Efficient Method of Visual Feature Extraction for Facial Image Detection and Retrieval

Ahmed Abdu Alattab, Sameem Abdul Kareem
Faculty of Computer Science & Information Technology
University of Malaya
50603 Kuala Lumpur, Malaysia
ahmed_alattab@yahoo.com, sameem@um.edu.my

Abstract—Due to the significant increase in the already huge collection of digital images that we have today, it has become imperative to find efficient methods for the archival and retrieval of these images. In this research, a content-based-human facial image detection and retrieval model is proposed for retrieving facial images of humans based on their visual content from an image database. The research proposes a technique of face segmentation based on which a new method of features extraction from the human face is devised. The capability and effectiveness of the color space models (RGB, HSV, and HSI) on facial image retrieval technique are also investigated. Eigenfaces features are used as a domain specific visual content to extract the characteristic feature images of the human facial images, while the color histogram of the facial image is used as a general visual content. Viola-Jones face detection method is employed to obtain the location, extent and dimensions of each face. Moreover, for the measurement of distance and classification purposes, Euclidean distance is utilized. The sample image database consists of 1500 local facial images of one hundred and fifty participants from the University of Malaya (UM), Kuala Lumpur, and some of their friends and families outside the UM. Several experiments based on precision and recall approach were conducted to evaluate the proposed methods. The retrieval result of the facial image given by the proposed method showed excellent improvement comparing to those achieved when using the traditional method of visual features extraction.

Keywords—face retrieval; face recognition; face detection; eigenfaces; color space; segmentation.

I. INTRODUCTION

In recent years, there has been a significant increase in the number of digital images and videos that we need to store and manage, owing to the immense popularity and use of image capturing devices such as digital cameras and image scanners. As such, it has become essential to find new and efficient means for archiving and retrieving such massive amount of digital content. Amongst these images, facial images have gained its importance due to its use in various aspects of life such as in airports, law enforcement applications and security systems.

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, which make the exact retrieval of the target face difficult or 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.

On the other hand, content-based image retrieval is a technique to find and retrieve images from a database using the visual contents of the image. Most of the basic image retrieval systems utilize low-level visual features extracted automatically using image-processing methods to represent the raw content of the image [1]. These features can be classified into general features or domain specific features. General visual content consist of the application independent features such as color, texture, shape, spatial relationship, and so on. On the other hand, domain specific visual content includes application dependent features such as human faces, fingerprint, etc. Retrieving image based on color usually yields images with similar colors, while image retrieval based on shape yields images that have clearly the same shape, and so on. Thus, such a system when utilized for the purpose of image retrieval using the general visual content is not effective with facial images.

Moreover, most of the features in the human facial image are domain specific. The features for a facial image are extracted using two methods. In the first method, the information theory concepts are employed by preparing a computational model that gives the best description of a face. This is done by deriving the related information included in the face. Eigenfaces method is one such method where information for the best description of a face is extracted from the facial image. The second is the components based method, where deformable templates and active contour models with excessive geometry and mathematics are used to extract the feature vectors of the facial parts such as nose, mouth, eyes and chin. Efficient retrieval system requires a robust method for extracting the features extraction that has the ability to achieve satisfactory retrieval performance. However, there are many intrinsic and extrinsic factors that can degrade the performance of facial image retrieval, with intrinsic factors such as facial expressions and aging that mostly affects facial appearance, and extrinsic factors such as illumination variation and pose change that affect the images.

The objective of this research is to take the above facts in consideration and investigate methods that can improve the performance of the current CBFIR system for the purpose of facial image retrieval.
II. RELATED WORKS

The term content-based image retrieval was first used by T. Kato in 1992 [2]. He was interested in how shape and color information could be used to query a database of images. The term has been used to describe any technique that uses the visual content for search and retrieval purpose. QBIC, from IBM [3], was one of the earliest CBIR systems that used the colors to retrieve images from a database. Wang and other researchers [4] used a segmentation model to divide the image into regions that ideally correspond to different objects and use those regions for retrieval. However, Leibe and his team [5] used a segmentation method as a way of integrating individual image cues showing that this combination scheme increases detection performance compared to any cue in isolation. In a joint study, Navarrete and Ruiz-Del-Solar [6] organized facial images in a tree structured self-organizing map. Projections of the principal component analysis (PCA) were used to form the map for features representation of the facial image in the image space. Each facial image represents a cluster in the whole image space. The objective of the study was a quick search and retrieval of the images in the database. In [7], a unified framework of structural information and statistical aspects of pattern description is proposed for pattern retrieval based on local features. However, the method is not limited to facial images. More in depth reviews on image retrieval are provided in [1, 8]. The currently used systems still have many limitations, especially in domain-specific applications, such as facial image retrieval. In this research, content-based human facial image detection and retrieval model is proposed for retrieval human facial images from the database based on a new proposed method of human facial image segmentation.

