Freckles	Convex	Blue		Big	Thin
Rectangle				Concave	Green			
				Rounded				

Figure 5.15: Form sample of facial image annotation.

The prototype interface allows participants to view many samples of the annotated image. Participants were directed to describe the face by selecting the suitable semantic description for each feature based on their own perception.

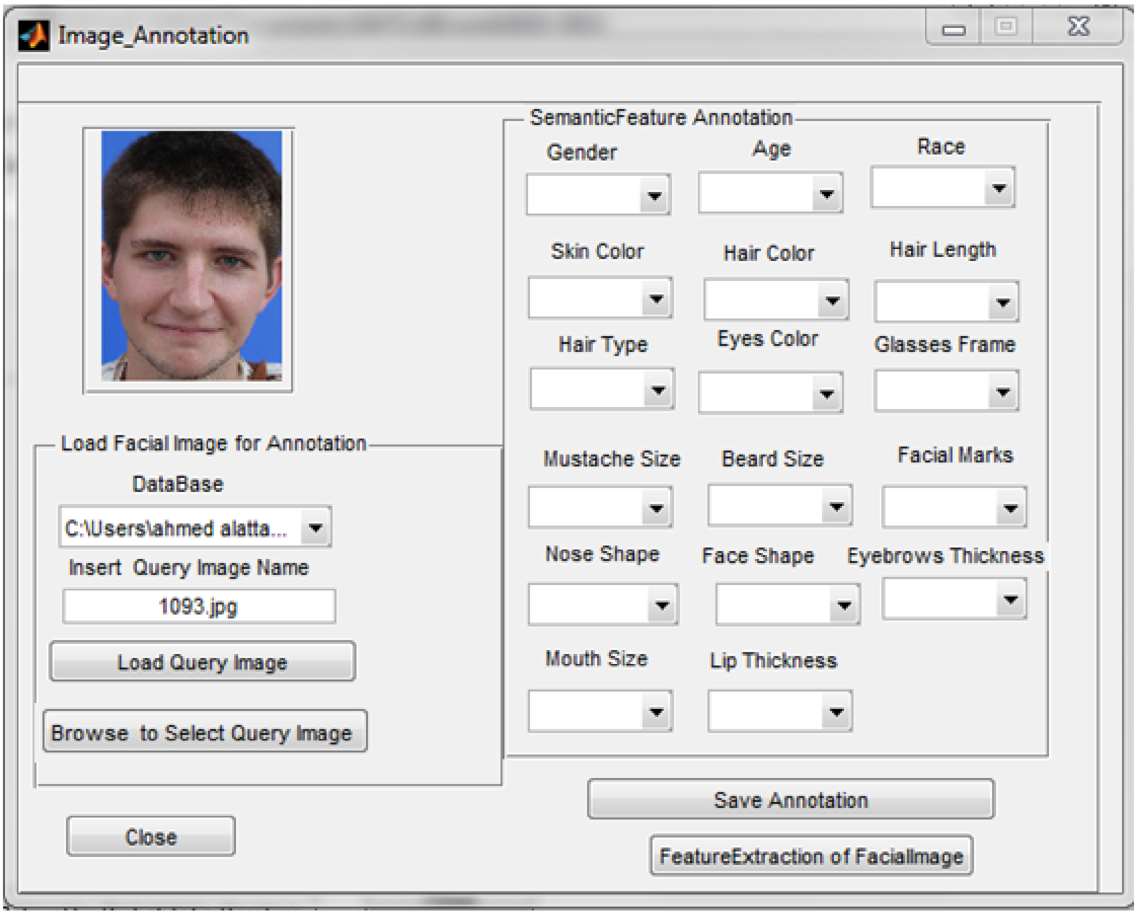

Figure 5.16: The prototype interface for image annotation.

5.8.2.4 Semantic Feature Representation

Images are annotated to simplify their access using metadata. Images are described by textual information, and a text search technique can be used to search for images. All keywords were assigned based on the visual features of individual face images. Expressions of characters, symbolized by keywords, are intuitive, simple, conceptual, easy to understand and manipulate, and more effective than exact specifications using visual features.

After the annotation process, each facial image was associated with a n-dimensional term-vector, where n is the concepts number. Each element in the term-vector is attached a keyword that describes the element semantically. In this context, every image comprises a 17-dimensional vector of the semantic features, which represented the corresponding semantic concepts of gender, age, race, skin color, hair color, hair length, hair type, eyes color, glasses shape, moustache size, beard size, facial marks, nose shape, face shape, eyebrow thickness, mouth size, and lip thickness.

The concepts vector is not stored together with the image so that the retrieval process can be performed more efficiently and reading and comparing the description vector of the facial image can be done directly without extracting it from the desired image, simultaneously maintaining the size of the annotated facial image as small as possible. The size of the output data from the image annotation process is relatively big. Therefore, as an alternative measure, a reference between the facial image and the corresponding concept vector can be kept.

The semantic description cannot be directly interpreted by a classifier. There is also a need to merge semantic features descriptions with other facial features that are extracted automatically by the system, and represented as numeric data. This method will make image searches using content-based image retrieval more effective. Because of this, an indexing procedure that maps a concept C_i into a compact representation of its content needs to be applied. The choice of a representation for text depends on what one regards as the meaningful units of concept (1 denoting presence and 0 absence of the concept in the description vector). In the case of non-binary indexing, for determining the weight w_i of concept C_i any style indexing technique that represents the face description as a vector of weighted concepts may be used. The standard term frequency–inverse document frequency(TF-IDF) function is used (Salton & Buckley, 1988; Sebastiani, 2002), modulated as

$$CfIFf\left(C_i, F_j\right) = \mathrm{Cf}\left(C_i, F_j\right).\log\left(\frac{|\mathrm{TF}|}{Ff\left(C_i, F_j\right)}\right) \quad (5.2)$$

where | TF | denotes the number of facial images in the training set, $Cf(C_i, F_j)$ denotes the number of times concept C_i occurs in F_j, and $Ff(C_i, F_j)$ denotes the facial images frequency of concept C_i, that is, the number of facial images in which C_i occurs. In order for the weights to fall in the [0, 1] interval, the length normalization is applied as follows

$$w_{ij} = \frac{CfIFf(C_i,F_j)}{\sqrt{\sum_{k=1}^{|TF|}(CfIFf(C_k,F_j))^2}}. \tag{5.3}$$

In this research the single normalized numerical value of each semantic feature was adjusted based on its rank from the semantic feature weighting and selection. Due to differences in perception and viewpoint of the users pertaining to semantic attributes, descriptions of some semantic features have resulted in some subjectivity and uncertainty. For example, one user may consider the mouth as big. On the other hand, another user may consider the same mouth as medium. For this reason, the representation should be done carefully to enable the assignment of intermediate values with the same term and for different terms, as shown in Figure 5.17.

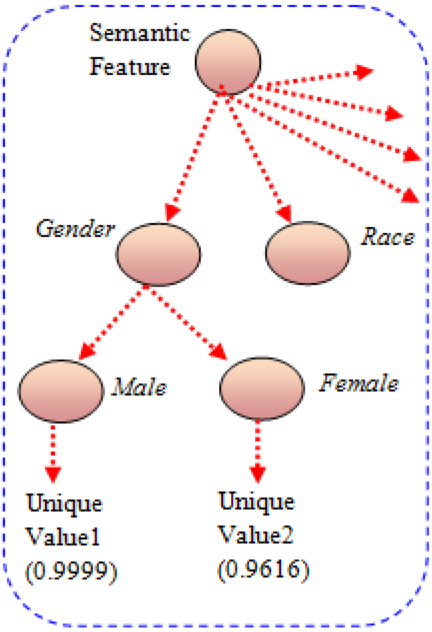

Figure 5.17 : Semantic features representation.

For example, the trait ‘gender’ can be given a value of 0.9999, after normalization to represent the semantic term ‘male’ and a different value of 0.9616 to represent the semantic term ‘female’. If the trait ‘age’ is given a value of 0.909 then the terms ‘infant’, ‘child’, ‘adolescent’, ‘young adult’, ‘mid-adult’, and ‘senior’ should be given values such as 0.913784, 0.918568, 0.925745, 0.937705, 0.947274, and 0.952058, after normalization. The complete listing of the numerical values of the semantic features is provided in Appendix A.

After the facial images were annotated and the annotated concepts were numerically represented, each facial image would be associated with two vectors of 17-dimensions. The first vector covers the semantic concepts, while the second vector comprises the corresponding numerical representation of the semantic concepts.

The semantic features of the facial image *i,* is represented in the database in the following form:

$$Fk_i = (k_{1,} k_{2,}, \ldots, k_n) . \tag{5.4}$$

$$Fv_i = (v_{1,} v_{2,}, \ldots, v_n) . \tag{5.5}$$

Where, k refers to the keyword or the semantic concept, v the numerical value of the corresponding semantic concept and n is the number of the semantic features used. If the semantic concept is assigned to the semantic concept vector, then its representation value is assigned to the semantic concept weight vector, otherwise, it is given the value of zero. These vectors with other vectors of the facial image features are stored in the image database.

In the query process as shown in Figure 5.18, the user will specify the suitable attributes of the queried facial image based on visual features of individual images (face color, race, gender, etc.). The retrieval mechanism will map the individual concepts to the predefined weights in the matrix of semantic concept values, previously built, to generate the query value vectors. Comparison of the query value vectors with other vectors in the database would provide respective vector values based on the probability value of each vector, while comparison of the query concepts vectors with others concept vectors in the database would yield a logical value of each concept defined for an image. The final output for these comparisons is an array of 1's or 0's with the same length as the number of concepts in the vector, indicating whether the corresponding query features are defined or not present in the facial images stored in the database.

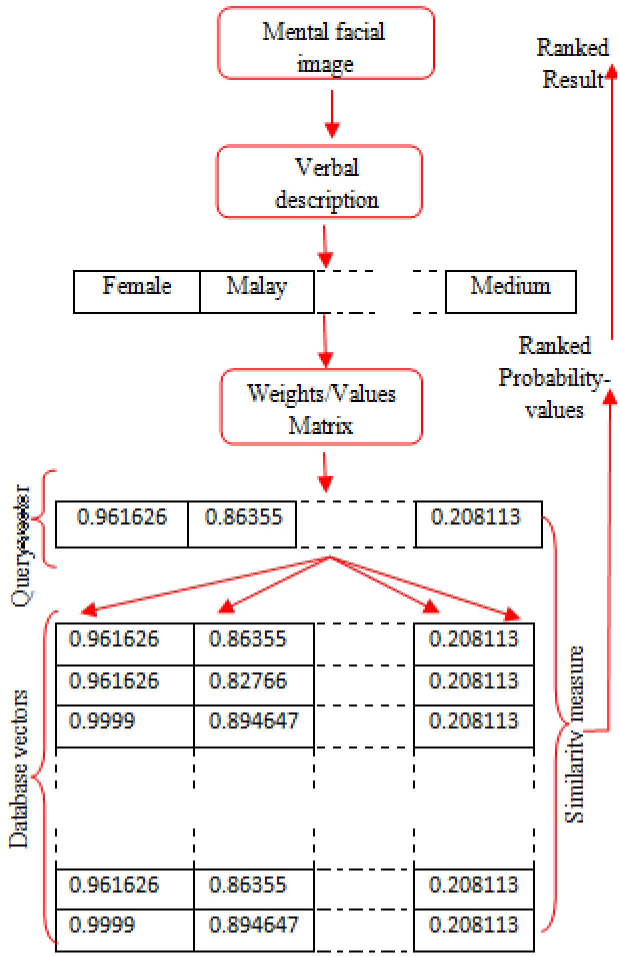

Figure 5.18: Process of features description and representation for facial image retrieval.

5.9 Probabilistic Approach

Most image retrieval methods have focussed on image-based matching and retrieval and subsequently display the top images. In our research, the objective is to match verbal queries to the corresponding values of semantic features found in the metadata.

Specifying semantic attributes could be subjective as users provide different opinions of various semantic domains. It is not easy to measure and specify image features and errors can occur leading to inaccuracy. Retrieval requirements of a user can be vague leading to uncertainty. Therefore, in a retrieval system, where semantic features are utilised in a query process, the result would inevitably be subjective, inaccurate and uncertain.

Pruning has been the most obvious method in verbal query retrieval. For instance, if there were information that the targeted face has moustache, then the system would prune away faces without moustaches in the database. This pruning process for retrieval could be fatal in the proposed system for two reasons – (i) errors in semantic tagging during the annotation process would result in potential image targets being pruned away and (ii) element of subjectivity due to different user descriptions of a feature, for instance different definitions of a long nose.

Given this scenario, the proposed approach has introduced the innovative probabilistic approach of image ranking to replace image pruning and has facilitated more proactive user interactions in the image attribute specification process in reducing subjectivity, uncertainty and inaccuracy. The aim is to improve the differences in observations based on human perception and the viewpoint that may appear during image annotation and/or query processing, where the system sorts images according to their probabilities of being the desired image, given the user descriptions. The system then retrieves and displays the top images.

The matching query formation is adjusted for effective retrieval. Though no completely automated system for verbal query based face retrieval exists, probabilistic retrieval can handle these issues where each facial image in the search space is given initial probability. The system uses each description by the user to update the probability of each facial image, using Bayesian learning. "Bayesian inference is a method of statistical inference in which evidence is used to update the state of uncertainty over competing probability models". Facial images are sorted according to their probability of being the targeted face.

In the following section, we show how probabilistic value can be assigned for each facial image and how it can be used to find the desired facial image from the search space. Our proposed method is an improvement to the weakness of the method in the work by (Sridharan, Nayak, Chikkerur, & Govindaraju, 2005).

5.9.1 Proposed Approach

In the probabilistic model, we assume the semantic features of images in the database, $F = [f_1, f_2, \dots, f_Z]$,where z is the number of the semantic features. The description of the each feature is $D = [d_1, d_2, \dots, d_j]$, where $j <= m$, m is the maximum number of the features description concepts.

Given a user-specified text query, the system first assigns the normalized values of these description features. Based on the values, the system computes the Bayesian score (will be discussed in the next section) for each facial image in the database corresponding to that query. Facial images are then ranked using this score and the top images are returned.

Each facial image in the database is assigned an initial probability. Suppose the facial image database size is n, the initial probability of each facial image is equal to *1/n*. At each stage, the system updates the facial images probability of being the face searched for, based on the description given by the user.

The updating process is achieved using Bayesian learning as following:

$$P(FacialImage_{\mathrm{i}}|F_f = D_d) = \frac{P(F_f = D_d|FacialImage_i) * P(FacialImage_i)}{\sum_{i=1}^{n} P(F_f = D_d|FacialImage_{\mathrm{i}}) * P(FacialImage_i)}$$

where $P(FacialImage_{\mathrm{i}}|F_f = D_d)$ is the probability of the facial image i with feature f and description d, and $P(F_f = D_d|FacialImage_{\mathrm{i}})$ is the probability of the feature f with description d.

As discussed in section 3.4, an attempt has been made to avoid pruning the image from the search space. This is based on the fact that features that do not match the descriptions given by user may result in elimination of desired images from the search space through pruning. The limitation confronted in previous methods is that if two features - query feature and database feature are not exactly matched, the latter will be ignored. Although these methods avoid eliminating a face from the search space, mismatched features are pruned away. However, the user may describe features close to the "correct" annotation of the image features in the database. For example, the user may describe the lip size as medium whereas the image feature description in the database is small. Using the previous method would give the same probabilities for different facial images with different feature descriptions, and would then appear in the top ranked results. To address this problem we have proposed to use of the function expressed below.

$$P(F_f = D_d|DFacialImage_i) = exp\left(\frac{-\,||(F_f = D_d|QFacialImage) - (F_f = D_d|DFacialImage_i)||\,)}{\sigma}\right) * \mathrm{W}$$

where $\mathrm{P(F_f = D_d|DFacialImage_i)}$ is the probability of the current described features of the database facial image I, $F_f = D_d|QFacialImage$ is the description of the features of the query facial image and ($\mathrm{F_f} = \mathrm{D_d}|\mathrm{DFacialImage}_i$) is the description of the feature of the database facial image i, σ is the standard deviation of the Gaussian function, which is supposed to be equal to one, and w is the weight of the current described feature.

The formula above is applied to search for the similarities between the features from the query vector and the database vector. The degree of similarity reflects the probability of the described features, rather than their matching that may lead to pruning them away from the image vector features. The probability of the features are measured based on the distance of the feature descriptions of the face in the database from those of the query features. This probability is computed in the above formula, where the output value ranges from 0 to 1. If the query feature and the database feature are closed to each other, the output value of the function will be close to 1. However, if the two features are far apart, the output value of the function will be close to 0. The output of the function will be greater when the two features are similar. In this method, there is no pruning of features. The probability value of each feature will be weighted using the weights computed in the case study processing. The final output is the summation of the probabilities of the facial images in the database that is being tested.

The summary of the proposed method is as the following:

1. Suppose we have the features description of the query Q and the database facial image x as the following :

$$\mathrm{F}= [f_1, f_2, \dots, f_z], \tag{5.6}$$

where z is the number of the semantic description.

2. The initial probability for each facial image x_j in the database is assigned such that :

$$P\ (x_j) = 1/n, \tag{5.7}$$

where n is the number of the images in the database.

3. The probability of each feature of the matching facial image is calculated using the formula :

$$P(f_i|x_j) = exp\left(\frac{-||(f_i|x_j) - (f_i|_Q)||)}{\sigma}\right) * w, \tag{5.8}$$

 where w is the weight of the feature f_i.

4. The probability of the facial image x_j is updated using the Bayesian learning:

$$P(x_j|f_i) = \frac{P(f_i = D_d|x_i) * P(x_j)}{\sum_{j=1}^{n} P(f_i = D_d|x_j) * P(x_j)} . \tag{5.9}$$

5. Step 3 and 4 are repeated until features F is achieved.

Faces will be displayed to the user, ranked from highest to lowest probability.

5.10 Features Integrations and Classification

Image retrieval using a single image attribute certainly lacks discriminatory information and does not really look into the large variation in image orientation. To improve classification and retrieval accuracy, integration of multiple features is obligatory. The problem confronting the use of heterogeneous set of features is how to integrate these features in a classification engine as well as to integrate the similarity results between the query and the database features to generate the integrated ranking of each image in the database. Suppose x is the query image and y is a database image, and $D_1(x,y)$, $D_2(x,y)$, and $D_n(x,y)$ are the similarity indices between x and y based on n different feature vectors (example, color, eigenfaces and semantic feature,) - Figure 5.19, then defining an integrated similarity index is the issue to be addressed.

To merge different features of the images together in an efficient and distributed manner requires an innovative solution. The proposed addressing in this research is to find a similarity metric with a suitable weight parameter that is directly applicable to the input data in the machine learning. The main idea is to use a function with a suitable parameter that maps input patterns into target space.

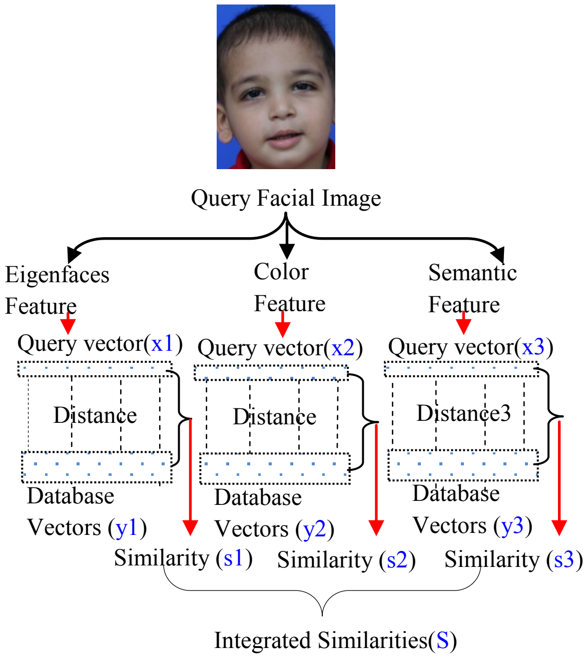

Figure 5.19 : Different similarity for different features.

The more precise idea is, a functions $f(x)$ parameterized by w_i and have a number of different features i, we have to find a value for the parameter w_i such that the distance between x_i and y_i , f $(||x_i$ -$y_i||)* w_i$, is small enough if x_i and y_i belong to the same category or class and large if they belong to different classes.

The proposed method aims to a learned similarity metric through the RBFN machine learning technique.

5.10.1 Proposed Approach

In this research, we have proposed that for constructing a RBF network, have been described in section 4.7.2, the number of input nodes in the input layer of neural network are set equal to the number of feature vector elements. The number of neurons in the hidden layer is set equal to the number of features classes. In addition, the center vector length of each RBF unit is set equal to the number of feature vector elements of each feature class as expressed in the following equation:

$$length(C_i) = length(x_i\ ),\ i = 1,2,..,L. \tag{5.10}$$

Where x_i is the feature vector of the features class i, C_i is the RBF centre of vector *i,* and L is the number of features classes. The first training vector is input to the RBFN centre vectors as a query vector, while the other training vector is applied to the network. The output of each neuron is then computed. To compute the error (target output minus actual output), the sum squared error (SSE) is used. Once the error is calculated, the learning rule would adjust the weight based on the *learning rate* value, which has the effect of adjusting the weights to reduce the output error. The weight is adjusted for each training vector following each input vector to the RBFN centre. Repetition of this process continues until the mean squared error (MSE) is less than an acceptable value.

The proposed method is described as follows:

1) Select η , ϵ and training vectors M that are a pair of the form *(x,d)* where *x* is the vector of input values, *d* is the target output, η is the *learning rate and* ϵ *is* the target error.
2) Initialize w_{ij} with random values. Initialize the sum square error (*SSE)* and the mean square error (MSE) with zero value.

3) Set the number of hidden layer neurons *I* equal to number of features classes.

4) Initialize the centers vectors c with the vector values of training vector x_q where each center vector c_i is initialized by one features class vector values of that training vector x_q :

$$c_i = x_{qi} \text{ where } i \in \{1,2,\dots,I\}, q \in \{1,2,\dots,M\} \tag{5.11}$$

5) Compute the initial response:

$$h_{i,k} = (g(\|x_k - c_i\|^2)^{\frac{1}{2}} \ , \ \forall k, i, \tag{5.12}$$

$$\mathrm{h_k} = [\mathrm{h_{1,k}}, \dots \mathrm{h_{I,k}}]^{\mathrm{T}}, \forall \mathrm{k}. \tag{5.13}$$

$$y_{j,k} = w_{i,j} h_k \quad \forall k, j. \tag{5.14}$$

6) Compute

$$\mathrm{SSE} = \sum_{k=1}^{M} \sum_{j=1}^{n} (d_{j,k} - y_{j,k})^2. \tag{5.15}$$

7) Update the adjustable parameters

$$\text{Set} \qquad c_i = x_{qi} \ . \tag{5.16}$$

$$\varepsilon_{\mathrm{j,k}} = \mathrm{d_{j,k}} - \mathrm{y_{j,k}} \ \forall \mathrm{k, j} \ , \tag{5.17}$$

$$w_{ij} \leftarrow w_{ij} + \eta \sum_{k=1}^{M} \varepsilon_{j,k} \, h_k, . \tag{5.18}$$

8) Compute the current response $h_{i,k}$, h_k, and $y_{j,k}$using equations (5.12), (5.13), and (5.14).

9) Compute the *SSE* using the equation (5.15).

10) Compute the mean squared error : $MSE = SSE/M$

11) If: $(MSE > \epsilon)$ then go to step 7.

where n is the number of the neuron in output layer, $j \in \{1,2,\dots,n\}$, M the number of training vector and $k \in \{1,2,\dots,M\}$.

The training computes the respective weights with the Gaussian function, that maps the input patterns into the target space. An appropriate transformation is applied to the data to emphasize on the most discriminative direction of each features class.

Our proposed method was based on injecting the query vector of class i to the center C_i of the RBF as depicted in Figure 5.20. The RBF with the Gaussian function are conducted as the similarity metric. The trained weight from the RBFN training stage represents the weight parameter for the similarity metric. During the query process, the proposed similarity metric computes the distance between the query and the database vectors. The output is then weighted using the respective weights. Figure 5.21 shows the proposed similarity metric while Figure 5.22 shows the overall network architecture based on the proposed method.

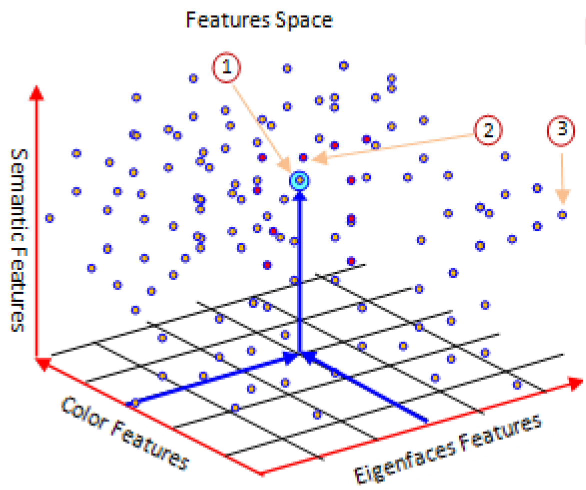

Figure 5.20: The query vector is fed to the RBFN center: (1) is the center, (2) vector near to the center and (3) vector far from the center.

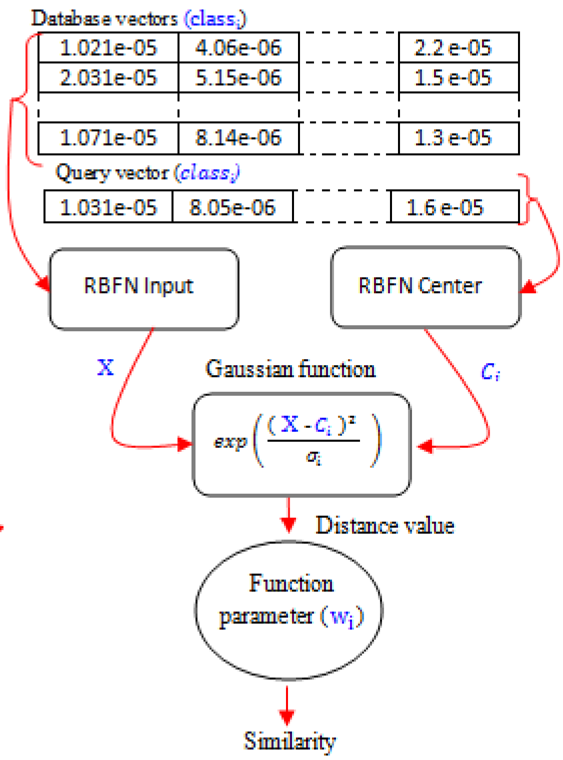

Figure 5.21 : The purposed learned similarity metric to overcome the problem of integration heterogeneous features

.

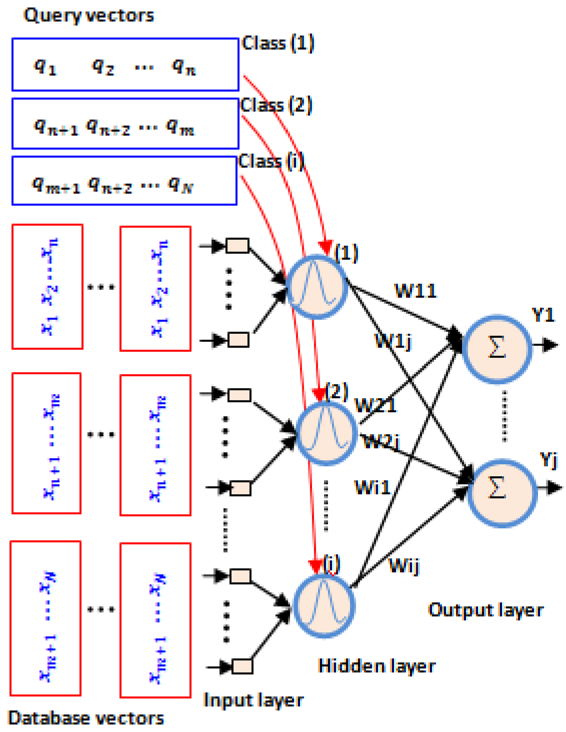

Figure 5.22: RBFN architecture with the proposed method.

5.11 Query and Retrieval Process

In this research, a prototype system was designed based on the combination of face detection, CBIR, FERET techniques, and semantic feature descriptions as shown in Figure 5. 23.

In the query by example, the user provides an initial image to the system or selects one from the image database. This query image looks similar to the required facial image. During the retrieval process, the candidate facial image is segmented using the proposed method based on the eyes and mouth levels. The eigenfaces and color histogram features are then extracted from each segment.

