Journal of Theoretical and Applied Information Technology
30th April 2020. Vol.98. No 08
© 2005 – ongoing JATIT & LLS

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

EFFICIENT INTEGRATION METHOD FOR HUMAN FACIAL IMAGES RETRIEVAL BASED ON VISUAL CONTENT AND SEMANTIC DESCRIPTION

[1]AHMED ABDU ALATTAB, [2]SAMEEM ABDUL KAREEM, [3]IBRAHEEM M.G. ALWAYLE,
[4]ANWAR ALI YAHYA, [5]KHALED M.A. ALALAYAH

[1,3,5]Dept. of Comp. Science, Faculty of Science and Arts, Sharurah, Najran University, Saudi Arabia.
[2]Dept. of Artificial Intelligence, Faculty of Comp. Science and Info. Technology, University of Malaya, Kuala Lumpur, Malaysia.
[4]Faculty of Computer Science and Information Systems, Najran University, Najran, Saudi Arabia.
[1,4]Dept. of Comp. Science, Faculty of Comp. Science and Info. Systems, Thamar University, Yemen.
[3]Dept. of Math. Science, Faculty of Science and Arts, Amrran University, Yemen.
[5]Dept. of Math. & Comp. Science, Faculty of Science, IBB University, Yemen.
E-mail: [1]ahmed_alattab@yahoo.com, [2]sameem@um.edu.my

ABSTRACT

A semantic-content based facial image retrieval (SCBFIR) technique that incorporates multiple visual and semantic features to improve the accuracy of the facial image retrieval is proposed. The proposed technique based on reducing the semantic gap between the high-level query requirement and the low-level facial features of the human facial image. Visual features and semantic features are extracted by different methods, moreover, some features may be considered more important than others, so features weighting is used to distinguish the importance of the various features. This research proposed a model that links the high-level query requirement and the low-level features of the human facial image. A newly proposed method based on radial basis function network is introduced for measuring the distance between the query vectors and the database vectors of the different features for finding, weighting, and combining the similarities. The proposed system of SCBFIR is trained and tested on the 'ORL Database of Faces' from AT&T Laboratories, Cambridge, and a local database consisting of local facial images from the University of Malaya (UM), Kuala Lumpur. The results of the experiments show that, as compared to the current content-based facial image retrieval technique (CBFIR), the proposed methods of SCBFIR achieve the best performance. More precisely the CBFIR achieves 84.0% and 92.41% 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.

Keywords: *Image Retrieval, Face Retrieval, Semantic Features, RBFN, Eigenfaces, Color*

1. INTRODUCTION

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 for (i) security reasons, (ii) law enforcement applications, (iii) human-computer interaction applications, and other public places for automated surveillance applications.

Image retrieval systems are developed in order to search the target image more easily, speedily, and at a lower cost of retrieval. In the current systems[1, 2] [3-5] [6-11] the visual features (low-level features) 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 use the representation of these features that reflect a global description of images to calculate the similarity and matching between images without considering the contents of the images. This leads to the failure to consider the implicit semantics of an image. Humans compare and measure the similarities between images, and the semantic contents 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, eigenfaces, 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. As such, the CBFIR 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 CBFIR techniques.

The suitable ways of describing image semantic content are through high-level semantic concepts. This is because humans understand and expressed things through concepts represented in keywords more easily. Users express 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.

The heart component for image retrieval is the similarity measure. A common approach to computing a similarity metric among the patterns to be classified is by using the distance-based method. By and large, the Euclidean distance has been widely used as a similarity measure. Yang & Jin [12] conducted a comprehensive survey of distance matrices. In spite of many successful works on distance matrices[13], 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. For applications, where different algorithms and techniques extract the feature attributes, the above methods would be inefficient. This is also applicable to the visual and 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. To merge different features of the images together in an efficient and distributed manner requires an innovative solution.

Based on the aforementioned, the problem confronting the use of heterogeneous set of features is how to integrate them 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 $D_1(x_1, y_1)$, D_2 (x_2, y_2), and D_n (x_n, y_n) are the similarity indices between x and y based on n different feature vectors (example: color, eigenfaces and semantic feature) – Figure 1, defining an integrated similarity index, is then the issue to be addressed. The introduced proposed solution to this problem in this research is to form a new similarity metric with a suitable weight parameter that is directly applicable to the input data in the machine learning and maps input patterns into the target space. 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 , y_i belong to the same class and large if they belong to different classes.

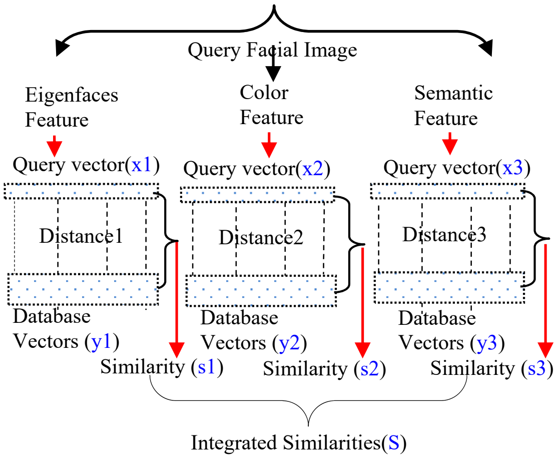


Figure.1: Different Similarity for Different Features.

The proposed addressing is based on learned similarity metric through the radial basis function neural network (RBFN) machine learning technique.

The proposed (SCBFIR) involves the retrieval of facial images based on their visual content and the semantic content including the semantic features such as race, age, gender, face shape, etc. SCBFIR based on the combination of content-based image retrieval (CBIR) and face recognition techniques, face detection, face segmentation, semantic features model, and the newly proposed integration method RBFN.

2. RELATED WORKS

(CBFIR) is a computer-based vision technique that is applied to the problem of facial 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 [14]. An early attempt to automate such systems was by [15, 16], who developed Compusketch, 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.

A mechanism where query images are submitted as color sketches was applied in the works by[15] [16] [17] [18] [19], 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. 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. 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. In a similar work by [20], the synthesizing process involves choosing similar faces and combining them. 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. In the works by [1, 2] [3-5] [6-10] [18] [21] facial retrieval systems deal only with the low level features and face model such as structural information and the connectivity of the edge points of the face objects characteristics[22], the landmarks of the face[23], Haar-like features [24, 25], statistical and structural information of the local feature sets of the face[26], PCA[27], the edges of a face image and its corresponding face components [28], identity based quantization[6], facial marks[4, 29], the geometric face attributes[30], LDA[31] and LBP (local binary patterns) feature[32]. The retrieval objective in most of these approaches is simply to match images and display top images. They do not capture the face's semantic aspects, especially when the query is some kind of user description. Facial image shape and texture features are fused to make the learned representation for convolutional neural networks for fast large-scale face retrieval [33]. However, a general description of the face is semantic (verbal) in nature.

