Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Content-Based Face Color Image Retrieval using Multi-Feature Fusion Extractor Methods

Syaidatus Syahira Ahmad Tarmizi[1], Nor Surayahani Suriani[1,*], Nik Shahidah Afifi Md. Taujuddin[1], Nan Md Sahar[1], Xin Wang[2]

[1] Department of Electronic Engineering, Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia
[2] Sichuan Institute of Industrial Technology China

ABSTRACT

The retrieval of visual image content has been the most active research in various applications. In this paper, the benchmark datasets have been used as a fast screening process for extracting representative facial features. Despite extracting relevant features information from the entire face, local features focused on the segmented regional area have proved to be more effective as suggested in the literature search. In doing so, four labeled region area was chosen before it was combined together to create a new sample image as input data for further analysis. To enhance the image representation, variation color spaces conversion is used and the first four color moments are selected for acquiring color information about the image. Also, the main five texture features are concatenated later with the color moments to analyze the complementary effects of color features in texture. In total, the nine selections of feature fusion methods have been presented, whereas the high dimensional space has been through the dimensional reduction process. The experimental result demonstrates that higher image content retrieval accuracy can be obtained by applying the CM+BSIF feature for YCbCr thermal image (0.4688 ± 0.1481) and CM+BSIF+Tamura for HSV visible image (0.4631 ± 0.1512).

Keywords:
Thermal image; visible image; multi-feature fusion; color moments; statistical based; transform based

1. Introduction

Content-based face image retrieval has received a significant interest as the image search and retrieval technique in which the image would be indexed according their visual content either from thermal and visible datasets. The study for retrieving the visual facial image content has begins with the effort to extract relevant features information that can be used to analyze pathophysiological abnormalities, emotional expression classification, attractiveness, gender and aging detection [1-3]. It can be a challenging problem in such cases on the basis in finding the similarity measures based on shape signatures, color signatures or texture signatures. Also, prior studies have shown that extracting global feature from the entire face is ineffective for this cases. There are several local facial

* *Corresponding author.*
E-mail address: nsuraya@uthm.edu.my

https://doi.org/10.37934/araset.58.2.2840

features that possess valuable information and frequently included with the inclusion of the forehead, eyes, nose, cheek and mouth [4].

Several criteria must be considered in selecting the optimum extractor methods, this should include the quality measure such as sharpness, tone and color that may be used to assess the quality image score [5]. Because of this, the exploration of feature fusion approaches has shown a variety combination of single feature based methods in existing research. This single feature has their own complementary effects on the shape and textures, shape and color, and color features in texture. Humans have distinctive facial feature signatures that vary from person to person with different size and shape even it has the standard anatomical positions of regional areas. Therefore, the implementation of shape based is necessary with an attempt to monitor a range of heads postures by employing the active appearance model (AAM) [6], particle filter-based object tracker [7], Kanade Lucas Tomasi (KLT) tracker [8], and Harris corner detector [9].

In related work, the color image analysis has become an alternative to grayscale images analysis. It has achieved outstanding performance in color texture analysis, which is currently being used to transform grayscale texture into color texture to improve the classification tasks using RGB (red, green, blue), HSV (hue, saturation, value), YCbCr (luminance, chrominance blue, chrominance red), CIELAB (luminosity, chrominance red-green, chrominance blue-yellow) and color gamut (six color centroid) [10-12]. In general, the changes in facial skin textures may have positive association between aging and growth of disease process that could explain why there is textural variability among them. It is helpful in study to detect subtle changes in the face such as wrinkles and furrows, loose skin, increased spots and coarseness [2]. For precisely analyzing skin textures, varied texture families have been suggested by extending them to extract texture from a color image [13].

It's interesting to observe how researchers have proposing different feature fusion approaches for different tasks. In this paper, the complementary role of multi-feature based methods in obtaining shape, color and textural information using sample thermal and visible images in the form of grayscale or color image representation was studied. The imperfections of data are only discovered after the image analysis process has started. Therefore, several preprocessing methods have been utilized as possible solution in addressing facial features problems as the initial process before the multi-feature based methods is adopted. The rest of this paper is structured as follows: Section 2 discusses the detailed of the proposed designated works. Section 3 presents experimental results and limitation in current study. Finally, section 4 draws a conclusion about the study.

2. Experimental Design

This section presented the proposed framework for extracting facial feature information from two different image spectrum as shown in Figure 1. Preprocessed image may require the selection of preprocessing procedures to improve the visual image quality. By using the colored sample image, the analysis is done to gain insight into which color transformation can yields better results in improving the color texture information. The multi-feature fusion extraction is used to studies the benefits of combining several extractors methods for effectively extracting the image content. The performance evaluation is based on the final feature vector obtained from each feature fusion.

2.1 Image Acquisition

The experiment is conducted on benchmark dataset, which is a collection of facial thermal and visible image groups for extracting meaningful facial feature information. Overall, 2400 images were chosen from thermal faces in the wild (TFW) datasets [14]. These images are simultaneously acquired

in indoor settings at a resolution of 464x348 pixels. In the present case, the inclusion of gender-balanced subjects (n=50) with an average of 24 image per subjects was selected. Subjects with glasses, beards, hairy parts on the forehead and excessive non-frontal head postures were excluded to alleviate the risk of inaccurate localization on the selected region of measurements.