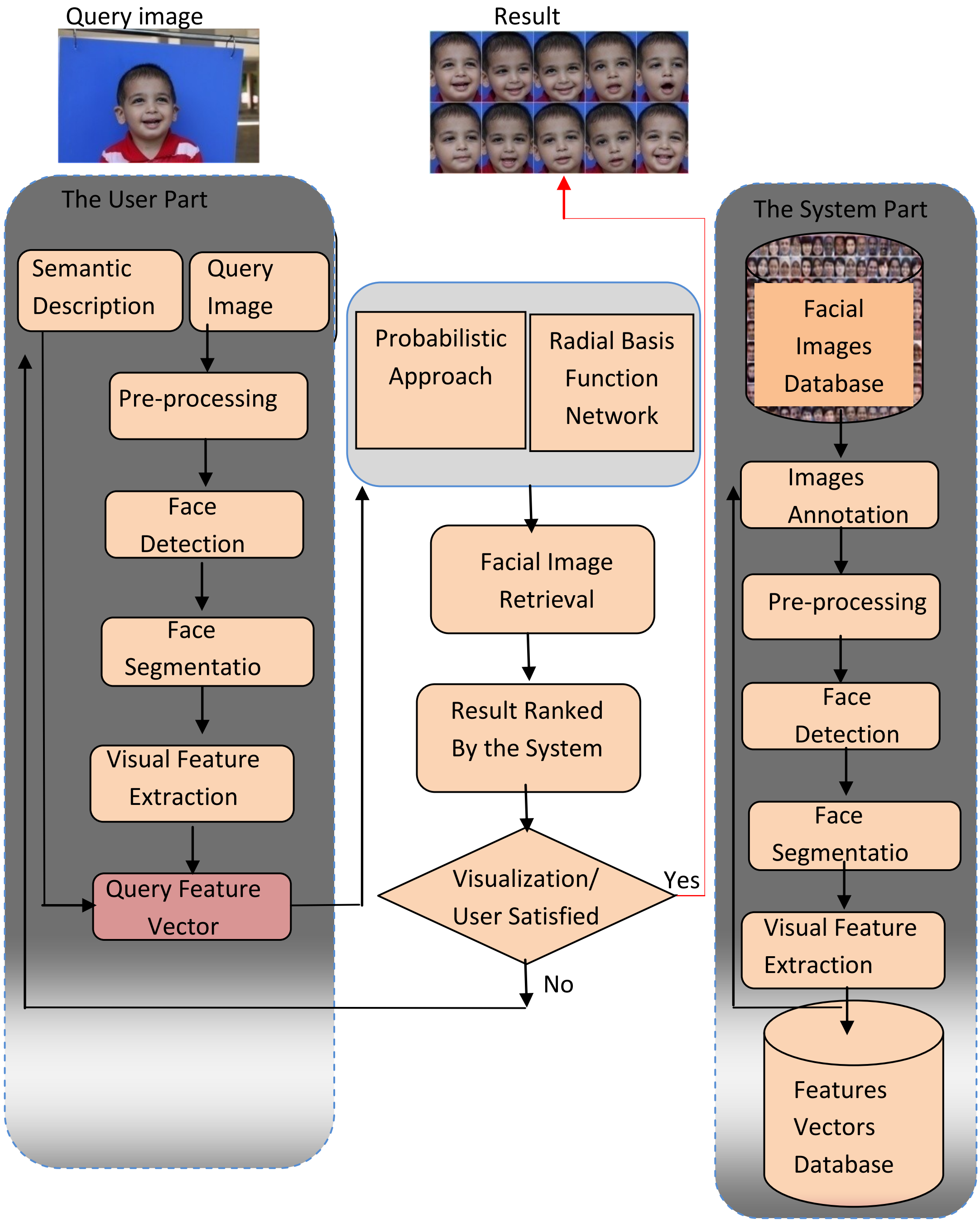

Figure 5.23: Semantic-content-based facial image retrieval model.

The combination of these features is used to identify and retrieve similar faces from the database to the query face.

In the query by verbal description, the user specifies the semantic features of the face. Individual concepts contained in the verbal description are then converted to predefined weights in the matrix of semantic concept values, previously built to generate the query vector values. Next, the system matches the verbally described query with the feature descriptions of faces contained in the database. Based on the probabilistic approach the image with identical descriptions is retrieved and displayed on top. Subsequently, the query by example process is initiated, based on the initial facial image (instant image) selected from the query through the description result pool. This instant image looks similar to the required facial image. The system then automatically extracts the visual vector features of the query image. The user can alternatively query directly by selecting the image and semantic feature descriptions that correspond to the desired face or the most discriminative image in a subset of the database faces. The system uses the proposed function through RBFN for classification and distance measurement to compute the distances between this query vector feature and vector features found in the database. Faces with the least distance are retrieved and displayed on top. If, after the retrieval, the user is not satisfied with the output, he or she is prompted by the system to update the verbal descriptions interactively based on the previous description and its displayed image and then selects again the example image from the resultant pool images.

5.12 Performance Measurement

Precision and recall methods were applied to measure the performance efficiency of the retrieval methods (Deselaers, Keysers, & Ney, 2008). Recall is the ratio of the relevant facial images of the retrieved facial images to the total number of relevant facial images in the database. Precision is the ratio of the relevant facial images of the retrieved facial images to the total number of irrelevant and relevant facial images retrieved. These definitions are represented in the equations below as well as in Figure 5.24.

$$Recall \quad = \frac{\boldsymbol{Relevant\ Faces\ of\ The\ Retrieved\ Faces}}{\boldsymbol{Total\ Relevant\ Faces}} \ . \tag{5.19}$$

$$Precision = \frac{\boldsymbol{Relevant\ Faces\ of\ The\ Retrieved\ Faces}}{\boldsymbol{Total\ Retrieved\ Faces}} \ . \tag{5.20}$$

$$F - score = \frac{\boldsymbol{2 * Precision * Recall}}{\boldsymbol{Precision + Recalll}} \ . \tag{5.21}$$

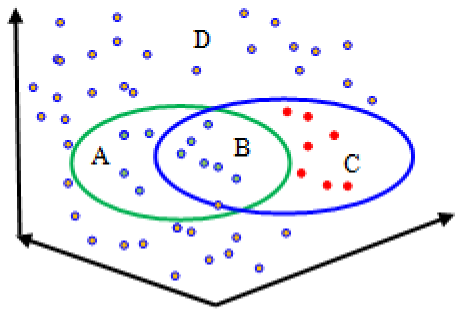

Figure 5.24 : A is un-retrieved relevant faces, B is retrieved relevant faces, C is retrieved irrelevant faces, D is Un-retrieved irrelevant faces.

5.13 Summary

In this chapter, we described and explained the research methodology, including the proposed methods, procedures adopted, and data that are used to design the current research.

Two databases are used in this research for training and testing the prototype performance. A local facial images database was developed in the University of Malaya, Malaysia to be used in the current research, and Olivetti Research Laboratory (ORL) Database from the AT&T Laboratories Cambridge website. Image normalization and filtering were implemented in this research for reducing the complexity computations of the facial image, to enhance the images through noise reduction, and to emphasize some details of the facial image. Different filters were experimented; the eigenfaces features showed the best performance with Prewitt filter while color histogram features showed the best performance with Unsharp filter. The face detection algorithms of Viola and Jones method and skin color detection method were applied in this research. The Viola and Jones technique is fast and accurate dealing with frontal view faces but it has not the same capability with the non-frontal faces. Skin color detection and the Viola and Jones method are combined to reduce the limitations of the Viola and Jones method. A new method for facial imaged segmentation was proposed. The idea was based on segmentation of the facial image into four partitions based on the human eyes and mouth and the ratio of their respective heights to face height. The aim is to improve the accuracy of the facial image retrieval. Eigenfaces features and color histogram features were used as visual features. Semantic features were selected and ranked based on a case study that we conducted. The proposed method was used for weighting and representation of the semantic features based on a formal statistic formula.

The proposed method was introduced based on a probabilistic approach using Bayesian learning and Gaussian function. The aim is to improve the differences in observations based on human perception and the viewpoint that may appear during image annotation and/or query processing.

The proposed method was introduced to integrate the heterogeneous features; visual features and the semantic features of the facial image. The idea is based on using the RBF and its Gaussian function as a learner similarity metric.

The experimental results and discussion of the proposed methods are discussed in chapter six.

CHAPTER 6
EXPERIMENTAL RESULTS
AND DISCUSSION

6.1 Introduction

Testing the performance of an image retrieval system is essentially measuring how well the system has retrieved similar facial images to the queried image. In this research, numerous experiments were conducted to assess and evaluate the proposed methods of semantic content-facial image retrieval. Two databases were used. The first database is the ORL database. It is well known, publicly available and has been used as a standard database in many face recognition systems. The second database is a local database, which is huge in size and contains color images with heterogeneous contents and a variety of semantic features such as gender, race, and age. This local database would meet all the requirements in evaluating our proposed semantic content-based facial image retrieval system.

The ORL and local databases consist of 400 facial images from 40 participants and 1500 facial images from 150 participants, respectively. 200 and 750 images (5 images for each participant) representing 50% the two databases respectively were randomly selected for training, and the remaining images were used for the experiments.

To evaluate the system performance during the retrieval process, a threshold to determine the level of the retrieval is not set but rather the number of the images to be retrieved is subjected to a certain pre-determined values. Hence, in both methods of precision and recall that has been described in section 5.12, cut-off levels are considered as necessary. Therefore, the experiments were performed with different cut-off levels -

10, 16, and 25. However, the images considered for performance analysis were the images within the top 10, 16, and 25 of the displayed results. The analysis rests on the decision whether the appropriate images were returned as the topmost result or otherwise. In the analysis, each queried image was matched visually with groups of images from the database and the database images were ranked according to how similar they were perceived to the queried image.

The number of queries means the number of system runs. With the ORL database the number of queries is 200, which is equivalent to the number of testing image sets, while with local database the number of queries is 750, which is also equivalent to the number of testing image sets.

When the query is run and the resultant images are retrieved, the user is required to count how relevant images of the retrieved images are similar to the queried image. This is a standard way where the similarity determination between two images is subjected to the user's perception.

Measurement of the performance of the retrieval system is highly dependent on the determination of the 'expected relevant results'. Human perception can easily notice the similarity between two images semantically or visually, while in some cases different users can give different opinions. Inevitably, defining the expected results would include some subjective conclusions and difficulties particularly with facial images, where the visual similarities between some 1500 images in the database with the queried image are to be determined.

To avoid these subjective conclusions during the performance measurement we have defined the 'expected relevant results' by tens of relevant images to each queried image, in which each person has ten images in the database. Consequently, the expected relevant results of 200 queried images would be 2000 relevant images and the expected relevant results of 750 queried images would be 7500 relevant images.

Precision and recall ratio values are computed in a way that reflects the actual measurement of system performance as a user sees it. In our experimental results, the recall value reflects the average recall of all system runs, while the precision value reflects the average precision of all system runs at specified cut-off levels within the database and retrieval method.

The results were also analyzed using precision versus recall graphs. The graph is commonly used for comparing system performance. The plots of different runs are superimposed on the same graph to determine which run is better. It is used to characterize precision and recall performance. The precision-recall curve depicts the performance of a system in terms of precision (y-axis) and recall (x-axis). In our experimental results, most of the graph slope downwards from left to right, indicating that as images that are more relevant are retrieved, the recall would increase. Positions of curves closed to the upper right-hand corner of the graph are indicative of maximized recall and precision meaning high performance. This appears more frequently with the cut-off level of the top 10 results, where the retrieved images equal the relevant images.

It was observed from all experimental results that the recall values increased for the top 10, 16 and 25 cut off levels, while most of the precision values decreased. This was due to the way-the expected relevant result was defined. As mentioned above, the expected relevant results were limited to 10 relative images to each query. While recall indicates the number of relevant images in the database that are retrieved in response to a query, precision refers to the proportion of the retrieved images that are relevant to the query. Consequently, expanding the measured results with defined expected relevant results would lead mostly to decreasing proportion of the retrieved images that are relevant to the query.

6.2 Color Space Models Experiments and Results

The choice of colors space is a significant issue. Three-color space models - RGB, HSV, and HIS were experimented to find out which color space in the facial image retrieval system has provided the best performance. Several experiments were carried out on both databases.

Tables 6.1, 6.2 and Figures 6.1, 6.2 show the results of the image retrieval system for the different color space models using the eigenfaces features and color histogram features with the ORL database.

Considering the F-score measurements at the top 10 of the retrieved images, it is observed that the eigenfaces-based facial image retrieval in RGB color model has shown better performance than the HSV and HIS color space models, achieving 68.55% accuracy in comparison to 57.4% and 60.45%. Considering the recall measurements only, the best performance of the eigenfaces was 85.15% accuracy within the top 25 retrieved images in the RGB color space model.

Table 6.1: Eigenfaces-based face retrieval in different color space models without segmentation method on the ORL database.

Color Model	Top Faces	Query Faces	Expected Faces	Retrieved Faces	Relevant Faces	Recall	precision	F-score
RGB	10	200	2000	2000	1371	0.6855	0.6855	0.6855
	16	200	2000	3200	1563	0.7815	0.4884	0.6011
	25	200	2000	5000	1703	0.8515	0.3406	0.4866
HSV	10	200	2000	2000	1148	0.574	0.574	0.574
	16	200	2000	3200	1355	0.6775	0.4234	0.5211
	25	200	2000	5000	1549	0.7745	0.3098	0.4426
HSI	10	200	2000	2000	1209	0.6045	0.6045	0.6045
	16	200	2000	3200	1465	0.7325	0.4578	0.5635
	25	200	2000	5000	1674	0.837	0.3348	0.4783

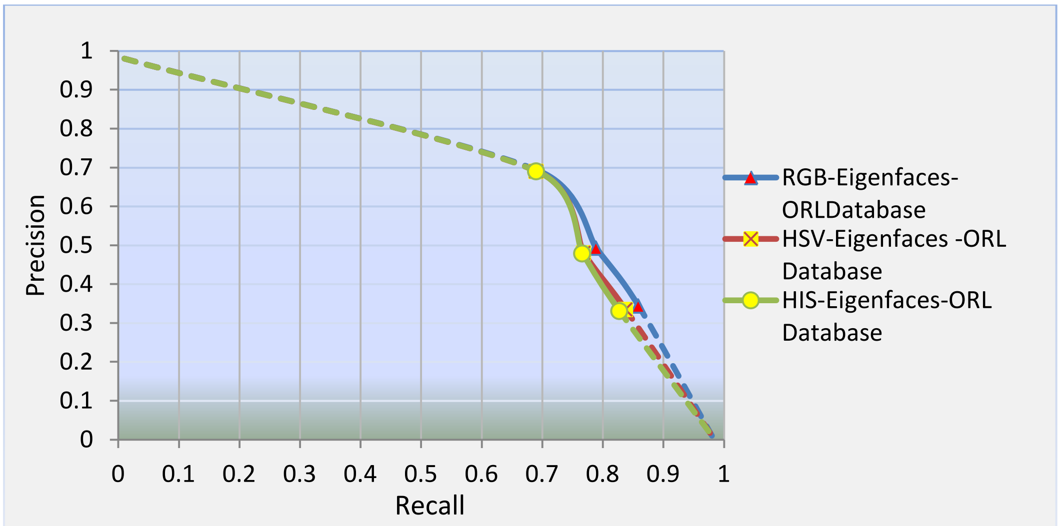

Figure 6.1: Eigenfaces-based face retrieval in different color space models without segmentation method on the ORL database.

Color histogram-based facial image retrieval in RGB color space showed the best performance achieving 62.25% accuracy in comparison to 52.25% and 52.75% for the HSV and HSI respectively. Considering the recall measurements only, the best performance of the color histogram is observed in the RGB color space model with 77.45% accuracy within the top 25 retrieved images.

Table 6.2 : Color histogram-based face retrieval in different color space models without segmentation method on the ORL database.

Color Model	Top Faces	Query Faces	Expected Faces	Retrieved Faces	Relevant Faces	Recall	precision	F-score
RGB	10	200	2000	2000	1245	0.6225	0.6225	0.6225
	16	200	2000	3200	1414	0.707	0.4419	0.5439
	25	200	2000	5000	1549	0.7745	0.3098	0.4426
HSV	10	200	2000	2000	1045	0.5225	0.5225	0.5225
	16	200	2000	3200	1221	0.6105	0.3816	0.4696
	25	200	2000	5000	1413	0.7065	0.2826	0.4037
HSI	10	200	2000	2000	1055	0.5275	0.5275	0.5275
	16	200	2000	3200	1231	0.6155	0.3847	0.4735
	25	200	2000	5000	1416	0.708	0.2832	0.4046

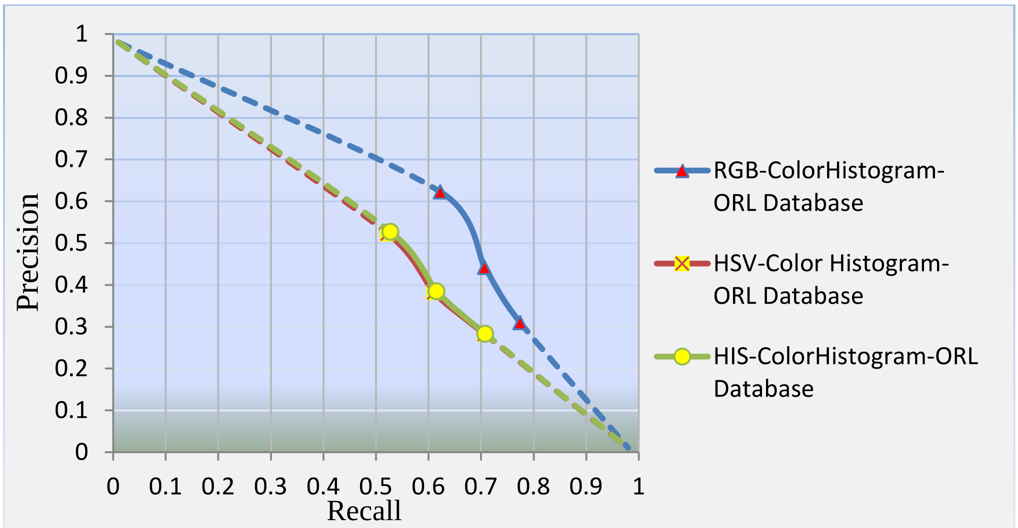

Figure 6.2 : Color histogram-based face retrieval in different color space models without segmentation method on the ORL database.

Tables 6.3, 6.4 and Figures 6.3, 6.4 show the results of the image retrieval system for the different color space models using the eigenfaces and color histogram features with the local database.

Considering the F-score measurements at the top 10 of the retrieved images, eigenfaces-based facial image retrieval in the HSV color model has shown better performance than in the RGB and HIS color space models achieving 73.52% accuracy in comparison to 63.57%, and 60.16% respectively. Considering the recall measurements only, the best performance of the eigenfaces achieved is 86.43% accuracy within the top 25 retrieved images in the HSV color space model.

Table 6.3 : Eigenfaces-based face retrieval in different color space models without segmentation method on the local database.

Color Space	Query Faces	Expected Faces	Top Faces	Retrieved Faces	Relevant Faces	Recall	Precision	F-score
RGB	750	7500	10	7500	4768	0.6357	0.6357	0.6357
	750	7500	16	12000	5379	0.7172	0.4483	0.5517
	750	7500	25	18750	5781	0.7708	0.3083	0.4404
HSV	750	7500	10	7500	5514	0.7352	0.7352	0.7352
	750	7500	16	12000	6096	0.8128	0.508	0.6252
	750	7500	25	18750	6482	0.8643	0.3457	0.4939
HSI	750	7500	10	7500	4512	0.6016	0.6016	0.6016
	750	7500	16	12000	5104	0.6805	0.4253	0.5235
	750	7500	25	18750	5542	0.7389	0.2956	0.4223

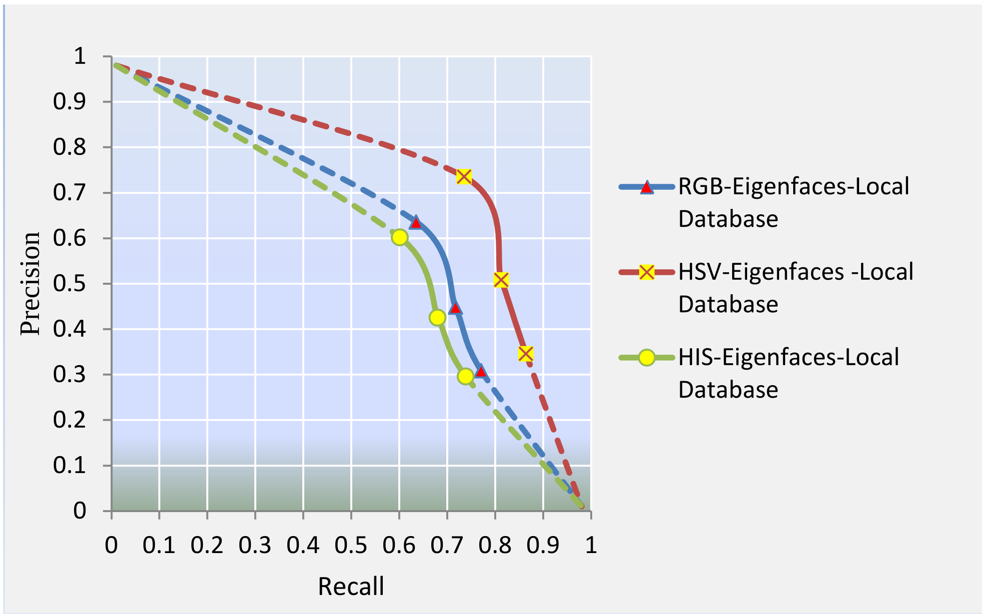

Figure 6.3: Eigenfaces-based face retrieval in different color space models without segmentation method on the local database.

Color histogram-based facial image retrieval in the RGB color space model gives the best performance among the other color space models achieving 79.55% accuracy in comparison to 77.39% and 76.39% of the HSV and HIS models respectively. Considering the recall measurement only, the best performance of the color histogram is observed in the HSV color space model with 85.72% accuracy within the top 25 retrieved images.

Table 6.4 : Color histogram-based face retrieval in different color space models without segmentation method on the local database.

Color Space	Query Faces	Expected Faces	Top Faces	Retrieved Faces	Relevant Faces	Recall	Precision	F-score
RGB	750	7500	10	7500	5966	0.7955	0.7955	0.7955
	750	7500	16	12000	5966	0.7955	0.4972	0.6119
	750	7500	25	18750	5966	0.7955	0.3182	0.4546
HSV	750	7500	10	7500	5804	0.7739	0.7739	0.7739
	750	7500	16	12000	6198	0.8264	0.5165	0.6357
	750	7500	25	18750	6429	0.8572	0.3429	0.4898
HSI	750	7500	10	7500	5729	0.7639	0.7639	0.7639
	750	7500	16	12000	6077	0.8103	0.5064	0.6233
	750	7500	25	18750	6311	0.8415	0.3366	0.4809

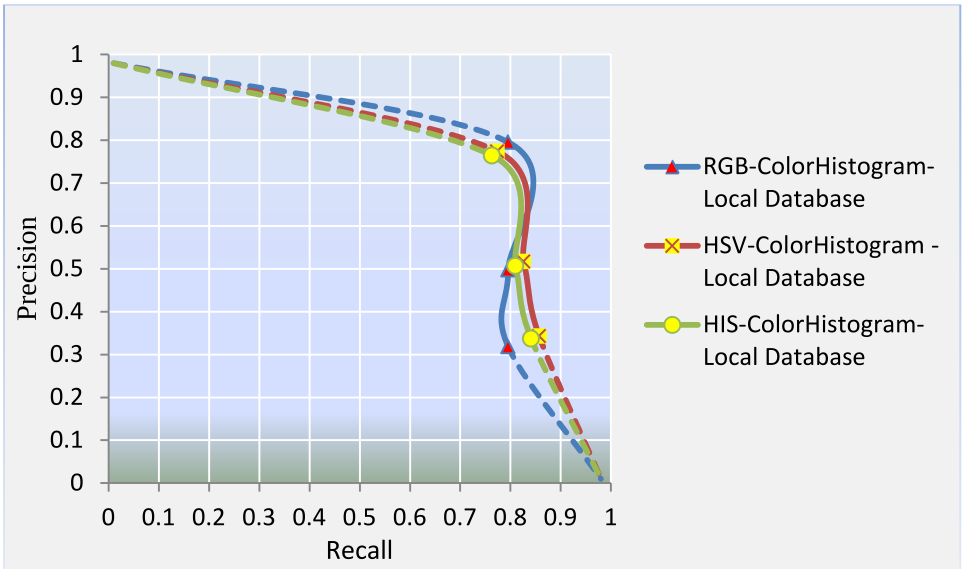

Figure 6.4 : Color histogram-based face retrieval in different color space models without segmentation method on the local database.

With the local database, eigenfaces-based facial image retrieval in HSV color model has achieved the best performance. The color histogram-based facial image retrieval in RGB color space has also achieved the best performance. With the ORL database, both eigenfaces and color histogram features have achieved the best performance of the retrieval in the RGB color space model. This is most probably because the Red, Green, and Blue channels in a grayscale image would contain the same information, if converted to HSV or HIS color spaces.

6.3 Face Segmentation Model Experiments and Results

The three methods of features extraction were experimented based on the entire facial image - three segments of the facial image partitioned at two levels: eyes and mouth; and four segments including as well the center portion of the facial image. The first method is a traditional method, while the others are the proposed methods. All experiments were conducted using eigenfaces and color histogram features, separately as well as in combination. Table 6.5, 6.6, and Figure 6.5, 6.6 show the results of the analysis using the above mentioned methods. Considering the F-score measurements within the top 10 retrieved images, eigenfaces-based facial image retrieval achieved accuracies of 0.6855%, 0.7745%, and 0.7665% respectively for the traditional, 3-segment and 4-segment methods using the ORL database. Slightly higher accuracies of 73.52 %, 81.53%, and 81.97% were achieved respectively by the three methods using the local database. The reasons for getting these results will be discussed in the end of the current section.

Table 6.5 : Eigenfaces-based face retrieval using different segments and extraction methods of human face with RGB color space and ORL database.

Extraction Method	Query Faces	Expected Faces	Top Faces	Retrieved Faces	Relevant Faces	Recall	Precision	F-score
Traditional Method	200	2000	10	2000	1371	0.6855	0.6855	0.6855
	200	2000	16	3200	1563	0.7815	0.4884	0.6011
	200	2000	25	5000	1703	0.8515	0.3406	0.4866
Three Segments	200	2000	10	2000	1549	0.7745	0.7745	0.7745
	200	2000	16	3200	1712	0.856	0.535	0.6585
	200	2000	25	5000	1832	0.916	0.3664	0.5234
Four Segments	200	2000	10	2000	1533	0.7665	0.7665	0.7665
	200	2000	16	3200	1687	0.8435	0.5272	0.6489
	200	2000	25	5000	1797	0.8985	0.3594	0.5134

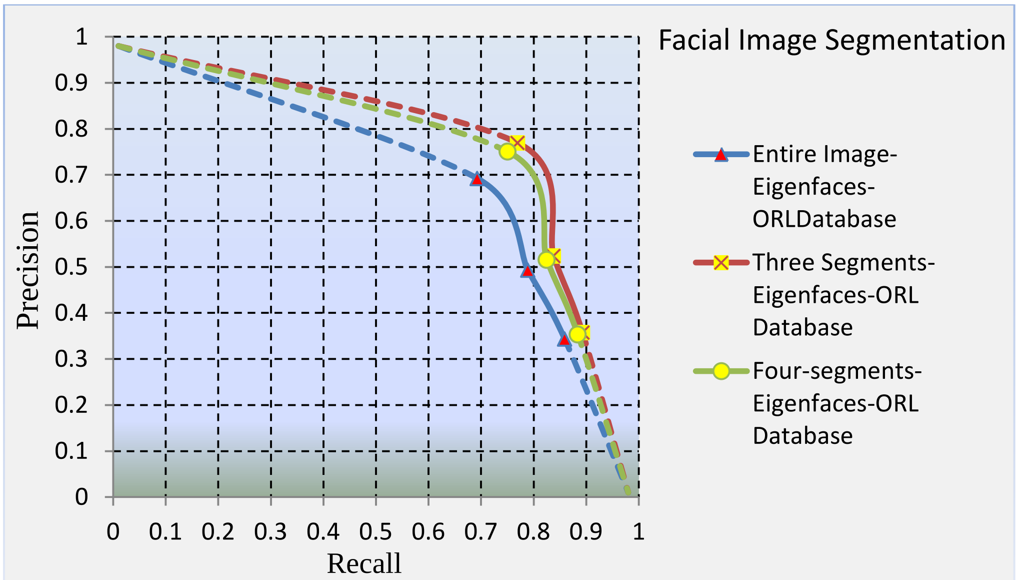

Figure 6.5: Eigenfaces-based face retrieval using different segments and extraction methods of human face with RGB color space and ORL database.

Table 6.6: Eigenfaces-based face retrieval using different segments and extraction methods of human face with HSV color space and local database.

Extraction Method	Query Faces	Expected Faces	Top Face	Retrieved Faces	Relevant Faces	Recall	Precision	F-score
Traditional Method	750	7500	10	7500	5514	0.7352	0.7352	0.7352
	750	7500	16	12000	6096	0.8128	0.508	0.6252
	750	7500	25	18750	6482	0.8643	0.3457	0.4939
Three Segments	750	7500	10	7500	6115	0.8153	0.8153	0.8153
	750	7500	16	12000	6585	0.878	0.5488	0.6754
	750	7500	25	18750	6871	0.9161	0.3665	0.5235
Four Segments	750	7500	10	7500	6148	0.8197	0.8197	0.8197
	750	7500	16	12000	6602	0.8803	0.5502	0.6772
	750	7500	25	18750	6890	0.9187	0.3675	0.525

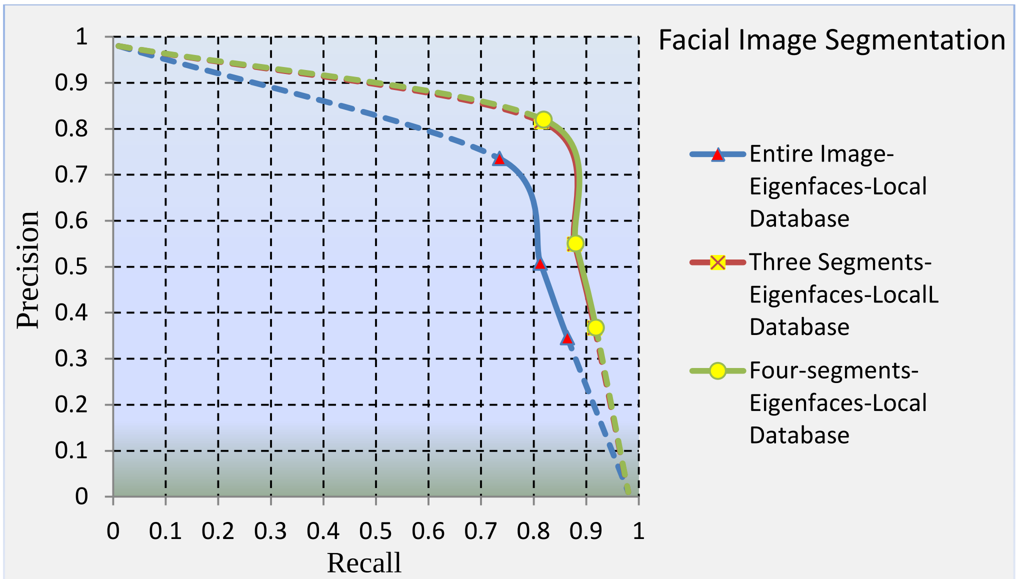

Figure 6.6: Eigenfaces-based face retrieval using different segments and extraction methods of human face with HSV color space and local database

The results of the analysis for the color histogram-based facial image retrieval are given in Tables 6.7, 6.8 and Figures 6.7, 6.8. Respective accuracies of 62.25%, 72.3%, and 71.5% were attained for the traditional, 3-segment and 4-segment methods with the ORL database. The accuracies were also achieved for the three methods -79.55%, 86.53%, and 85.44% with the local database.