A relevance feedback technique was suggested in the works by [34] for the semantic problem. 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, the appropriate user's feedback will not be offered. Relevance feedback does not provide semantic retrieval functionality for users.

Deep learning applies multiple processing layers to learn representations of face with multiple levels of feature extraction achieved through a series of works in recent years[35-39]

Some current image semantic retrieval systems applied for the general domain purpose such as in [40], who classified images into one of the many predefined categories such as 'indoor' or ' outdoor' images. Evidently, the description ability for such a 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 image's features. However, the image semantics should be defined in a more complete manner. In [41], a facial image retrieval model was developed, while the semantic features were only represented through symbols to be used for pruning as well as narrowing down the search space during the retrieval process. In [42] a probabilistic approach was proposed for face retrieval. Hybrid Markov Chain sampling model was applied to perform the localization of the facial features. In [43], they employed convolutional neural network (CNN) features for face image retrieval. They constructed a compact yet discriminative subset of raw deep transferred CNN descriptors to add scalability to the task of face image retrieval. while CNN unable to capture the most semantic features of face.

Most early efforts in facial image retrieval problem focused on solving this problem completely within a query and retrieval based on image content [4, 44]. Previous works on facial image retrieval have not addressed the semantic facial image retrieval problem. While facial image retrieval systems should offer maximum support in removing the gap between the low-level visual features and the depth of human semantics. In our research, a new proposed model that integrates the facial image low-level features with identified semantic facial features (high-level features) is introduced. Visual features are extracted automatically while facial images are annotated with semantic terms, enabling a user to specify his or her queries through natural language descriptions together with a query by example.

3. FEATURES EXTRACTION

Retrieval and recognition systems are based on features extraction, which maps the image from the image space to the feature space as depicted in Figure 2. The image content may include visual content called the low-level features and semantic content called the high-level features. Visual content can be classified into general or domain-specific types. General visual contents are application-independent features, such as color and texture. Domain-specific visual contents are application - dependent features, such as the features of the 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 such as the PCA-eigenfaces method [45]. 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[44].

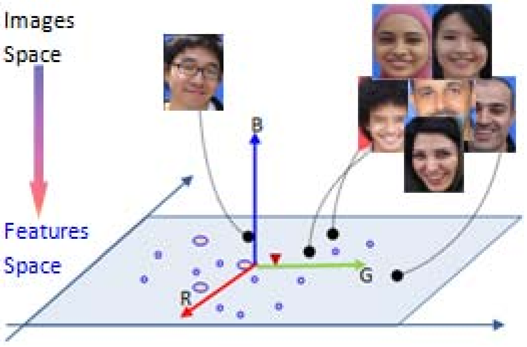


Figure 2: Feature Extraction Transfers The Images Into Various Content Features.

3.1 Visual Features Extraction

To improve the performance of the features extraction method, the position and size of a face in the entire image is detected and segmented before the extraction task. The face detection and segmentation methods described in the works by [2] are used. Each detected facial image is segmented into four partitions based on human eyes and mouth and the ratio of their respective heights to face height through a template-matching technique proposed. A combination of the features vectors of each sub-image, independently extracted represents the feature vector of the facial image. In this research, a color histogram is used as general visual content, while eigenfaces features are employed as domain-specific visual content [46]. Eigenfaces 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)[45]. Each eigenface 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. The "best" eigenface 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. Figure 3 shows the cycle of eigenfaces extraction.

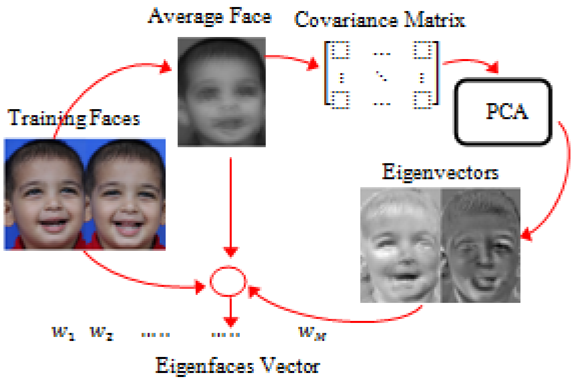


Figure 3: Example of Eigenfaces Extraction.

The dimension of the eigenfaces vector is based on the size of the training set. Based on the experimental results, the suitable size of the eigenfaces vector used in this research for facial image retrieval is 10. The use of principal components to represent human faces was originated by Sirovich and Kirby [47] and used by Turk and Pentland [48] for face detection and recognition.

The second visual feature used in this research is the color histogram. Its an important feature that has enabled recognition of images by humans. As Human face contains unique characteristics of color distribution, in this research, color histograms are used to capture the special relations of these unique regions characteristics.

Generally, 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 comparison 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 by counting the number of 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. Color quantization method is used to reduce the number of colors available in an image. 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.

$$h(i) = \frac{h(j)}{N}, j \in [0, \dots, M]. \qquad (1)$$

Figure 4 shows an illustration of color histogram extraction.

3.2 Semantic Features Model

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 [49] (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). Human facial image includes a variety of these semantic attributes that represent the features of the face and the visual impressions such as Demographic features (such as age, race, and gender), Skin color of a face and face parts, face parts descriptions, Description of other components and accessories of a face such as hairstyle, moles, a pair of glasses etc., and impressions implied from a face image, using descriptive keywords for character or personalities, such as "serious," and "happy", these features are used for recognizing faces and characterizing them. Semantic features of the human face are expressed by semantic description concepts. Each description consists of keywords and specification of sizes and lengths of face parts. Using these semantic description 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. Human beings measure the similarity among faces using these semantic concepts, so the machine should be ready to meet human natural tendencies and needs. In our proposed technique, these description features are later organized into a vector. Each description is

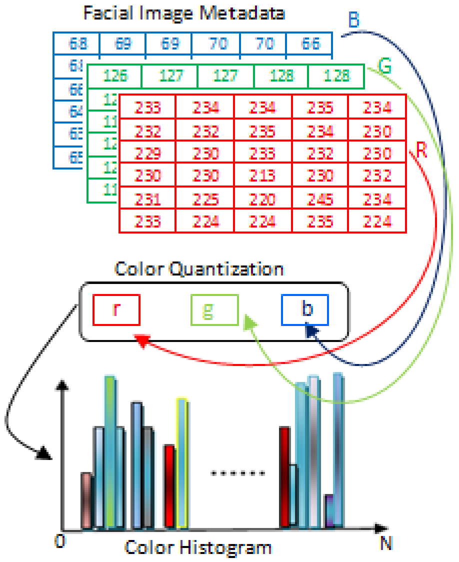


Figure 4: Illustration of Color Histogram Extraction.

called a face description vector and later used in the search and comparison process to retrieve the desired facial images in the images database.

