Adaptive Multiple Layer Retinex-enabled Color Face Enhancement for Deep Learning-based Recognition

Duong Binh Giap¹, Tuyen Ngoc Le², Jing-Wein Wang³, Chia-Nan Wang⁴

¹Department of Electronics Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan.
²Department of Electronic Engineering, Ming Chi University of Technology, New Taipei City, 24301, Taiwan.
³Institute of Photonics Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan.
⁴Department of Industrial Engineering and Management, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan.

Corresponding author: Tuyen Ngoc Le (e-mail: tuyennt75@gmail.com), Jing-Wein Wang (e-mail: jwwang@nkust.edu.tw).

This work was supported in part by Taiwan, ROC, Ministry of Science and Technology under Grant 110-2221-E-992-076.

ABSTRACT Face image captured under uncontrolled illumination conditions is one of the most significant challenges for real-world human face recognition systems. To overcome this problem, we proposed a novel method called adaptive multiple-layer retinex-based color face enhancement (AMRF) to enhance the face images. Firstly, we use an associative filter to decompose a color face image into illumination and reflectance components at multiple layers. Then, the illumination components in each layer are adjusted automatically by multiplying with corresponded illumination compensation coefficients calculated through a referenced Gaussian template. The enhanced color face image is finally obtained by composing the compensated illumination components with the integrated reflectance component based on the Retinex theory. The experiment was performed on four popular color face datasets: LFW, IJB-C, CMU Multi-PIE, and CMU-PIE. Our proposed method makes face images more precise, natural, and smooth. The experiment results show that AMRF’s image quality assessment scores are significantly better than the original and other enhanced methods’ images. Furthermore, AMRF considerably improves the recognition accuracy of deep learning-based face recognition models, such as FaceNet, and ArcFace. Finally, our proposed method also saves computational time comparing the other techniques.

INDEX TERMS Retinex model, color face image enhancement, illumination compensation, face recognition.

I. INTRODUCTION

In recent years, face recognition has become a popular research topic of computer vision, pattern recognition, and machine learning due to its essential role in real-world applications. Achievements accomplished in recent researches on face recognition are significant. However, many challenges remain that affect the high performance of a human face recognition system [1]. These challenges are caused mainly by the significant variations in facial expression, viewpoint, illumination conditions, noise, etc. Meanwhile, an exemplary face recognition system should ideally provide a capability to recognize a face, in which this face image is independent as possible of the variations [2]. Among mentioned factors, the illumination variation is acknowledged as the most complex and challenging challenge in face recognition. Fig. 1. shows an example of one human face under different illumination conditions. As seen, the quality of the observed face image is severely affected by the illumination variation. Even it is nearly impossible for human visual to judge the identity of the face images under harsh illumination conditions. Unfortunately, this problem is unavoidable in human face recognition systems in real-world applications, such as public security, social multimedia, or intelligent systems. Therefore, illumination compensation of the face image is a practical approach to eliminate lighting variations before face recognition. [3].

FIGURE 1. An example of one human face image taken under different illumination conditions
In the past decades, numerous methods [4]-[19] have been proposed to overcome the problem of illumination variation, which can be mainly grouped into three categories: methods using a transform in the spatial domain [4]-[7], methods using transform to the frequency domain [8]-[11], and methods based on deep learning [17]-[19]. Here, we briefly overview the corresponding methods of these three categories.

The first category involves methods employed in the spatial domain to overcome the illumination variations problem. Most methods in this category try to estimate the spatial low-frequency component, which is contained the illumination information of the face image. Liu et al. [4] proposed methods based on Local Histogram Specification (LHS) to pre-process face images under different illumination conditions by using a high-pass filter on a face image to remove the low-frequency illumination signals, as well as enhance the feature of the face images. Shan et al. [5] use Gamma Intensity Correction (GIC) to normalize the entire image intensity. Then the histogram equalization (HE) and Quotient Illumination Relighting (QIR) methods are applied to synthesize face images under a pre-defined normal lighting condition. Shim et al. [6] proposed a face relighting method by estimating the pose, reflectance, and lighting from a set image of a face. In this way, they train a pose and pixel-dependent subspace model of the reflectance function using a face dataset containing samples of pose and illumination variations. In a different approach, Yang et al. [7] use a physical model named retinex decomposition to extract the invariant component considered the reflectance of the face image by using the bilateral filter. Although the results of the mentioned methods show that it can partly resolve the illumination variation problem, there are remaining problems when using these methods. Firstly, the improvement of recognition accuracy by using these methods is relatively low. In addition, the result of face images is not natural and smooth when a huge amount of information of the face images are significantly removed, especially the color information demonstrated as one of the most important factors for face recognition in many studies [8]-[11].