Table 6.7 : Color histogram-based face retrieval using different segments and extraction methods of human face with RGB color space and ORL database.

Extraction Method	Query Faces	Expected Faces	Top Faces	Retrieved Faces	Relevant Faces	Recall	Prec-ision	F-score
Traditional Method	200	2000	10	2000	1245	0.6225	0.6225	0.6225
	200	2000	16	3200	1414	0.707	0.4419	0.5439
	200	2000	25	5000	1549	0.7745	0.3098	0.4426
Three Segments	200	2000	10	2000	1446	0.723	0.723	0.723
	200	2000	16	3200	1600	0.8	0.5	0.6154
	200	2000	25	5000	1737	0.8685	0.3474	0.4963
Four Segments	200	2000	10	2000	1430	0.715	0.715	0.715
	200	2000	16	3200	1575	0.7875	0.4922	0.6058
	200	2000	25	5000	1702	0.851	0.3404	0.4863

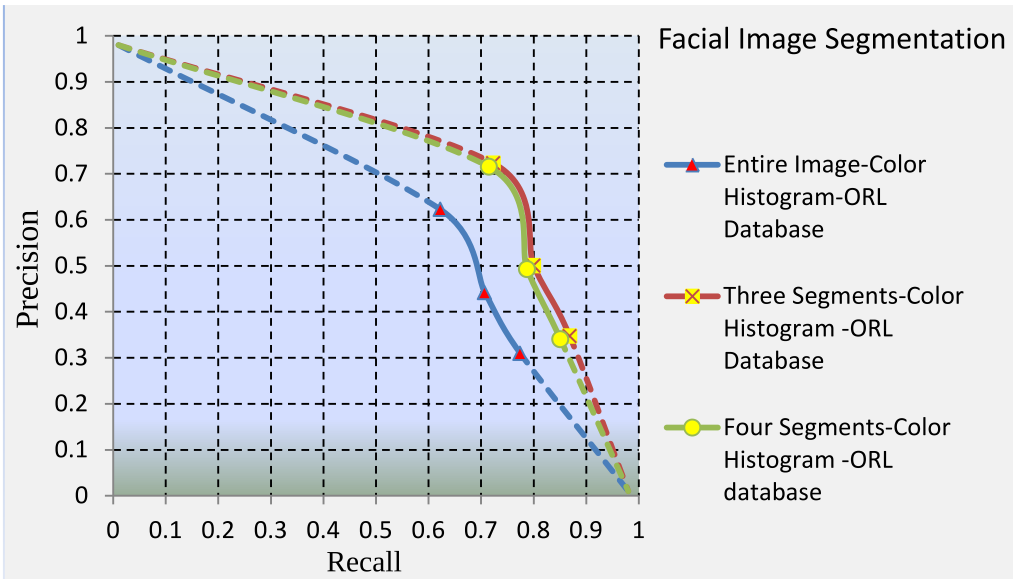

Figure 6.7: Color histogram-based face retrieval using different segments and extraction methods of human face with RGB color space and ORL database.

Table 6.8 : Color histogram-based face retrieval using different segments and extraction methods of human face with RGB color space and local database.

Extraction Method	Query Faces	Expected Faces	Top Face	Retrieved Faces	Relevant	Recall	Preci-sion	F-score
Traditional Method	750	7500	10	7500	5966	0.7955	0.7955	0.7955
	750	7500	16	12000	5966	0.7955	0.4972	0.6119
	750	7500	25	18750	5966	0.7955	0.3182	0.4546
Three Segments	750	7500	10	7500	6490	0.8653	0.8653	0.8653
	750	7500	16	12000	6690	0.892	0.5575	0.6862
	750	7500	25	18750	6809	0.9079	0.3631	0.5187
Four Segments	750	7500	10	7500	6408	0.8544	0.8544	0.8544
	750	7500	16	12000	6635	0.8847	0.5529	0.6805
	750	7500	25	18750	6770	0.9027	0.3611	0.5158

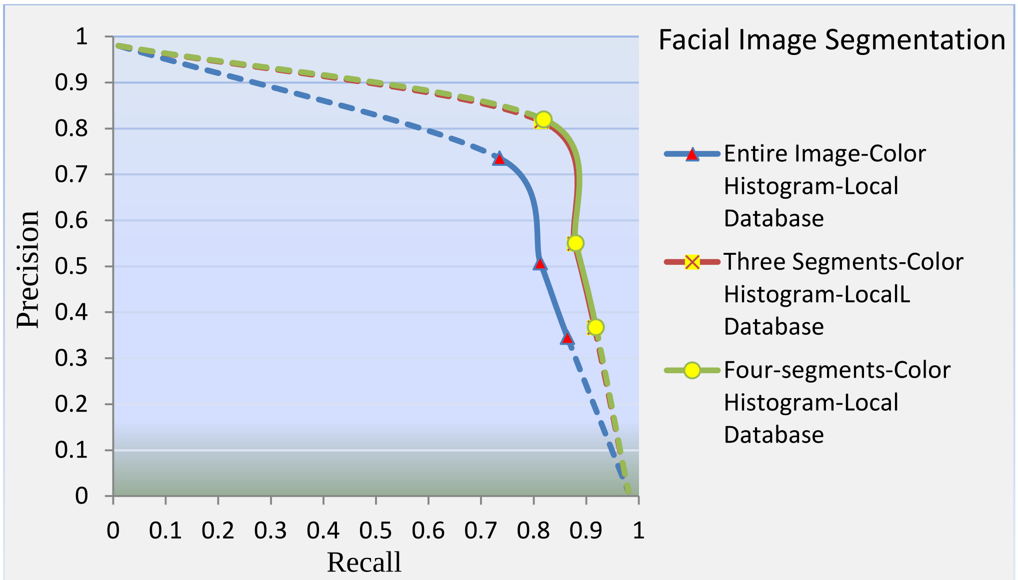

Figure 6.8 : Color histogram-based face retrieval using different segments and extraction methods of human face with RGB color space and local database.

As shown in Table 6.9 and Figure 6.9 it is observed that the combination of color histogram and eigenfaces-based facial image retrieval has achieved accuracies of 72.25%, 83.5%, and 77.55% respectively for the traditional, 3-segment and 4-segment methods using the ORL database. Likewise, Table 6.10 and Figure 6.10 show again higher accuracies at 79.49%, 89.51%, and 88.24% for the respectively mentioned method and the Local database.

Table 6.9 : Eigenfaces and color histogram-based face retrieval on different segments and extraction methods of human face with RGB color space and ORL database.

Extraction Method	Query Faces	Expected Faces	Top Faces	Retrieved Faces	Relevant Faces	Recall	Precision	F-score
Traditional Method	200	2000	10	2000	1445	0.7225	0.7225	0.7225
	200	2000	16	3200	1586	0.793	0.4956	0.61
	200	2000	25	5000	1687	0.8435	0.3374	0.482
Three Segments	200	2000	10	2000	1670	0.835	0.835	0.835
	200	2000	16	3200	1737	0.8685	0.5428	0.6681
	200	2000	25	5000	1784	0.892	0.3568	0.5097
Four Segments	200	2000	10	2000	1551	0.7755	0.7755	0.7755
	200	2000	16	3200	1676	0.838	0.5238	0.6447
	200	2000	25	5000	1757	0.8785	0.3514	0.502

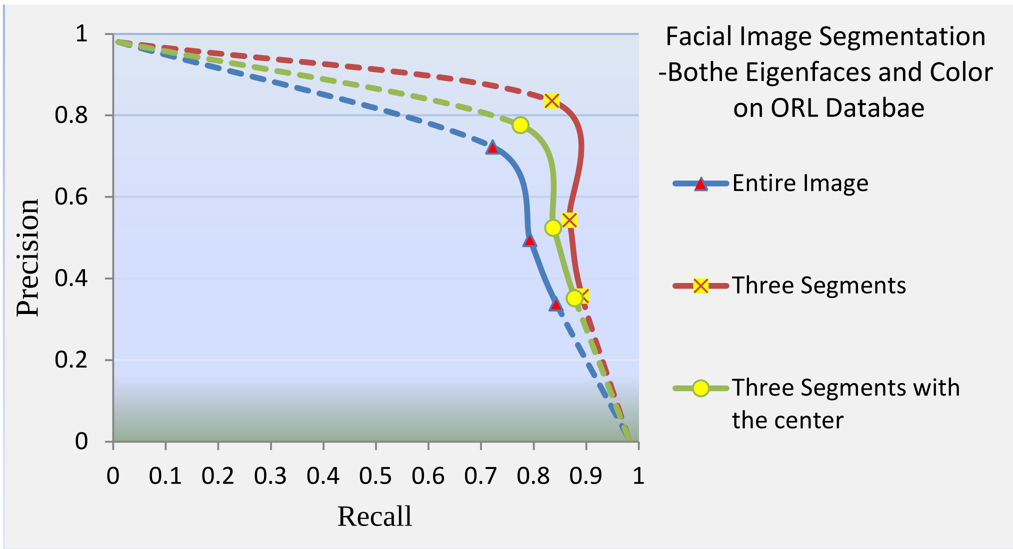

Figure 6.9: Eigenfaces and color histogram-based face retrieval
on different segments and extraction methods of human face
with RGB color space and ORL database.

Table 6.10 : Eigenfaces and color histogram-based face retrieval using
different segments and extraction methods of human face
with HSV and RGB color space ,and local database.

Extraction Method	Query Faces	Expected Faces	Top Face	Retrieved Faces	Relevant	Recall	Preci-sion	F-score
Traditional Method	750	7500	10	7500	5962	0.7949	0.7949	0.7949
	750	7500	16	12000	6332	0.8443	0.5277	0.6495
	750	7500	25	18750	6559	0.8745	0.3498	0.4997
Three Segments	750	7500	10	7500	6713	0.8951	0.8951	0.8951
	750	7500	16	12000	6915	0.922	0.5763	0.7093
	750	7500	25	18750	7028	0.9371	0.3748	0.5354
Four Segments	750	7500	10	7500	6618	0.8824	0.8824	0.8824
	750	7500	16	12000	6839	0.9119	0.5699	0.7014
	750	7500	25	18750	6968	0.9291	0.3716	0.5309

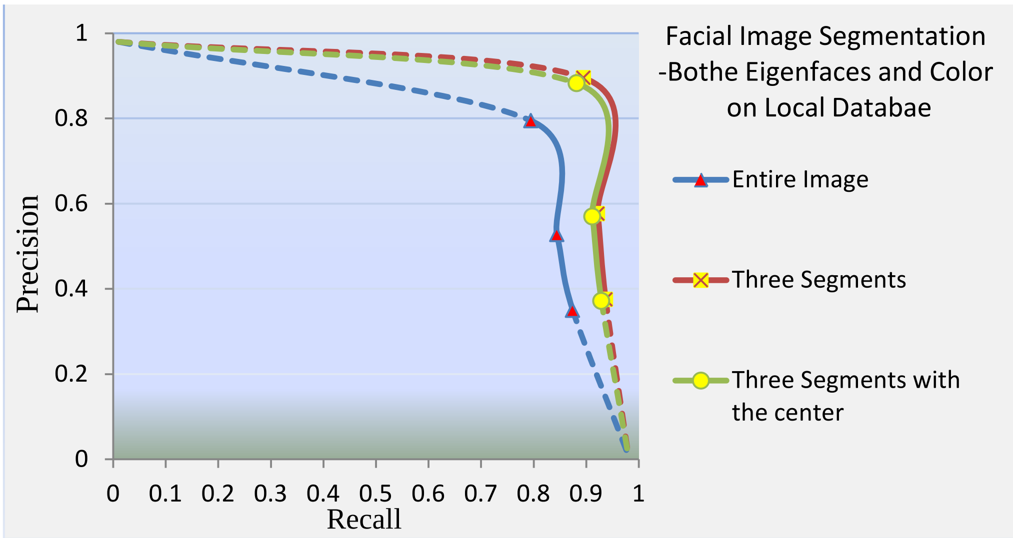

Figure 6.10: Eigenfaces and color histogram-based face retrieval using different segments and extraction methods of human face with HSV and RGB color space, and local database .

The results of the experiments show that the 3-segments extraction method performs better as compared to the traditional and the 4-segments methods. Considering the recall measurement itself, it is apparent that the best performance of the system with the 3-segment method is found within the top 25 retrieved images of both database using eigenfaces features, color histogram features as well as their combination. This is clearly substantiated in Tables 6.5; 6.6; 6.7; 6.8; 6.9 and 6.10, where accuracies of 91.6%, 86.8%, 89.2% have been achieved on the ORL database and 91.6%, 90.9%, 93.1% on the local database.

With our proposed method of facial segmentation, querying is simply done through global image using the global descriptors of the whole image and the processing in the system to extract the features vectors based on local descriptors, which includes 3 local descriptors in the center region of the face.

By using the local descriptors with facial image, it is obvious that the performance of the algorithms has improved. Most importantly, there is a significant performance improvement of some 10% in the proposed methods over the traditional method. This could be attributed to the fact that in the proposed method, the local area features of the face are extracted and compared separately to the same locality of the other faces. This has resulted in more focus and details in finding more important differences between the faces.

In the proposed method, algorithms have been developed to extract the features semantically such as the color histogram. In the traditional method, the histogram is calculated for the whole image without considering the position, focus, and distribution of colored spots on the skin of the face, for instance, not realizing that the colors in the center of the image are more important than those at the periphery. However, in the proposed method this is considered, such as the area of the eyes and forehead having more dark and white spots as well as a color gradient and configuration. These head features can be extracted and compared to other faces on the same facial localities. With this proposed method, querying over the whole image is based on the number of the objects found on the face, which will then be compared to the same number of objects on other faces.

Between the two proposed methods of face segmentation, namely, the three segmentation of the face and three segmentation of the face plus face center, the former has achieved better performance of the system. This suggests that adding the features of the face center to the features extracted from the three segments could result as noise, which in turn has degraded the system performance, leading to significantly lower accuracy.

Comparing the accuracy results of eigenfaces, color histogram, and eigenfaces-color integration using the traditional method of features extraction with those achieved by the proposed method of facial image segmentation and extraction, the results show that there is a significant degree of enhancement in the performance of the latter method.

6.3.1 Eigenfaces Features

Eigenfaces features have the capability to provide the significant features for face recognition. The advantages of these features are that processing is fast and no heavy storage of data is required. However, there existent factors originating from the facial image itself, which could affect the performance of the eigenfaces processing. These include the facial hair, skin scarring and face multiple view. Actually, this has been a long-standing problem of most features extraction methods specially those depending on face parts modeling. Figure 6.11 shows an example of a query dialogue.

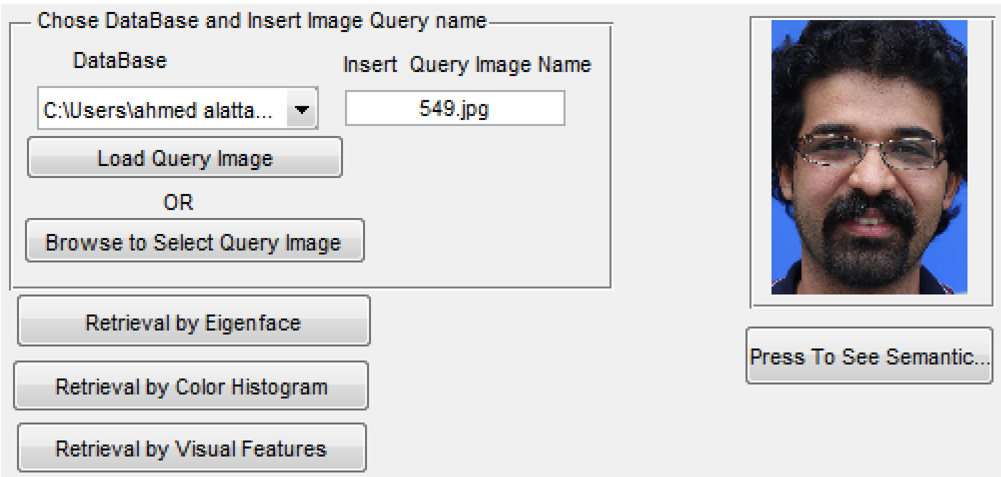

Figure 6.11: Example of facial image query based eigenfaces features.

Using the recall method of performance measure, Figure 6.12 shows that 70% accuracy was achieved within the top 10 cut-off level; 70% within top 16 and 100% within top 25. It is noted that the relevant images in row 4 are slightly orientated to the left, while those in row 5 are slightly orientated downwards. These results correspond to the fact that in face recognition different face images with same postures are considered similar rather than those of the same face images with different postures. The results of this example are considered the worst-case scenario of the system performance based on eigenfaces, as in some other runs the achieved accuracy was 100% within the first top 10 images.

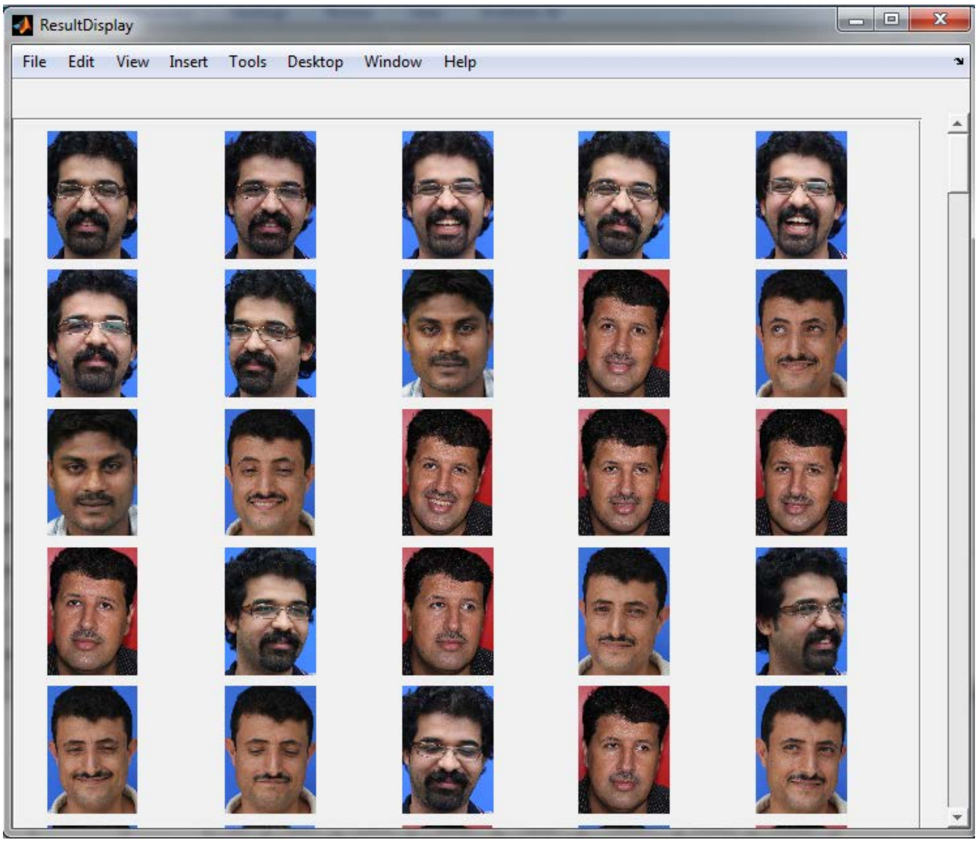

Figure 6.12: Example results of eigenfaces
based facial image retrieval.

One of the factors considered is the vector dimension of the eigenfaces features. The maximum number of eigenfaces that can be used per vector equals the size of the training vectors. For instance, if the training vectors set contain 750 images, the eigenfaces vector dimension would contain a maximum of 750 values. The eigenvalues of the first 50 training eigenvectors were plotted in a descending order as depicted in Figure 6.13. It is observed that at the beginning, the eigenvalues are high, then sloping downwards to significantly lower values. Larger eigenvalues are indicative that the corresponding eigenvectors contain more information for high-level discrimination. Conversely, much less information will be found in eigenvectors with low eigenvalues. Consequently, the vectors with small eigenvalues were omitted in our research. In this example, the first eigenfaces that corresponded to the first eigenvalues position was chosen for eigenfaces features vector. This did not affect the results significantly.

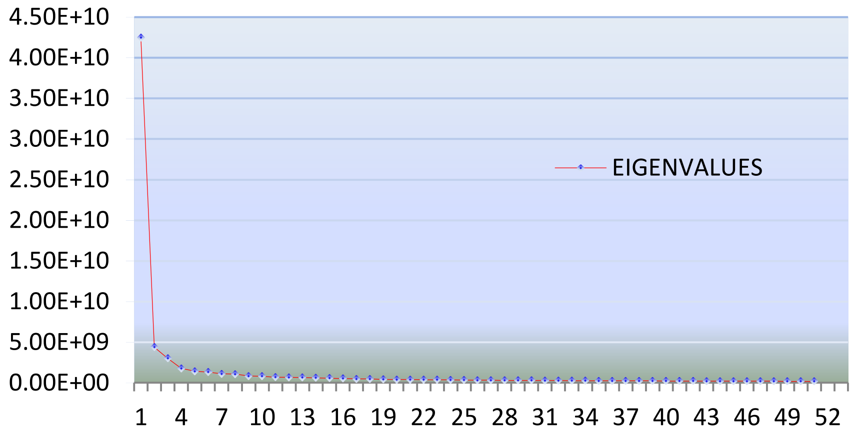

Figure 6.13 : The first 50 eigenvalues of the training vectors.

For testing purpose, the system was trained and tested on different dimensions of eigenfaces vector features. The vector dimension was phased from 1 to 200 eigenfaces. Figure 6.14 shows that after 20 eigenfaces per vector the recall accuracy does not improve significantly. This is because of the lack of discriminative information in weak eigenvalues.

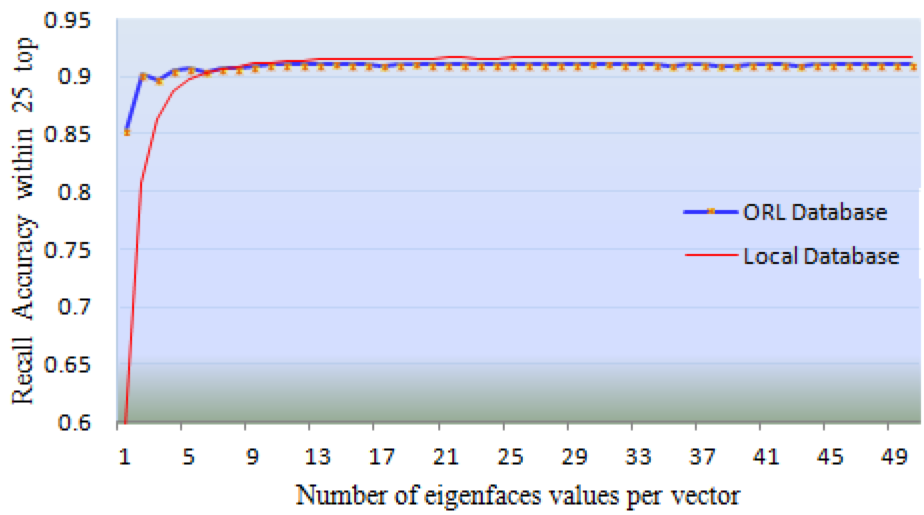

Figure 6.14: Facial image retrieval with different eigenfaces vector dimension.

Increasing the dimension of the vectors will not necessarily result in higher performance. Moreover, the existence of some trivial information may be consider as noise and will degrade the system performance, especially when the eigenfaces are combined with other features, let alone the complexity and time needed for selecting and processing each vector.

6.3.2 Color Histogram Features

Facial image retrieval based on the color histogram algorithm has produced some excellent results in that it was able to retrieve most of the relevant images to the queried image. Unlike the eigenfaces features, performance of the color histogram algorithm somewhat depends on the dimension of the features vectors. With color histogram, increasing the size of the bines, would result in slight increase in the dimension of the features vectors, leading to improved retrieval performance. Results in Table 6.11 and Figure 6.15 on ORL database, and Table 6.12 and Figure 6.16 on the local database show the accuracies of the facial image retrieval system based on color histogram features with different sizes of bins.

Table 6.11: Color histogram-based face retrieval with different size of bins on the ORL database .

Bins Size	Query Faces	Expected Faces	Top Faces	Retrieved Faces	Relevant Faces	Recall	Precision	F-score
4*4*4	200	2000	10	2000	1322	0.661	0.661	0.661
	200	2000	16	3200	1531	0.7655	0.4784	0.5888
	200	2000	25	5000	1709	0.8545	0.3418	0.4883
16*4*4	200	2000	10	2000	1446	0.723	0.723	0.723
	200	2000	16	3200	1600	0.8	0.5	0.6154
	200	2000	25	5000	1737	0.8685	0.3474	0.4963
8*8*8	200	2000	10	2000	1442	0.721	0.721	0.721
	200	2000	16	3200	1608	0.804	0.5025	0.6185
	200	2000	25	5000	1750	0.875	0.35	0.5

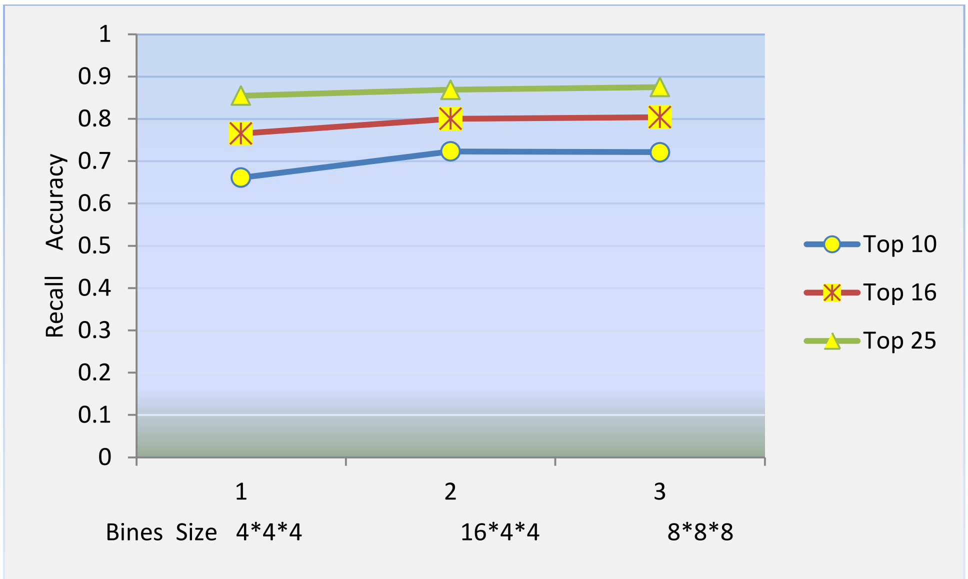

Figure 6.15: Color histogram-based face retrieval with different size of bins on the ORL database.

Table 6.12: Color histogram-based face retrieval with different size of bins on the local database .

Bins Size	Query Faces	Expected Faces	Top Faces	Retrieved Faces	Relevant Faces	Recall	Precision	F-score
4*4*4	750	7500	10	7500	6355	0.8473	0.8473	0.8473
	750	7500	16	12000	6615	0.882	0.5513	0.6785
	750	7500	25	18750	6754	0.9005	0.3602	0.5146
16*4*4	750	7500	10	7500	6490	0.8653	0.8653	0.8653
	750	7500	16	12000	6690	0.892	0.5575	0.6862
	750	7500	25	18750	6809	0.9079	0.3631	0.5187
8*8*8	750	7500	10	7500	6783	0.9044	0.9044	0.9044
	750	7500	16	12000	7040	0.9387	0.5867	0.7221
	750	7500	25	18750	7175	0.9567	0.3827	0.5467

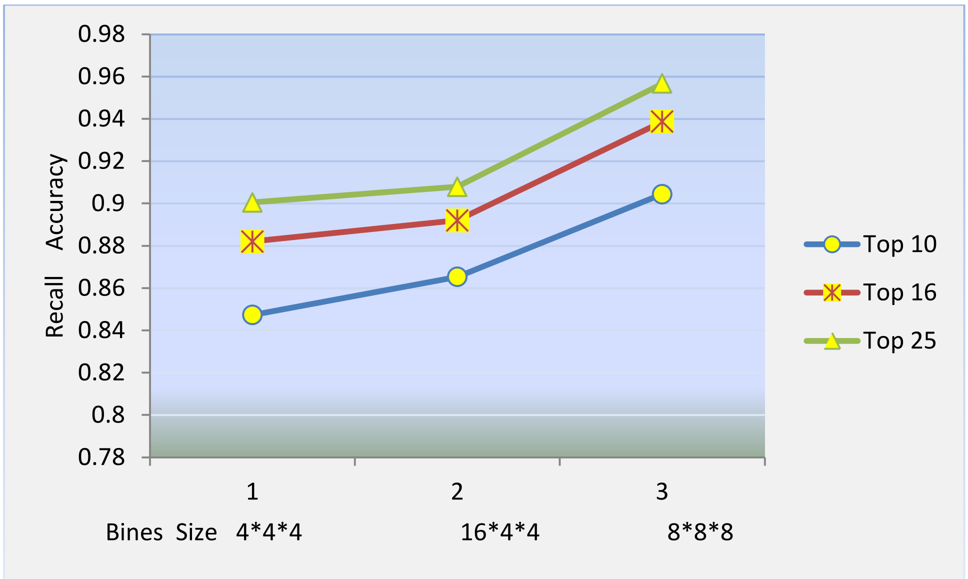

Figure 6.16: Color histogram-based face retrieval with different size of bins on the local database**.**

It was noted that choosing the distribution values of the color space coordinates to represent the bines size during the quantization processing, has influenced the color based facial image retrieval results. The accuracy will be based on the distribution of the colors in the image. For instance, if the red color distribution in the image is better than the other colors, then the red color should be given more focus than the other colors and thus will lead to better result.

While this influence is clear on color images database, as it is shown in Table 6.14 and Figure 6.18 the distribution of the color space coordinate has no influence on gray level image database as shown in Table 6.13 and Figure 6.17. This is because the three channels of the gray image carry the same information.

Table 6.13: Color histogram-based face retrieval with different distribution of the color space coordinate on ORL database.