III. FACE DETECTION

The aim of face detection in this research is to give the position and size of the face in the entire image. Learning-based face detection techniques are the most successful methods in terms of detection accuracy and speed. One of the most popular methods employed in this research is Viola & Jones method [9]. The technique relies on a set of simple rectangular features, known as Haar-like features, to detect the intended face. These features are resemblance of Haar basis functions, which have been used by Papageorgiou and other researchers [10]. Integral image concept was first introduced as a new image representation technique. By utilizing this technique, rectangular features can be computed quickly. Integral image of any position is the sum of the pixels above and to the left of the position. For instance, the integral image at position x and y is the sum of the pixels above and to the left of x and y, then

\[ ii(x, y) = \sum_{m=0}^{x} \sum_{n=0}^{y} i(m, n). \]  

(1)

Here i(x,y) represents the original image and ii(x,y) is the integral image as shown in Fig.1. Four kinds of rectangles features are used with varying numbers of sub-rectangle: a tow horizontal rectangle features, a two vertical rectangle feature, a three-rectangle feature and a four-rectangle feature, as shown in Fig.1. Within any image sub-window, the total number of Haar-like features, that will be generated based on the integral image method and the four rectangles features, is very huge. The feature set should be reduced to a small number of important features. AdaBoost learning method [11] was successfully employed to select a restricted number of critical features, which are used to create very efficient classifiers. Boosting is used for performing supervised learning. The key idea is to use a set of weak learners (classifiers) in order to create a single strong learner. Such as, a weak classifier \( h_j(x) \) of a feature f, a threshold \( \theta_j \) and a parity \( p_j \) indicating the direction of the inequality sign [12]:

\[ h_j(x) = \begin{cases} 1, & \text{if } p_j f_j(x) < p_j \theta_j \\ 0, & \text{otherwise} \end{cases} \]  

(2)

The strong classifier is shown in Eq. (3) where \( h_l(x) \) is the weak classifier, \( a_l \) is the coefficient for \( h_l \). Given a test sub-window x, it will be classified as a face if only the output is one.

\[ h(x) = \begin{cases} 1, & \text{if } \sum_{l=1}^{L} a_l h_i(x) > \theta \\ 0, & \text{otherwise} \end{cases} \]  

(3)

Viola & Jones arranged the classifiers in cascade architecture as shown in Fig. 1.

![Figure 1. Face detection algorithm: four-rectangular features (a, b, c, and d).](image-url)
In the cascade architecture, a series of classifiers are applied to every sub-window. Negative sub-windows will be rejected and the positive ones sub-windows will be detected and selected in the beginning stages by the initial classifier with fewer features and less computational time. The cascade classifiers in the final stages with more features classify only the sub-windows that have passed the initial classifier. After several stages of processing the background region will be eliminated, while the focus will be more on the face-like region. The window that is used for scanning can be scaled to detect faces at multiple scales, as well as the features evaluated. The detector is applied on gray-scale images.

IV. FACIAL IMAGE SEGMENTATION

Facial image segmentation plays an essential role in some face detection systems in helping to extract only face part of a given large image based on skin-colors and non skin-colors classification [13].

In this study, a new facial image segmentation technique is proposed to improve the accuracy of the facial image retrieval performance. The proposed method is based on the fact that every sub facial image has its spatial information of orientation and specific scale relevant to this sub-image. Combination of the features vectors of each sub-image, independently extracted, is expected to produce more robust features vector. The research suggests the facial image to be segmented into four partitions based on human eyes and mouth level and the ratio of their height to face height, based on the assumption that image will always have at least one face. Each detected facial image is scaled to fixed size beginning by face detection step to optimize the candidate for face segmentation. In order to segment candidate’s faces in the image, a template matching technique needs to be employed. The template consists of four sub-templates as shown in Fig. 2. The first sub-template is used for matching the upper region of the face until the eyes level, the second for matching the middle region of the face between the eyes level and mouth level, the third for matching the lower region of the face starting from mouth level and finally, the fourth sub-template is used for matching the region of the facial image center. The sub templates are scaled based on intensity 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 feature extraction algorithm to extract the features in each segment separately.