The semantic features used in this research were weighted and selected based on a case study of 100 volunteers of different gender, race, and age. Each volunteer was required to rank 20 traits, based on their significance and ability to distinguish among individuals.

Each feature was given different rankings by different participants. For example, skin color was ranked as *R1* by 3 participants, *R2* by 20 participants, *R3* by 15 participants and so on. To compute the ranking and weight of each feature, the following proposed statistical analysis was applied[50] :

$$\mathrm{w(F)} = \sum_{i=1}^{n} x_i . \frac{1}{i} \qquad (2)$$

Where *w* is the weight of feature *F* given by the participants. The 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, *R1* is weighted heavier than the feature in the second position, *R2*, which is similarly weighted heavier than the feature in the third position *R3*, 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. Based on the results of the case study, 17 relevant features involving 70 different concepts were used in this research, as shown in Table 1.

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 [51, 52], modulated as

$$\mathrm{CfIFf}(C_i, F_j) = \mathrm{Cf}(C_i, F_j).\log\left(\frac{|\mathrm{TF}|}{\mathrm{Ff}(C_i, F_j)}\right) \qquad (3)$$

where | TF | denotes the number of facial images in the training set, $\mathrm{Cf}(C_i, F_j)$ denotes the number of times concept C_i occurs in F_j, and $\mathrm{Ff}(C_i, F_j)$ denotes the frequency of the facial images 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

$$w_{ij} = \frac{\mathrm{CfIFf}(C_i, F_j)}{\sqrt{\sum_{k=1}^{|TF|}(\mathrm{CfIFf}(C_k, F_j))^2}} . \qquad (4)$$

Table 1: 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 Color	Black, Brown, Tan, White
Hair Color	Black, Brown, Blond, Red, Gray, Covering
Hair Length	Short, Medium, Long, Bald, Covering Head
Hair Type	Curly, Wavy, Straight, Covering Head
Eye Color	Black, Brown, Blue, Green
Glasses Shape	Oval, Circular, Square, Rectangle
Moustache Size	Short, Medium, Long
Beard Size	Short, Medium, Long
Facial	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

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 as the following form:

$$\mathrm{Fk}_i = (k_1, k_2, \ldots, k_n) \qquad (5)$$

$$\mathrm{Fv}_i = (v_1, v_2, \ldots, v_n) \qquad (6)$$

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.

During the query process, the user specifies the suitable attributes of the queried facial image based on visual features of individual images (face color, race, gender, etc.). The retrieval mechanism maps the individual concepts to the predefined weights in the matrix of semantic concept values, previously built, to generate the query value vectors, as shown in Figure 5.

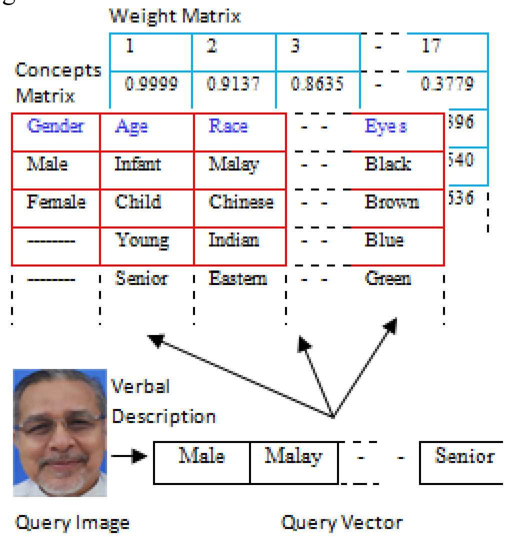


Figure 5: Generation of The Query Value Vectors of The Semantic Description.

4. HETEROGENEOUS FEATURES INTEGRATION

This section introduces a new innovated approach, through RBFN, to address the issue of features integration and similarity measurement.

An RBFN is an artificial neural network that uses radial basis functions as activation functions. Its theory is based on the function of approximation theory. It has the architecture of a traditional three-layer feed-forward neural network. In its most basic design form, it consists of three separate layers with respective feed-forward architecture as shown in Figure 6. 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 activation function of the RBFN is expressed as follows:

$$h_{i,k} = (g(\|x_k - c_i\|^2)^{\frac{1}{2}} \quad , \quad k \in \ \{1,2,\dots,M\}, \mathrm{i} = 1,2,\dots.\mathrm{r} \qquad (7)$$

where *x* is an *n*-dimensional input feature vector, *c* is an n-dimensional vector called the center of the RBF unit, *M* the number of training vector, and *r* is the number of the RBF units.

In this research, we proposed a learned similarity metric through RBFN. To construct an RBF network, the number of input nodes in the input layer of the neural network is 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:

$$\mathrm{length}(\mathrm{C_i}) = \mathrm{length}(\mathrm{x_i})\,, \mathrm{i} = 1,2,..,\mathrm{L} \qquad (8)$$

Where, $\mathrm{x_i}$ is the feature vector of the features class i, $\mathrm{C_i}$ is the RBF center of vector i, and L is the number of features classes. The first training vector is fed to the RBFN center vectors as a query vector, while the other training vector is input to the network. The output of each neuron is then computed. To compute the error (target output minus actual output), the Sum Square Error (SSE) is used. Once the error is calculated, the learning rule would adjust the weights 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 center. This process is repetitively continued until the mean square error (MSE) is less than an acceptable value.

The proposed method is described as follows:

1) Select η , Є 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* Є *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 the 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 :

$\boldsymbol{c_i = x_{qi}}$, $\boldsymbol{i} \in \{\mathbf{1, 2, .., I}\}, \boldsymbol{q} \in \{\mathbf{1, 2, .. M}\}$, $\boldsymbol{x_q}$ *is* the query vector. (9)

5) Compute the initial response:

$$h_{i,k} = (g(\|x_k - c_i\|^2)^{\frac{1}{2}} \quad , \quad \forall k, i. \qquad (10)$$

$$\mathbf{h_k} = \left[\mathbf{h_{1,k}}, \ldots \mathbf{h_{I,k}}\right]^{\mathbf{T}}, \forall \mathbf{k}. \qquad (11)$$

$$\boldsymbol{y_{j,k} = w_{i,j} h_k \; \forall k, j}. \qquad (12)$$

6) Compute SSE=$\sum_{k=1}^{M} \sum_{j=1}^{n} (d_{j,k} - y_{j,k})^2$. (13)

7) Update the adjustable parameters
Set $\boldsymbol{c_i = x_{qi}}$. (14)

8) $\boldsymbol{\varepsilon_{j,k} = d_{j,k} - y_{j,k} \; \forall k, j}$. (15)

$$\boldsymbol{w_{ij}} \leftarrow \boldsymbol{w_{ij}} + \eta \sum_{\boldsymbol{k=1}}^{\boldsymbol{M}} \boldsymbol{\varepsilon_{j,k} \, h_k}, \forall. \qquad (16)$$

9) Compute the current response $h_{i,k}$, h_k and $y_{j,k}$ using equations (10), (11), and (12).