Bins Size	Query Faces	Expected Faces	Top Faces	Retrieved Faces	Relevant Faces	Recall	Precision	F-score
8*8*8	200	2000	10	2000	1442	0.721	0.721	0.721
	200	2000	16	3200	1608	0.804	0.5025	0.6185
	200	2000	25	5000	1750	0.875	0.35	0.5
16*8*4	200	2000	10	2000	1446	0.723	0.723	0.723
	200	2000	16	3200	1600	0.8	0.5	0.6154
	200	2000	25	5000	1737	0.8685	0.3474	0.4963
8*16*4	200	2000	10	2000	1446	0.723	0.723	0.723
	200	2000	16	3200	1600	0.8	0.5	0.6154
	200	2000	25	5000	1737	0.8685	0.3474	0.4963
8*4*16	200	2000	10	2000	1446	0.723	0.723	0.723
	200	2000	16	3200	1600	0.8	0.5	0.6154
	200	2000	25	5000	1737	0.8685	0.3474	0.4963

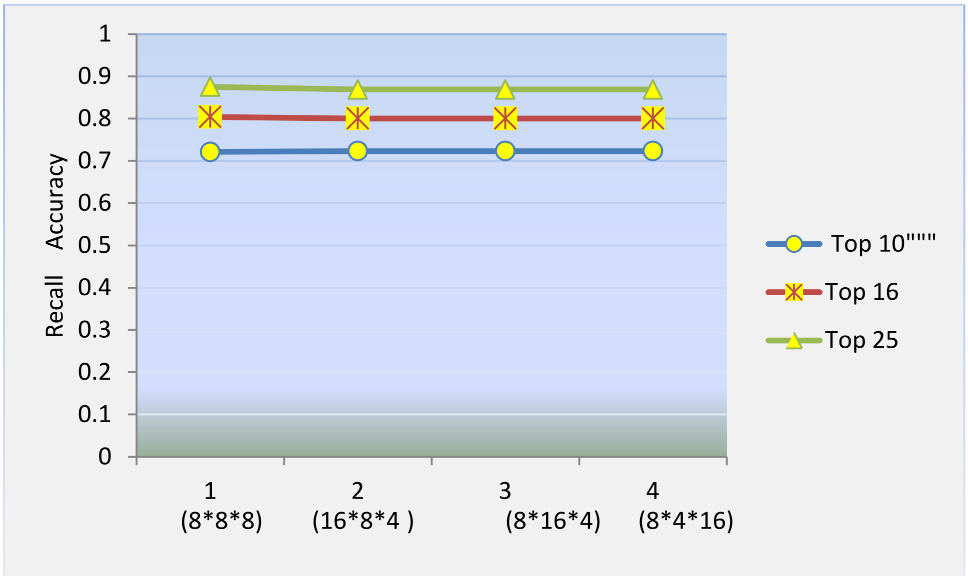

Figure 6.17: Color histogram-based face retrieval with different distribution of the color space coordinate on ORL database.

Table 6.14: Color histogram-based face retrieval with different distribution of the color space coordinate on local database.

Bins Size	Query Faces	Expected Faces	Top Faces	Retrieved Faces	Relevant Faces	Recall	Precision	F-score
8*8*8	750	7500	10	7500	6783	0.9044	0.9044	0.9044
	750	7500	16	12000	7040	0.9387	0.5867	0.7221
	750	7500	25	18750	7175	0.9567	0.3827	0.5467
16*8*4	750	7500	10	7500	6854	0.9139	0.9139	0.9139
	750	7500	16	12000	7065	0.942	0.5888	0.7247
	750	7500	25	18750	7190	0.9587	0.3835	0.5478
8*16*4	750	7500	10	7500	6833	0.9111	0.9111	0.9111
	750	7500	16	12000	7060	0.9413	0.5883	0.7241
	750	7500	25	18750	7180	0.9573	0.3829	0.547
8*4*16	750	7500	10	7500	6854	0.9139	0.9139	0.9139
	750	7500	16	12000	7086	0.9448	0.5905	0.7268
	750	7500	25	18750	7208	0.9611	0.3844	0.5492

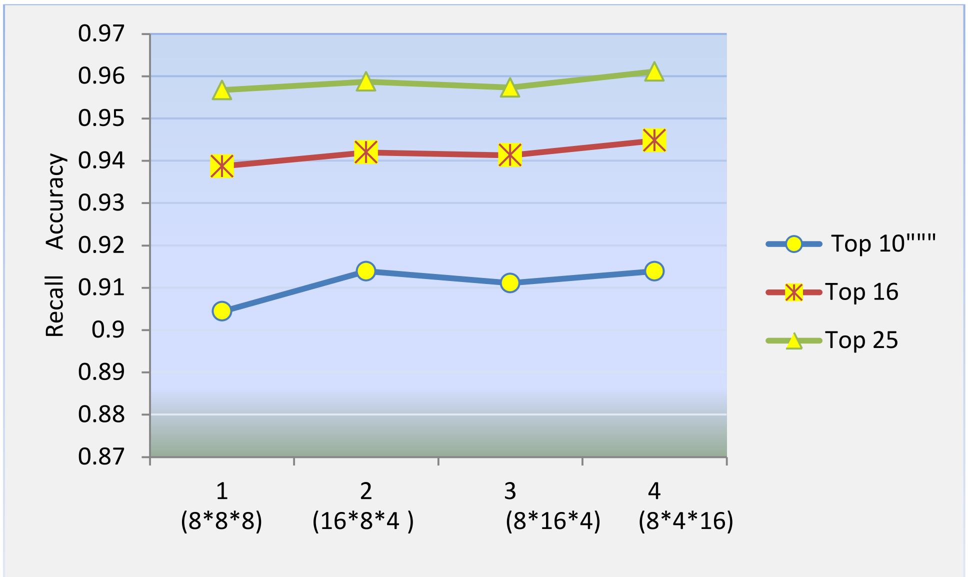

Figure 6.18: Color histogram-based face retrieval with different distribution of the color space coordinate on local database.

The colour histogram algorithm conducts image colour analysis without consideration for locations of colour components in the image. Consequently, object location information is obviously left out. In addition, this colour analysis generates similarity of colours as seen by the computer and may necessarily differ from those visualized by the human eyes. Semantically, this is a major weakness in the colour histogram analysis.

The proposed method has addressed this weakness through the employment of the facial image segmentation algorithm.

Visual examples of facial retrieval based on the color histogram algorithm using the proposed method of facial image segmentation have been provided. Figures 6.19 and 6.20 respectively show the results of visual query and image retrieval of the color histogram using the local database. The recall method of performance measure shows that 100% accuracy was achieved for the top 10, 16, and 25 cut-off levels, where all the 10 images related to the query image were retrieved in the first and second rows of the results frame. The results given in this example constitute the best achievement of face retrieval based on color histogram. It is also considered the best result of system performance based on color histogram, as in some other runs the achieved accuracies fall below 100% for the topmost 10 images provided in results tabulated earlier.

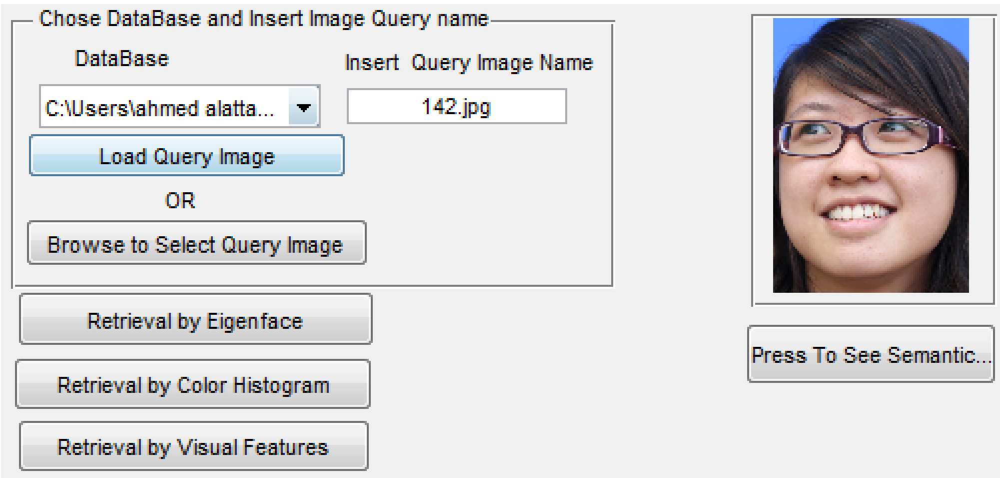

Figure 6.19: Example of facial image query based color histogram features.

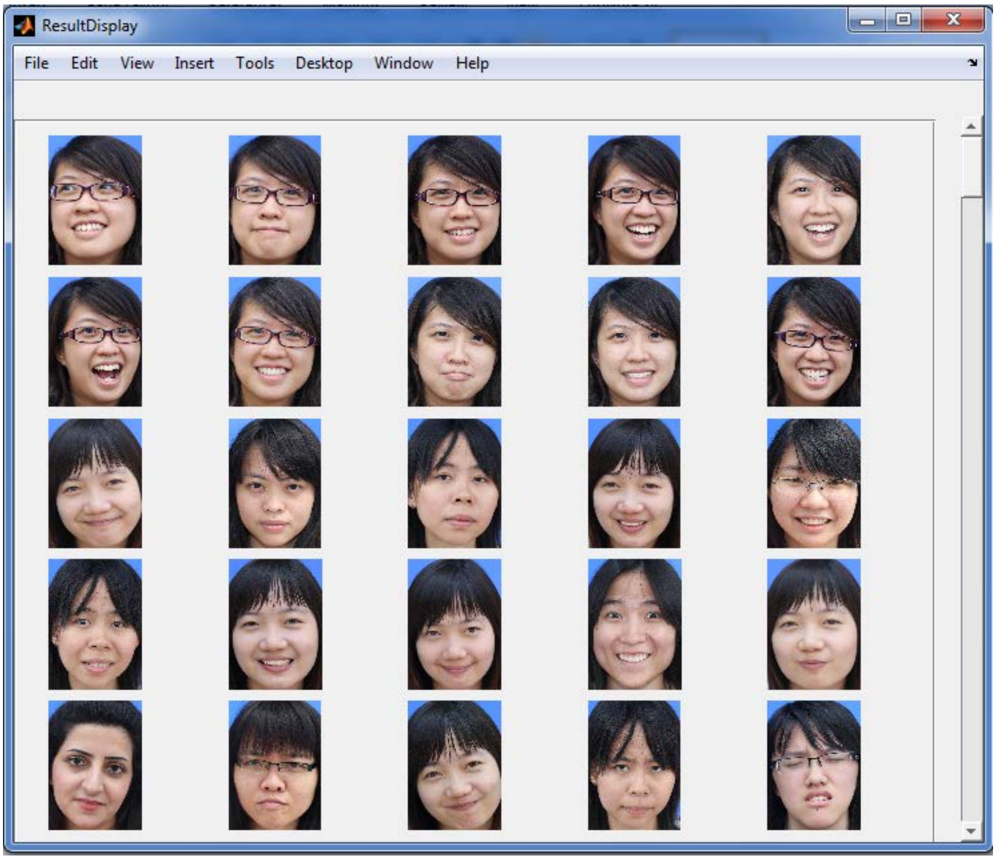

Figure 6.20: Example results using color histogram
based on image segmentation method.

6.4 Probabilistic Approach Experiments and Results

The proposed method of probabilistic approach was tested for the facial image retrieval system using the facial image semantic features. Table 6.15 and Figure 6.21 show the results of the system testing on both the ORL and local databases.

Table 6.15: Semantic features-based face retrieval using probabilistic
approach on ORL and local database .

The databases	Query Faces	Desired Faces	Top Faces	Retrieved Faces	Relevant Faces	Recall	Precision	F-score
ORL	200	2000	10	2000	1952	0.976	0.976	0.976
	200	2000	16	3200	1975	0.9875	0.6172	0.7596
	200	2000	25	5000	2000	100.0	0.400	0.5714
Local	750	7500	10	7500	7154	0.9539	0.9539	0.9539
	750	7500	16	12000	7300	0.9733	0.6083	0.7487
	750	7500	25	18750	7384	0.9845	0.3938	0.7876

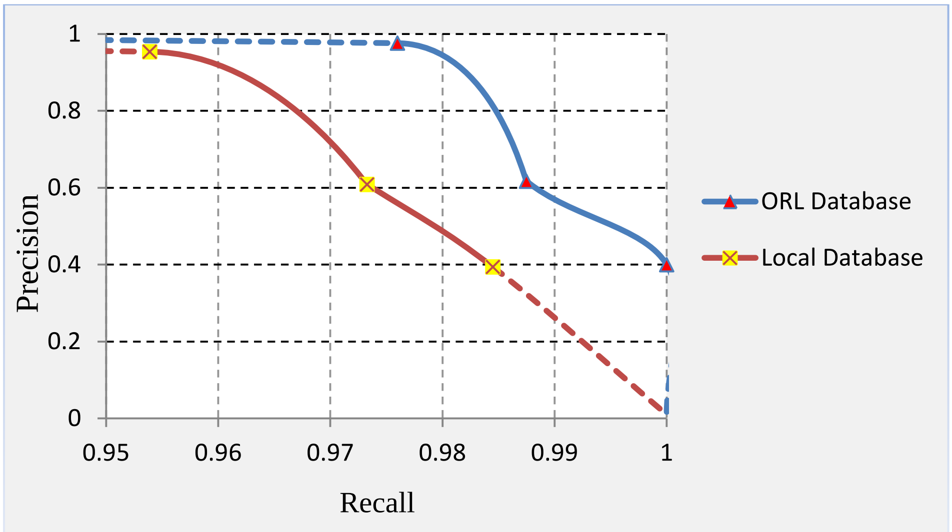

Figure 6.21: Semantic features-based face retrieval using probabilistic approach on ORL and local database .

High accuracies for the top 10 faces are observed in both the recall and precision methods for the ORL and local databases, respectively 97.6%, and 95.39%. Higher accuracies are also observed in the recall method for the top 25 results - 100 % and 98.45%.

Although retrieval by semantic description was not expected to retrieve all the relative images of the person concerned in the top range, as there were many people in the database having the same semantic descriptions such as race and gender, the results achieved based on the proposed method were more than 90% of the accuracies which can be considered as excellent.

The results of the ORL database were better than the local database using the same method. In the context of image retrieval the relationship between the result accuracy and size of the database is somewhat inversed. However, if the desired face has a unique semantic feature used during the search process, the search space will be narrowed down, thus increasing the accuracy. An example of such unique feature is the facial mark.

The results in the tables above constitute the whole semantic features vector testing. If we use some of these features in the query stage, the advantage of the proposed method and the weakness of the previous methods will be conspicuous.

Suppose, the query vector includes only the features 'Gender' with description 'male', 'Race' with description 'Malay', and 'Glasses Frame' with description 'Rectangle ' to be submitted to the system , the finalized query vector is shown in Figure 6.22.

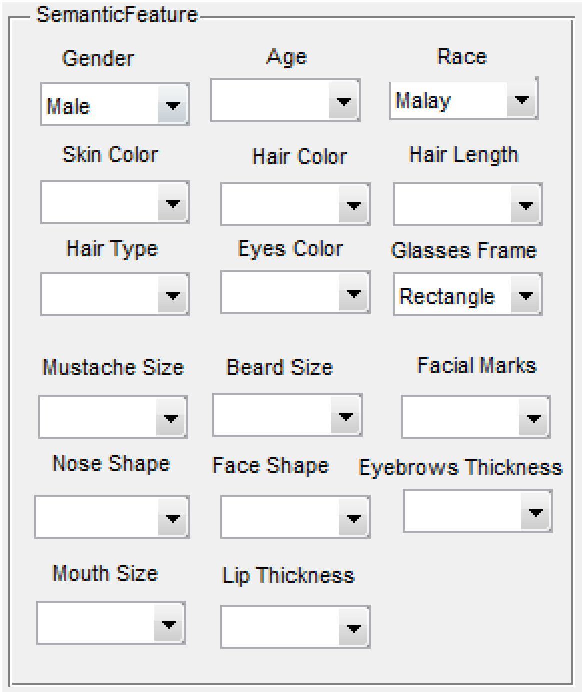

Figure 6.22: Query vector includes some semantic features description.

Three methods are discussed, namely, (i) the traditional method based on pruning the image from the search space, (ii) the features pruning method based on pruning the non-matching features from the image description, and (iii) our proposed method.

The traditional method is based on the exact symbolic matching of the description value of the query vector against the database vectors. For the above query vector the system will search the images in the database to detect features that match the query feature description - 'male' ,'Malay' and 'rectangle' exactly. If one or more features from the images features do not match the corresponding features in the query vector, the system will prune these images from the database search space. Such technique may lead to pruning desired faces from the search space if one or more of their features do not match the query features attributed to different viewpoint annotations. Figure 6.23 shows the result of the system performance based on the traditional method.

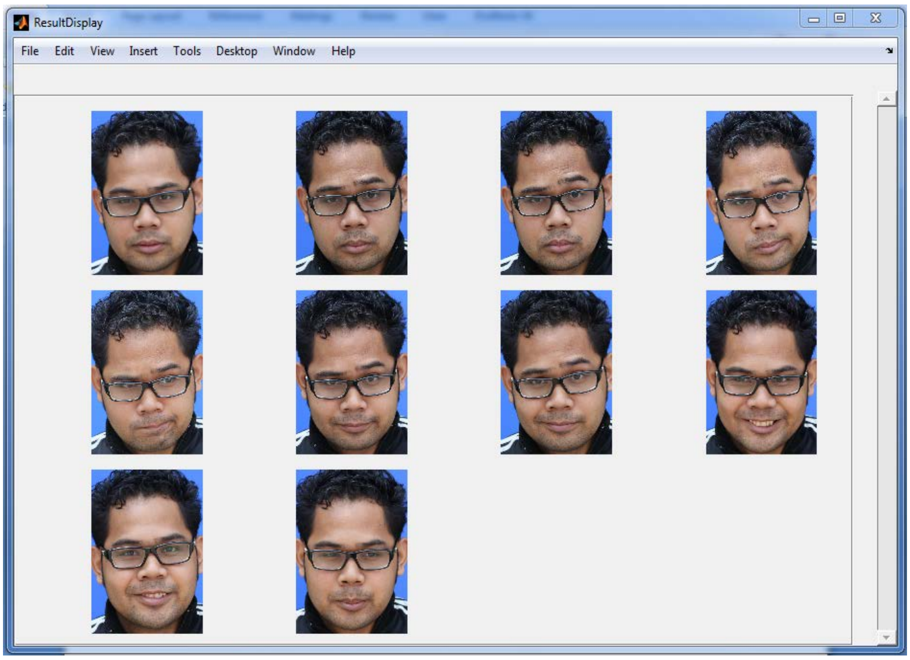

Figure 6.23: Retrieval using the symbolic matching technique.

The second method lies in the features pruning method. The system will compare the query vector with the database vectors. If one or more of the image features do not match the queried features, the images will not be pruned from the search space. Instead, these non-matching features will be ignored. Figure 6.24 show the results using the same queried vectors in Figure 6.22.

The weakness of this features pruning method is that by ignoring the feature from the image database means that the image is not pruned from the database and this procedure will decrease the probability of the ignored images to be displayed in the top for users. For example, there are images with glasses feature descriptions as square or circular. These descriptions do not match the rectangular descriptions. Therefore, these features are ignored and the probability of the associated images will be similar to those images having feature descriptions of glasses as ‘none’, meaning face without glasses.

From the Figure 6.24 frame (1), frame (2) and frame (3) that shows the results based on the pruning feature method. It is clear that the system has retrieved the exact corresponding images in the first and second rows of the displayed result. In the third and fourth rows it has retrieved the images based on the features ‘gender’ and ‘race’ and have ignored the features ‘glasses’ that did not correspond to the description of the features ‘glasses’ of the queried vector, and so on. The results in the frames show that mostly the features race and glasses were pruned and ignored.

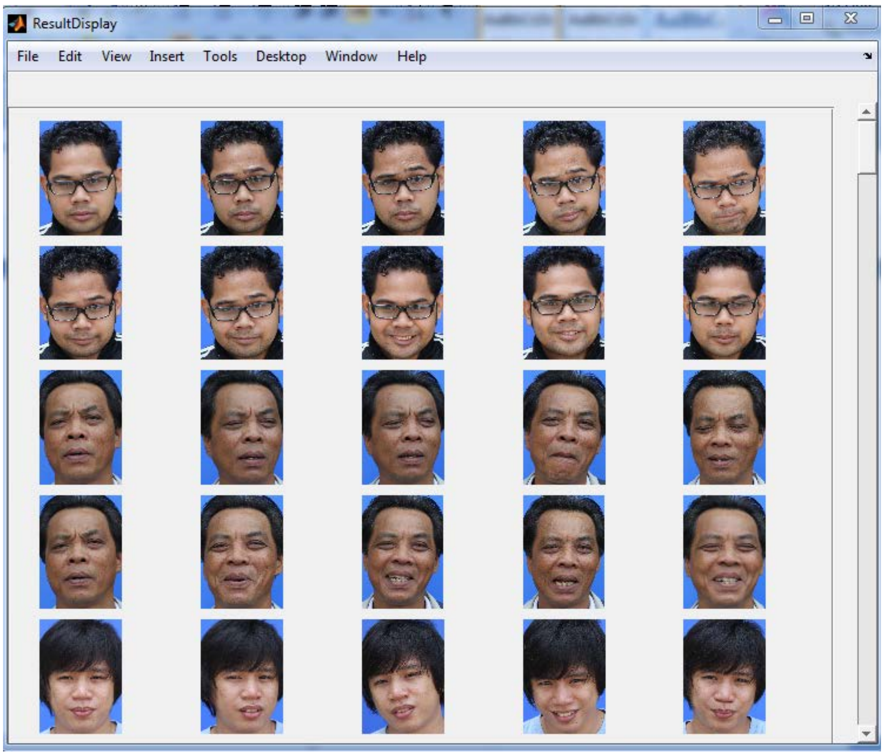

Figure 6.24: Frame (1).

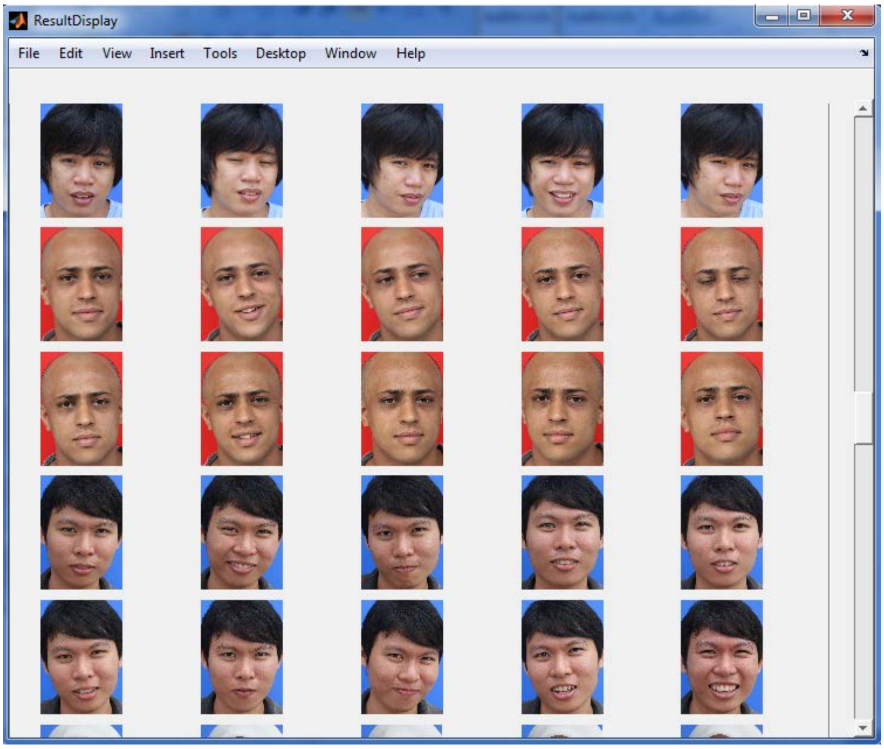

Figure 6.24: Frame (2).

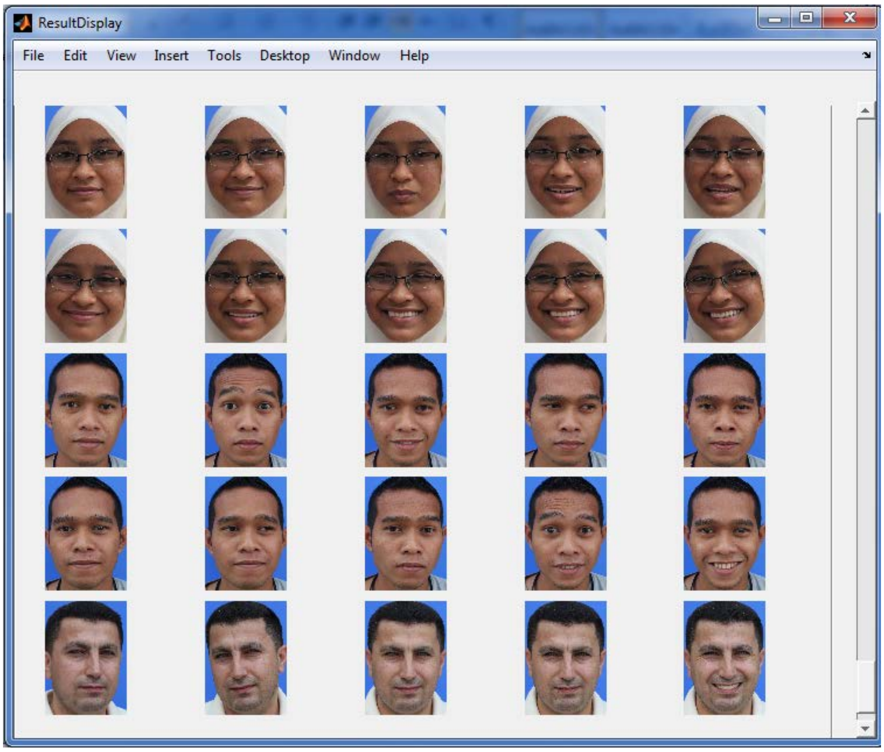

Figure 6.24: Frame (3).

Figure 6.24 Frame (1), Frame (2), and Frame (3): Retrieval using the pruning features method.

Other example is the description of the nose, mouth, and face shape. It could be that the same features correspond but are not considered so due as the reflected in the description between small and medium or large and medium.

In our proposed method, the weaknesses found in the previous methods were improved. The system will not prune the images from the search space and will not ignore the features from the image features vector. The system will search for the images that correspond to the query vector and display them on top. If an image features does not match the query features, the system will measure its similarity distance from the query features and will compute a probability value based on this distance.

Figure 6.25 frame (1), frame (2), and frame (3) shows the system performance using the proposed method for the same queried vectors in Figure 6.22. The system has retrieved in the first and second rows the exact available images in the database that matches the descriptions ‘Male’, ‘Malay’, and ‘Rectangle’ glass frames. In the next rows it has retrieved the images that were close to but did not much exactly the query features - In addition to the features ‘Male’, ‘Malay’, and ‘Square’ glass frames, the features ‘Male’, ‘Malay’, and ‘Oval’ or ‘Circular’ glass frames and so on are included.

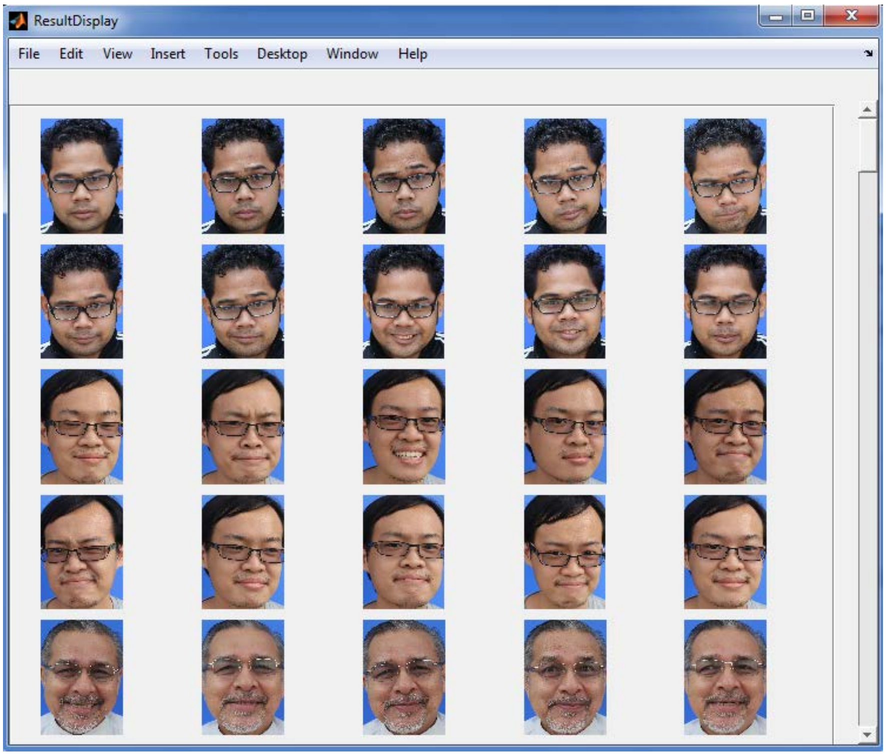

Figure 6.25: Frame (1).

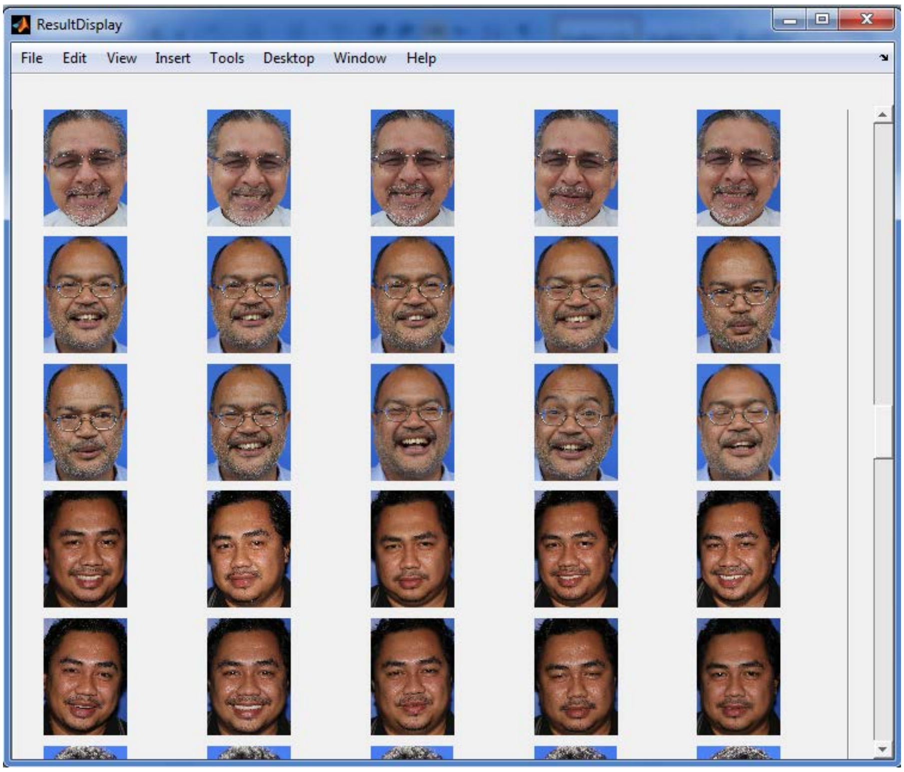

Figure 6.25: Frame (2).