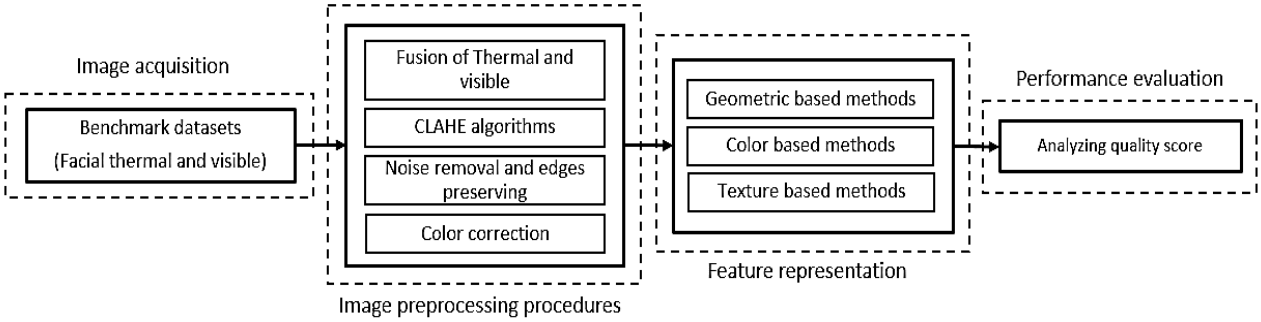

Fig. 1. The framework for extracting facial feature information

2.2 Image Preprocessing Procedures

The preprocessing procedures are performed on facial thermal and visible images to enhance the quality of the image representation. Because high-resolution images can produce finer images, the original size of the input images was used in this experiment. All the sampled images have been converted into grayscale color images with the intent to lessen the effect of illumination variation and minimize computational complexity. For thermal images, fusion methods were applied to compensate for the detailed loss of facial thermal features by fusing them with visible images. To enhance the intensity of an image, the CLAHE technique is used to enhance the image contrast based on the frequent intensity values as illustrated in Eq. (1). Where M is the pixel size and the histogram h is presented as the gray value of N for each sub-region. In this case, the clip limit threshold is set to 2.0 with an 8x8 tile grid size. Through this process, the contrast for the desired foreground is increased while the background contrast is decreased.

$$f_{i,j}(n) = \frac{N-1}{M} . \sum_{k=0}^{n} h_{i,j}(k) \tag{1}$$

$$BF[I]_p = \frac{1}{W_p} \sum_{q \in s} G_{\sigma_S}(\|p-q\|) G_{\sigma_r}(|I_P - I_q|) I_q \tag{2}$$

While preserving the edges detail of facial features, the image was denoised using a non-linear bilateral filter (BF) as written in Eq. (2). Given the normalized weighted W_P and the Gaussian kernel parameters of G_{σ_s} and G_{σ_r} are the spatial distance pixels and intensity values for the filtering image I. The low intensity values of G_{σ_r}can help in image smoothing, but its high intensity values may result in blurring effects. The normalization is applied to the filtered images before it was converted using inferno colormaps to visualize facial image variations on thermal and visible for extracting any relevant features information.

2.3 Multi-Feature Extraction

The state-of-the-art extractor methods such as geometric, color and texture based method have been proposed to improve the facial features representation. Through this process, the influence of each extracted method in providing precise findings by using two different types of images were studied. At the same time, it has also been attempted for introducing various color conversion enhancements for color image analysis. Although, it was a time-consuming operation, the proposed multi-feature fusion can be utilized to retrieve the image content within a large collection of data.

2.3.1 Geometric based methods

Geometric based methods have been utilized to appropriately shape facial image with a range of head postures. In doing so, Dlib landmark detectors were applied to precisely localize both global and local facial regions [15]. The Dlib detector is easier to use for an automatic process of face allocation with up to 81 positional points. By choosing four labelled region of measurements with 50x50 pixels, an additional face recognition library was included and it was utilized mainly to improve localization at the forehead region above between the eyebrow point. The cheek region is extracted using references at the right and left sides of the eyes and nose coordinate. Then, the nose regions were automatically choosing from the Dlib landmark point. All labelled regions are concatenated together to form a new sample facial block images with 100x100 pixels.

2.3.2 Color based methods

The color based methods have been performed as color correction procedure to extract color information from thermal and visible images. In this paper, a non-uniformity RGB color has been transformed into HSV, CIELAB, YCbCr and grayscale. The problem occurs when the color images is separated into three color channel $x = (x_1, x_2, x_3)$, in which there is possibility of color channels to have values close to zero saturation and at a singularity of $x_n = 0$. Therefore, the color space enhancement was performed only for HSV, CIELAB, YCbCr before extracting their color information. To ensure non-overlapping color exists in an attempt to enhance visual representations, color moments (CM) will be utilized to distinguish images based on their features of color [16, 17].