To overcome the shortcomings of the methods in the first category, numerous methods employed in the frequency domain have been proposed. Goh et al. [12] used the wavelet transform to eliminate the variant illumination existing on the low-frequency subband and enhance the face image by using local binary pattern histogram (LBP). Then, a face image with invariant illumination is obtained by adopting the wavelet fusion technique. Chen et al. [13] proposed a method using a discrete cosine transform (DCT) to enhance the illumination in the logarithm domain. Accordingly, the low-frequency band mainly contains variations of illumination. Therefore, an appropriate number of the DCT coefficients are eliminated to minimize the illumination variations of face images. On the other hand, Wang et al. [14] proposed a novel method named adaptive singular value decomposition (SVD) in the two-dimensional discrete Fourier domain (ASVDF). They adopt the SVD and Fourier transform to enhance the face images to eliminate the influence of the environmental lighting of the face image under the variant illumination conditions. Accordingly, the proposed method first transforms a color face image into the two-dimensional discrete Fourier domain to obtain the magnitudes of three color channels. Then, the illumination of the face image can be compensated by multiplying the singular value matrix of the magnitudes with their compensation weight coefficients. Similarly, Wang et al. [15] proposed adaptive SVD in the 2D discrete wavelet domain (ASVDW) to enhance a color face image. This proposed method is implemented by computing the distributions of the brightness pixels in the three color channels of the color face image and the correlations among its wavelet subbands. Although these methods can produce clear and natural face images and relatively improve the recognition rate of face recognition systems, there are still problems in the efficiency of face image enhancement and computational time consumption.

In recent years, deep learning-based enhancement methods have been proposed gradually in the computer vision domain. In particular, generative adversarial network (GAN) [16] is a strong and powerful model, which has achieved many successes in solving image-to-image problems. Therefore, most of the methods in the third category are proposed by using GAN to resolve the illumination variation problem. Han et al. [17] proposed the asymmetric joint GAN (AJGAN) to normalize face images under different illumination conditions. Zhang et al. [18] proposed the IL-GAN framework, which adopts the variational auto-encoder and conditional GAN (cGAN) to produce the illumination variation descriptors for face recognition. Hu et al. [19] proposed a FIN-GAN framework combining the retinex theory and conditional GAN for face illumination normalization. Firstly, they use a self-supervised learning model to learn the illumination variation under different lighting conditions of a pair of face images. Then, the conditional GAN is adopted to generate the illumination normalization. However, the deep learning-based methods have a common characteristic that the illumination condition is directly adjusted by image-to-image translation. Therefore, these methods require a set of face images used for the training process of the network model. As a result, it leads to inefficiency since it cannot provide high performance and adaptability in practical applications.

To resolve the problems from the mentioned methods, in this paper, we proposed a novel method named adaptive multiple layer retinex-based color face enhancement (AMRF). First, we adopt the retinex theory [20] to represent a face image in combinations of multiple illumination and reflectance components by using an associative filter in the spatial domain. Then, a Gaussian template corresponding to
the given face image is generated as an adjustment reference template for illumination compensation. Next, the illumination components in each layer are enhanced adaptively by multiplying with corresponded compensation weight coefficients. The enhanced color face image is finally obtained by composing the enhanced illumination with the integrated reflectance components. The experiments performed on four of the most famous color face datasets, namely, LFW [28], IJB-C [29], CMU Multi-PIE [30], and CMU-PIE [31], show that the proposed AMRF method cannot only enhance color face images to become more natural, clearer, and smoother, but can also dramatically improve the recognition accuracy and computational time consumption of human face recognition systems.

The rest of this paper is organized as follows. Section II presents the multiple layer representation of the color face image by using the retinex model. Section III details our proposed method. Section IV discusses the experimental results, and Section V presents the conclusions.

II. COLOR FACE MULTIPLE LAYER REPRESENTATION

In this section, the representation of a color face image in multiple layers by using the retinex model based on the associative filter is discussed in detail. Firstly, the associative filter-based retinex model used for color face decomposition is presented. At last, the method used to represent a color face by combinations of multiple layers of illumination and reflectance component is introduced.