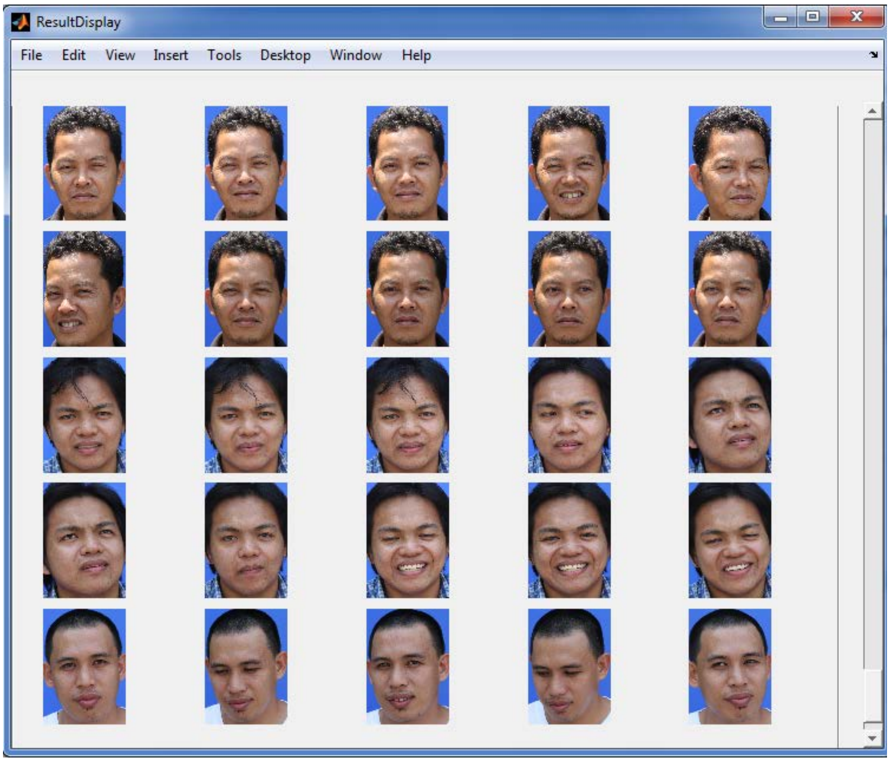

Figure 6.25: Frame (3).

Figure 6.25 Frame (1), frame (2), and frame (3): Retrieval using the proposed method.

Glasses frame features are chosen in these examples because these are visually conspicuous to the reader. Nevertheless, the same problems was discussed above could be occurred with descriptions of other face features due to varying human viewpoints. The aim of the proposed method is to reduce the side effects of these problems, making the facial retrieval more accurate than previous methods. The retrieved images in the above examples were based on their descriptions that have been annotated by the volunteers.

More experiments were also carried out to evaluate the proposed method using queries of 10 test vectors. The Recall and Precision methods were used to evaluate the achievements made using the local database and the results are indicated in Table 6.16 and Figure 6.26. The evaluation was based on two frames (2*25 images) of results for each query vector, including the images of the top five persons relative to the query vector, where the targeted images were not the images of a particular person.

Table 6.16: Semantic features-based face retrieval using probabilistic approach on local database.

The technique	Query vectors	Top Faces	Expected Faces	Retrieved Faces	Relevant Faces	Recall	Precision	F-score
Image pruning	10	25*2	500	250	250	0.50	1.00	0.6667
Features Pruning	10	25*2	500	500	360	0.72	0.72	0.7200
Proposed Method	10	25*2	500	500	480	0.96	0.96	0.9600

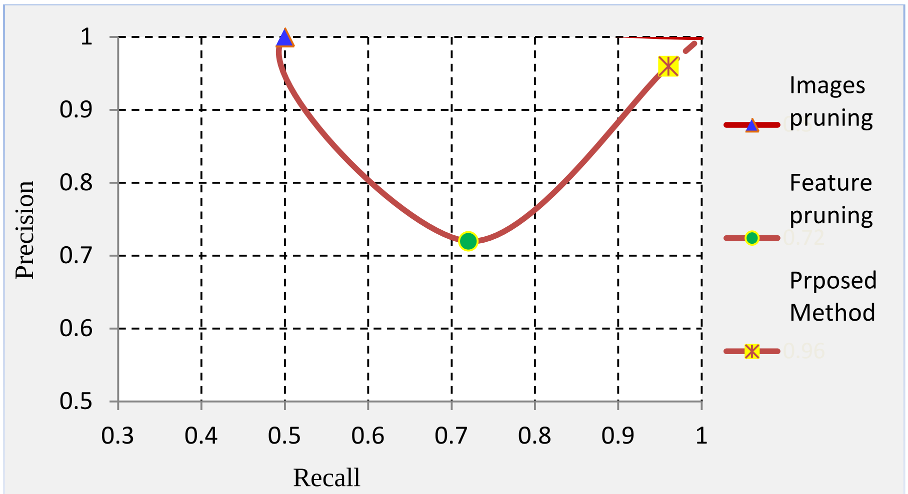

Figure 6.26: Semantic features-based face retrieval using probabilistic approach on local database.

Semantic features with probabilistic approach can be used for the first query without visual image, in case the user has no available image to compare the nearest faces to the required face. The result would be a number of facial images their semantic description fit the query features. It is supposed to be the closest to the mental image in the user mind. From the pool of the result user can pick an example image for his/her visual query. The user can then utilize the query image alone to search the database and proceed to associate it with the description that he/she thinks is closest to the desired image.

6.4.1 Semantic Features Weighing

Without weighting and assigning priority for the semantic features, it will be difficult to direct the retrieval process to retrieve what the user exactly needs using the semantic features and hence the retrieval will be somewhat random.

The estimated vector weights of the semantic features used in this research are shown in Table 6.17.

Table 6.17: The weights values of the semantic features.

The features	The weights Values
Gender	0.836700
Age	0.666900
Race	0.496200
Skin Color	0.291000
Hair Color	0.183500
Hair Length	0.156600
Hair Type	0.154500
Eyes Color	0.152900
Glasses Shape	0.121000
Moustache Size	0.113000
Beard Size	0.111100
Facial Marks	0.110500
Nose Shape	0.106500
Face Shape	0.105500
Eyebrows Thickness	0.084400
Mouth Size	0.081000
Lip Thickness	0.079700
Eyes Size	0.074200
Ears Size	0.057800
Forehead Length	0.057600

We have adjusted the weights of the features - race and age to ensure the more balanced distribution of the weights among the features and to avoid the domination of one feature. Of course, each feature was given its priority and importance based on the weights obtained from the case study process as discussed in chapter five, sections 5.8.2.2 and 5.8.2.4.

The weights were not assigned directly according to the descriptive terms of the semantic features, because firstly, every term has more than one semantic description. Secondly, the output of the distance measurement of two pair of features could be the same. Thus, we would not benefit from weighting the features. Consequently, numerical sequential values were assigned for each term description. Close values were given to descriptions that were considered close. For instance, we gave South East Asia races close values whereas the Africans or Europeans were assigned very far values. Another example is that the descriptions of shape - square and rectangle, and cycle and oval were given close values.

The similarity distance outputs of pairs of descriptions derived from the query and database vectors were then weighted by the corresponding weights of their respective descriptive terms.

For instance, the similarity distance between the description of the semantic features – gender from the query vector and the descriptions of the semantic features – gender from the database vectors will be calculated and the output will be weighted through the weight value 0.836700 of the gender term from Table 6.17.

We provide an illustration in the next section to show the effectiveness of weighting the features. Suppose that the query vectors included the semantic features - gender ‘female’, hair length ‘covered head’, mustache size ‘short‘, beard size ‘short‘, and facial marks ‘mole’ as shown in electronic form of Figure 6.27, the system will then seek the images that include these features of the query vectors. If some features are not found, the images with the remaining features will be the targeted.

Considering that the current database used in this research has no female with mustache or beard, unless there are errors during the annotation process, the image with features ‘female’, ‘covered head’, and ‘mole’ are then searched and retrieved. The images with features mustache size ‘short’, beard size ‘short’, and facial marks ‘mole’ are considered as well. Therefore, the first group includes three features and the second group includes three features.

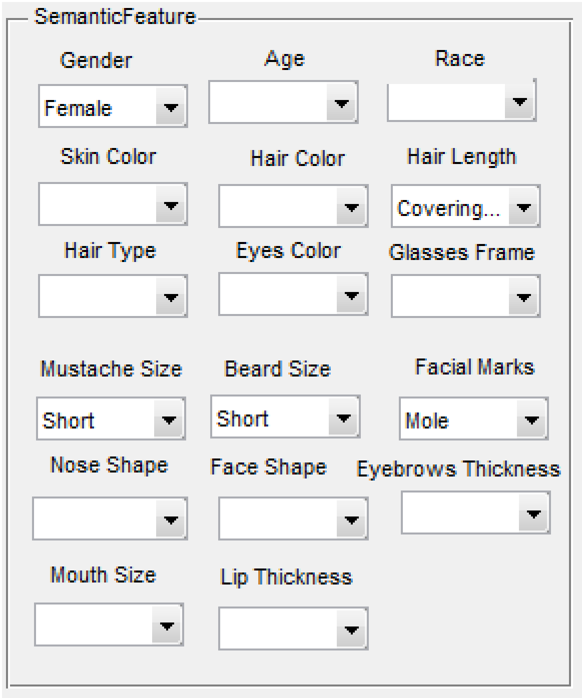

Figure 6.27: Query vector includes some semantic features description.

As these features are not weights, the probability of each image in both groups is equal. There is therefore no prioritizing based on weights. Images are retrieved and displayed for the user on the top range based on its priority position in the database. In other words, images are assessed early if they meet requirements of some of the query features and then displayed to the user on the top range. This procedure can lead to inaccurate retrieval results. Figure 6.28 frame (1), frame (2), and frame (3) shows the results of query without weighting features. The system ranks the results based on the probability of each image. As the individual image probabilities are equal, the display is then based on the position of the images in the database -”first come first serve”.

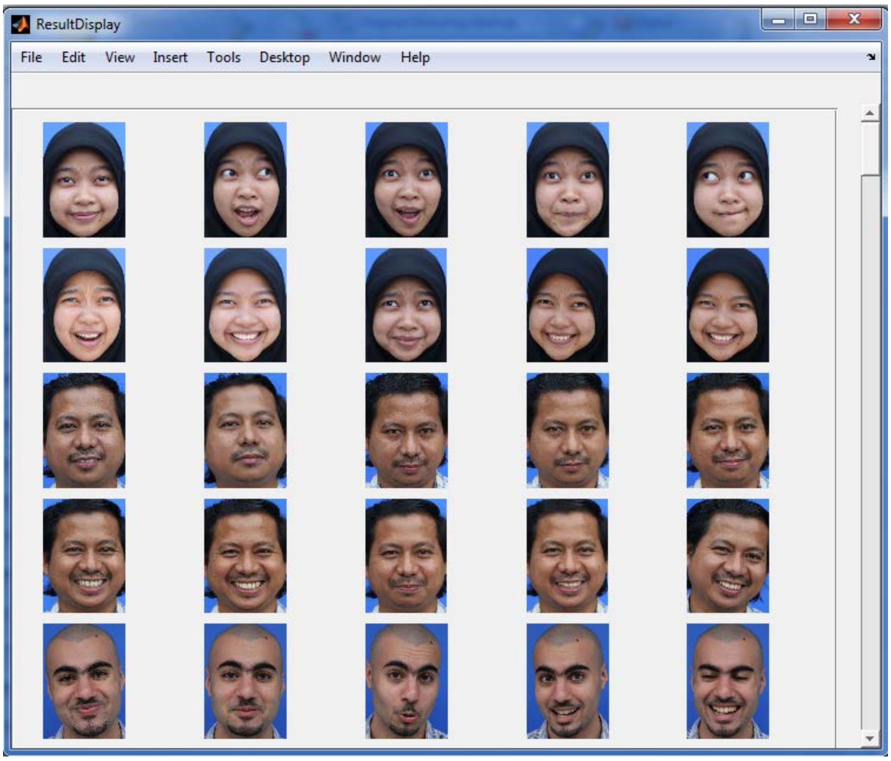

Figure 6.28: Frame (1).

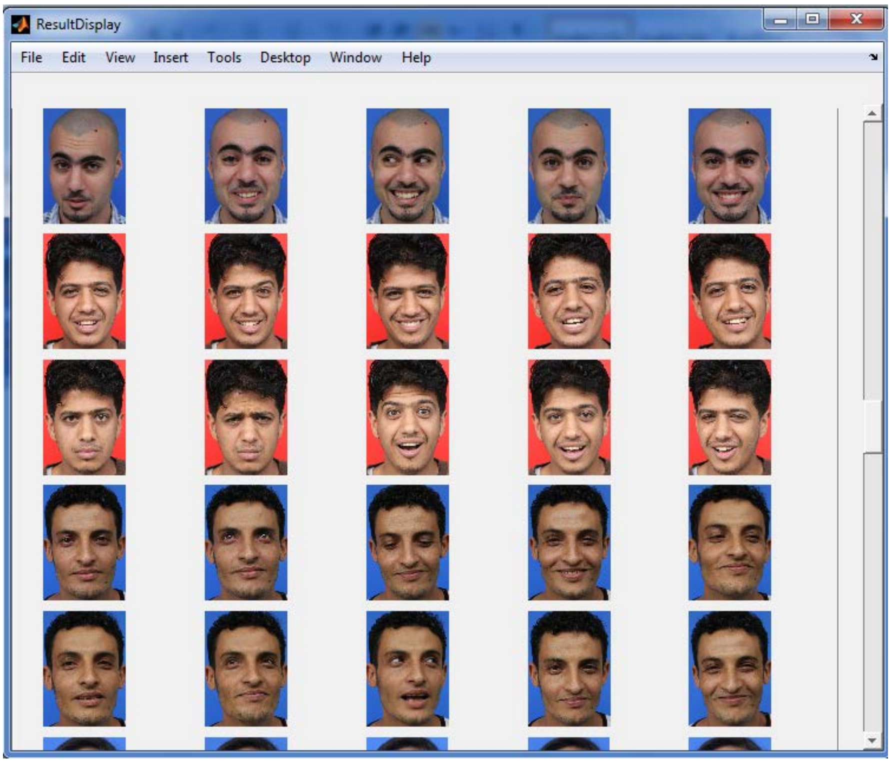

Figure 6.28: Frame (2).

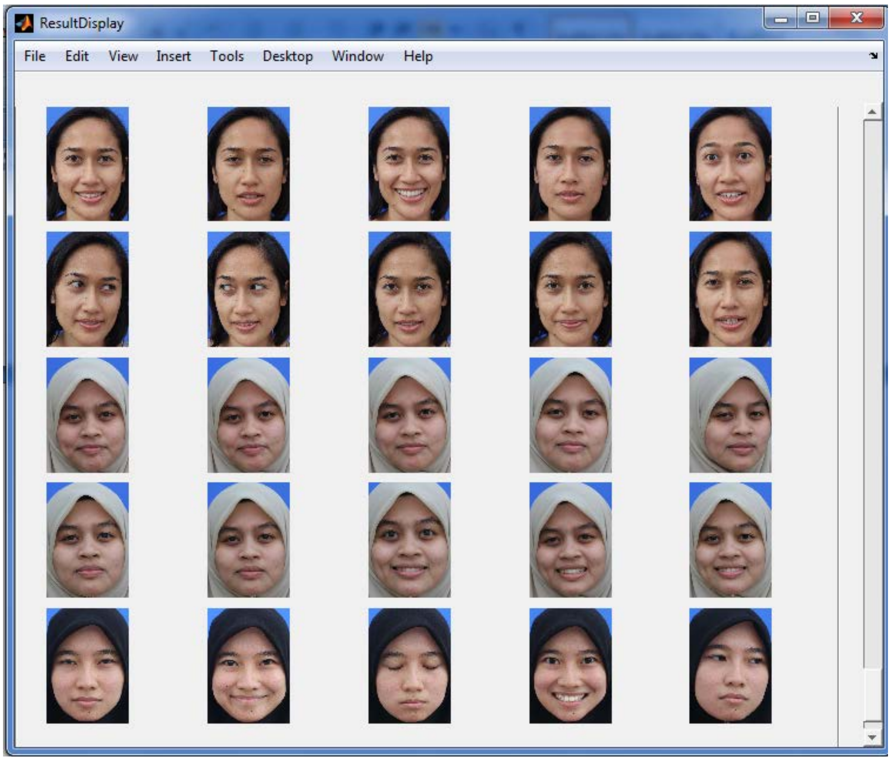

Figure 6.28: Frame (3).

Figure 6.28 Frame (1), Frame (2), and Frame (3): System performance without weighted features.

With our proposed method of feature weighting, the retrieval is directed to the specific requirements of the user. The probability of each image will be based on the weights of its features, which correspond to the queried features. Features like gender and hair length will have weights heavier than features like mustache and beard size. The system will emphasize on 'gender' as more than other features. For example, if the system finds the face with the same gender but does not match the other features while simultaneously finds another face which does not match the gender but matches the other features, it will then consider the face with 'gender' more similar to the query and retrieves it in the top range. Figure 6.29 frame (1), frame (2), and frame (3) shows the results of the query with weighted features. It is apparent that most of the retrieved images in the top row have corresponded to the features of the first group - gender, hair length, and facial marks. It is observed that the images with the other group features were also retrieved but below the top range and the image with the least similar features to the query feature are retrieved and displayed in the final windows frame (3). The features in this example are specifically selected for illustration. Problems are also confronted in other features if they were assigned with unsuitable weights.

If two faces are described as similar in the ‘gender’, ‘race’ or ‘age’ features, they are semantically given more importance over other similar isolated face parts features, because the former represents an overall visually perception convening information that is captured and semantically interpreted by the human brain. For instance, when you inform that you met someone with concave nose and red hair color, the information that the listener’s mind would have perceived are the shape of the nose and the color of the hair. However, if you also inform that the person was a ‘Malay’ or ‘Chinese’, an overall visual perception would reached the listener’s brain and be interpreted regarding the race, face shape, nose shape, face color, hair type and color, the behavior, and the emotion. Consequently, when features like gender, race, and age are used for face annotation or retrieval, semantically all their corresponding information will be annotated or retrieved at the same time.

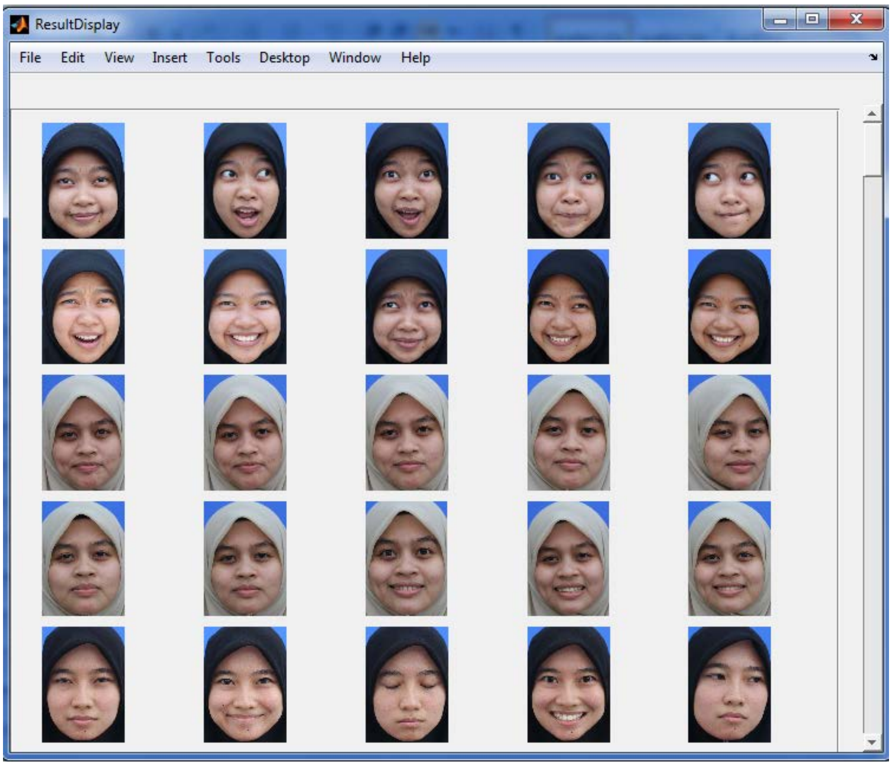

Figure 6.29: Frame (1).

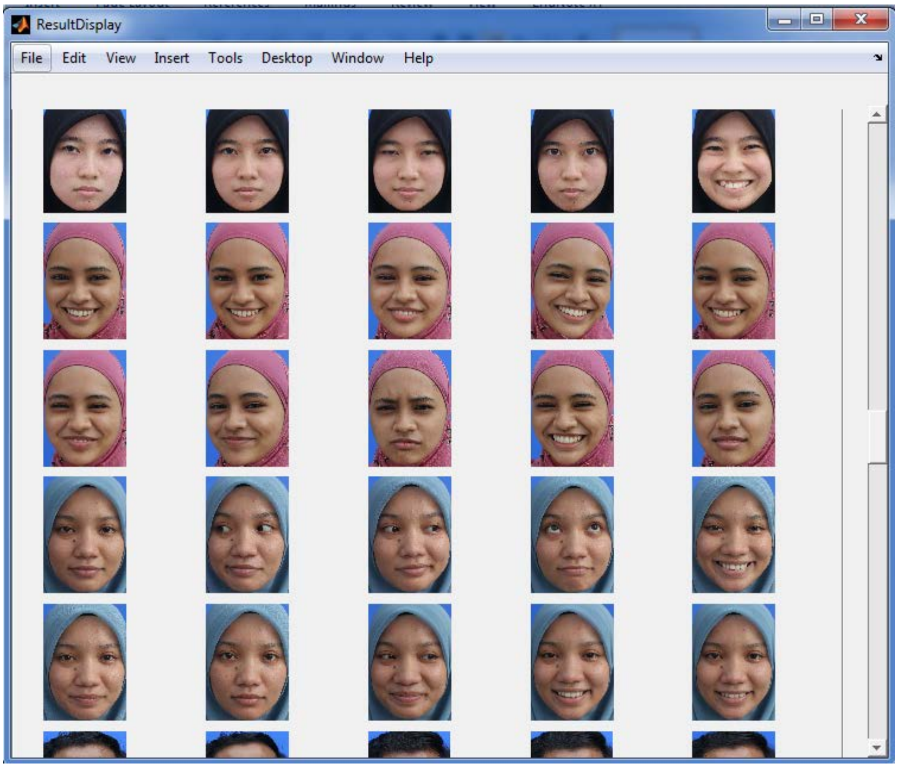

Figure 6.29: Frame (2).

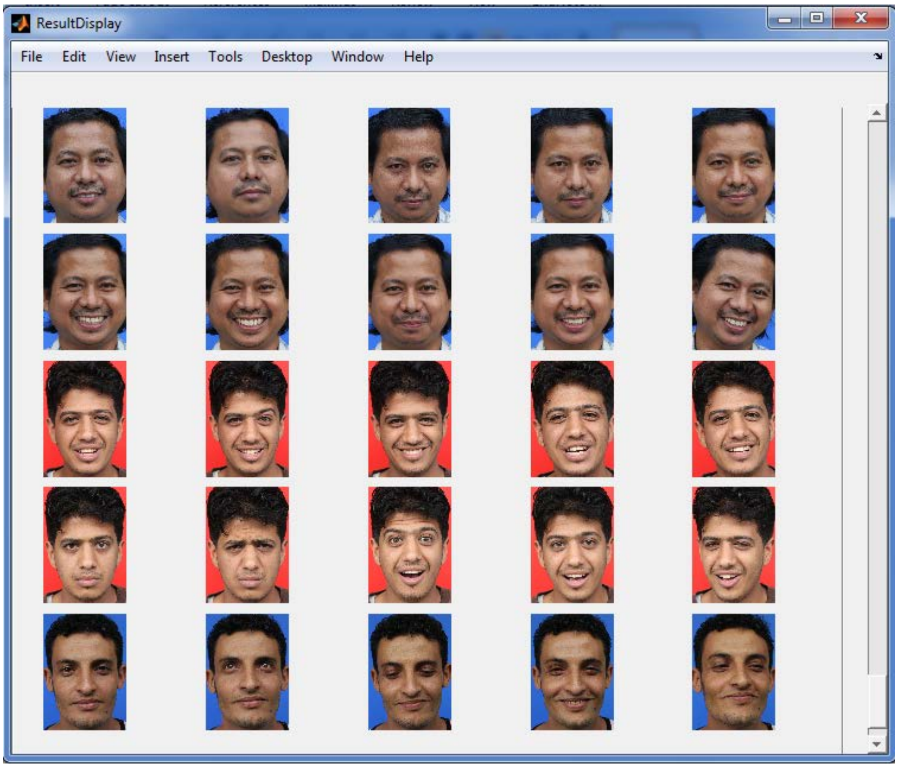

Figure 6.29: Frame (3).

Figure 6.29 Frame (1), Frame (2), Frame (3): System performance with weighted features.

Human evaluates and weights the face features naturally. If these features were extracted automatically, it would be impossible for the system to prioritize the importance of features without human intervention.

Comparison of the results of the above query example with those of the test experiments shows that the proposed method of feature weightings is obviously effective in the retrieval processing. This has not only increased retrieval accuracy but also enhanced the performance of the image retrieval system, thus meeting current user demands.

6.5 Integration of Heterogeneous Features Vectors

Retrieving the facial image based on low-level features has achieved good performance in some system runs and excellence performance in other system runs. Integrating the low-level features with high-level features is expected to improve the performance of the overall system.

Retrieval by semantic description is very important to help the user to express his/her query while simultaneously reduce the search space and direct the retrieval process to the desired images. However, the user will be confronted with problems when searching for identical faces with incomplete information. A semantic description retrieval system reduces the search space and displays the results to the maximum extent based on the descriptions given. What if the database is huge and the given information is insufficient? Many images that carry the same query information are then candidates to be displayed to the user. From this argument, the integration between the low-level feature and the high-level features is indispensable.

The advantages inherent in individual feature classes are integrated so that the retrieval more accurate is the current issue confronting researches. Directly combining the features may risk combining their respective weaknesses, which in turn will have a negative effect on the retrieval, constituting therefore a setback to the combination. To address this issue, two methods were proposed in this research - Euclidean distance approach and an innovative approach based on RBFN.

6.5.1 Euclidean Distance Metric Approach

The results below were generated from a facial image retrieval system using integrated features incorporated with a Euclidean distance (ED) method. The visual features were extracted in a similar manner following the methods proposed earlier except that the eigenfaces vector dimension was 10 and the color histogram was quantized based on the color space coordinates distribution of 4-Red, 4-Green and 4-Blue.

Tables 6.18 ; 6.19 and Figures 6.30 ; 6.31 show the results of the ED based facial image retrieval respectively for the ORL and local databases. The results shown were generated from the integrations of eigenfaces-color histogram, eigenfaces-semantic features, and color histogram-semantic features, as well as the integration of the 3 above mentioned features.

Considering the top 10 results of each integration of Table 6.18 of the ORL database, the system has achieved accuracies of 80.60%, 80.00%, 72.15% , and 83.10% in both the recall and precision methods. However, for the top 25 results, only the recall method has achieved high accuracies for the four integrations, respectively 89.80%, 90.45%, 89.65%, and 95.4%.

Table 6.18: Integration of features classes using Euclidean distance on the ORL database.

Features	Query Faces	Expected Faces	Top Faces	Retrieved Faces	Relevant Faces	Recall	Precision	F-score
Eigenfaces with Color	200	2000	10	2000	1612	0.806	0.806	0.806
	200	2000	16	3200	1715	0.8575	0.5359	0.6596
	200	2000	25	5000	1796	0.898	0.3592	0.5131
Eigenfaces with Semantic	200	2000	10	2000	1600	0.8	0.8	0.8
	200	2000	16	3200	1729	0.8645	0.5403	0.665
	200	2000	25	5000	1809	0.9045	0.3618	0.5169
Color with Semantic	200	2000	10	2000	1443	0.7215	0.7215	0.7215
	200	2000	16	3200	1641	0.8205	0.5128	0.6311
	200	2000	25	5000	1793	0.8965	0.3586	0.5123
Eigenfaces and Color with semantic	200	2000	10	2000	1662	0.831	0.831	0.831
	200	2000	16	3200	1817	0.9085	0.5678	0.6988
	200	2000	25	5000	1908	0.954	0.3816	0.5451

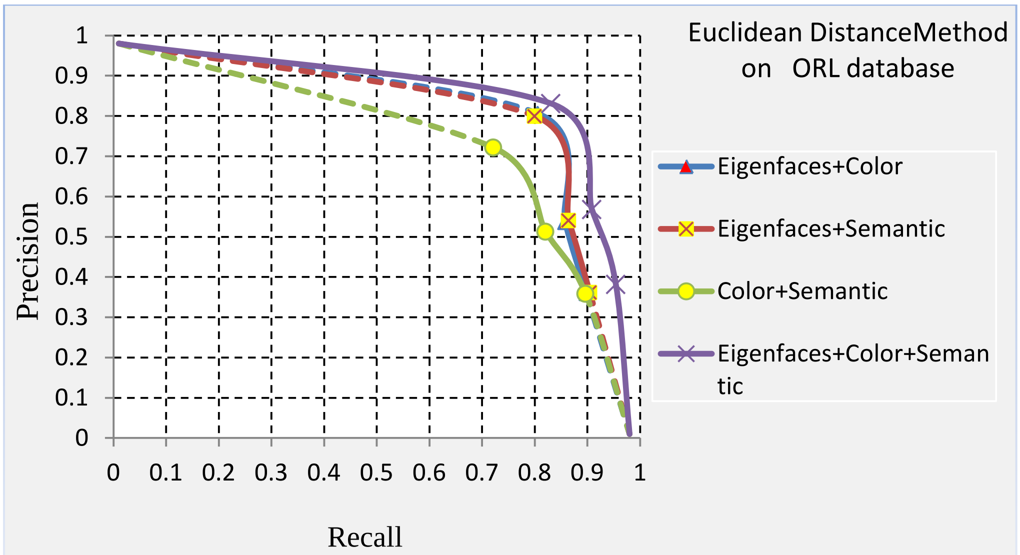

Figure 6.30: System performance with integration of features classes using Euclidean distance on ORL database.