$$E_i = \frac{\sum_{j=1}^{N} p_{i,j}}{N} \tag{3}$$

$$\sigma_i = \sqrt{\frac{\sum_{j=1}^{N} \left(P_{i,j} - E_i\right)^2}{N}} \tag{4}$$

$$S_i = \sqrt[3]{\frac{\sum_{j=1}^{N} \left(P_{i,j} - E_i\right)^3}{N}} \tag{5}$$

$$K_i = \sqrt[4]{\frac{\sum_{j=1}^{N} \left(P_{i,j} - E_i\right)^4}{N}} \tag{6}$$

It this study, the color moment can be defined as: (i) mean is computed using Eq. (3) to calculate the average color value in the image; (ii) standard deviation is obtained using Eq. (4) by finding the squared root of the variance of the distribution; (iii) skewness is used to measure the degree of asymmetry in the distribution and provides useful information about the shape of the color distribution which is represented in Eq. (5); (iv) kurtosis is used to measures how flat or tall the color distribution was in relation to the normal distribution using Eq. (6). Given the P_{ij} is the value of the i^{th} color channel of the image pixel j, and N is the number of pixels in the sample image. Then, E_i is presented as the mean value for the i^{th} color channel of the image.

2.3.3 Texture based methods

The human face has a broad range of appearances which could explain why there is textural variability among them. Five different texture based methods have been applied to eliminate any noise or weak related images that could degrade the sharpness of the texture image [13]. In this experiment, the proposed color moments will be combined with texture based methods to construct the two groups of feature sets. Here, the proposed texture based has been through the modification process that makes it possible to extract texture from a color based image.

The gray level co-occurrence matrix (GLCM) algorithm utilized the graycomatrix function to transform an image into co-occurrence matrices. Each element P_{ij} which is the result of GLCM, is calculating how often a pixel with the intensity value i occurred in a particular spatial relationship to a pixel with the value j in the input image. GLCM can generate various feature pattern sets with distinct set of directions before being used to extract the sixth statistical GLCM features [18-20].

Contrast is employed to measure the intensity contrast between two neighboring pixels through Eq. (7). It is also used to measure the number of local variations present in an image. Dissimilarity is applied to calculate the distance between two neighboring pixels in an image using Eq. (8). Correlation is computed using Eq. (9) to determine the correlation between two neighboring pixels over the entire image. Homogeneity can be expressed using Eq. (10) to determine the proximity of the conveyance of segments in the GLCM to the GLCM corner to corner. Energy is applied to measure the textural uniformity using Eq. (11). Angular second moment (ASM) is used to measure the homogeneity of an image using Eq. (12).

$$Contrast = \sum_{i,j=0}^{N-1} p_{i,j}\left(i - j^2\right) \tag{7}$$

$$Dissimilarity = \sum_{i,j=0}^{N-1} p_{i,j}\left|i - j\right| \tag{8}$$

$$Correlation = \sum_{i,j=0}^{N-1} \frac{p_{i,j}(i-\mu)(j-\mu)}{\sigma^2} \tag{9}$$

$$Homogeneity = \sum_{i,j=0}^{N-1} \frac{p_{i,j}}{\left(1+\left(i-j\right)^2\right)} \tag{10}$$

$$Energy = \sqrt{\sum_{i,j=0}^{N-1} p_{i,j}^2} \tag{11}$$

$$ASM = \sum_{i,j=0}^{N-1} p_{i,j}^2 \tag{12}$$

Local binary Patterns (LBP) is a texture analysis approach that combines statistical and local structural texture. This study used a uniform LBP, where an 8-bit LBP vector was considered uniform when there were no more than 2 transitions of the LBP code (0/1 or 1/0) ranging from 0 to $2^P - 1$ with radius values of 2 from the center pixel of the circle [21]. The LBP can be expressed using Eq. (13), where (x_c, y_c) are the coordinates of the center pixel, p is the p^{th} pixel of the neighborhood that will be calculated clockwise, P is the total number of neighborhood pixels and R is the radius for neighborhood pixels between the center pixel coordinates. Then, $x = i_p - i_c$, where i_p denotes the intensity value of the neighborhood pixels and i_c is the center pixel within the circular neighborhood.

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{p-1} s(x) 2^p \tag{13}$$

The binarized statistical image feature (BSIF) algorithms were employed to enhance the accuracy of feature tasks through the use of visual quality assessment which is classified in accordance with reference information [22]. The BSIF is calculated using an image patch X and a linear filter of W_i on the same size with the filter response s_i in Eq. (14). Here w_i and x are vectors that hold the pixels of W_i and X. The binarized feature b_i is obtained by setting with 1 and 0. The independent component analysis (ICA) algorithms is adopted to effectively analyze the BSIF descriptor by choosing the most appropriate filters kernel size of l x l and bit string length n. For the best accuracy, only the kernel size of l x l $= 13$x13 and $n = 6$ was considered.

$$s_i = \sum_{u,v} W_i(u,v) X(u,v) \tag{14}$$

Tamura feature is based on descriptions of human visual perception of appearance. The inclusion of four Tamura features like coarseness, contrast, directionality and roughness were selected in the present studies [23]. The coarseness is used to measure the granularity of the image by finding any texture elements at various scales. The coarseness elements are represented by Eq. (15), where w and h are the width and height of an image, and $S_{best}(i,j) = 2^k$ is the neighborhood size that is used to generate the highest similarity of averaging intensity of the neighborhood centered at (i,j) and $k = 5$ is the maximum value for either the horizontal or vertical direction.