A. ASSOCIATIVE FILTER-BASED RETINEX MODEL FOR COLOR FACE IMAGE

The retinex theory was early proposed by Land and McCann [20], [21] as a model of human vision’s lightness and color perception. In particular, the light coming to the human eye is essentially the combinational product of the illumination and reflectance components. These two components always maintain a dependent correlation from each other. The human eye cannot determine the illumination unless the reflectance is uniform, and the eye cannot determine the reflectance unless the lighting is uniform. In the digital image processing domain, the correlation of illumination and reflectance are mathematically modeled as follows:

\[ I^c(x, y) = R^c(x, y) \cdot L^c(x, y), \]  

where \( I(x, y) \) indicates an image, \( R(x, y) \) is the reflectance component, \( L(x, y) \) is the illumination component of the given image \( I \), and \( c \in \{b, g, r\} \) is the color channel index. The operator \((\cdot)\) denotes the element-wise multiplication of two matrices.

The estimation of the illumination and reflectance components is based on the separation of spatial-frequency subbands of the given image. In particular, the illumination component resides in the low-frequency subband, and the reflectance component resides in the high-frequency subband of the image. Therefore, the illumination represents the various lightness of the object. In contrast, the reflectance shows the natural properties of the captured object, which is considered to be consistent under the different lightness conditions. Many studies attempted to propose algorithms for separating illumination and reflectance components based on this idea. However, most algorithms are based on using a low-pass filter, such as Gaussian [22], bilateral [23], or bright-pass filter [24], to estimate the illumination component. Then, the reflectance component can be estimated by the ratio of the original image over the estimated illumination component.

In general, filters used for the retinex model to estimate the spatial low-frequency component of an image play an important role. It is the primary key to deciding the pureness of estimated components. Several typical low-pass filters, such as Gaussian, bilateral, bright-pass filters, were used to estimate the illumination in the previous studies. However, these filters cannot effectively remove the fine details because of the global filtering implements of these filters. This paper utilizes a low-pass filter called associative filter [25] to represent a color face image through combinations of multiple layers of illumination and reflectance components. The associative filter outperforms to remove fine details of an image than other typical low-pass filters used in [22]-[24] by its capability to calculate the local maximum’s weighted average. The weight is negatively related to the difference of the pixels between the input image and its local maximum. In addition, the complexity of the associative filter is much lower than other filters. Finally, the product of the associative filter application is a pure and uniform low-frequency component.

The illumination component \( I^c(x, y) \) of color channel \( c \) in Eq. (1) can be obtained by using the associative filter defined in [25] for a color channel \( I^c(x, y) \) of the given image \( I(x, y) \) as follows:

\[ I^c(x, y) = \frac{1}{W(x, y)} \sum_{(i,j) \in \Omega(x,y)} \left[ e^{-\left(\frac{\left(I^c(i,j) - I^c(x,y)\right)^2}{2\sigma^2}\right)} I^c(i,j) \right] \]  

(2)

where the local patch \( \Omega(x, y) \) and \( \sigma \) are set empirically. \( I^c(x, y) \) is the given color channel image, \( I^{cs}(x, y) \) is the coarse image and is defined as the local maximum of \( I^c(x, y) \) as follows:

\[ I^{cs}(x, y) = \max_{(u,v) \in \Omega(x,y)} I^c(u, v). \]  

(3)

where \( \Omega(x, y) \) is a sliding window, and centered at \( (x, y) \). \( W(x, y) \) denotes the normalization factor to ensure pixel weights sum to 1.0 as determined as follows:

\[ W(x, y) = \sum_{(i,j) \in \Omega(x,y)} e^{-\left(\frac{\left(I^c(i,j) - I^c(x,y)\right)^2}{2\sigma^2}\right)}. \]  

(4)

Using the associative filter, the high-frequency spatial component, which contains the details of the color face image, is effectively removed. Besides, the low-frequency component considered as illumination of the given color face image is preserved. Therefore, the reflectance component corresponding to such estimated illumination can be
determined as the ratio of the given image over its corresponding estimated illumination as follows:

\[ R^c(x, y) = \frac{I^c(x, y)}{E(x, y)}. \]  

(5)

An illustration of the retinex decomposition based on the associative filter on color face images is shown in Fig. 2. As seen, the shape and details of the face image in Fig. 2(a) are considerably removed in the illumination component depicted in Fig. 2(b). In contrast, this relevant information is mainly contained in the reflectance component, as shown in Fig. 2(c). Besides, although these face images were taken under different illumination conditions, the reflectance, which is the invariant component can be effectively separated from the illumination component. Therefore, the color face image can be decomposed into the illumination and reflectance components effectively by using the retinex model based on the associative filter.