The results on the local database (Table 6.19) also show high accuracies for the top 10 results in both the recall and precision methods. The respective accuracies for the four integrations are 88.97%, 83.52%, 90.09%, and 92.75%. High accuracies are also observed in the recall method for the top 25 results - 93.52%, 91.97%, 94.69%, and 95.51%.

Table 6.19: Integration of features classes using Euclidean distance on local database.

Features	Query Faces	Expected Faces	Top Faces	Retrieved Faces	Relevant Faces	Recall	Precision	F-score
Eigenfaces with Color	750	7500	10	7500	6673	0.8897	0.8897	0.8897
	750	7500	16	12000	6900	0.92	0.575	0.7077
	750	7500	25	18750	7014	0.9352	0.3741	0.5344
Eigenfaces with Semantic	750	7500	10	7500	6264	0.8352	0.8352	0.8352
	750	7500	16	12000	6665	0.8887	0.5554	0.6836
	750	7500	25	18750	6898	0.9197	0.3679	0.5256
Color with Semantic	750	7500	10	7500	6757	0.9009	0.9009	0.9009
	750	7500	16	12000	6975	0.93	0.5813	0.7154
	750	7500	25	18750	7102	0.9469	0.3788	0.5411
Eigenfaces and Color with semantic	750	7500	10	7500	6956	0.9275	0.9275	0.9275
	750	7500	16	12000	7085	0.9447	0.5904	0.7267
	750	7500	25	18750	7163	0.9551	0.382	0.5457

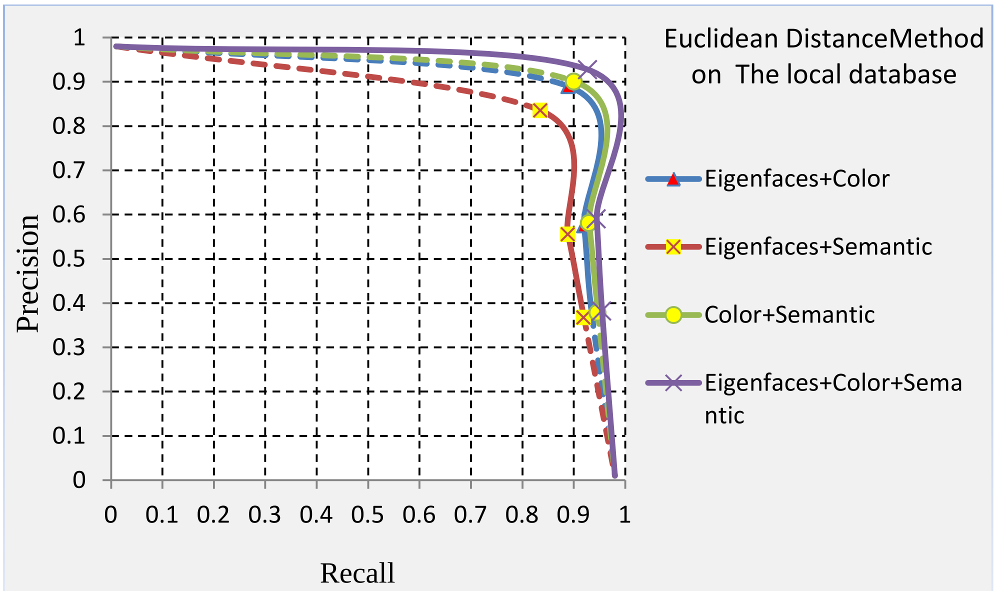

Figure 6.31: System performance with integration of features classes using Euclidean distance on local database.

Although the results of the low-level features and high-level features integration using Euclidean distance were more than 80% of the accuracies which can be considered as satisfactory, nevertheless it did not meet the expectation of the researchers involved. The advantage of each feature class should be integrated in a more effective way, which would enhance retrieval accuracy. Combining them directly may lead to actually integrating their weaknesses, resulting therefore in lower accuracies than expected.

6.5.2 The Proposed Approach of RBFN

Our proposed method was based on using the RBFN for the task of integrating the varied features. In this method, weights are generally assigned to each class of information extracted from an image and an overall similarity is computed. Images are then ranked based on this similarity computation.

6.5.2.1 Proposed Method Training

The training stage results of the RBFN are shown in Figure 6.32 and 6.33, which are essentially the sum squared errors (SSE) of all training vectors in all cycles of training on the ORL and local databases respectively. The SSE of each vector was computed based on its output with the other vectors to the target outputs during each cycle, which encompassed all training vectors. Each vector of the training vectors was feed to the network center as the query vector. The remaining training vectors were input to the networks as the training vectors. Their similarities were measured and the SSE computed.

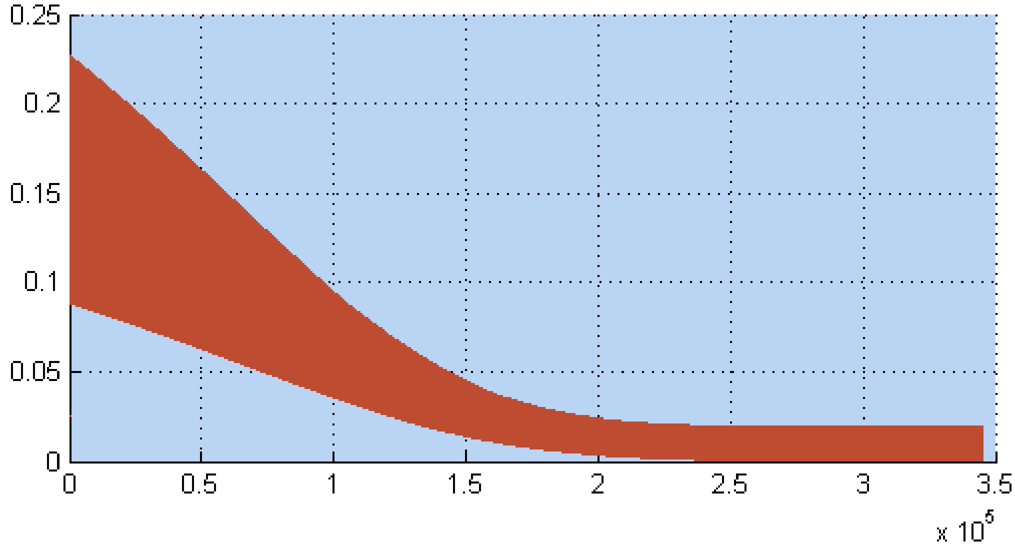

Figure 6.32 : Sum squared errors of all input training vectors on the ORL database.

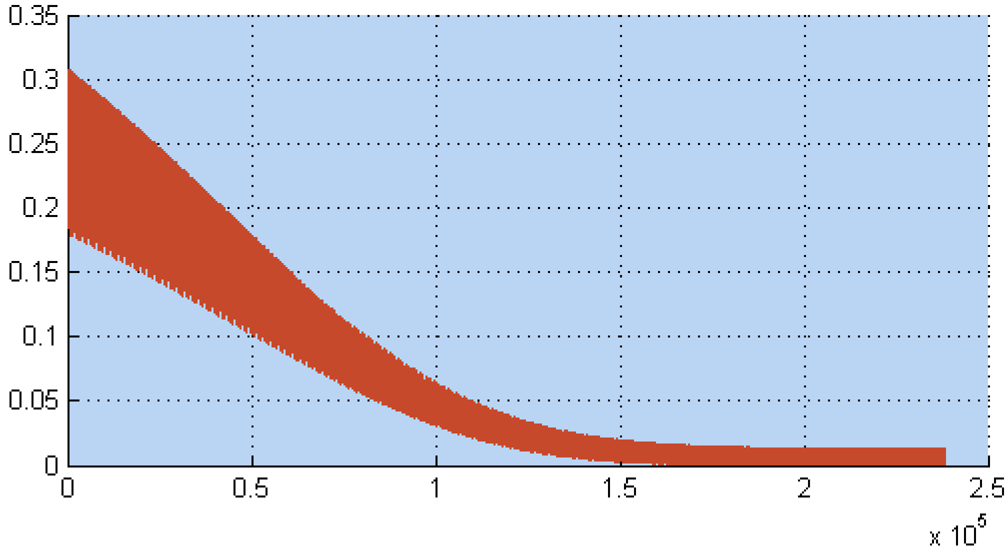

Figure 6.33: Sum squared errors of all input training vectors on the local database.

Figures 6.34 and 6.35 show the SSEs of the last training cycle on the ORL and local database respectively. It is observed that most of the SSE of the vectors are approaching or equal to zero.

SSE measures the discrepancy between the target data and the neural networks model. A small SSE indicates a tight fit of the model to the data. The SSE of each vector is then used to adjust the networks weights.

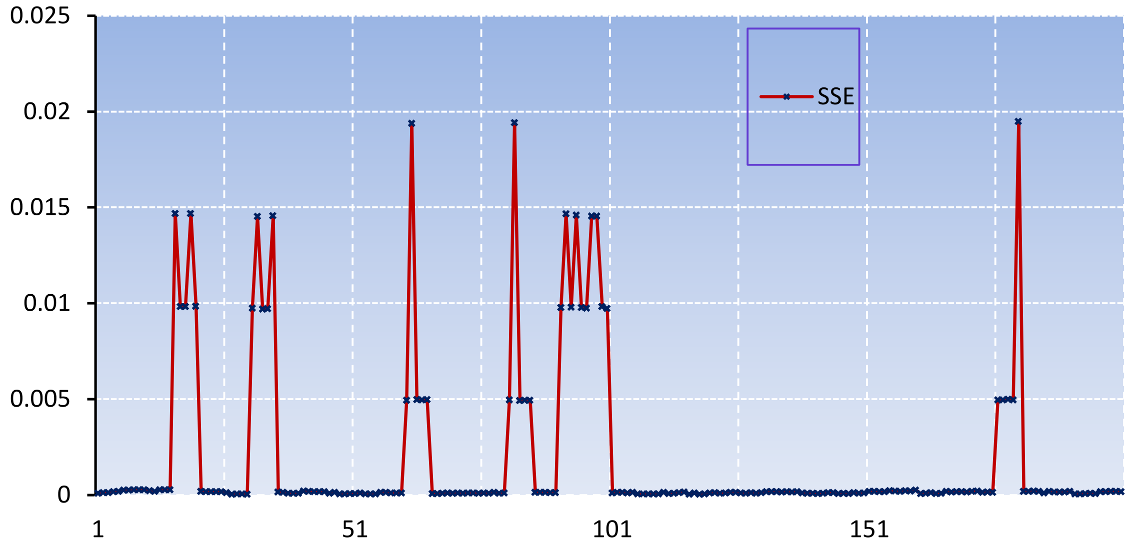

Figure 6.34: Sum squared errors of the last cycle of the network training on the ORL database.

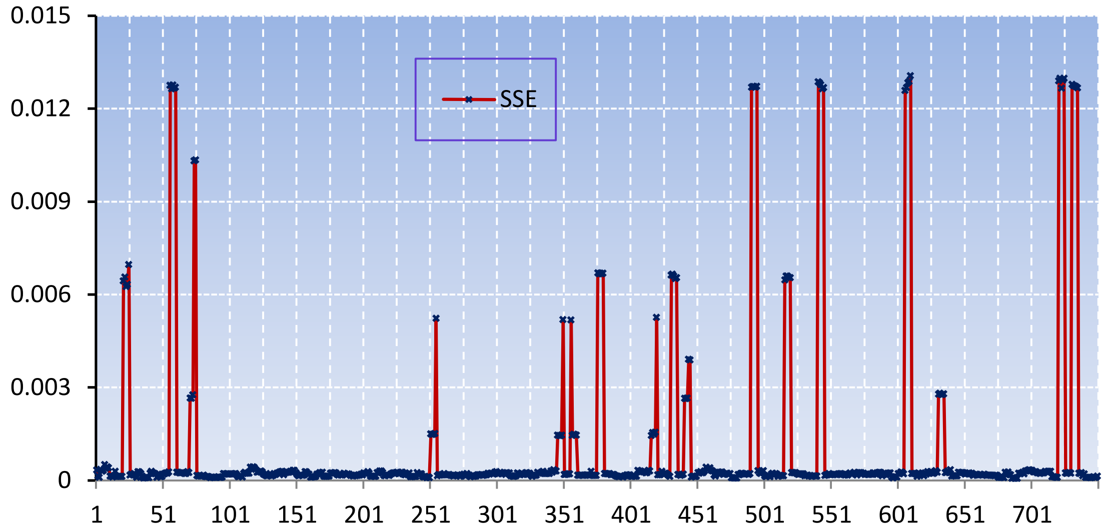

Figure 6.35: Sum squared errors of the last cycle of the network training on the Local database.

Figures 6.36; 6.37 show the mean squared errors (MSE) of the networks training on the ORL and local databases respectively. The MSE is computed to monitor and measure the performance of the network training. It is the most common measure of network accuracy during training. It is calculated by the summation of the sum-squared errors of each training cycle vector and the summation is then divided by the number of vectors in the cycle. The network training should be stopped, when the MSE is less than the network error target. In our research, the error target was 0.005.

It is observed that network learning with the local database is faster than that of the ORL. This is attributed to the variety, size, and color of the local database images.

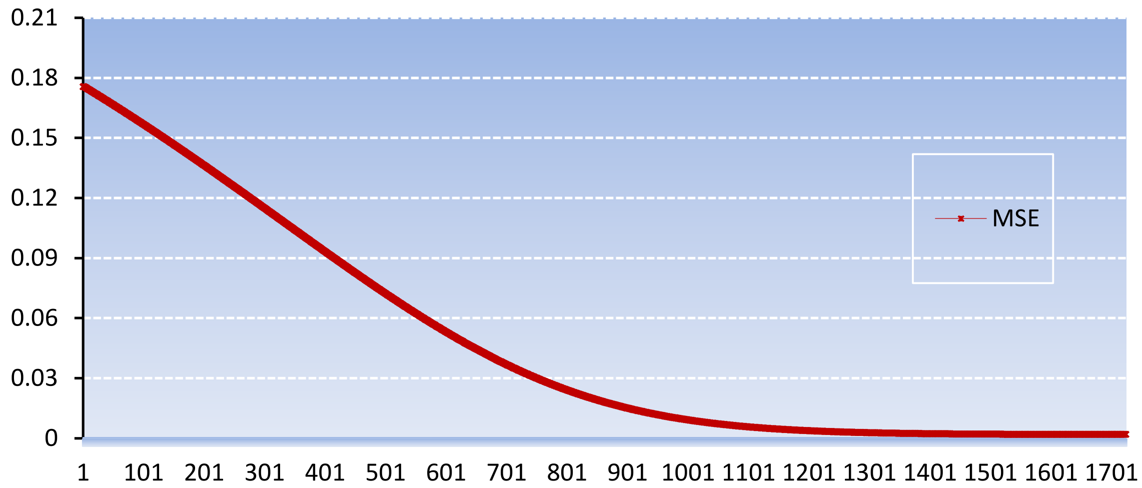

Figure 6.36: Mean squared error of the network training
on the ORL database.

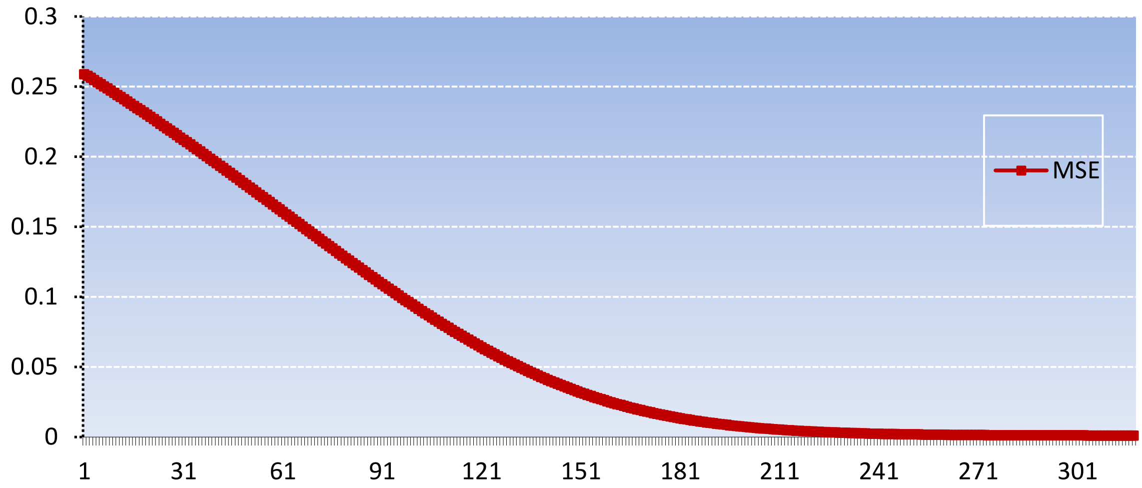

Figure 6.37: Mean Squared Error of the network training on the Local database.

The two databases were used for training the networks to study the network response using the proposed method. The network weights acquired through the ORL database showed results more than 90% of the accuracies which can be considered as excellent performance with testing image from the ORL database and showed results less than 80% of the accuracies which can be considered as good performance with the testing image from the local database. On the other hand, the network weights acquired through the local database showed results more than 90% of the accuracies which can be considered as excellent performance with the testing image from both local and ORL databases. The final testing of the proposed method was then based on the weight acquisition from the network training through the local database.

6.5.2.2 Proposed Method Testing

Tables 6.20 ; 6.21 and Figures 6.38; 6.39 show the experimental results of the facial image retrieval respectively for the ORL and local databases. The results shown were generated from the integration of eigenfaces-color histogram, eigenfaces-semantic features, color histogram-semantic features, as well as the integration of the 3 above mentioned features.

Considering the top 10 results of each integration of Table 6.20 - ORL database, the system has achieved accuracies of 84%, 95.05%, 93.9%, and 97.85% in both the recall and precision methods. However, for the top 25 results, only the recall method has achieved high accuracies for the four integrations, respectively 93.25%, 95.65%, 96.1%, and 99.65%.

Table 6.20: Integration of features classes using proposed method through RBFN on ORL database.

Features	Query Faces	Expected Faces	Top Faces	Retrieved Faces	Relevant Faces	Recall	Precision	F-score
Eigenfaces with Color	200	2000	10	2000	1680	0.84	0.84	0.84
	200	2000	16	3200	1800	0.9	0.5625	0.6923
	200	2000	25	5000	1865	0.9325	0.373	0.5329
Eigenfaces with Semantic	200	2000	10	2000	1901	0.9505	0.9505	0.9505
	200	2000	16	3200	1905	0.9525	0.5953	0.7327
	200	2000	25	5000	1913	0.9565	0.3826	0.5466
Color with Semantic	200	2000	10	2000	1878	0.939	0.939	0.939
	200	2000	16	3200	1905	0.9525	0.5953	0.7327
	200	2000	25	5000	1922	0.961	0.3844	0.5491
Eigenfaces and Color with semantic	200	2000	10	2000	1957	0.9785	0.9785	0.9785
	200	2000	16	3200	1987	0.9935	0.6209	0.7642
	200	2000	25	5000	1993	0.9965	0.3986	0.5694

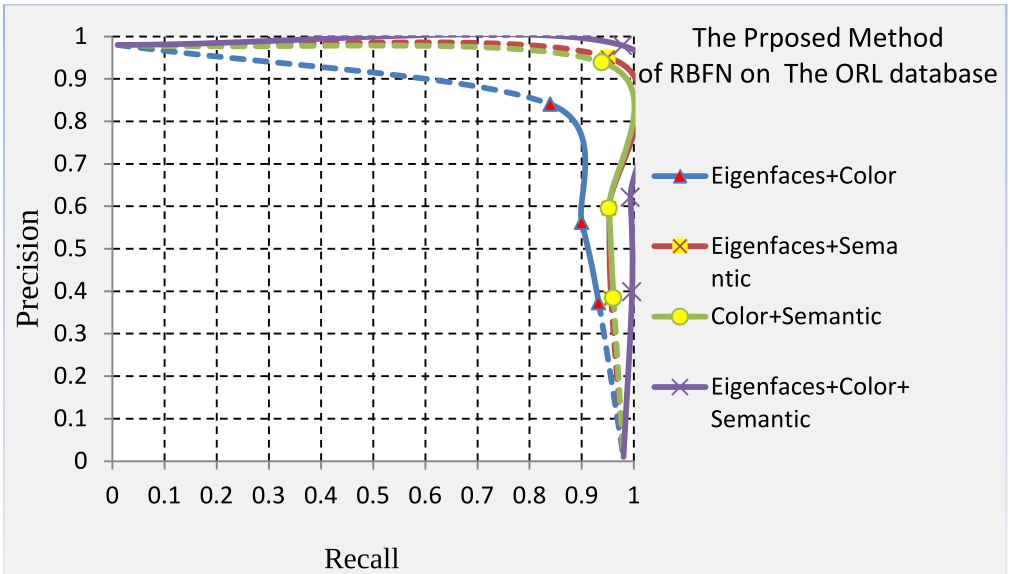

Figure 6.38: Facial image retrieval performance using proposed method through RBFN on ORL database.

The results on the local database (Table 6.21) show higher accuracies for the top 10 results in both the recall and precision methods. The respective accuracies for the four integrations are 92.41%, 95.36%, 96.37%, and 99.39%. High accuracies are observed in the recall method for the top 25 results – 97.75%, 96.75%, 96.91%, and 99.99%.

Table 6.21: Integration of features classes using proposed method through RBFN on local database.

Features	Query Faces	Expected Faces	Top Faces	Retrieved Faces	Relevant Faces	Recall	Precision	F-score
Eigenfaces with Color	750	7500	10	7500	6931	0.9241	0.9241	0.9241
	750	7500	16	12000	7215	0.962	0.6013	0.74
	750	7500	25	18750	7331	0.9775	0.391	0.5586
Eigenfaces with Semantic	750	7500	10	7500	7152	0.9536	0.9536	0.9536
	750	7500	16	12000	7212	0.9616	0.601	0.7397
	750	7500	25	18750	7256	0.9675	0.387	0.5529
Color with Semantic	750	7500	10	7500	7228	0.9637	0.9637	0.9637
	750	7500	16	12000	7250	0.9667	0.6042	0.7436
	750	7500	25	18750	7268	0.9691	0.3876	0.5537
Eigenfaces ,Color with semantic	750	7500	10	7500	7454	0.9939	0.9939	0.9939
	750	7500	16	12000	7493	0.9991	0.6244	0.7685
	750	7500	25	18750	7499	0.9999	0.3999	0.5713

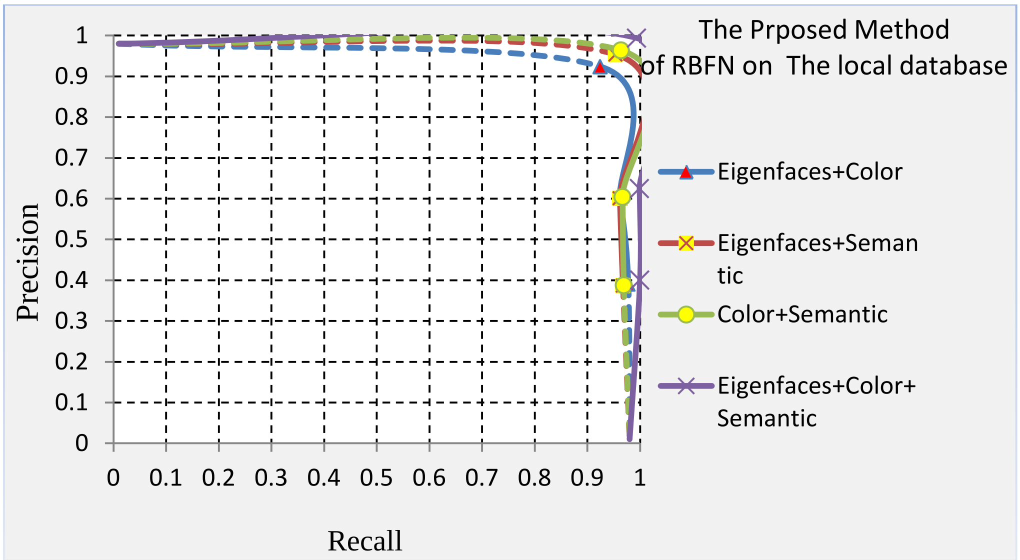

Figure 6.39: Facial image retrieval performance using proposed method through RBFN on local database.

In the next section, the example taken to illustrate results of visual experiments was chosen randomly. It shows clearly the improvement of the facial image retrieval method based on the integration of visual and semantic features using the proposed RBFN method, over both the facial image retrieval method based on visual features and the method integrating visual and semantic features using Euclidean distance algorithm.
Take for instance, (i) the semantic query vectors have included semantic features such as gender 'Female', race 'Middle Eastern', and face-shape 'Long` and (ii) the visual query is taken from Figure 6.40**.**

As shown in Figure 6.41, the recall method of performance gives an accuracy of 60% within the top 10 cut off level based on eigenfaces features of 10-dimention vectors. Figure 6.42 shows an accuracy of 70% within the top 10 cut-off level based on color histogram features comprising 4-Red, 4-Green, and 4-Blue color space distribution. Integration of the two features has resulted in a higher accuracy of 80% within the top 10 cut-off level as similarity measured by Euclidean distance measurement (Figure 6.43). Facial image retrieval based on the integration of eigenfaces, color histogram and, semantic features using Euclidean distance has also attained 80% accuracy within the top 10 cut-off level as depicted in Figure 6.44.

Comparing the results of Figure 6.43 and Figure 6.44 it is evident that there is no improvement in performance based on features integration using the Euclidean distance method. This is attributed to the dominance of visual features over the semantic features, resulting in no significant impact of the latter in improving the accuracy within the top 10 cut-off level.

However, when the proposed method of RBFN was used on the same visual-semantic query (Figure 6.40) for facial image retrieval based on (i) the integration of the eigenfaces and color histogram and (ii) the integration of the three features - eigenfaces , color histogram, and semantic features, the accuracies achieved are respectively 80% and 100% within the top 10 cut-off level. The detail results are given in Figures 6.45 and 6.46 respectively. With these achieved results, it is apparent that the improvement is significant using the proposed RBFN method.

Table 6.22 summarizes the performances of the Euclidean distance method and the proposed method through RBFN against the integrations of (i) eigenfaces-color and (ii) eigenfaces-color-semantic features. It is apparent that there is significant improvement in the accuracies using the new proposed method, where the majority of the relevant images were returned to the top ten results.

Table 6.22: Results comparison of features-classes integration using the proposed method through RBFN and Euclidean distance on ORL and local database (Recall on top 10).

Features	**Euclidean Distance**		**Proposed Method**	
	ORL	***Local***	***ORL***	***Local***
Eigenfaces with Color	80.60%	89.51%	84.00%	92.41%
Eigenfaces and Color with semantic	83.10%	93.71%	97.85%	99.39%

With our proposed method through RBFN, a certain weight was acquired during the training stage for each feature class of the vector features. In this manner, every weight has maintained the existence of its associated vector properties. Integrating the feature classes through their respective weights has strengthened digital coverage in each class of properties enhancing discrimination and segregation between two different faces. In addition, each particular class covers the weaknesses inherent in the other classes, thus resulting in higher performance in the classification and retrieval processes.

The integration of the semantic based facial image retrieval technique and the visual features based technique using appropriate integration methods has achieved the best results. The benefits of the individual techniques were essentially merged and enhanced. By integration the two techniques, the best results benefits were combined.

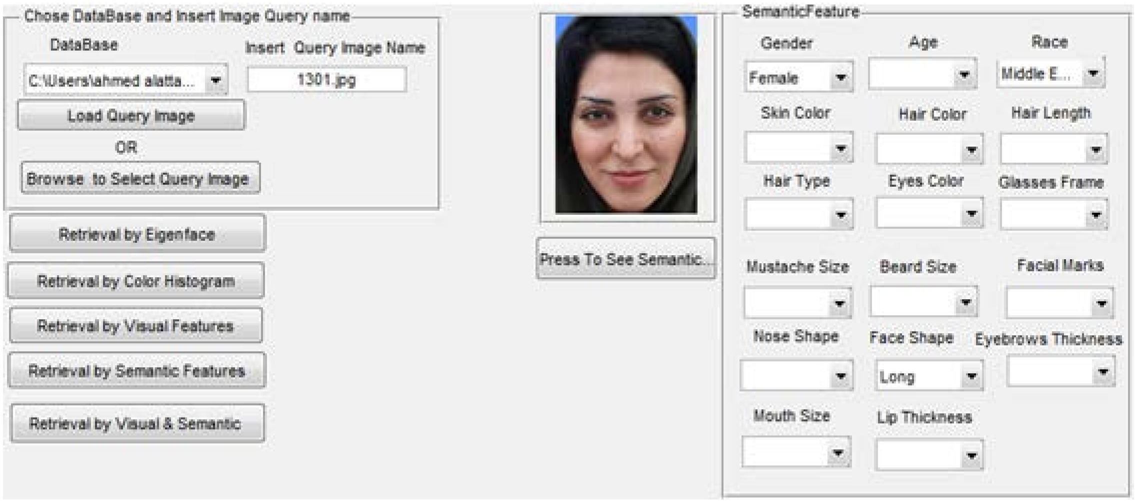

Figure 6.40: Example of facial image retrieval based on semantic and visual query vector.