Contrast can be influenced by the distribution polarization of separating color channels and the dynamic ranges of color levels between two texture patterns with different structural elements. It can be computed using Eq. (16), where σ is the std. deviation and α_4 corresponds to the fourth moment of the mean. Directionality is used to analyze any directional pattern in an image along particular orientations. Here, directionality was calculated using Eq. (17), where H_D is the direction histogram, the number of peaks n_P and Φ was the p^{th} peak position in the H_D. Then, w_p is the range of p^{th} peak between valleys and Φ used to quantize the direction angle. The r is a normalizing factor related to the number of quantization levels of Φ. Roughness is related to the std. deviation of the normalized color levels, which was described in Eq. (18) as the fusion of coarseness and contrast elements.

$$F_{crs} = \frac{1}{w \times h} \sum_{i=0}^{w} \sum_{i=0}^{h} s_{best}(i,j) \tag{15}$$

$$F_{con} = \frac{\sigma}{(\alpha_4)^n} \tag{16}$$

$$F_{dir} = 1 - rn_p \sum_p^{n_p} \sum_{\phi \in w_p} (\phi - \phi_p)^2 H_D(\phi) \tag{17}$$

$$F_{rgh} = F_{crs} + F_{con} \tag{18}$$

$$G_{\sigma, f_0, \phi}(x, y) = \exp\left(-\frac{1}{2}\left[\frac{x'^2}{\sigma_x^2} + \frac{y'^2}{\sigma_x^2} \right] \right) \cos(2\pi f_0 x' + \phi) \tag{19}$$

The Gabor filter is an edge detection algorithm from the transform based texture family in which the output of this filter method is the Gabor wavelet transform (GWT) [13]. Here, the Gabor filter parameters for kernel size 64x64 modulated by a sinusoidal wave, the std. deviation of σ is set to eight with distinct eight orientations θ. A 2D Gabor function $G(\mathrm{x}, \mathrm{y})$ can be measured using Eq. (19), where $\acute{x} = x \cos\theta + y \sin\theta$ and $\acute{y} = -x \sin\theta + y \cos\theta$. The f_0 and $\emptyset$ are respectively the frequency and phase of the sinusoidal function.

2.4 Dimensionality Reduction Process

The feature fusion sets based on color texture may contain a low- or high-level dimensional data with varying feature vector values. The output from feature selection generally has dimension that correspond to the number of feature channels that were extracted from three or single color channels of the image. Extracting high level dimensions in the present study is not necessarily effective and will also have an impact on the processing speed to analyze huge number of features, which might result in redundant data that not necessarily useful to improve the visual image quality. Therefore, the dimensionality reduction will be considered with an attempt to reduce the original high dimensional space by using principle component analysis (PCA) algorithms [21, 24]. The PCA is applied to obtain a new orthogonal variable that called as the principal component, which were used to minimize high dimensional space. The selection of feature fusion with high features value needs to be first normalized using Standard Scalar operations before to applying PCA operation.

3. Results

The present study demonstrates the significant role of preprocessed method in improving the representation of distorted images. Results from the segmented local facial images were obtained using smaller uniform size and different color transformation were used to shows the effectiveness of color features in improving the properties of texture features. In addition, image content retrieval depends highly on the proposed feature fusion for construct a feature vector as the final output to deliver the results. A way to increase the accuracy results is address by reducing their high dimensional space. Then, the following subsection presents the statistical measurement of feature fusion with an attempt to analyze the complementary effects between color texture performance. The evaluation is done in figuring out the optimal feature fusion according to their color conversion image used in progress to enhance visual perception respectively.

3.1 Result on the Proposed Preprocessing Methods

Initially, the obtained thermal and visible image will through computation of each proposed preprocessing methods to correct their visibility image quality. It appeared that new image colorization using inferno colormaps may use the correct combination of the primary red, green and blue color channels. As can see, Figure 2 presents the thermal and visible face subjects that randomly choose to illustrate the distribution of pixel intensities in the preprocessed image. The pixel is the elements of an image that also contain information about color in each of a defined list of color range that considerable differ for consecutive frames as well as for an individual subject. It can be concluded that most of the darker images can be found on thermal images as compared to the visible images with a variety of lighter shades of colors appearance due to illumination effects.

From the physiological point of view, literature search has emphasized the important factor for monitoring a common facial region that predominates in vital sign measurements to measure the human physiological responses of the body with faster processing time in a noninvasive way. Concerning on how image content retrieval process is conducted on specific regional area, the reference of facial positioning points from Dlib landmark detectors is used to shape the facial image as illustrates in Figure 3. Through this process is a bounding box containing the identified facial image. Next, using a bounding box, the forehead, nose, left cheek and right cheek area can be allocated. These regional areas were resized into smaller blocks of size 50x50 pixels, before being concatenated into non-overlapping facial blocks with a size of 100x100 pixels, which were then used for color and texture extraction.