\section*{B. MULTIPLE LAYER REPRESENTATION OF COLOR FACE IMAGE}

Since the high-frequency spatial information, such as shape and details, may still exist in the whole range of the spatial frequency bands, decomposing a face image at the first level by using the retinex model based on a low-pass filter might not be enough to separate the illumination and reflectance components effectively, as shown in Fig. 2. Therefore, multiple layer decomposition is proposed to resolve this problem to represent the information of a color face image in multiple spatial frequency subbands. In this way, the information of the face image can be represented in more detail, and valuable information can be further used to enhance the quality of the face image.

To implement the multiple layer decomposition, the estimated illumination at the \(l\)th level of color channel \(c \in \{b, g, r\}\), denoted as \(L^c_l(x, y)\), is further decomposed into a low-frequency component \(L^c_{l+1}(x, y)\) and a high-frequency \(R^c_{l+1}(x, y)\). Finally, the illumination component \(L^c_{l+1}(x, y)\) is estimated by filtering the illumination component \(L^c_l(x, y)\) at the previous level with the associative filter as expressed in Eq. (2). Therefore, the decomposition is implemented to satisfy:

\[ L^c_l(x, y) = L^c_{l+1}(x, y) \cdot R^c_{l+1}(x, y), \]

where \(L^c_l(x, y)\) is obtained by filtering the color channel \(I^c(x, y)\) by the associative filter in Eq. (2). The reflectance component \(R^c_{l+1}(x, y)\) is computed by

\[ R^c_{l+1}(x, y) = \frac{L^c_l(x, y)}{L^c_{l+1}(x, y)}. \]

Consequently, an image \(I^c(x, y)\) is composed of multiple spatial frequency bands as follows:

\[ I^c(x, y) = L^c_n(x, y) \prod_{i=1}^{n} R^c_i(x, y), \]

or

\[ I^c(x, y) = L^c_n(x, y) \prod_{i=0}^{n} \frac{L^c_i(x, y)}{L^c_{i+1}(x, y)}. \]

where \(n\) is the number of layers. Since the reflectance component is calculated as the ratio of the previous and its present illumination components, the dynamic range or the lightness range of the low-frequency component after one decomposition operation becomes smaller or the variation of \(L^c_{l+1}(x, y)\) is less than \(L^c_l(x, y)\). Once the number of layers reaches an enormous enough value, the variation of the illumination \(L^c_n\) is tiny, and its lightness will be completely uniform.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig2.png}
\caption{The associative filter-based retinex decomposition of color face images under different illumination conditions. (a) Cropped face images in the CMU Multi-PIE database; (b) The illumination component of (a); (c) The reflectance component of (a).}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig3.png}
\caption{The multiple layer representation process of a color face image using the retinex model based on the associative filter. \(I(x, y)\) denotes a given color face image, \(L_i(x, y)\) indicates the illumination component, \(R_i(x, y)\) indicates the reflectance component, and the circled \(D\) symbol indicates the retinex decomposition operation using the associative filter.}
\end{figure}

Fig. 3 illustrates the multiple layer decomposition process of the retinex model based on the associative filter to obtain illumination and reflectance components in different spatial frequency subbands of a color face image. First, the estimated illumination component obtained at the first level is further used as the following decomposition process input. Then, in the second level decomposition, the estimated illumination at the first level is filtered by the associative filter one more time to obtain the second level illumination and reflectance components, and so on. In this way, the decomposition process is implemented \(n\) times iteratively.
until obtaining an utterly uniform illumination component denoted as \( L_{n}^{c}(x,y) \).