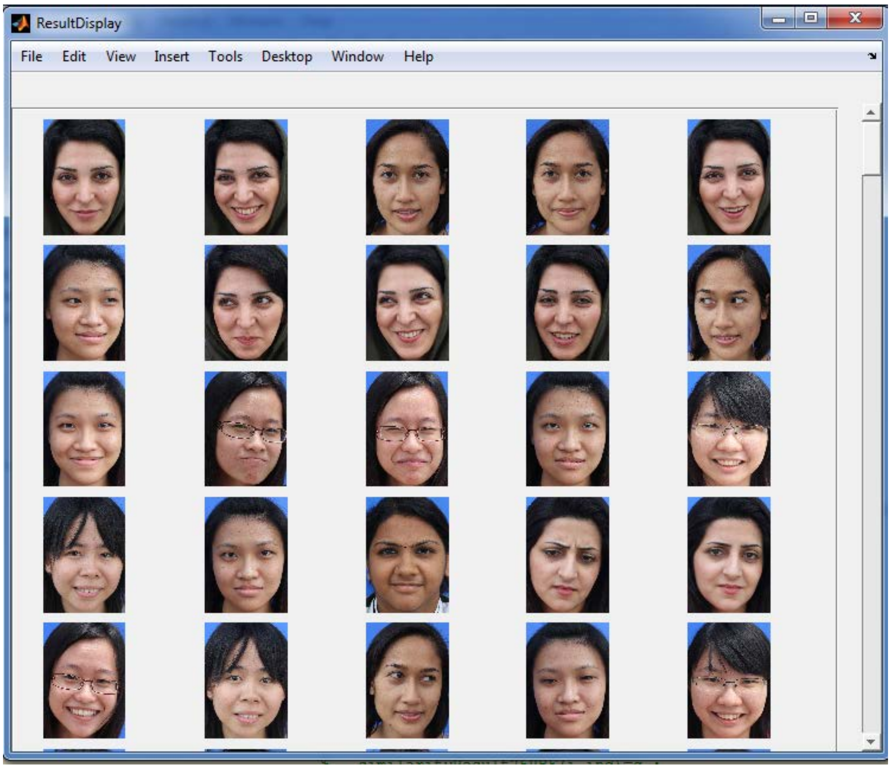

Figure 6.41: Facial image retrieval based on eigenfaces of 10-dimention vector.

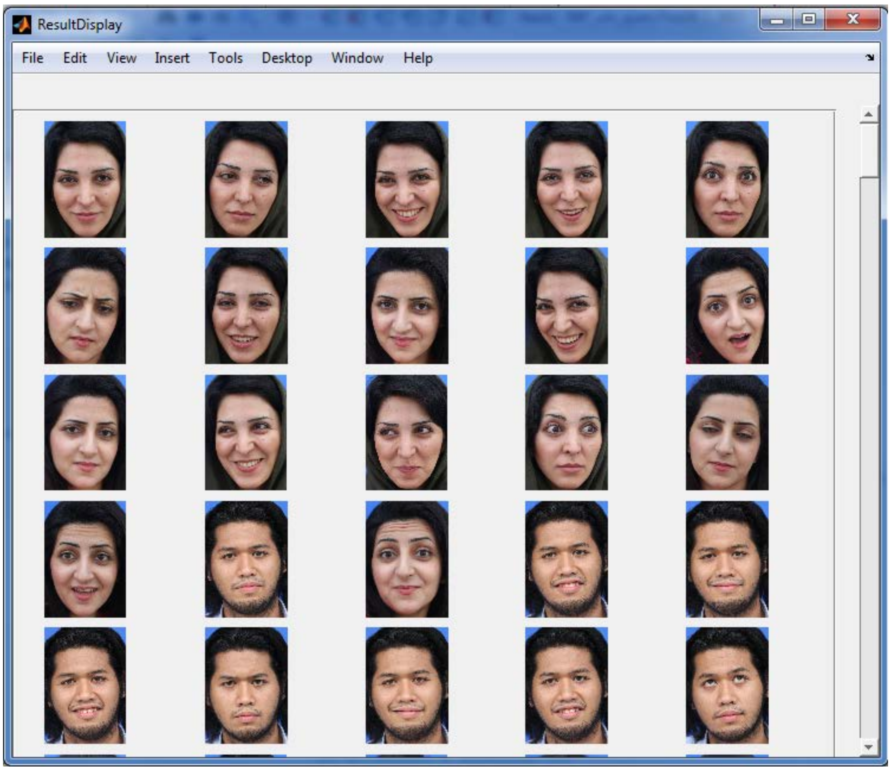

Figure 6.42: Facial image retrieval based on color histogram of (4-Red, 4-Green, 4-Blue) color space distribution.

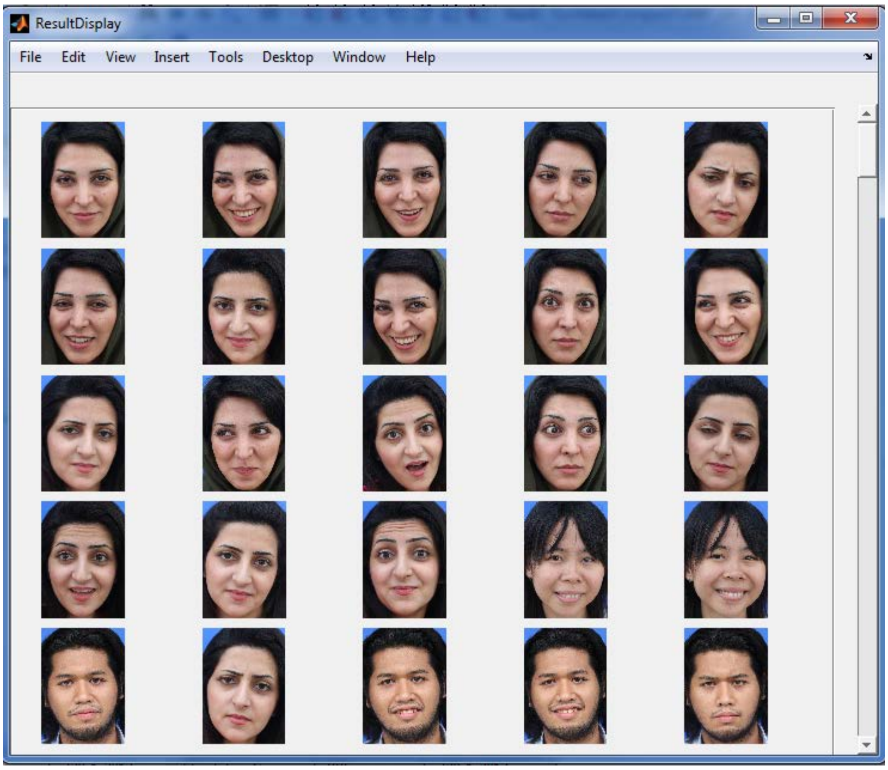

Figure 6.43: Facial image retrieval based on integration of color and eigenfaces using Euclidean distance method.

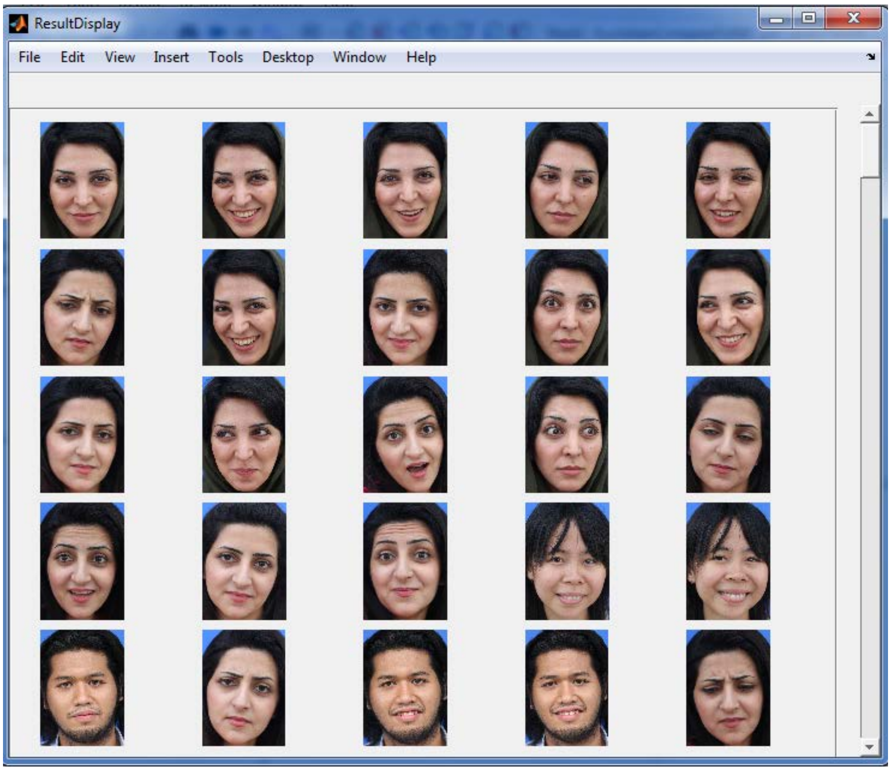

Figure 6.44: Facial image retrieval based on integration of eigenfaces, color and semantic features using Euclidean distance method.

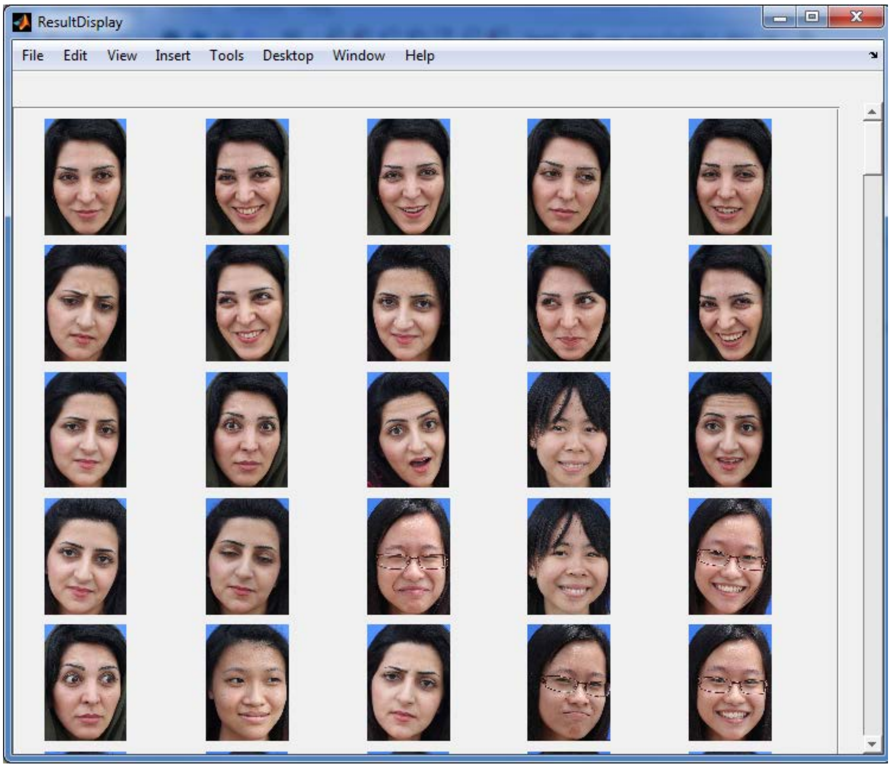

Figure 6.45: Facial image retrieval based on integration of eigenfaces and color using the proposed method through RBFN.

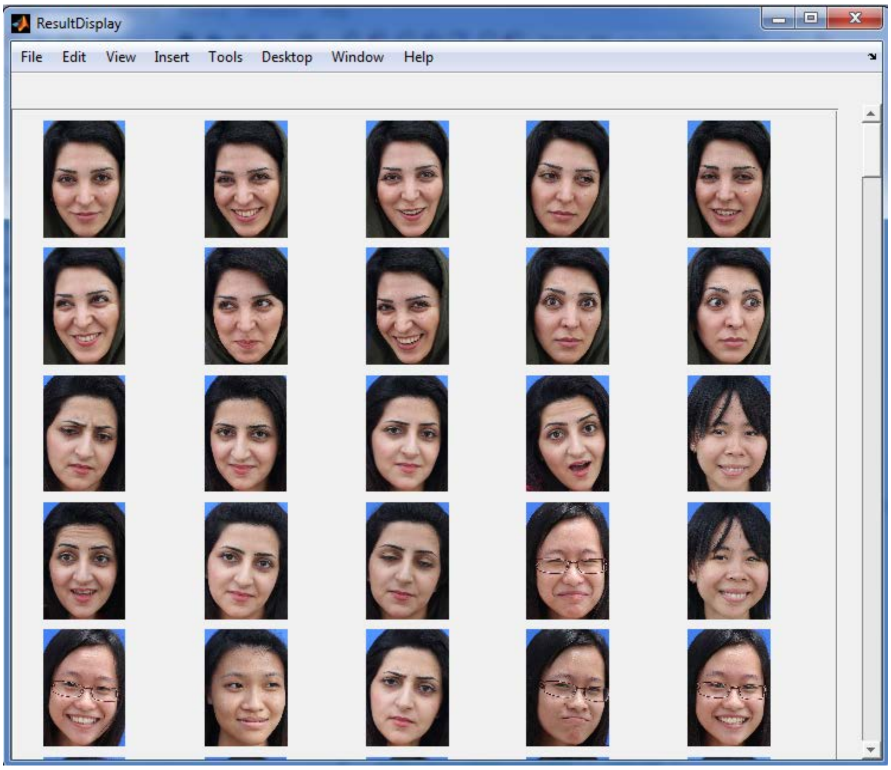

Figure 6.46: Facial image retrieval based on integration of eigenfaces, color and semantic features using the proposed method through RBFN.

CHAPTER 7
CONCLUSION AND FUTURE WORKS

7.1 Conclusion

7.1.1 Summary of Thesis Achievements

In this research, a new method for human facial image retrieval model was proposed based on the integration of face recognition techniques and the traditional content-based image retrieval technique with the human face semantic features. Eigenfaces features and color histogram features were used as low-level features, whilst the proposed method is used with different visual features.

A prototype was built to facilitate the retrieval performance testing for the user and to further verify the results of the algorithm investigation.

Numerous experiments have been conducted to assess and evaluate the proposed methods of facial image retrieval based on the selected and represented semantic features. Two databases were used; the Olivetti Research Lab (ORL) database, which consists of 400 facial images. The second database is a local database consisting of 1500 local facial images of 150 participants from the University of Malaya (UM) in Kuala Lumpur, Malaysia. Precision and recall methods were applied to measure the performance efficiency of the retrieval methods. The results were also analyzed using precision versus recall graphs.

Color space models RGB, HSV, and HSI were used to investigate in which color space the facial image retrieval technique shows the best performance. Experimental results on the local database showed that eigenfaces-based facial image retrieval in the HSV model yields the best accuracy among the other models and the color histogram-based facial image retrieval in RGB color space showed the best performance among the other models. With the ORL database (grayscale images), both eigenfaces and color histogram features have achieved the best performance of the retrieval in the RGB color space model. This is because the Red, Green, and Blue channels in a grayscale image would contain the same information, if converted to HSV or HIS color spaces.

A proposed features extraction method based on segmented facial images was introduced in order to improve the performance of the facial image retrieval. Three methods were experimented based on (i) the entire facial image – (ii) three segments of the facial image partitioned at two levels: eyes and mouth; (iii) and four segments including the center portion of the facial image. The first method is a traditional method, while the others are the subject of this research. All experiments were conducted using eigenfaces and color histogram features, separately as well as in combination. The results of the experiments show that the extraction of features based on the 3- segments method has enhanced the performance of the facial image retrieval technique compared to the traditional and the 4-segments methods. With the proposed method of facial segmentation, querying is simply done through the global image using the global descriptors of the whole image and the processing in the system to extract the features vectors are based on local descriptors, which includes 3 local descriptors of the face. In the proposed method, features are extracted semantically where features can be extracted and compared to other faces on the same facial localities.

Between the two proposed methods of face segmentation, namely, the three segmentation of the face and the three segmentation of the face plus face center, the former have achieved a better system performance. This suggests that adding the face center to the features extracted from the three segments could have resulted in noise, which in turn has degraded the system performance, leading to significantly lower accuracy.

Eigenfaces features have the capability to provide the significant features for face retrieval. The advantages of these features are that processing is fast and no heavy storage of data is required. However, there exist factors originating from the facial image itself, which could affect the performance of the eigenfaces processing. These include the facial hair, skin scarring, and face multiple view. Experimental results also showed that increasing the dimension of the vectors to more than 20 will not necessarily result in higher performance. Moreover, the existence of some trivial information may be considered as noise and will degrade the system performance, especially when the eigenfaces are combined with other features.

With color histogram features, it was noted that choosing the distribution values of the color space coordinates to represent the bines size during the quantization process, has influenced the color based facial image retrieval results. The accuracy is based on the distribution of the colors of the image. While this influence is clear on color images database, the distribution of the color space coordinate has no influence on the gray level image database. This is because the three channels of the gray images carry the same information.

Semantically, the main weakness of the color histogram method is it does not necessarily allow the relevant images as seen by machines to be the same as those relevant images visualized by the human eyes. Applying the proposed method of facial image segmentation has reduced this weakness.

Human face semantic features were selected and represented in the proposed method of facial image retrieval based on a case study. The semantic features were annotated to each facial image, enabling the user to state the query through natural language descriptions. Retrieval by semantic features based on verbal description helps the user to express the query, reduce the search space, and direct the retrieval towards reducing the semantic gap. Three methods are discussed, namely, (i) the traditional method based on pruning the image from the search space, (ii) the features pruning method based on pruning the non-matching features from the image description, and (iii) the proposed method based on probabilistic approach. The proposed method reduced the side effects of the subjectivity of the human perception problems in facial image retrieval. Experimental results showed excellent improvement in the accuracy based on the proposed method as compared to the other methods in (i) and (ii).

Several experiments of the facial image retrieval were carried out based on the integration of (i) eigenfaces and color histogram, (ii) eigenfaces and semantic features, (ii) color histogram and semantic features, and (iv) eigenfaces, color histogram and semantic features.

Compared to the low-level feature facial image retrieval system, the results of this research have reflected a significant improvement in the facial image retrieval performance using the integration of low-level and proposed semantic features vectors.

Two methods were used to integrate the different classes of facial image features.

(i) Euclidean distance, The experimental results of the integration on facial retrieval were more than 80% of the accuracies which can be considered as satisfactory, nevertheless it did not meet the expectation of the researchers where the advantage of each feature class should be integrated in a more effective way, which would enhance the retrieval accuracy while directly combining them may lead to actually integrating their weaknesses, resulting therefore in lower accuracies than expected.

(ii) The experimental results of a new proposed method based on RBFN, showed that using the proposed method for the integration of the semantic based facial image retrieval technique and the visual features based technique has achieved the best results as compared to the Euclidean method. This may be because the benefits of the individual techniques were essentially merged and enhanced. In addition, each particular class covers the weaknesses inherent in the other classes, thus resulting in higher performance in the classification and retrieval processes.

The core contributions of our research can be summarized as following:

1) The proposed method that was followed for semantic features selection, weighting, and representation to be described by the user direct as a verbal description using the natural language concepts, however, in the machine, these descriptions are represented symbolically and numerically and are integrated with the low-level features for facial image searching and retrieval in an accurate way. The aim was to develop a model that links the high-level query requirement and the low-level facial features of the human facial image towards reducing the semantic gap between them, and enabling the system to meet human natural tendencies and needs in the description and retrieval of facial images. The proposed method based on the interactive system matches the verbal query of the user to the corresponding represented semantics features of the image in the database.

2) The proposed method of using the probabilistic approach on the verbal descriptions of the represented and weighted semantic-features. The aim was to avoid the problem of subjectivity, imprecision, and/or uncertainty involved in the specified semantic-attributes and to improve the differences observed based on humans 'perception and the viewpoint that may appear during image annotation and/or query process. The proposed method is based on the concept that each facial image-features gains probability according to its distance from the description given by the user and the faces in the database are ranked based on these probabilities; the top images are displayed to the user as retrieved images.

3) The proposed method of the facial image segmentation and extraction. The aim was to improve the accuracy of the facial image retrieval performance based on visual features. The idea is based on the fact that every sub facial image has its spatial information of orientation and specific scale relevant to this sub-image. Combining the features vectors of each sub-image which were independently extracted, produced more robust features vector. The results of the research recommend the facial image to be segmented into three and four partitions based on human eyes and mouth level and the ratio of their height to the face height.

4) The proposed method through RBFN for similarity measure and classification problems of facial image retrieval. The aim was to address the problem of combining the heterogeneous attributes of visual features and semantic features using efficient and accurate method for improving the performance accuracy of the facial image retrieval and enabling the user to specify his/her query through the query by example together with the natural language descriptions. These features were extracted through different methods, and characterized by different distribution, importance, and as well as dominance of one over the other. Combining the features using the proposed method combines their benefits, produces a unique value of similarity between the query vectors and the database vectors of these features, and merged ranking for each image in the database, where the same image receives different rankings from the different features. The proposed method based on a learning similarity metric through the RBFN machine learning technique.

7.2 Futures Works

The future enhancement of the current system is to integrate the proposed system with the relevance feedback technique that could be used to adjust the annotations of the facial images in the database.

Other research options that could be done in the future are:

- To develop a method that will improve the correlation between human and machine perceptions of facial images, especially pertaining to the ways of measuring similarities between images.

- Eigenfaces and color histogram features are effective and useful methods for facial image retrieval and their contribution is evident in our research. While the color histogram features are used as a general visual content, the Eigenfaces are used as a domain specific visual content. Both features possess sufficient image information, are easily computed, and facilitate large image collection and rapid retrieval. Although the efficiency of these used features is more than 90% of the accuracies, which can be considered as satisfactory, nevertheless, investigation and development of new visual features for integrating with the current visual and semantic features for facial image retrieval system should be one of our future options.

- To investigate the semantic feature keywords for describing the human body such as body shape, weight, height, length of legs and arms, and gait properties. The system can then be extended to work in video applications such as monitoring mechanism in airports. In this context, the system should be able to automatically capture the descriptions of human face and/or body under scanning to retrieve similar images.

We conclude our research with the assertion that the proposed method of semantic-content based facial image retrieval (SCBFIR), based on the integration of verbal descriptions of the human face with visual features achieved excellent results in the retrieval of facial images compared with the content-based facial image retrieval technique that is based on retrieval by image content. The results of the experiments show that, the content-based facial image retrieval technique achieves 80.60% and 89.51% accuracy, while the SCBFIR achieves 97.85% and 99.39% accuracy for the ORL and local database respectively within the top 10 retrieved facial images. Combining the two methods of query by description and query by image example automatically improves the accuracy of the retrieval process, reduces the required time to find the desired faces, and reduces the semantic gaps between the high-level query requirements represented by the user's verbal descriptions and the low-level facial features represented by the image content features.

The proposed method of semantic-content based facial image retrieval could be used in law enforcement applications, where the verbal description of the witness is used to retrieve the similar facial images to the suspect's face from the criminal's mug shot database.

REFERENCES

Agarwal, M., Jain, N., Kumar, M. M., & Agrawal, H. (2010). Face recognition using eigenfaces and artificial neural network. *International Journal of Computer Theory and Engineering, 2*(4).

Aiping, C., Lian, P., Yaobin, T., & Ning, N. (2010). *Face detection technology based on skin color segmentation and template matching.* Paper presented at the Second International Workshop on Education Technology and Computer Science (ETCS 2010), Wuhan, China.

Alfalou, A., Brosseau, C., Katz, P., & Alam, M. (2012). Decision optimization for face recognition based on an alternate correlation plane quantification metric. *Optics Letters 37*(9), 1562-1564.

Alsamman, A., & Alam, M. S. (2002). Invariant face recognition using a neural network based on the fringe-adjusted joint transform correlator. *Optical Engineering 41*(11), 2768-2777.

Alsamman, A., & Alam, M. S. (2005). Comparative study of face recognition techniques that use joint transform correlation and principal component analysis. *Applied Optics 44*(5), 688-692.

Araki, Y., Shimada, N., & Shirai, Y. (2002). *Detection of faces of various directions in complex backgrounds.* Paper presented at the 16th International Conference on Pattern Recognition.

Asadi, S., Rao, D., & Saikrishna, V. (2010). A comparative study of face recognition with principal component analysis and cross-correlation technique. *International Journal of Computer Applications 10*(8), 17-21.

Atalay , I. (1996). *Face recognition using eigenfaces* Istanbul Technical University, Istanbul.

Bagdi, V., Patil, S., & Dharaskar, R. (2013). Analysis Of Various CBIR Techniques for Retriving Forensically Viable Information From The Images. *International Journal of Engineering, 2*(1).

Barnard, K., & Shirahatti, N. V. (2003). *A method for comparing content based image retrieval methods.* Paper presented at the Proceedings of the SPIE 2003, Santa Clara, CA, USA.

Bau-Cheng, S., Chu-Song, C., & Hui-Huang, H. (2008). *Face image retrieval by using haar features.* Paper presented at the 19th International Conference on Pattern Recognition (ICPR 2008), Tampa, Florida.

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools,* Retrieved March, 2011, from http://www.drdobbs.com/open-source/184404319 (2000).

Brunelli, R., & Mich, O. (1996). SpotIt : An Interactive Identikit System. *Graphical Models and Image Processing, 58*(5), 399-404.

Brunelli, R., & Poggio, T. (1993). Face recognition: Features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence 15*(10), 1042-1052.

Caldwell, C., & Johnston, V. (1997). *Tracking a criminal suspect through face-space with a genetic algorithm* Paper presented at the 4th international Conference on Genetic Algorithm,(ICGA'91), San Diego, CA, US.

Carson, C., Belongie, S., Greenspan, H., & Malik, J. (2002). Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence 24*(8), 1026-1038.

Carson, C., Thomas, M., Belongie, S., Hellerstein, J., & Malik, J. (1999). *Blobworld : A system for region-based image indexing and retrieval.* Paper presented at the 3rd International Conference on Visual Information Systems (VISUAL'99), Amsterdam, Netherlands.

Chakravarti, R., & Meng, X. (2009). *A study of color histogram based image retrieval.* Paper presented at the Sixth International Conference on Information Technology : New Generations (ITNG'09), Las Veegas,Nevada.

Chang, S. K., & Hsu, A. (1992). Image information systems: where do we go from here? *IEEE Transactions on Knowledge and Data Engineering 4*(5), 431-442.

Chen, C., Gagaudakis, G., & Rosin, P. (2000). *Similarity-based image browsing.* Paper presented at the Proceedings of the 16th IFIP World Computer Congress(International Conference on Intelligent Information Processing), Beijing, China.

Chen, Y., & Wang, J. Z. (2004). Image categorization by learning and reasoning with regions. *The Journal of Machine Learning Research, 5*, 913-939.

Choras, R. S. (2007). Image feature extraction techniques and their applications for CBIR and biometrics systems. *International Journal of Biology and Biomedical Engineering, 1*(1), 6–16.

Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence 24*(5), 603-619.

da Silva Torres, R., & Falcão, A. X. (2006). Content-based image retrieval: theory and applications. *Revista de Informática Teórica e Aplicada, 2*(13), 161-185.

Daidi, Z., Defee, I., & LEFEVRE, S. (2008). Face Retrieval Based on Robust Local Features and Statistical-Structural Learning Approach. *EURASIP Journal on Advances in Signal Processing, 2008*(18).

Datta, R., Joshi, D., Li, J., & Wang, j. (2008). Image retrieval: ideas, influences, and trends of the new age. *ACM Computing Surveys (CSUR), 40*(2), 1-60.

Deng, Y., Manjunath, B., Kenney, C., Moore, M. S., & Shin, H. (2001). An efficient color representation for image retrieval. *IEEE Transactions on Image Processing 10*(1), 140-147.

Depalov, D., Pappas, T., Li, D., & Gandhi, B. (2006). *Perceptually based techniques for semantic image classification and retrieval.* Paper presented at the Proceeding of SPIE, San Jose, CA.

Deselaers, T., Keysers, D., & Ney, H. (2008). Features for image retrieval: an experimental comparison. *Information Retrieval, 11*(2), 77-107.

Deselaers, T., Rybach, D., Dreuw, P., Keysers, D., & Ney, H. (2005). Face-based image retrieval-one step toward object-based image retrieval. In H. M. u. a. A. Hanbury (Ed.), *First International Workshop on Evaluation for Image Retrieval* (pp. 25-32): Vienna, Austria.

Deselaers, T., Weyand, T., Keysers, D., Macherey, W., & Ney, H. (2006). *FIRE in ImageCLEF 2005 : Combining content-based image retrieval with textual information retrieval.* Paper presented at the Working Notes of the CLEF Workshop, Vienna, Austria.

Djeraba, C. (2003). Association and content-based retrieval. *IEEE Transactions on Knowledge and Data Engineering, 15*(1), 118-135.

Dumais, S. T. (2004). Latent semantic analysis. *Annual review of information science and technology, 38*(1), 188-230.

Eakins, J. (2001). Retrieval of still images by content. Lecture Notes in Computer Science: Lectures on Information Retrieval (Vol. 1980/2001, pp. 111-138). Berlin.

Eakins, J., & Graham, M. (1999). Content-based Image Retrieval. *Technical Report, University of Northumbria at Newcastle*.

Enser, P., Sandom, C., & Lewis, P. (2005). Automatic annotation of images from the practitioner perspective. *Image and Video Retrieval*, 595-595.

Fang, Y., & Geman, D. (2005). Experiments in mental face retrieval. In T. Kanade, A. Jain & N. Ratha (Eds.), *Audio- and Video-Based Biometric Person Authentication* (Vol. 3546, pp. 637-646): Springer Berlin / Heidelberg.

Felzenszwalb, P. F., & Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision 59*(2), 167-181.

Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., et al. (1995). Query by image and video content : The QBIC system. *IEEE Computer, 28*(9), 23-32.

Ford, A., & Roberts, A. (1998). Colour space conversions. *Technical report, Westminster University, London*.

Fournier , J., Cord , M., & Philipp-Foliguet , S. (2001). *Back-propagation algorithm for relevance feedback in image retrieval.* Paper presented at the IEEE International Conference in Image Processing (ICIP'01), University of Cergy Pontoise, France.

Freund, Y., & Schapire, R. (1995). A desicion-theoretic generalization of on-line learning and an application to boosting Computational Learning Theory. In P. Vitányi (Ed.), (Vol. 904, pp. 23-37): Springer Berlin / Heidelberg.

Fromherz, T., Stucki, P., & Bichsel, M. (1997). *A survey of face recognition*. Technical Report ,Dept. of Computer Science, University of Zurich, Zurich MM, Switzerland.