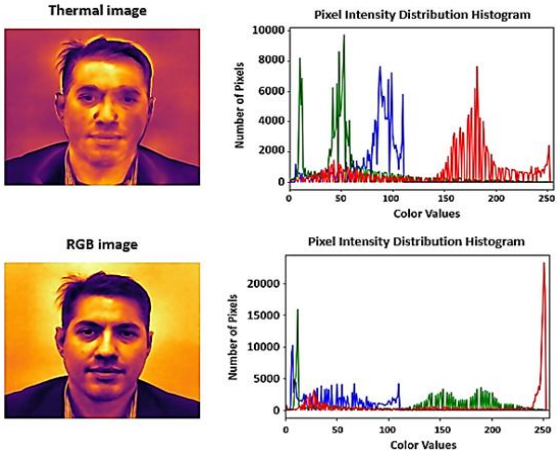

Fig. 2. The preprocessed images and its corresponding histogram

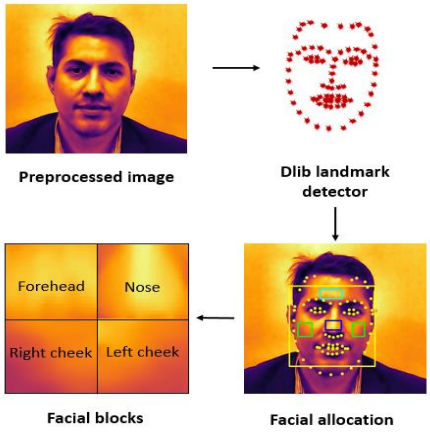

Fig. 3. Example of facial landmark detection

The existing color image may influence by a combination of RGB color space which has encouraged the investigation for optimal color space conversion for improving pathological facial skin tones as found in Figure 4. As a highlight, color contrast enhancement is performed for the HSV, CIELAB and YCbCr color spaces derived from the RGB color space for simulating visual perception. To obtain the desired color that is as realistic as the original RGB color version, the inverse of the intensity from any color channel is done followed by an increase in the brightness value of the darker color channel before enhancing using the CLAHE method. The fact that a variety of five distinct color spaces may provide a total of fifteen color channels in this experiment.

Figure 5 illustrates a schematic of the proposed feature fusion extraction, whereas each facial blocks that possesses details about the forehead, nose, left cheek and right cheek in RGB color space is separated into three color channels (red, green, blue). The color and texture histogram feature are

construct for each individual color channel before it was concatenating to form a global histogram to generate the final feature vector. This process is repeated for the HSV, CIELAB and YCbCr color space conversion. For grayscale color image, their color and texture histogram feature are computed directly on the individual color channels.

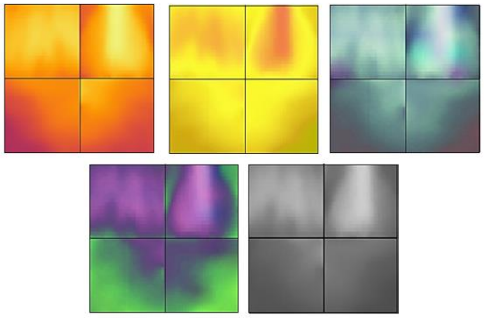

Fig. 4. An example of color space conversion for RGB, HSV, CIELAB, YCbCr and grayscale image

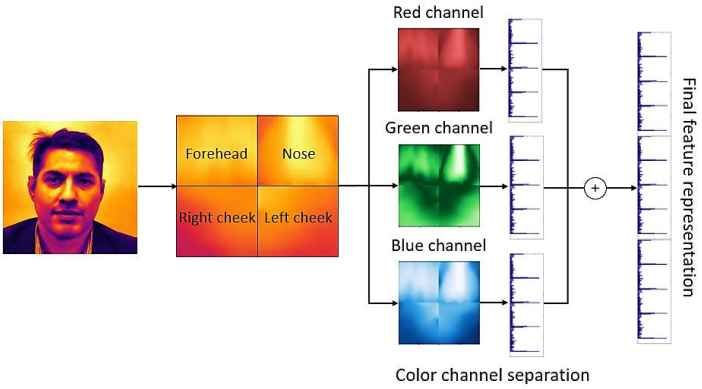

Fig. 5. Structure of the proposed feature fusion extraction

It can be seen that various colors and texture features have been utilized in this experiment to investigate the complementary effects that may lies on the color features in texture as tabulated in Table 1. The finding highlights the possible feature fusion sets that may contain at least one combination of color and texture features, while the other features fusion sets also include at least one color feature with an additional two combination of texture feature that consists of Tamura texture. From the observation, there is existing feature fusion that has a high dimensional space with higher feature values as a result of feature fusion selection. Alternatively, dimensionality reduction is introducing to reduce high dimensional space into low dimension space with a fixed standard dimensional space of 30. Although the dimensional space is similar after feature reduction, the final results will differ and be significantly better than the low dimensional space.