V. COLOR SPACE

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 HIS (hue, saturation, Intensity) also known as HSL (hue, saturation, lightness/luminance). Three coordinates are specified for each model. This coordinates describe color position within the corresponding color space. The HSV and HIS color space models are derived from the RGB space cube [14], where saturation S = 1 - min(R,G,B)/I, Intensity I = (R+G+B)/3, the value V = max(R,G,B), and the hue defined as:

\[
H = \cos^{-1}\left(\frac{1/2([-R+G]+[-R+B])}{\sqrt{([-R+G]+[-R+B])^2+([R-G]+[G-B])^2}}\right).
\]

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 models mentioned above with regard to the performance and accuracy of the facial image retrieval system.

VI. FEATURES EXTRACTION

Features extraction is the process of transforming the content of the images into various content features commonly known as feature vectors. In this research, color histogram is used as general visual content, while eigenfaces features are employed as domain specific visual content.

A. Eigenfaces Features

Eigenfaces is one of the most important methods used for face recognition. It 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. Sirovich and Kirby [15] represented human faces using PCA, and Turk and Pentland [16] developed a face recognition technique using eigenfaces. For facial image retrieval, eigenfaces are calculated using PCA, such as in Fig. 3, where the following steps based on [16] are applied:

- Suppose the facial image set is \( T_1, T_2, \ldots, T_m \), Then, the mean face of the set is defined by

\[
\Psi = \frac{1}{m} \sum T_M.
\]

- The mean face is subtracted from each original face vector

\[
\Phi_i = T_i - \Psi.
\]

- The covariance matrix is calculated as

\[
S = \frac{1}{m} \sum (T_i - \Psi) (T_i - \Psi)^T.
\]
where
\[ \begin{align*}
C &= \frac{1}{M} \sum_{n=1}^{M} \Phi_n \Phi_n^T = A A^T, \\
A^T A v_l &= u_l v_l, \\
A A^T v_l &= u_l A v_l.
\end{align*} \tag{7}
\]

where \( C \) is \( N^2 \times N^2 \) and \( A \) is \( N^2 \times M \).

- The eigenvectors and eigenvalues of the covariance matrix are calculated. Consider the eigenvectors \( v \) of \( A A^T \) such that
\[ A^T A v_l = u_l v_l, \tag{8} \]
\[ A A^T v_l = u_l A v_l. \tag{9} \]

The \( A v_l \) are the eigenvectors of \( C = A A^T \).

- From these analysis, \( M \times M \) matrix \( L = A A^T \), \( L_{mn} = \Phi_m^T \Phi_n \) is constructed, and \( M \) eigenvectors \( v_l \) of \( L \) can be found.

- The eigenfaces \( u_l \) are formed from \( M \) training facial images linear combinations, such that
\[ u_l = \sum_{k=1}^{M} v_{lk} \Phi_k, \quad l = 1, \ldots, M. \tag{10} \]

B. Color Histogram

A color histogram is a common approach used in image retrieval system [1]. The computing methodology of the color histogram is based on the fact that images are represented as a series of pixel values, each corresponding to a visible color and similar images contain similar proportions of certain colors. 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 histogram refers to the probability mass function (pmf) of the image intensities and can be defined [17] by:
\[ h_{A,B,C}(a,b,c) = N \times P(A = a, B = b, C = c). \tag{11} \]

Here, \( A, B \) and \( C \) are the three-color channels, and \( N \) is the number of pixels in the image. The probability mass function of the image are indicated by \( p(x) \), \( x = 0 \ldots X \), where \( X \) is the maximum luminance value in the image, the cumulative probability function is defined [18] as
\[ P(x) = \sum_{i=0}^{x} p(i). \tag{12} \]

One of the problems with color histogram-based retrieval is the high dimensionality of the color histograms. Color quantization method is used to reduce the number of colors available in an image. Histogram values are normalized by dividing the number of pixels in each histogram bin by the total number of pixels in the image.