10) Compute the *SSE* using the equation (13).

11) Compute the mean square error:
MSE = SSE/M
If: (*MSE* > Є then go to step 7. Where *n* is the number of the neuron in the output layer, $j \in \{ 1, 2, \ldots, n\}$.

In this research, the Gaussian function in equation (17) are used to compute the respective weights of each features class. This maps the input patterns into the target space. An appropriate transformation is applied to the data to emphasize the most discriminative direction of each features class.

$$h(\|X - C_i\|) = exp\left(\frac{(\mathrm{X} - c_i)^2}{\sigma_i} \right) \qquad (17)$$

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

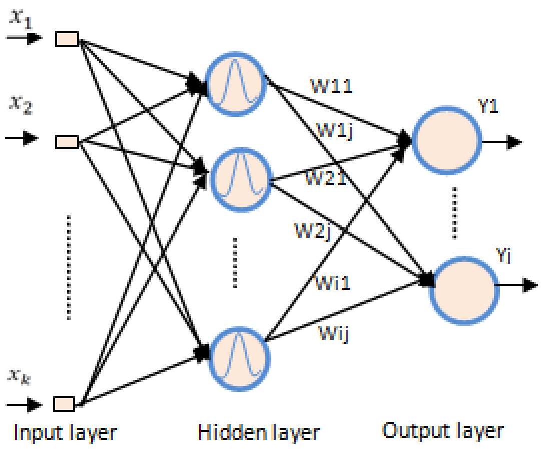


Figure 6: The RBFN Architecture.

Our proposed method is based on injecting the query vector of class *i* to the center C_i of the RBF as shown in Figure 7.

The Gaussian function is conducted as the similarity metric. The trained weight from the RBFN training stage represents the weight parameter for the similarity metric. Figure 8 shows the proposed similarity metric. During the query process, the proposed similarity metric computes the distance between the query and the database vectors.

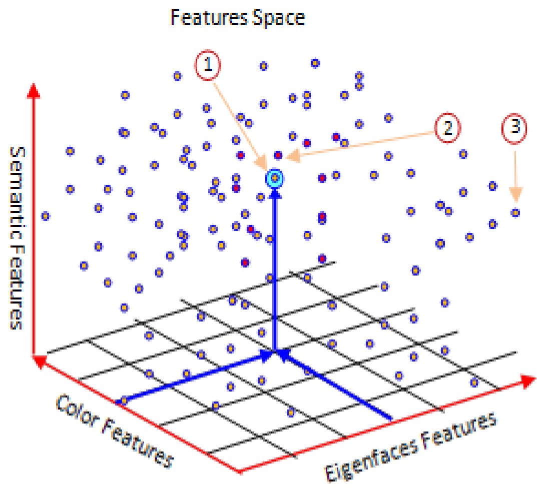


Figure 7:The Query Vector Is Fed to The RBFN Center: (1) Is the Center, (2) Vector Near to The Center and (3) Vector Far from The Center.

The output is then weighted using the respective weights. Figure 9 shows the overall network architecture based on the proposed method.

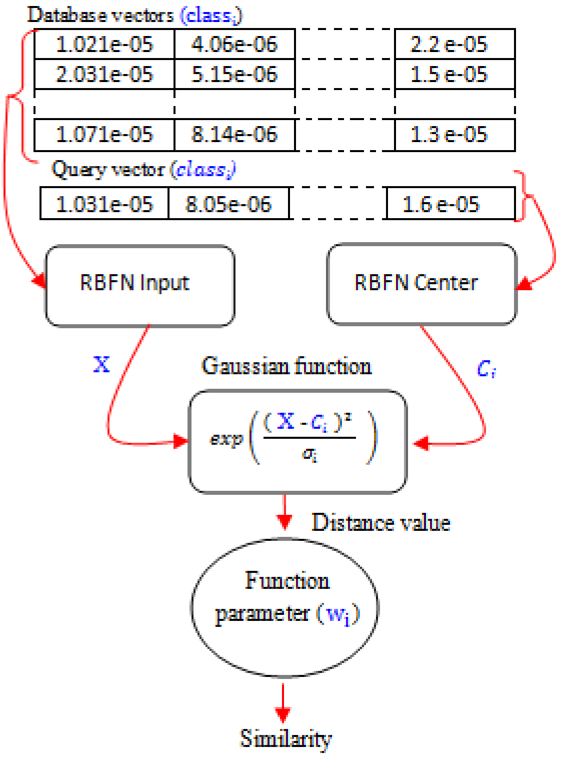


Figure 8: The Proposed Learned Similarity Metric To Overcome the Problem Of Integrating Heterogeneous Features.

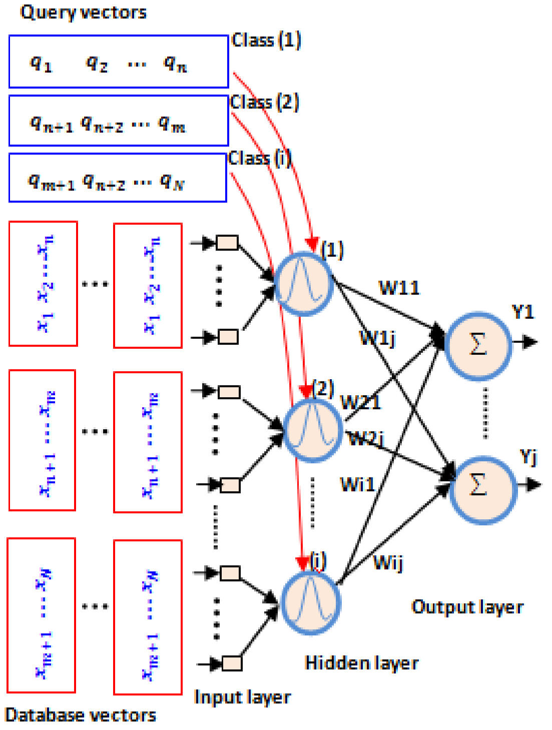


Figure 9: The RBFN Architecture with The Proposed Method.