Table 1
Dimensional reduction

Multi-feature based	Feature dimension					
	Color based	PCA	New dimension	Gray based	PCA	New dimension
CM + Tamura	24	-	24	8	-	8
CM + GLCM	30	-	30	10	-	10
CM + LBP	42	30	30	14	-	14
CM + Gabor	15	-	15	5	-	5
CM + BSIF	204	30	30	68	30	30
CM + GLCM + Tamura	42	30	30	14	-	14
CM + LBP + Tamura	54	30	30	18	-	18
CM + Gabor + Tamura	27	-	24	9	-	9
CM + BSIF + Tamura	216	30	30	72	30	30

3.2 Comparative Analysis with Multi-Feature Based Methods

The prominence results for the nine feature fusions that were tested on five distinct facial color block transformations are summarized in Table 2. Here, the histogram analysis of the sample image was used to compute statistical parameters like mean and std. deviation (STD). The results were

thoroughly analyzed and it turns out that the STD value may be statistically lower, although their corresponding mean value is considered high. For thermal image, there are two favorable feature fusion that can be observed. For instance, the optimal CM+BSIF feature extraction appears to be effective for color image transformation in RGB (0.4534 ± 0.1496) and YCbCr (0.4688 ± 0.1481), while CM+BSIF+Tamura is suitable to obtain the temporal information about the image content from HSV (0.4446 ± 0.1456), CIELAB (0.4600 ± 0.1497) and grayscale (0.4482 ± 0.1405) color space.

Table 2
Evaluation of thermal color based

Multi-feature based	RGB		HSV		CIELAB		YCbCr	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD
CM + Tamura	0.4156	0.2400	0.4072	0.2464	0.4412	0.2247	0.4004	0.2219
CM + GLCM	0.3981	0.2534	0.4012	0.2577	0.3922	0.2291	0.3904	0.2289
CM + LBP	0.4273	0.1491	0.4257	0.1429	0.4525	0.1468	0.4547	0.1484
CM + Gabor	0.3781	0.2493	0.3915	0.2618	0.3790	0.2049	0.3520	0.1981
CM + BSIF	**0.4534**	**0.1496**	0.4414	0.1489	0.4579	0.1493	**0.4688**	**0.1481**
CM + GLCM + Tamura	0.3950	0.1351	0.3907	0.1493	0.4302	0.1316	0.4416	0.1343
CM + LBP + Tamura	0.4429	0.1434	0.4301	0.1457	0.4552	0.1474	0.4570	0.1465
CM + Gabor + Tamura	0.3902	0.2396	0.3878	0.2434	0.4180	0.2240	0.3862	0.2165
CM + BSIF + Tamura	0.4494	0.1497	**0.4446**	**0.1456**	**0.4600**	**0.1497**	0.4572	0.1500

Table 3
Evaluation of visible color based

Multi-feature based	RGB		HSV		CIELAB		YCbCr	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD
CM + Tamura	0.3911	0.2451	0.3928	0.2522	0.3906	0.2440	0.3566	0.2217
CM + GLCM	0.3827	0.2691	0.3810	0.2736	0.3612	0.2485	0.3663	0.2426
CM + LBP	0.4353	0.1456	0.4406	0.1467	**0.4511**	**0.1487**	**0.4618**	**0.1463**
CM + Gabor	0.3721	0.2550	0.3887	0.2619	0.3543	0.2157	0.3436	0.2041
CM + BSIF	0.4275	0.1586	0.4570	0.1497	0.4453	0.1537	0.4553	0.1479
CM + GLCM + Tamura	0.4104	0.1337	0.4053	0.1236	0.4118	0.1385	0.4250	0.1333
CM + LBP + Tamura	**0.4493**	**0.1426**	0.4176	0.1425	0.4309	0.1497	0.4465	0.1507
CM + Gabor + Tamura	0.3713	0.2409	0.3825	0.2458	0.3829	0.2351	0.3511	0.2142
CM + BSIF + Tamura	0.4272	0.1585	**0.4631**	**0.1512**	0.4484	0.1526	0.4574	0.1516

Table 4
Evaluation of thermal and visible gray based image

Multi-feature based	Thermal		Visible	
	Mean	STD	Mean	STD
CM + Tamura	0.4465	0.1401	0.3761	0.2307
CM + GLCM	0.4140	0.2301	0.3659	0.2377
CM + LBP	0.3794	0.1813	0.3638	0.2004
CM + Gabor	0.3691	0.1980	0.3503	0.2038
CM + BSIF	0.4322	0.1362	**0.4385**	**0.1468**
CM + GLCM + Tamura	0.4218	0.2339	0.3699	0.2406
CM + LBP + Tamura	0.3932	0.1981	0.3674	0.2118
CM + Gabor + Tamura	0.4012	0.2215	0.3634	0.2247
CM + BSIF + Tamura	**0.4482**	**0.1405**	0.4367	0.1441

Contrarily, for the visible image, it might be difficult to determine which feature fusion is the most appropriate to use in order to improve image representation because the feature selection used here may produce difference higher accuracy results depending on the selection of facial color block

transformations. The CM+LBP+Tamura is the best selective features for RGB (0.4493 ± 0.1426) color image and the CM+BSIF+Tamura is better suited for HSV (0.4631 ± 0.1512) color enhancements. It is unlikely for CM+LBP feature that outperforms for the CIELAB (0.4511 ± 0.1487) and YCbCr (0.4618 ± 0.1463) color space. However, the CM+BSIF feature consistently achieves the best validation accuracy, even for the visible grayscale (0.4385 ± 0.1468) color image. As expected, the results confirm that CM+BSIF feature can perform better for YCbCr thermal image and CM+BSIF+Tamura is the best for HSV visible color image.