VII. Query and Retrieval Process

In this research, a prototype system was designed based on the combination of face detection, CBIR and FERET techniques, with the proposed method of facial image segmentation as shown in Fig. 4. In the query processing, the user provides an initial image to the system or selects one from the images database pool. This query image looks similar to the required facial image. During the retrieval process, the system automatically detects the facial image from the query image. The candidate facial image is segmented using the proposed method based on the eyes and mouth level, and eigenfaces features and color histogram features are extracted from each segment. Combinations of these features are used to identify and retrieve the similar faces to the query face from the database.
Euclidian distance approach is employed to compute the distance between the query features vector \( Q \) and the features vectors \( D \) in the database. Faces with the least distance are retrieved and displayed on top.

\[
\text{Faces Similarity Distance} (Q, D) = (\sum_{i=1}^{n} (Q_i - D_i)^2)^{\frac{1}{2}}.
\]  

VIII. RESULTS AND DISCUSSION

Numerous experiments have been conducted to assess and evaluate the proposed method of facial image retrieval. The database that is used consists of 1500 local facial images database of 150 participants from the University of Malaya (UM) in Kuala Lumpur. Ten different images were taken of each participant of different race, gender, age, skin color, and so on. Facial images were taken at different times with different facial expressions (e.g., happy, sad, smiling, angry, etc.) and facial details (e.g., glasses, beard, mustache, and facial marks). A total of 750 images, half of the database, were utilized for training, and the rest was employed for the experiments. Precision and recall methods were applied to measurement of the performance efficiency of the retrieval methods, as defined:

\[
\text{Recall} = \frac{\text{Relevant Faces of The Retrieved Faces}}{\text{Total Relevant Faces}}.
\]

\[
\text{Precision} = \frac{\text{Relevant Faces of The Retrieved Faces}}{\text{Total Retrieved Faces}}.
\]

\[
F - \text{score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.
\]

The facial image retrieval system differs from the face recognition method in which the system does not look for an identical image but only for similar images. Therefore, to evaluate the system performance during the retrieval process, a threshold to determine the level of the retrieval was not set but rather the number of the images to be retrieved was subject to a certain pre-determined value. Hence, following the method of precision and recall cut-off rank was considered as necessity. Therefore, the experiments were performed with different cut-off levels, that is, 10, 16, and 25. Conversely, the images taken into account are the images that are on the top (10, 16, and 25) of the displayed results.

The experimented three-color space models in this study were RGB, HSV, and HSI to find out which color space in the facial image retrieval system has the best performance. Considering the F-score measurement on the top 10 of the retrieved images, eigenfaces-based facial image retrieval in HSV color space model showed better performance than in RGB and HIS color space models achieving 73.52% accuracy in comparison to 63.57%, and 60.16% accuracy in RGB and HSI respectively. Considering the recall measurement only, the best performance of the color histogram belonged to HSV color space model with 85.72% accuracy within the top 25 retrieved images as shown in Table I (b).

The three methods of features extraction were experimented based on the entire facial image, three segments of the facial image based on the eyes level and mouth level, and four segments including the center of the facial image. The first method is a traditional method, while the others are the proposed methods. Experiments were conducted using eigenfaces and color histogram features, separately and then using a combination of them. Considering the F-score measurement within the top 10 retrieved images, eigenfaces-based facial image retrieval achieved 73.52%, 81.53%, and 81.97% accuracy using the traditional method, the first method, and the second method of feature extraction respectively, as displayed in Table II (a).

### Table I. (a) Eigenface-based face retrieval in different color space models with out segmentation.

<table>
<thead>
<tr>
<th>Color Space</th>
<th>Top Faces</th>
<th>Retrievd Faces</th>
<th>Relevant Faces</th>
<th>Recall</th>
<th>Precision</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGB</td>
<td>10</td>
<td>7500</td>
<td>4768</td>
<td>0.6357</td>
<td>0.6357</td>
<td>0.6357</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>12000</td>
<td>5379</td>
<td>0.7172</td>
<td>0.4483</td>
<td>0.5517</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>18750</td>
<td>5804</td>
<td>0.7739</td>
<td>0.3083</td>
<td>0.4404</td>
</tr>
<tr>
<td>HSV</td>
<td>10</td>
<td>7500</td>
<td>5014</td>
<td>0.7352</td>
<td>0.3752</td>
<td>0.4939</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>12000</td>
<td>6096</td>
<td>0.8128</td>
<td>0.5080</td>
<td>0.6252</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>18750</td>
<td>6482</td>
<td>0.8643</td>
<td>0.3457</td>
<td>0.4939</td>
</tr>
<tr>
<td>HSI</td>
<td>10</td>
<td>7500</td>
<td>4512</td>
<td>0.6016</td>
<td>0.6016</td>
<td>0.6016</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>12000</td>
<td>5104</td>
<td>0.6705</td>
<td>0.4255</td>
<td>0.5235</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>18750</td>
<td>5542</td>
<td>0.7339</td>
<td>0.2306</td>
<td>0.4223</td>
</tr>
</tbody>
</table>