5. EXPERIMENTAL RESULTS AND DISCUSSION

In this research, a number of experiments were conducted to assess and evaluate the proposed methods of semantic content-based facial image retrieval. In doing so, two databases were used. The first database is the Olivetti Research Laboratory (ORL) database of faces available at the AT&T Laboratories Cambridge website. This is a well-known, publicly available and has been used as a standard database in many face recognition systems. The second database is a local facial images database that was developed in the University of Malaya (UM) in Kuala Lumpur to be used in the current research. The database is huge in size and contains color images with heterogeneous contents, different facial expressions (e.g., happy, sad, smiling, angry, etc.) and facial details (e.g., glasses, beard, mustache, and facial marks), and a variety of semantic features such as gender, race, and age.
The ORL and local databases consist of 400 facial images from 40 participants and 1500 facial images from 150 participants, respectively. In our experiments, 200 and 750 images (5 images for each participant) representing 50 % of the two databases respectively were randomly selected for training, and the remaining images were used for testing. Precision and recall methods were applied to measure the performance efficiency of the retrieval methods. The definitions of precision and recall are represented in the equations below as well as in Figure 10.

$$\text{Recall} = \frac{\text{Relevant Faces of The Retrieved Faces}}{\text{Total Relevant Faces}}, \quad \text{Precision} = \frac{\text{Relevant Faces of The Retrieved Faces}}{\text{Total Retrieved Faces}}. \quad (18)$$

In order to evaluate the system performance during the retrieval process, the number of images to be retrieved is subjected to certain pre-determined values. Hence, in both methods of precision and recall, 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 number of queries means the number of the system runs. With the ORL database, the number of queries is 200, which is equivalent to the number of testing image sets, while with the local database the number of queries is 750, which is also equivalent to the number of testing image sets.

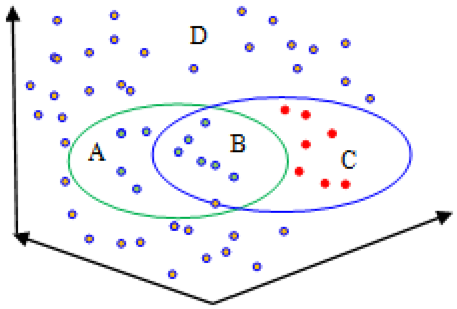


Figure 10: A Is Un-Retrieved Relevant Faces, B Is Retrieved Relevant Faces, C Is Retrieved Irrelevant Faces, and D Is Un-Retrieved Irrelevant Faces.

5.1 Model Training

In our proposed method, weights are generally assigned through the RBFN to each class of information extracted from an image and an overall similarity is then computed. Images are then ranked based on this similarity computation.

The training stage results of the RBFN are shown in Figures 11 and 12, 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 fed 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 SSEs computed.

Figures 13 and 14 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 weights of the network.

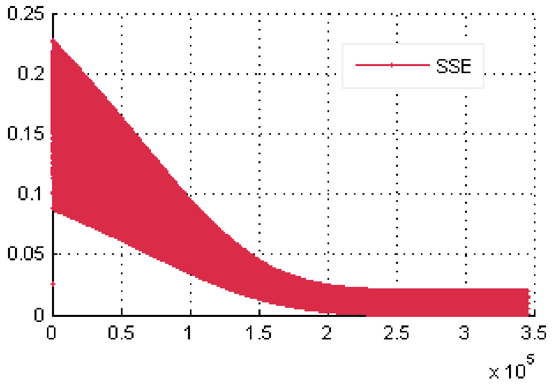


Figure 11: Sum Squared Errors of All Input Training Vectors on The ORL Database.

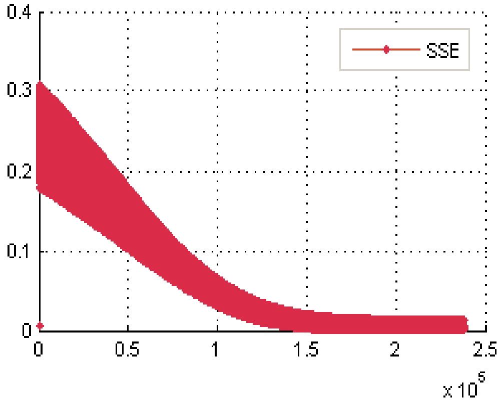


Figure 12: Sum Squared Errors of All Input Training Vectors on The Local

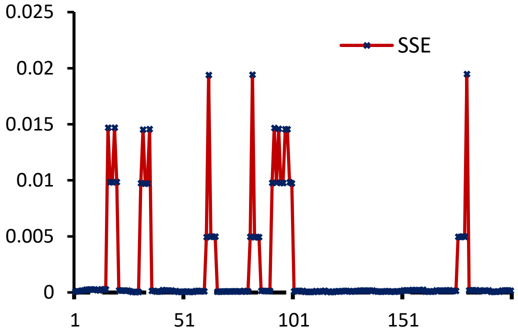


Figure 13: Sum Squared Errors of The Last Cycle of The Network Training on The ORL Database.

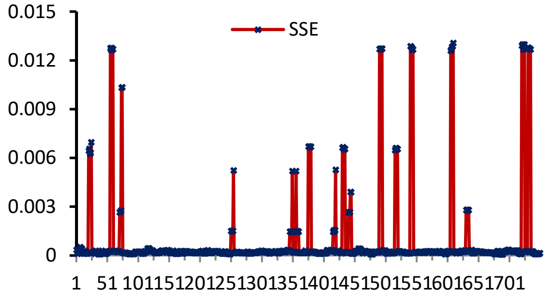


Figure 14: Sum Squared Errors of The Last Cycle of The Network Training on The Local Database

Figures. 15; 16 show the mean squared errors (MSE) of the network's training on the ORL and local databases respectively. The MSE is computed to monitor and measure the performance of the network's training. The network's training should be stopped when the MSE is less than the network's 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 database. This is attributed to the variety, size, and color of the local database images.

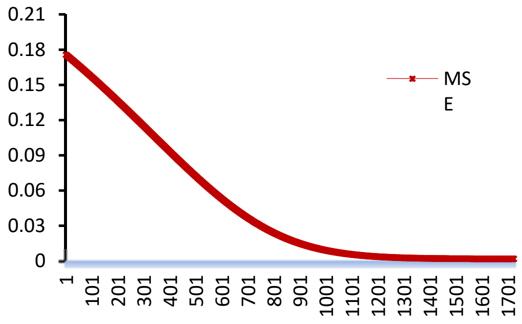


Figure 15: Mean Squared Error of The Network Training on the ORL Database.

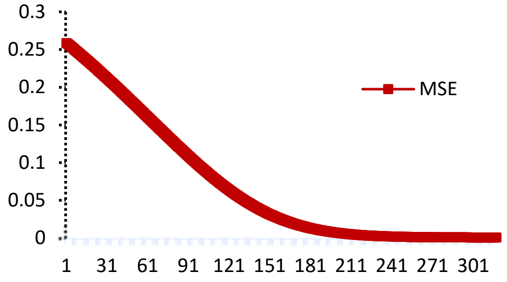


Figure 16: Mean Squared Error of The Network Training on The Local Database.