While designing of this experimental study, some features relatively may have good performance but the implementation of the algorithm itself is challenging that may require large amount of data to verify the effectiveness of the proposed features. For this reason, it is worth highlighting the important factors that can enhance the accuracy of results: (1) The literature search has an emphasis on comparing the type of image used to quantify the image content; (2) the interest in small pixel size of the input image and the selection of facial regions that needs to be taken due to the consequences of processing time; (3) the consideration of color space conversion used for simulating visual perception; (4) the selection of state-of-the-art feature descriptor and the comparator of feature fusion to extract the image representation; and (5) the minimization of features vector values that may offer superior performance after dimensionality reduction.

As a limitation of the study, the proposed multi-feature extraction could be relatively limited in an effort to retrieve the facial image content to improve image representation. That being said, the variety of feature fusion with the inclusion of the color and textural variability could be extended to achieve the best accuracy results. For color features, the selection of the global color histogram, color difference histogram and color coherence vector can be potentially applied with texture based method like structural based, model based or graph based that have been suggested in the existing experiments. In order to improve the accuracy of feature extraction, it may also be required to include data on emotional expression, aging related and face skin diseases from unhealthy datasets.

Instead of the used the statistical analysis, the implementation of machine learning and deep learning are among the best possible solutions that can be recommended to accelerate the facial feature extraction process. Although machine learning is the most popular classifier that is frequently being used, the current literature search has shown a highly good performance of deep learning classifiers in addressing the problem with computational complexity and processing time [25, 26]. Here, it is worth mentioning that despite the fact the similar datasets being used in several research papers, further research is need to carried out to analyse a promising combination of feature fusion and classifiers that may not necessarily improve facial feature representation in certain cases [3].

4. Conclusions

This paper examined the important criteria that were included to analyze the effectiveness of the proposed methods based on the trade-off between processing time and accuracy in obtaining precise information about the visual image content. Hence, various experimental works have been conducted in an attempt to get more insights into the benefits of color texture within a large collection image. In this case, color space conversion has been designed to mimic human visual perception of colors and brightness sensitivity to observe the typical pathological change of facial skin tones between consecutive frames. The results were thoroughly analyzed and it was confirmed that the color image gave the most perceptually reasonable results for improving the color texture performance. By taking advantage of color and texture features, the new formulation of feature fusion based has shown a complementary role in representing color texture information. It became clear that the improvement brought by dimensionality reduction as an important measure in

reducing high dimensional space was significantly helpful in selecting the best feature fusion methods.

Acknowledgement
This research was supported by Ministry of Higher Education (MOHE) through Fundamental Research Grant Scheme (FRGS/1/2021/TK0/UTHM/02/12).