### Table I. (b) Color histogram-based face retrieval in different color space models with out segmentation.

<table>
<thead>
<tr>
<th>Color Space</th>
<th>Top Faces</th>
<th>Retrieved Faces</th>
<th>Relevant Faces</th>
<th>Recall</th>
<th>Precision</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGB</td>
<td>10</td>
<td>7500</td>
<td>5966</td>
<td>0.7955</td>
<td>0.7955</td>
<td>0.7955</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>12000</td>
<td>5966</td>
<td>0.7955</td>
<td>0.4972</td>
<td>0.6119</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>18750</td>
<td>5966</td>
<td>0.7955</td>
<td>0.3182</td>
<td>0.4546</td>
</tr>
<tr>
<td>HSV</td>
<td>10</td>
<td>7500</td>
<td>5804</td>
<td>0.7739</td>
<td>0.7739</td>
<td>0.7739</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>12000</td>
<td>6198</td>
<td>0.8264</td>
<td>0.5165</td>
<td>0.6357</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>18750</td>
<td>6429</td>
<td>0.8572</td>
<td>0.3429</td>
<td>0.4898</td>
</tr>
<tr>
<td>HSI</td>
<td>10</td>
<td>7500</td>
<td>5729</td>
<td>0.7639</td>
<td>0.7639</td>
<td>0.7639</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>12000</td>
<td>6077</td>
<td>0.8103</td>
<td>0.5064</td>
<td>0.6233</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>18750</td>
<td>6311</td>
<td>0.8415</td>
<td>0.3366</td>
<td>0.4809</td>
</tr>
</tbody>
</table>

### Table II. (a) Color histogram-based face retrieval in different color space models with out segmentation.

<table>
<thead>
<tr>
<th>Extraction Method</th>
<th>Top Face</th>
<th>Retrieved Faces</th>
<th>Relevant</th>
<th>Recall</th>
<th>Precision</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional Method</td>
<td>10</td>
<td>7500</td>
<td>5514</td>
<td>0.7352</td>
<td>0.7352</td>
<td>0.7352</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>12000</td>
<td>6096</td>
<td>0.8126</td>
<td>0.5080</td>
<td>0.6252</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>18750</td>
<td>6482</td>
<td>0.8643</td>
<td>0.3457</td>
<td>0.4939</td>
</tr>
<tr>
<td>Three Segments</td>
<td>10</td>
<td>7500</td>
<td>6115</td>
<td>0.8135</td>
<td>0.8135</td>
<td>0.8135</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>12000</td>
<td>6585</td>
<td>0.8787</td>
<td>0.5488</td>
<td>0.6754</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>18750</td>
<td>6871</td>
<td>0.9161</td>
<td>0.3665</td>
<td>0.5235</td>
</tr>
<tr>
<td>Four Segments</td>
<td>10</td>
<td>7500</td>
<td>6148</td>
<td>0.8197</td>
<td>0.8197</td>
<td>0.8197</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>12000</td>
<td>6602</td>
<td>0.8803</td>
<td>0.5502</td>
<td>0.6772</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>18750</td>
<td>6890</td>
<td>0.9187</td>
<td>0.3675</td>
<td>0.525</td>
</tr>
</tbody>
</table>
The color histogram-based facial image retrieval achieved 79.55%, 86.53%, and 85.44% of accuracy, using the traditional method, the first method, and the second method of feature extraction respectively, as shown in Table II (b). Integration color histogram and eigenfaces-based facial image retrieval achieved 79.49%, 89.51%, and 88.24% of accuracy, using the traditional method, the first method, and the second method of feature extraction respectively, as shown in Table II (c). The results of the experiments show that extraction the features based on the three segments of the facial image in the first method improved the performance of the facial image retrieval technique comparing to the traditional method and the other methods. Considering the recall measurement only, the best performances of the system with this method are 91.61%, 90.9%, 93.1% of accuracy, using the traditional method, the first method, and the second method of feature extraction respectively, as shown in Table II (a, b, and c).

IX. CONCLUSION

In this research, a new method for content-based human facial image detection and retrieval model is proposed based on integration of the face detection, face recognition technique and the traditional content-based image retrieval technique. In addition, a new method of facial image segmentation for improving the performance of the features extraction process is also suggested. Integration of eigenfaces features and color histogram features were used as low level features, whilst the proposed method is applicable with different visual features.

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 showed that eigenfaces-based facial image retrieval in 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. Comparing to the traditional method of visual features extraction, the results reflect an excellent improvement in the facial image retrieval achieved based using the proposed method of visual features extraction.

REFERENCES