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 excellent performance with the testing image from the ORL database and good performance with the testing image from the local database. On the other hand, the network weights acquired through the local database showed 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.

5.2. Model Testing

Tables 2; 3 and Figures. 17; 18 show the experimental results of the facial image retrieval testing based on the proposed method, 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 2 - 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%.

The results on the local database (Table 3) 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 %.

In the next section, the example taken in Figure 19 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 method

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 shown in Figure 19.

As shown in Figure 20, the recall method of performance gives an accuracy of 60 % within the top 10 cut-off levels based on eigenfaces features of 10-dimension vectors. Figure 21 shows an accuracy of 70 % within the top 10 cut-off levels based on color histogram features comprising 4-Red, 4-Green, and 4-Blue color space distribution. Figure 22 shows an accuracy of 80 % within the top 10 cut-off levels based on the integration of the eigenfaces and color histogram. However, when the proposed method was used on the same visual-semantic query (Figure 19) for facial image retrieval based on the integration of the three features - eigenfaces, color histogram and semantic features, the accuracy achieved is

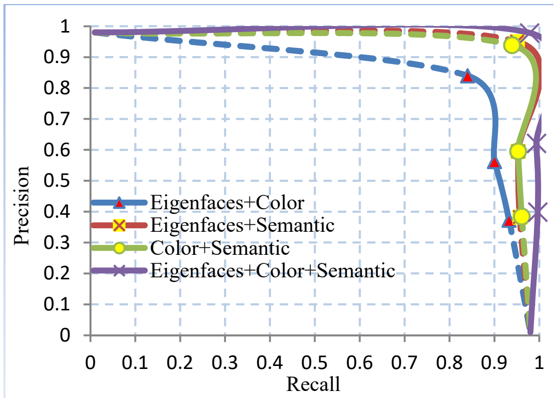


Figure17: Facial Image Retrieval Performance Using the Proposed Method on ORL Database.

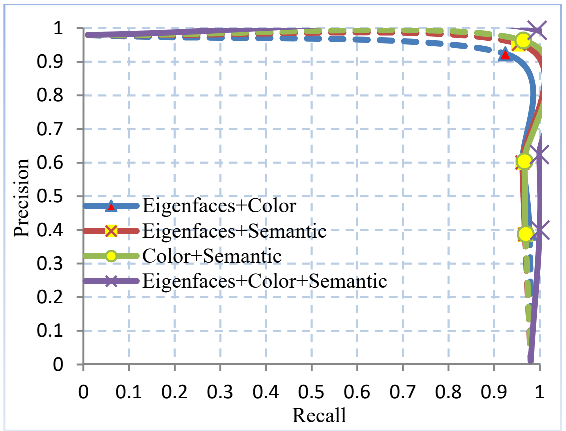


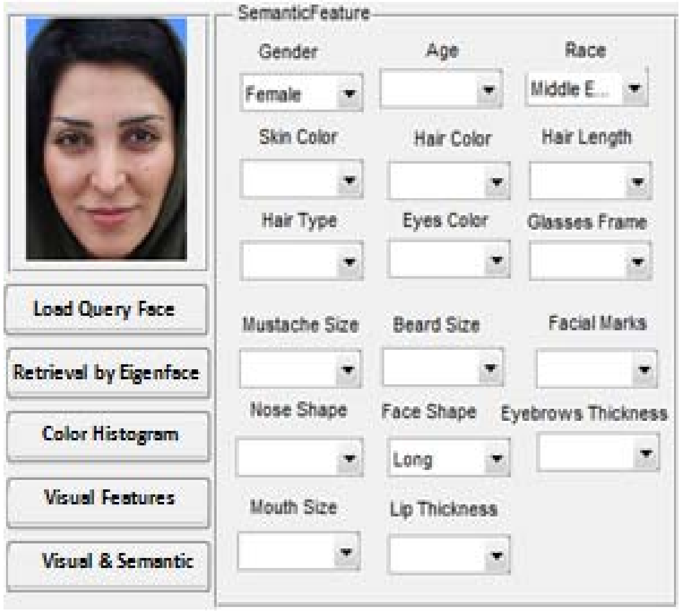
Figure 18: Facial Image Retrieval Performance Using the Proposed Method On Local Database.

Table 2: Integration of Features Classes using The Proposed Method on ORL Database.

Features	Top Faces	Retrieved Faces	Relevant Faces	Recall	Precision
Eigenface with Color	10	2000	1680	0.840	0.840
	16	3200	1800	0.9	0.5625
	25	5000	1865	0.9325	0.373
Eigenface with Semantic	10	2000	1901	0.9505	0.9505
	16	3200	1905	0.9525	0.5953
	25	5000	1913	0.9565	0.3826
Color With Semantic	10	2000	1878	0.939	0.939
	16	3200	1905	0.9525	0.5953
	25	5000	1922	0.961	0.3844
Eigenface, Color and Semantic	10	2000	1957	0.9785	0.9785
	16	3200	1987	0.9935	0.6209
	25	5000	1993	0.9965	0.3986

Table 3: Integration of Features Classes Using the Proposed Method on the Local Database.

Features	Top Faces	Retrieved Faces	Relevant Faces	Recall	Precision
Eigenface with Color	10	7500	6931	0.9241	0.9241
	16	12000	7215	0.962	0.6013
	25	18750	7331	0.9775	0.391
Eigenface with Semantic	10	7500	7152	0.9536	0.9536
	16	12000	7212	0.9616	0.601
	25	18750	7256	0.9675	0.387
Color With Semantic	10	7500	7228	0.9637	0.9637
	16	12000	7250	0.9667	0.6042
	25	18750	7268	0.9691	0.3876
Eigenface Color and Semantic	10	7500	7454	0.9939	0.9939
	16	12000	7493	0.9991	0.6244
	25	18750	7499	0.9999	0.3999



Gender Female

Race Middle E...

Face Shape Long

Figure 19: Semantic and Visual Query Example.

respectively 100% within the top 10 cut-off levels. The detail results are given in Figure 23. It is apparent that there is a significant improvement in the accuracies, where the majority of the relevant images were returned to the top ten results.

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

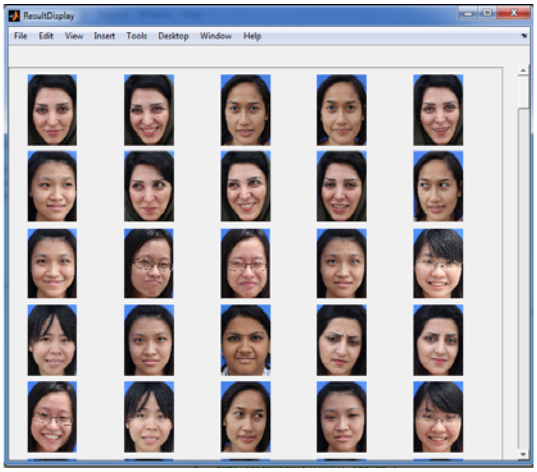


Figure 20: Facial Image Retrieval Based on Eigenfaces with 10-Dimention Vectors.