References
[1] Boughida, Adil, Mohamed Nadjib Kouahla, and Yacine Lafifi. "A novel approach for facial expression recognition based on Gabor filters and genetic algorithm." *Evolving Systems* 13, no. 2 (2022): 331-345. https://doi.org/10.1007/s12530-021-09393-2
[2] Liu, Xinhua, Yao Zou, Hailan Kuang, and Xiaolin Ma. "Face image age estimation based on data augmentation and lightweight convolutional neural network." *Symmetry* 12, no. 1 (2020): 146. https://doi.org/10.3390/sym12010146
[3] Tarmizi, Syaidatus Syahira Ahmad, and Nor Surayahani Suriani. "A review of facial thermography assessment for vital signs estimation." *IEEE Access* 10 (2022): 115583-115602. https://doi.org/10.1109/ACCESS.2022.3217904
[4] Zhang, Qi, Jianhang Zhou, Bob Zhang, and Enhua Wu. "DsNet: Dual stack network for detecting diabetes mellitus and chronic kidney disease." *Information Sciences* 547 (2021): 945-962. https://doi.org/10.1016/j.ins.2020.08.074
[5] Yang, Dan, Veli-Tapani Peltoketo, and Joni-Kristian Kamarainen. "CNN-based cross-dataset no-reference image quality assessment." In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pp. 0-0. 2019. https://doi.org/10.1109/ICCVW.2019.00485
[6] Kopaczka, Marcin, Raphael Kolk, Justus Schock, Felix Burkhard, and Dorit Merhof. "A thermal infrared face database with facial landmarks and emotion labels." *IEEE Transactions on Instrumentation and Measurement* 68, no. 5 (2018): 1389-1401. https://doi.org/10.1109/TIM.2018.2884364
[7] Hochhausen, Nadine, Carina Barbosa Pereira, Steffen Leonhardt, Rolf Rossaint, and Michael Czaplik. "Estimating respiratory rate in post-anesthesia care unit patients using infrared thermography: an observational study." *Sensors* 18, no. 5 (2018): 1618. https://doi.org/10.3390/s18051618
[8] Majtner, Tomáš, Esmaeil S. Nadimi, Knud B. Yderstræde, and Victoria Blanes-Vidal. "Non-invasive detection of diabetic complications via pattern analysis of temporal facial colour variations." *Computer Methods and Programs in Biomedicine* 196 (2020): 105619. https://doi.org/10.1016/j.cmpb.2020.105619
[9] Kwaśniewska, Alicja, and Jacek Rumiński. "Real-time facial feature tracking in poor quality thermal imagery." In *2016 9th International Conference on Human System Interactions (HSI)*, pp. 504-510. IEEE, 2016. https://doi.org/10.1109/HSI.2016.7529681
[10] You, Hojoon, Kunyoung Lee, Jaemu Oh, and Eui Chul Lee. "Efficient and low color information dependency skin segmentation model." *Mathematics* 11, no. 9 (2023): 2057. https://doi.org/10.3390/math11092057
[11] Shu, Ting, Bob Zhang, and Yuan Yan Tang. "An improved noninvasive method to detect diabetes mellitus using the probabilistic collaborative representation based classifier." *Information Sciences* 467 (2018): 477-488. https://doi.org/10.1016/j.ins.2018.08.011
[12] Liu, Caixia, and Mingyong Pang. "One-dimensional image surface blur algorithm based on wavelet transform and bilateral filtering." *Multimedia Tools and Applications* 80, no. 19 (2021): 28697-28711. https://doi.org/10.1007/s11042-021-10754-x
[13] Humeau-Heurtier, Anne. "Texture feature extraction methods: A survey." *IEEE access* 7 (2019): 8975-9000. https://doi.org/10.1109/ACCESS.2018.2890743
[14] Kuzdeuov, Askat, Dana Aubakirova, Darina Koishigarina, and Huseyin Atakan Varol. "TFW: Annotated thermal faces in the wild dataset." *IEEE Transactions on Information Forensics and Security* 17 (2022): 2084-2094. https://doi.org/10.36227/techrxiv.17004538.v2
[15] King, Davis E. "Dlib-ml: A machine learning toolkit." *The Journal of Machine Learning Research* 10 (2009): 1755-1758.
[16] Indi, Trupti S., and Dipti D. Patil. "Nail feature analysis and classification techniques for disease detection." *International Journal. Computer Science Engineering* 7, no. 5 (2019): 1376-1383. https://doi.org/10.26438/ijcse/v7i5.13761383
[17] Wang, Zhi-Hao, Gwo-Jiun Horng, Tz-Heng Hsu, Chao-Chun Chen, and Gwo-Jia Jong. "A novel facial thermal feature extraction method for non-contact healthcare system." *IEEE Access* 8 (2020): 86545-86553. https://doi.org/10.1109/ACCESS.2020.2992908

[18] Anita, John Nisha, and Sujatha Kumaran. "Detection and segmentation of meningioma tumors using the proposed MENCNN model." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 32, no. 2 (2023): 1-13. https://doi.org/10.37934/araset.32.2.113

[19] Sarode, K., R. Savdekar, and T. Chaudhari. "Texture feature analysis of an image using gray level co-occurrence matrix." *International Journal of Novel Research and Development* 7, no. 2 (2022): 139-143.

[20] Nasharuddin, Nurul Amelina, and Nurul Shuhada Zamri. "Non-parametric machine learning for pollinator image classification: A comparative study." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 34, no. 1 (2024): 106-115. https://doi.org/10.37934/araset.34.1.106115

[21] Yang, Xiaoxia, Yisheng Gao, Shuhua Zhang, Zhedong Ge, and Yucheng Zhou. "Research on rosewood micro image classification method based on feature fusion and ELM." *Journal of Renewable Materials* 10, no. 12 (2022): 3587. https://doi.org/10.32604/jrm.2022.022300

[22] Adjabi, Insaf, Abdeldjalil Ouahabi, Amir Benzaoui, and Sébastien Jacques. "Multi-block color-binarized statistical images for single-sample face recognition." *Sensors* 21, no. 3 (2021): 728. https://doi.org/10.3390/s21030728

[23] Chi, Jianning, Xiaosheng Yu, Yifei Zhang, and Huan Wang. "A novel local human visual perceptual texture description with key feature selection for texture classification." *Mathematical Problems in Engineering* 2019, no. 1 (2019): 3756048. https://doi.org/10.1155/2019/3756048

[24] Khatri, Manvi, and Ajay Sharma. "A computational approach for score-level fusion decision-making of multi-biometric recognition system using ant colony optimisation." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 36, no. 1 (2023): 147-158. https://doi.org/10.37934/araset.36.1.147158

[25] Ching, Wong Pui, Shahrum Shah Abdullah, Mohd Ibrahim Shapiai, and AKM Muzahidul Islam. "Performance enhancement of alzheimer's disease diagnosis using generative adversarial network." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 45, no. 2 (2025): 191-201. https://doi.org/10.37934/araset.45.2.191201

[26] Duodu, Nana Yaw, Warish D. Patel, and Hakan Koyuncu. "Advancements in Telehealth: Enhancing breast cancer detection and health automation through smart integration of iot and cnn deep learning in residential and healthcare settings." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 45, no. 2 (2025): 214-226. https://doi.org/10.37934/araset.45.2.214226