Frowd, C. D., Hancock, P. J. B., & Carson, D. (2004). EvoFIT: A holistic, evolutionary facial imaging technique for creating composites. *ACM Transactions on Applied Perception (TAP), 1*(1), 19-39.

Gao , B., Liu , T. Y., Qin , T., Zheng , X., Cheng , Q. S., & Ma , W. Y. (2005). *Web image clustering by consistent utilization of visual features and surrounding texts.* Paper presented at the Proceedings of the 13th annual ACM international conference on Multimedia.

Gao , C., Sang , N., & Tang , Q. (2010). On selection and combination of weak learners in AdaBoost. *Pattern Recognition Letters, 31*(9), 991-1001.

Gao , Y., & Qi , Y. (2005). Robust visual similarity retrieval in single model face databases. *Pattern Recognition, 38*(7), 1009-1020.

Gevers , T., & Smeulders , A. W. M. (1997). *Pictoseek: A content-based image search engine for the WWW.* Paper presented at the International Conference on Visual Information Systems,VISUAL'97, San Diego, USA.

Gevers, T., & Smeulders, A. W. M. (2000). Pictoseek: Combining color and shape invariant features for image retrieval. *IEEE Transactions on Image Processing, 9*(1), 102-119.

Gimel'Farb, G. L., & Jain, A. K. (1996). On retrieving textured images from an image database. *Pattern Recognition, 29*(9), 1461-1483.

Golland, P. (2005). Face Recognition Using Eigenfaces. *representation-and-modeling-for-image-analysis,* Retrieved June, 2010, from http://ocw.mit.edu/ courses/ electrical-engineering-and-computer-science/6-881-representation-and-modeling -for-image-analysis-spring-2005/lecture-notes/l02.pdf.

Grosky, W. I. (2011). Image Retrieval-Existing Techniques,Content-Based image retrieval Systems. Retrieved June, 2011, from http://encyclopedia .jrank.org /articles/pages/6763/Image-Retrieval.html.

Gudivada, V. N., Raghavan, V. V., & Vanapipat, K. (1994). A unified approach to data modeling for a class of image database applications. *IEEE Transactions on Data and Knowledge Discovery*.

Gupta, A., & Jain, R. (1997). Visual information retrieval. *Communications of the ACM, 40*(5), 70-79.

Gupta, P., Saxena, S., Singh, S., Dhami, S., & Singh, V. (2012). Color Image Segmentation: A State of the Art Survey. *International Journal of Computational Intelligence Research, 8*(1), 17-25.

Haddadnia, J., Ahmadi, M., & Faez, K. (2002). *A hybrid learning RBF neural network for human face recognition with pseudo Zernike moment invariant.* Paper presented at the In Proceedings of the IEEE International Joint Conference on Neural Networks, Honolulu, USA.

Han , J.-h., Huang , D.-s., Lok , T.-m., & Lyu, M. R. (2005). *A novel image retrieval system based on BP neural network.* Paper presented at the International Joint Conference on neural Networks (IJCNN '05), Inst. of Intelligent Machines, Chinese Acad. of Sci., Hefei, China.

Hanbury, A. (2008). A survey of methods for image annotation. *Journal of Visual Languages & Computing, 19*(5), 617-627.

Hartvedt , C. (2007). *Utilizing context in ranking results from distributed image retrieval :The CAIRANK Prototype.* Bergensis University,Bergen, Norway.

Heesch, D. (2008). A survey of browsing models for content based image retrieval. *Multimedia Tools and Applications, 40*(2), 261-284.

Hongjiang , Z., Wenyin, L., & Hu, C. (2000). *iFind: a system for semantics and feature based image retrieval over Internet.* Paper presented at the the Proceedings of the eighth ACM international conference on Multimedia.

Hove, L. J. (2004). *Improving image retrieval with a thesaurus for shapes.* University of Bergen, Bergen.

Hsu, R.-L., Abdel-Mottaleb, M., & Jain, A. K. (2002). Face detection in color images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(5), 696-706.

Hsu, R. L., & Jain, A. K. (2002). *Semantic face matching.* Paper presented at the IEEE International Conference on Multimedia and Expo(ICME 02).

Hu, M. K. (1962). Visual pattern recognition by moment invariants. *Information Theory, IRE Transactions on, 8*(2), 179-187.

Ito, H., & Koshimizu, H. (2004). *Keyword and face image retrieval based on latent semantic indexing.* Paper presented at the IEEE International Conference on Systems, Man and Cybernetics, Hague, Netherlands.

Jacobs, C. E., Finkelstein, A., & Salesin, D. H. (1995). *Fast multiresolution image querying.* Paper presented at the Proceediags of SIGGRAPH Los Angels, CA

Jain, A., Hong, L., & Pankanti, S. (2000). Biometric Identification. *Communications of the ACM, 43*(2), 90-98.

Jain, K. (1989). *Fundamental of digital image processing* (Vol. 569): Prentice-Hall, New York.

Jeon, J., Lavrenko, V., & Manmatha, R. (2003). *Automatic image annotation and retrieval using cross-media relevance models.* Paper presented at the Proceedings of the 26th annual international ACM SIGIR conference on research and development in informaion retrieval, Toronto, Canada.

Jiang, X., Binkert, M., Achermann, B., & Bunke, H. (2000). Towards detection of glasses in facial images. *Pattern Analysis & Applications, 3*(1), 9-18.

Jörgensen, C. (1998). Attributes of images in describing tasks. *Information Processing & Management, 34*(2), 161-174.

Kam-Art, R., Raicharoen, T., & Khera, V. (2009). *Face recognition using feature extraction based on descriptive statistics of a face image.* Paper presented at the Proceedings of the 8th International Conference on Signal processing.

Karayiannis, N. B. (1999). Reformulated radial basis neural networks trained by gradient descent. *IEEE Transactions on Neural Networks, 10*(3), 657-671.

Kato, T. (1992). *Database architecture for content-based image retrieval.* Paper presented at the Proceedings of SPIE image storage and retrieval systems, San Jose, CA, USA.

Kaur , A., & Banga , V. (2011). *Color based image retrieval.* Paper presented at the ISEM2011 international conference, Bangkok, Thailand.

Kodituwakku, S., & Selvarajah, S. (2004). Comparison of color features for image retrieval. *Indian Journal of Computer Science and Engineering, 1*(3), 207-211.

Krishnapuram, R., Medasani, S., Jung, S. H., Choi, Y. S., & Balasubramaniam, R. (2004). Content-based image retrieval based on a fuzzy approach. *IEEE Transactions on Knowledge and Data Engineering, 16*(10), 1185-1199.

Kulkarni, S., & Verma, B. (2003). *Fuzzy logic based texture queries for CBIR.* Paper presented at the Fifth International Conference on Computational Intelligence and Multimedia Applications(ICCIMA 2003).

Kurban, T., & Beşdok, E. (2009). A comparison of RBF neural network training algorithms for inertial sensor based terrain classification. *Sensors, 9*(8), 6312-6329.

La Cascia , M., Sethi , S., & Sclaroff , S. (1998). *Combining textual and visual cues for content-based image retrieval on the world wide web.* Paper presented at the IEEE Workshop on Content-Based Access of Image and Video Libraries.

Lai, H. C., Marculescu, R., Savvides, M., & Chen, T. (2008). Communication-aware face detection using noc architecture, *Computer Vision Systems* (pp. 181-189): Springer-Verlag.

Lakshmi, H. V., & PatilKulakarni, S. (2010). Segmentation algorithm for multiple face detection in color images with skin tone regions using color spaces and edge detection techniques. *International Journal of Computer Theory and Engineering, 2*(4), 1793-8201.

Laughery, K. R., & Fowler, R. H. (1980). Sketch artist and Identi-kit procedures for recalling faces. *Journal of Applied Psychology, 65*(3), 307.

Le, D. D., Satoh, S., & Houle, M. E. (2007). *Boosting face retrieval by using relevant set correlation clustering.* Paper presented at the IEEE International Conference on Multimedia and Expo.

Lew, M. S., Sebe, N., Djeraba, C., & Jain, R. (2006). Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), 2*(1), 1-19.

Li, Q., Shi, Z., & Luo, S. (2007). *Image retrieval based on fuzzy color semantics.* Paper presented at the Fuzzy Systems Conference, FUZZ-IEEE 2007.

Li , Q., Shi , Z., & Luo, S. (2007). *A neural network approach for bridging the semantic gap in texture image retrieval* Paper presented at the International Joint Conference on Neural Networks(IJCNN).

Liu, Y., Zhang, D., Lu, G., & Ma, W. Y. (2007). A survey of content-based image retrieval with high-level semantics. *Pattern Recognition, 40*(1), 262-282.

Long , F., Zhang , H., & Feng , D. D. (2003). Fundamentals of content-based image retrieval. In W. S. D. Feng, and HJ Zhang (Ed.), *Multimedia Information Retrieval and Management-Technological Fundamentals and Applications*: Springer.

Lowe, D., & Broomhead, D. (1988). Multivariable functional interpolation and adaptive networks. *Complex systems, 2*, 321-355.

Luo, J., & Savakis, A. (2001). *Indoor vs outdoor classification of consumer photographs using low-level and semantic features.* Paper presented at the 2001 International Conference on Image Processing.

Ma, W. Y., & Zhang, H. J. (1998). *Benchmarking of image features for content-based retrieval.* Paper presented at the IEEE 32nd Asilomar Conference on Signals, Systems,and Computers.

Makadia, A., Pavlovic, V., & Kumar, S. (2010). Baselines for image annotation. *International Journal of Computer Vision, 90(1), 88-105.*

Manjunath, B., Salembier, P., & Sikora, T. (2002). *Introduction to MPEG-7: multimedia content description interface* (Vol. 1): John Wiley & Sons Inc.

Manjunath, B. S., & Ma, W. Y. (1996). Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence 18*(8), 837-842.

Meir, R., & Rätsch, G. (2003). An introduction to boosting and leveraging. *Advanced lectures on machine learning*, 118-183.

Mezaris, V., Kompatsiaris, I., & Strintzis, M. G. (2003). *An ontology approach to object-based image retrieval.* Paper presented at the ICIP2003.

Moody, J., & Darken, C. J. (1989). Fast learning in networks of locally-tuned processing units. *Neural computation, 1*(2), 281-294.

Moran, S. (2009). *Automatic Image Tagging.* University of Edinburgh, Edinburgh.

Navarrete, P., & Ruiz-del-Solar, J. (2002). *Interactive face retrieval using self-organizing maps.* Paper presented at the Proceedings of the 2002 International Joint Conference on Neural Networks(IJCNN 2002), Honolulu, USA.

Neruda, R., & Vidnerová, P. (2009). Learning errors by radial basis function neural networks and regularization networks. *International Journal of Grid and Distributed Computing, 1(2)*.

Pagare, R., & Shinde, A. (2012). A study on image annotation techniques. *International Journal of Computer Applications, 37*(6), 42-45.

Pai, Y. T., Ruan, S. J., Shie, M. C., & Liu, Y. C. (2006). *A Simple and accurate color face detection algorithm in complex background.* Paper presented at the IEEE International Conference on Multimedia and Expo (ICME2006).

Papageorgiou, C., Oren, M., & Poggio, T. (1998). *A general frame-work for object detection*. Paper presented at the International Conference on Computer Vision.

Park, S. B., Lee, J. W., & Kim, S. K. (2004). Content-based image classification using a neural network. *Pattern Recogn Letter, 25*, 287–300.

Pavlidis, T. (2008). *Limitations of content-based image retrieval.* Paper presented at the Invited Plenary Talk at the 19th Internationa Conference on Pattern Recognition, Tampa, Florida.

Pcyuen, Feng, G., & Dai, D. (1998). *Human face image retrieval system for large database.* Paper presented at the Fourteenth International Conference on Pattern Recognition.

Penry, J. (1974). Photo-fit. *Forensic Photography, 3*(7), 4-10.

Pentland, A., Picard, R. W., & Sclaroff, S. (1996). Photobook: Content-based manipulation of image databases. *International Journal of Computer Vision, 18*(3), 233-254.

Poggio, T., & Girosi, F. (1990). Regularization algorithms for learning that are equivalent to multilayer networks. *Science, 247*(4945), 978.

Powel, M. D. (1981,). *Approximation theory and methods*: Cambridge University Press, Cambridge.

Prasad, B., Gupta, A., Toong, H. M. D., & Madnick, S. E. (1987). A microcomputer-based image database management system. *IEEE Transactions on Industrial Electronics* (1), 83-88.

Puzicha, J., Buhmann, J. M., Rubner, Y., & Tomasi, C. (2001). Empirical evaluation of dissimilarity measures for color and texture. *Computer Vision and Image Understanding Journal, 84(1):25-43*.

Rasmussen, E. M. (1997). Indexing Images. *Annual Review of Information Science and Technology (ARIST), 32*, 169-196.

Rong, Z., & Grosky, W. I. (2002). Narrowing the semantic gap-improved text-based web document retrieval using visual features. *IEEE Transactions on Multimedia 4*(2), 189-200.

Rui, Y., Huang, T. S., & Chang, S. F. (1999). Image retrieval: current techniques, promising directions, and open Issues *Journal of visual communication and image representation, 10*(1), 39-62.

Sahin, F. (1997). *A radial basis function approach to a color image classification problem in a real time industrial application.* MSc thesis,Virginia Polytechnic Institute and State University , Blacksburg.

Salnn, F., Bay, JS,. (1998). *A radial basis function approach to a color image classification problem in a real time industrial application.* Paper presented at the First International Conference on Multisource - Multisensor Information Fusion, Las Vegas, USA.

Salton, G. (1989). *Automatic text processing : The transformation, analysis, and retrieval of information by computers*: Addison-Wesley Publishing Company.

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information processing & management, 24*(5), 513-523.

Samangooei, S., Guo, B., & Nixon, M. S. (2008). *The use of semantic human description as a soft biometric.* Paper presented at the 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems.

Schapire, R. E. (2003). The boosting approach to machine learning: An overview. *lecture notes in statistics* (pp. 149-172): New york-springer verlag.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM computing surveys (CSUR), 34*(1), 1-47.

Shih, P., & Liu, C. (2005). Comparative assessment of content based face image retrieval in different color spaces. *International Journal of Pattern Recognition and Artificial Intelligence 19*(7), 873-893.

Simon, D. (2002). Training radial basis neural networks with the extended Kalman filter. *Neurocomputing, 48*(1), 455-475.

Sinha, P., Balas, B., Ostrovsky, Y., & Russell, R. (2007). Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE, 94*(11), 1948-1962.

Sirovich, L., & Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A, 4*(3), 519-524.

Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(12), 1349-1380.

Smith, J. R., & Chang, S. F. (1996a). *Automated image retrieval using color and texture*: Technical Report, CU/CTR 408-95-14, Columbia University, New York, NY, USA.

Smith, J. R., & Chang, S. F. (1996b). *Searching for images and videos on the world-wide web*: Technical Report,459-96-25,Center for Communications Research, Columbia University.

Smith , J. R., & Chang , S. F. (1996). Tools and techniques for color image retrieval *Storage & Retrieval for Image and Video Databases IV* (Vol. 2670, pp. 426–437).

Smith, J. R., & Chang, S. F. (1997a). Visually searching the web for content. *IEEE Multimedia,4*(3), 12-20.

Smith, J. R., & Chang, S. F. (1997b). *VisualSEEk: a fully automated content-based image query system.* Paper presented at the Proceedings of the ACM Multimedia Conference, Boston, MA, USA.

Sridhar, V., Nascimento, M., & Li, X. (2002). *Region-based image retrieval using multiple-features.* Paper presented at the Proc. of the 5th International Conference on Recent Advances in Visual Information Systems. London, UK: Springer-Verlag, 2002: 61-75.

Sridharan, K., Nayak, S., Chikkerur, S., & Govindaraju, V. (2005). A probabilistic approach to semantic face retrieval system *Audio- and Video-Based Biometric Person Authentication* (Vol. 3546, pp. 85-100): Springer Berlin / Heidelberg.

Srihari, R. K. (1995). Automatic indexing and content-based retrieval of captioned images. *IEEE Comput., 28*(9), 49-56.

Stamou, G. (2006). *Image Annotation on the Semantic Web*: W3C Incubator Group Report,URL,http://www.w3.org/2005/Incubator/mmsem/XGR-image-annotation /,February 2012.

Stanchev, P. (1999). General image database model. In D. S. A. Huijsmans (Ed.), *Lecture Notes in Computer Science* (Vol. 1614, pp. 29-36).

Stricker, M., & Dimai, A. (1996). Color indexing with weak spatial constraints *Storage and Retrieval for Image and Video Databases IV* (Vol. 2670, pp. 29-40).

Stricker, M., & Orengo, M. (1995). *Similarity of color images.* Paper presented at the Proceeding of SPIE. Storage Retrieval Still Image Video Databases, San Diego/La Jolla, CA, USA.

Sumathi, T. (2011). An Overview of Automated Image Annotation Approaches. *International Journal of Research and Reviews in Information Sciences (IJRRIS), 1*(1).

Swain, M. J., & Ballard, D. H. (1991). Color indexing. *International Journal of Computer Vision, 7*(1), 11-32.

Tamura, H., & Yokoya, N. (1984). Image database systems: A survey. *Pattern Recognition, 17*(1), 29-43.

Tao, Y., & Grosky, W. I. (1998). *Image matching using the OBIR system with feature point histograms.* Paper presented at the Proceedings of 4th IFIP 2.6 Working Conference on Visual Database Systems (VDB4), L'Aquila, Italy.

Thang, N. D., Rasheed, T., Lee, Y. K., Lee, S., & Kim, T. S. (2011). Content-based facial image retrieval using constrained independent component analysis. *Information Sciences: an International Journal, 181*(15), 3162-3174.

Thilagamani, S. (2011). A Survey on Image Segmentation Through Clustering. *International Journal of Research and Reviews in Information Sciences (IJRRIS), 1*(1).

Tkalcic, M., & Tasic, J. F. (2003). *Colour spaces: perceptual, historical and applicational background.* Paper presented at the Proceeding of EUROCON 2003 : International Conference on " Computer as a tool".

Tolba, A., El-Baz, A., & El-Harby, A. (2005). Face recognition: A literature review. *International Journal of Signal Processing, 2*(2), 88–103.

Tong, S., & Chang, E. (2001). *Support vector machine active learning for image retrieval.* Paper presented at the Proceedings of the ninth ACM international conference on Multimedia.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: computation in neural systems, 14*(3), 391-412.

Tsai, C. F., & Hung, C. (2008). Automatically annotating images with keywords: A review of image annotation systems. *Recent Patents on Computer Science, 1*(1), 55-68.

Turk, M., & Pentland, A. (1991). *Face recognition using eigenfaces.* Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition.

Vailaya, A., Figueiredo, M. A. T., Jain, A. K., & Zhang, H. J. (2001). Image classification for content-based indexing. *IEEE Transactions on Image Processing 10*(1), 117-130.

Vailaya, A., Jain, A., & Zhang, H. J. (1998). On image classification: City images vs. landscapes. *Pattern Recognition, 31*(12), 1921-1935.

Vasylenko, E. N. (2010). *A Statistical Approach to Relevance Feedback for Content-Based Image Retrieval*.

Veltkamp, R. C., & Tanase, M. (2000). *Content-based image retrieval systems: A survey*: Technical Report UU-CS-2000-34,Department of Computing Science, Utrecht University ,Utrecht, The Netherlands.

Vezhnevets, V., Sazonov, V., & Andreeva, A. (2003). *A survey on pixel-based skin color detection techniques.* Paper presented at the Proceedings of Graphicon-2003, Moscow, Russia.

Vijaya Lata, Y., Tungathurthi, C. K. B., Rao, H. R. M., Govardhan, A., & Reddy, L. P. (2009). Facial recognition using eigenfaces by PCA. *International Journal of Recent Trends in Engineering, 1*(1), 587-590.

Vikram, T., Chidananda Gowda, K., Guru, D., & Urs, S. R. (2008). *Face indexing and retrieval by spatial similarity.* Paper presented at the International Congress on Image and Signal Processing(CISP'08).

Viola, P., & Jones, M. J. (2004). Robust real-time faced detection. *International journal of computer vision, 57*(2), 137-154.

Vu, K., Hua, K. A., & Tavanapong, W. (2003). Image retrieval based on regions of interest. *IEEE Transactions on Knowledge and Data Engineering 15*(4), 1045-1049.

Wang, C., Zhang, L., & Zhang, H. J. (2008). *Learning to reduce the semantic gap in web image retrieval and annotation.* Paper presented at the Proceedings of the 31st annual international ACM SIGIR conference on research and development in information retrieval.

Wang , D., & Ma , X. (2005). A hybird image retrieval system with user's relevance feedback using neurocomputing. *Informatica-Ljubljana, 29*(3), 271-279.

Wang, J. Z., Li, J., & Wiederhold, G. (2001). SIMPLIcity : Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 947-963.

Wenyin, L., Dumais, S., Sun, Y., Zhang, H. J., Czerwinski, M., & Field, B. (2001). *Semi-automatic image annotation.* Paper presented at the International Conference on Human-Computer Interaction, New Orleans, Louisiana, USA.

Wiskott , L., Fellous , J. M., Kuiger , N., & von der Malsburg , C. (1997). Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence 7*(19), 775-779.

Wu, B., Ai, H., & Liu, R. (2004). *Glasses detection by boosting simple wavelet features.* Paper presented at the 17th International Conference on Pattern Recognition (ICPR2004).

Xiang , Z. S., & Huang, T. S. (2000). *CBIR: from low-level features to high-level semantics.* Paper presented at the the Proceedings of SPIE,The International Society for Optical Engineering, San Jose, CA.

Xiangyu, J., & James, C. F. (2003). *Improving image retrieval effectiveness via multiple queries.* Paper presented at the The 1st ACM international workshop on Multimedia databases, New Orleans, LA, USA.

Yang, L., & Jin, R. (2006). *Distance metric learning: A comprehensive survey*: Technical Report ,Department of Computer Science and Engineering, Michigan State University.

Yang , M. H., Kriegman , D. J., & Ahuja , N. (2002). Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence 24*(1), 34-58.

Yoshitaka, A., & Ichikawa, T. (1999). A survey on content-based retrieval for multimedia databases. *IEEE Transactions on Knowledge and Data Engineering 11*(1), 81-93.

Yousef, R., & Hindi, K. (2005). Training radial basis function networks using reduced sets as center points. *International Journal of Information Technology, 2*(1), 21-35.

Yuille, A. L., Hallinan, P. W., & Cohen, D. S. (1992). Feature extraction from faces using deformable templates. *International Journal of Computer Vision, 8*(2), 99-111.

Zhang, L., Chen, L., Li, M., & Zhang, H. (2003). *Automated annotation of human faces in family albums.* Paper presented at the the Proceedings of the eleventh ACM international conference on Multimedia.

Zhao , R., & Grosky , W. I. (2002). Bridging the semantic gap in image retrieval. In T. Shih (Ed.), *Distributed Multimedia Databases: Techniques and Applications,* (pp. 14-36). Hershey, Pennsylvania Idea Group Publishing.

Zhao, W., Chellappa, R., Phillips, P., & Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM Computing Surveys (CSUR), 35*(4), 399-458.

Zhou, H., Sadka, A. H., Swash, M. R., Azizi, J., & Umar, A. S. (2009). Content Based Image Retrieval and Clustering: A Brief Survey. *Recent Patents on Electrical Engineering, 2009(2),* 187-199.

Zhou , H., Yuan , Y., & Sadka , A. H. (2008). Application of semantic features in face recognition. *Pattern Recognition, 41*(10), 3251-3256.

Zhou, X. S., & Huang, T. S. (2003). Relevance feedback in image retrieval: A comprehensive review. *Multimedia systems, 8*(6), 536-544.

APPENDIX A

TABLES OF SEMANTIC FEATURES REPRESENTATION

Table A.1: Gender and age, symbolic and numerical representation (before and after normalization)

Gender				**Age**		
Symbolic	**Numerical**			**Symbolic**	**Numerical**	
None	0	0		None	0	0
Male	2090	0.999900		Infant	1910	0.9137842
Female	2010	0.9616263		Child	1920	0.9185684
				Adolescent	1935	0.9257447
				Young Adult	1960	0.9377053
				Middle Adulthood	1980	0.9472737
				Senior	1990	0.9520578

Table A.2: Race and skin color representation

Race				**Skin Color**		
Symbolic	**Numerical**			**Symbolic**	**Numerical**	
None	0	0		None	0	0
Malay	1805	0.863550		Black	1710	0.818100
Chinese	1825	0.873118		Brown	1730	0.827668
Indian	1847	0.883643		Tan	1760	0.8420211
Middle Eastern	1865	0.892255		White	1790	0.8563737
European	1870	0.894647				
African	1895	0.906607				

Table A.3: Beard size and facial marks representation

Beard Size				Facial Marks		
Symbolic	**Numerical**			**Symbolic**	**Numerical**	
None	0	0		None	0	0
Medium	1040	0.4975579		Mole	910	0.4353632
Short	1010	0.3779526		Scar	990	0.3779526
Long	1090	0.5214789		Freckles	940	0.4497158

Table A.4: Hair color and hair length representation

Hair Color				Hair Length		
Symbolic	**Numerical**			**Symbolic**	**Numerical**	
None	0	0		None	0	0
Black	1610	0.7702580		Short	1522	0.7281568
Brown	1630	0.7893947		Medium	1544	0.7386820
Blond	1650	0.7941790		Long	1570	0.7511211
Red	1660	0.8037474		Bald	1510	0.7224158
Gray	1680	0.8085316		Covered Head	1599	0.7649953

Table A.5: Hair type and eye color representation

Hair Type					Eye Color	
Symbolic	**Numerical**			**Symbolic**	**Numerical**	
None	0	0		None	0	0
Curly	1410	0.6745737		Dark	790	0.3779526
Wavy	1450	0.6937105		Brown	710	0.3396789
Straight	1490	0.7128474		Blue	740	0.3540316
Covered Head	1470	0.7032790		Green	760	0.3636000

Table A.6: Glasses shape and mustache size representation

Glasses Shape				Mustache Size		
Symbolic	**Numerical**			**Symbolic**	**Numerical**	
None	0	0		None	0	0
Oval	1205	0.5764974		Medium	1140	0.545400
Circular	1220	0.3779526		Short	1110	0.3779526
Square	1270	0.6075947		Long	1190	0.5693211
Rectangle	1295	0.6195553				

Table A.7: Nose shape and face shape representation

Nose Shape					Face Shape	
Symbolic	**Numerical**			**Symbolic**	**Numerical**	
None	0	0		None	0	0
Flat	802	0.3836937		Oval	1310	0.6267316
Straight	860	0.3779526		Round	1315	0.3779526
Wide	815	0.3899132		Long	1380	0.6602211
Convex	899	0.4301005		Square	1365	0.6530447
Concave	875	0.4186184		Heart	1335	0.6386921

Table A.8: Eyebrows thickness and mouth size

Eyebrows Thickness				Mouth Size		
Symbolic	**Numerical**			**Symbolic**	**Numerical**	
None	0	0		None	0	0
Normal	610	0.2918368		Medium	542	0.2593042
Bushy	680	0.3779526		Short	510	0.3779526
				Long	590	0.2822684

Table A.9: Lip thickness representation

Lip Thickness		
Symbolic	**Numerical**	
None	0	0
Medium	435	0.2081132
Thick	495	0.3779526
Thin	402	0.1923253

APPENDIX B

TABLES OF SEMANTIC CONCEPTS FREQUENCY

Tables B.1, B.2, B.3, and B.4: Semantic concepts frequency on the local database based on the participants' annotation.

Table B.1

Gender		**Age**		**Race**		**Skin Color**		**Hair Color**	
Male	980	1-3	20	Malay	480	Black	70	Black	1200
Female	520	3-12	0	Chinese	310	Brown	80	Brown	60
		13-19	0	Indian	80	Tan	800	Blond	20
		20-40	1400	Middle Eastern	540	White	550	Red	0
		40-65	80	European	20			Gray	20
		65-	0	African	70			Covered Head	200

Table B.2

Glasses Shape		**Moustache Size**		**Beard Size**		**Facial Marks**		**Nose Shape**	
Oval	10	Medium	70	Medium	30	Mole	180	Straight	630
Circular	20	Short	560	Short	480	Scar	0	Wide/Flat	810
Square	247	Long	10	Long	40	Freckles	20	Convex	10
Rectangle	34							Concave	0
								Rounded	50

Table B.3

Eyes Color		**Eyebrows Thickness**		**Mouth Size**		**Lip Thickness**	
Black	1420	Normal	1390	Medium	1300	Medium	1360
Brown	40	Bushy	110	Small	110	Thick	100
Blue	20			Big	90	Thin	40
Green	20						

Table B.4

Hair Length		**Hair Type**		**Face Shape**	
Short	320	Curly	90	Oval	530
Medium	570	Wavy	100	Round	490
Long	380	Straight	1100	Long	200
Bald	30	Covered Head	200	Square	160
Covered Head	200			Heart	120

Tables B.5, B.6, B.7, and B.8: Semantic concepts frequency on the ORL database based on the participants' annotation.

Table B.5.

Gender		**Age**		**Race**		**Skin Color**		**Hair Color**	
Male	360	1-3	0	Malay	0	Black	10	Black	260
Female	40	3-12	0	Chinese	0	Brown	0	Brown	20
		13-19	0	Indian	0	Tan		Blond	100
		20-40	250	Middle Eastern	0	White	390	Red	0
		40-65	150	European	390		0	Gray	20
		65-	0	African	10			Covered Head	0

Table B.6.

Glasses Shape		**Moustache Size**		**Beard Size**		**Facial Marks**		**Nose Shape**	
Oval	0	Medium	40	Medium	40	Mole	0	Straight	160
Circular	119	Short	40	Short	10	Scar	0	Wide/Flat	10
Square	0	Long	40	Long	30	Freckles	0	Convex	170
Rectangle	0							Concave	20
								Rounded	40

Table B.7.

Eyes Color		**Eyebrows Thickness**		**Mouth Size**		**Lip Thickness**	
Black	390	Normal	380	Medium	350	Medium	290
Brown	0	Bushy	20	Small	20	Thick	10
Blue	10			Big	30	Thin	100
Green	0						

Table B.8.

Hair Length		**Hair Type**		**Face Shape**	
Short	10	Curly	10	Oval	40
Medium	310	Wavy	90	Round	140
Long	60	Straight	300	Long	40
Bald	20	Covered Head		Square	100
Covered Head				Heart	80