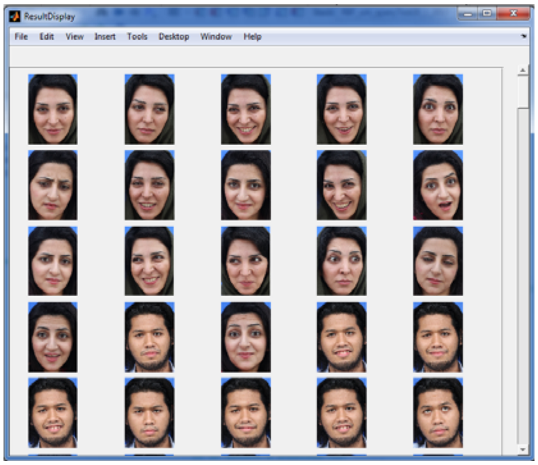


Figure. 21: Facial Image Retrieval Based on Color Histogram of (4-Red, 4-Green, 4-Blue) Color Space Distribution.

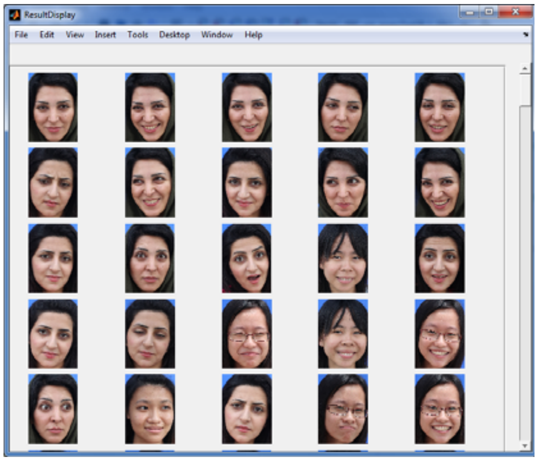


Figure 22: Facial Image Retrieval Based on Integration of Eigenfaces and Color features.

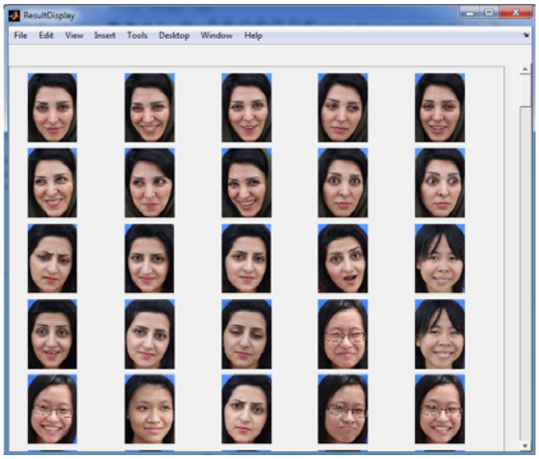


Figure 23: Facial Image Retrieval Based on Integration of Eigenfaces, Color, and Semantic Features

6. CONCLUSION

This research proposed a model that links the high-level query requirement and the low-level features of the human facial image. The proposed model improved the facial image retrieval by enabling the technique to meet human natural tendencies and needs in the description and retrieval of facial images. 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, (iii) color histogram and semantic features, and (iv) eigenfaces, color histogram and semantic features. The experimental results showed that as compared to the current content-based facial image retrieval technique, the proposed methods 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's content features. The current content-based facial image retrieval technique achieves 84.00% and 92.41% accuracy, while (SCBFIR) achieves 97.85 % and 99.39% accuracy for the ORL and local database respectively within the top 10 retrieved facial images. This because the benefits of the individual techniques were essentially merged and enhanced. In addition, each particular feature class covers the weaknesses inherent in the other classes, thus resulting in higher performance in the classification and retrieval processes.

The proposed model could be applied in law enforcement applications, where the verbal description of the witness is used to retrieve the similar facial images of the suspect's face from the criminal's mug shot database.

Future works are recommended for developing a method that will improve the correlation between human and machine perceptions of facial images.

ACKNOWLEDGMENT

Deanship of Scientific Research of Najran University has supported this work (Project no: NU/ESCI/15/082

REFERENCES:

[1] NGO TD, VU HT, Duy-Dinh L, SATOH Si. Face Retrieval in Large-Scale News Video Datasets. IEICE Transactions on Information and Systems. 2013;96(8):1811-25.

[2] Alattab AA, Kareem SA, editors. Efficient Method of Visual Feature Extraction for Facial Image Detection and Retrieval. Fourth International Conference on Computational Intelligence, Modelling and Simulation (CIMSiM) 2012: IEEE.

[3] Wang DH, Conilione P. Machine learning approach for face image retrieval. Neural Computing and Applications. 2012;21(4):683-94.

[4] Jain AK, Klare B, Park U. Face Matching & Retrieval: Applications in Forensics. IEEE Multimedia. 2012;19(1).

[5] Chen B, Chen Y, Kuo Y, Hsu W. Scalable Face Image Retrieval using Attribute-Enhanced Sparse Codewords. IEEE Transactions on Multimedia 2012.

[6] Wu Z, Ke Q, Sun J, Shum H-Y. Scalable face image retrieval with identity-based quantization and multireference reranking. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2011;33(10):1991-2001.

[7] Smith BM, Zhu S, Zhang L. Face Image Retrieval by Shape Manipulation. CVPR,20112011.

[8] Lee H, Chung Y, Kim J, Park D. Face image retrieval using sparse representation classifier with gabor-LBP histogram. Information Security Applications: Springer; 2011. p. 273-80.

[9] Kashani MAA, Ramezanpour M, editors. Face Image Retrieval Using Genetic Algorithm and Bags of Pixels. Eighth International Conference on Information Technology: New Generations (ITNG); 2011: IEEE.

[10] Shylaja S, Natarajan S, Balasubramanyamurthy K, Abhijit K, Diwakar J, Saifulla SM, editors. Aggregation of Gabor Wavelets And Curvelets With PCA For Efficient Retrieval of Face Images. Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India; 2010: ACM.

[11] Borovikov E, Vajda S, Gill M. Face Match for Family Reunification: Real-World Face Image Retrieval. Computer Vision: Concepts, Methodologies, Tools, and Applications: IGI Global; 2018. p. 735-53.

[12] Yang L JR, R. Distance Metric Learning: A comprehensive survey. Michigan State Universiy. 2006:1-51.

[13] Dokmanic I, Parhizkar R, Ranieri J, Vetterli M. Euclidean distance matrices: essential theory, algorithms, and applications. IEEE Signal Processing Magazine. 2015;32(6):12-30.

[14] Laughery KR, Fowler RH. Sketch artist and Identi-kit procedures for recalling faces. Journal of Applied Psychology. 1980;65(3):307.

[15] Penry J, Photography F, Kingdom U. Photo-fit. Forensic Photography. 1974;3(7):4-10.

[16] Caldwell C, Johnston V, editors. Tracking a criminal suspect through face-space with a genetic algorithm 4th international Conference on Genetic Algorithm,(ICGA'91); 1997; San Diego, CA, US: Morgan Kaufmann Publisher.

[17] Zhan S, Zhao J, Tang Y, Xie Z. Face image retrieval: super-resolution based on sketch-photo transformation. Soft Computing. 2018;22(4):1351-60.

[18] Gao X, Wang N, Tao D, Li X. Face Sketch–Photo Synthesis and Retrieval Using Sparse Representation. IEEE Transactions on Circuits and Systems for Video Technology. 2012;22(8):1213-26.

[19] Abdulbaqi HA, Sulong G, Hashem SH. A sketch based image retrieval: a review of literature. Journal of theoretical and applied

information technology. 2014;63(1):158-67.

[20] Frowd CD, Hancock PJB, Carson D. EvoFIT: A holistic, evolutionary facial imaging technique for creating composites. ACM Transactions on Applied Perception (TAP). 2004;1(1):19-39.

[21] Rakesh S, Atal K, Arora A, Purkait P, Chanda B. Face image retrieval based on probe sketch using SIFT feature descriptors. Perception and Machine Intelligence: Springer; 2012. p. 50-7.

[22] Gao Y, Qi Y. Robust visual similarity retrieval in single model face databases. Pattern Recognition. 2005;38(7):1009-20.

[23] Le DD, Satoh S, Houle ME, editors. Boosting face retrieval by using relevant set correlation clustering2007: IEEE.

[24] Bau-Cheng S, Chu-Song C, Hui-Huang H. Face Image Retrieval by Using Haar Features. 19th International Conference on Pattern Recognition (ICPR 2008) Tampa, Florida2008. p. 1-4.

[25] Zhang N, Jeong H-Y. A retrieval algorithm for specific face images in airport surveillance multimedia videos on cloud computing platform. Multimedia Tools and Applications. 2017;76(16):17129-43.

[26] Zhong D, Defee I, LEFEVRE S. Face Retrieval Based on Robust Local Features and Statistical-Structural Learning Approach. EURASIP Journal on Advances in Signal Processing. 2008;2008(18).

[27] Shih P, Liu C. Comparative assessment of content based face image retrieval in different color spaces. International Journal of Pattern Recognition and Artificial Intelligence - IJPRAI 2005;19(7):1039-48.

[28] Kam-Art R, Raicharoen T, Khera V, editors. Face recognition using feature extraction based on descriptive statistics of a face image2009.

[29] Park U, Jain AK. Face matching and retrieval using soft biometrics. IEEE Trans Inf Forensics Secur. 2010;5(3):406-15.

[30] Smith BM, Zhu S, Zhang L, editors. Face image retrieval by shape manipulation. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2011: IEEE.

[31] Kim T-K, Kim H, Hwang W, Kittler J. Component-based LDA face description for image retrieval and MPEG-7 standardisation. Image and Vision Computing 2005;23(7):631-42.

[32] Nguyen TN, Ngo TD, Le D-D, Satoh Si, Le BH, Duong DA, editors. An efficient method for face retrieval from large video datasets. Proceedings of the ACM International Conference on Image and Video Retrieval; 2010: ACM.

[33] Lu Z, Yang J, Liu Q. Face image retrieval based on shape and texture feature fusion. Computational Visual Media. 2017;3(4):359-68.

[34] Pagare R, Shinde A. A Study on Image Annotation Techniques. International Journal of Computer Applications. 2012;37(6):42-5.

[35] Schroff F, Kalenichenko D, Philbin J, editors. Facenet: A unified embedding for face recognition and clustering. Proceedings of the IEEE conference on computer vision and pattern recognition; 2015.

[36] Sun Y, Chen Y, Wang X, Tang X, editors. Deep learning face representation by joint identification-verification. Advances in neural information processing systems; 2014.

[37] Sun Y, Wang X, Tang X, editors. Hybrid deep learning for face verification. Proceedings of the IEEE international conference on computer vision; 2013.

[38] Taigman Y, Yang M, Ranzato M, Wolf L, editors. Closing the gap to human-level performance in face verification. deepface. IEEE Computer Vision and Pattern Recognition (CVPR); 2014.

[39] Wen Y, Li Z, Qiao Y, editors. Latent factor guided convolutional neural networks for age-invariant face recognition. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.

[40] Chen Y, Wang JZ. Image categorization by learning and reasoning with regions. The Journal of Machine Learning Research. 2004;5:913-39.

[41] Alattab AA, Abdul kareem S. Facial Image Retrieval based on Eigenfaces and Semantic Features. First International Conference on Advances in Computer and Information Technology; Kuala lumpure 2012.

[42] Sridharan K, Nayak S, Chikkerur S, Govindaraju V. A Probabilistic Approach to Semantic Face Retrieval System,. Audio- and Video-Based Biometric Person Authentication. Lecture Notes in Computer

Science. 3546: Springer Berlin / Heidelberg; 2005. p. 85-100.

[43] Banaeeyan R, Lye H, Fauzi MFA, Karim HA, See J. Semantic facial scores and compact deep transferred descriptors for scalable face image retrieval. Neurocomputing. 2018;308:111-28.

[44] Beham MP, Roomi SMM. A review of face recognition methods. International Journal of Pattern Recognition and Artificial Intelligence. 2013;27(04).

[45] Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2016;374(2065):20150202.

[46] Mishra S, Dubey A. Face recognition approaches: a survey. International Journal of Computing and Business Research (IJCBR). 2015;6(1).

[47] Sirovich L, Kirby M. Low-dimensional procedure for the characterization of human faces. Journal of the Optical Society of America A. 1987;4(3):519-24.

[48] Turk M, Pentland A. Face recognition using eigenfaces. IEEE Conf on Computer Vision and Pattern Recognition, 1991. p. 586-91.

[49] Alattab AAA. Retrieval of human facial images based on visual content and semantic description: University of Malaya; 2013.

[50] Alattab AA, Kareem SA, editors. Semantic Features Selection and Representation for Facial Image Retrieval System. 4th International Conference on Intelligent Systems Modelling & Simulation (ISMS); 2013: IEEE.

[51] Salton G, Buckley C. Term-weighting approaches in automatic text retrieval. Information processing & management. 1988;24(5):513-23.

[52] Sebastiani F. Machine learning in automated text categorization. ACM computing surveys (CSUR). 2002;34(1):1-47.