![Multiple layer representation uses the associative filter-based retinex decompositions of a color face image. (a) A color face image in the CMU Multi-PIE database. (b) The illumination components of (a). (c) The normalized reflectance components correspond to the illumination components in (b).](image)

An illustration of the multiple layer decomposition using the retinex model based on the associative filter of a color face image is shown in Fig. 4. As seen, the illumination component at the 1st level \((i = 1)\), as shown in the 1st column of Fig. 4(b), still contains facial primarily structures which are reflectance information. Similarly, the reflectance information remains considerable in the 2nd, 3rd, 4th, and 5th illumination layers. However, each filtering operation removes more detailed information to make the illumination component smoother. As a result, the variation of the illumination component is gradually decreased during the increase in \( i \). When the number of the decomposition operation reaches over 10, the detailed information existing on the illumination components is mostly eliminated since its corresponding reflectance components shown in Fig. 4(c) does not contain much detailed information of the color face image. In other words, the estimated illumination component becomes pure and more uniform when the details and shape of the color face image are mostly removed. Therefore, the color face image can separately decomposed into multiple layers of illumination and reflectance components by the associative filter-based retinex model. Based on multi-level estimated illumination components, we propose a method to adaptively enhance the illumination of the color face image at different frequency bands by multiplying the estimated illumination components by corresponding adaptive compensation weight coefficients. The detailed algorithm is discussed in the next Section.

III. MULTIPLE RETINEX-BASED COLOR FACE IMAGE ENHANCEMENT METHODOLOGY

Illumination is demonstrated as a vital factor in face recognition. To overcome the invariant illumination problem in face recognition, we propose a novel method called the adaptive multiple layer retinex-based color face enhancement (AMRF). This method involves three primary steps, as demonstrated in Fig. 5. In particular, we first adopt the retinex model using the associative filter to represent a color face image through combinations of multiple layers of illumination and reflectance components. Then, the estimated illuminations are enhanced with compensation weight coefficients generated by using a corresponding Gaussian template as the adjustment reference. To this end, the enhanced color face image is yielded by composing the enhanced illumination with the integrated reflectance component. The advantage of this approach is that the face image is more natural since it still retains both illumination and reflectance components.

![The illustration of the multiple layer color face image enhancement by using the associative filter-based retinex model.](image)

The distribution of pixel intensity values of an image can be described statistically by the probability density function (PDF). A normalized intensity image that does not suffer any illumination problem can be considered a matrix with a PDF having a Gaussian distribution with a mean of 0.5 and a standard deviation of 1.0 [26]. Therefore, the face illumination enhancement can be obtained by adopting a matrix with the PDF of Gaussian distribution as the reference.
Therefore, to enhance the color face image, we utilize a Gaussian template, denoted as $G_a$, is designed to have the same size as the given face image and PDF as follows:

$$
PDF(G_a(x, y)) = \frac{1}{\sigma \sqrt{2\pi}} e^{\frac{(G_a(x, y) - \mu)^2}{2\sigma^2}},
$$

where $\mu$ is the mean and $\sigma$ denotes the standard deviation. Due to the dynamic range of the pixel intensity value of an image which is between 0.0 to 255.0, the values of $\mu$ and $\sigma$ of the Gaussian template are linearly scaled to 128.0 and 32.0, respectively.

As discussed in section II.B, the illumination information exists in the whole range of spatial frequency bands. Therefore, to enhance the color face image, we utilize a Gaussian template as an adjustment reference to compensate for the illumination in the whole ranges of spatial frequency subbands of the face image. The objective is to obtain an enhanced color face image with a near normal distribution. Given a cropped color face image denoted as $I$ of size $M \times N$ and its corresponding Gaussian template, denoted as $G_a$, are first split into RGB color channels to obtain $I^c$ and $G_a^c$, $c \in \{r, g, b\}$, respectively. Then, $I^c$ and $G_a^c$ are simultaneously decomposed to obtain illumination and reflectance components by using the associative filter-based retinex model. Therefore, $I^c$ and $G_a^c$ can be expressed at the first layer ($i = 1$) as follows:

$$
I^c(x, y) = R^c_{i=1}(x, y) \cdot L^c_{i=1}(x, y),
$$

and

$$
G_a^c(x, y) = R^c_{Ga,a=1}(x, y) \cdot L^c_{Ga,a=1}(x, y),
$$

where $L^c_{i=1}(x, y)$ and $L^c_{Ga,a=1}(x, y)$ are the illumination components of color channel $c$ of the given color face image and its corresponding Gaussian template, respectively. $R^c_{i=1}(x, y)$ and $R^c_{Ga,a=1}(x, y)$ denote the reflectance components corresponding to $L^c_{i=1}(x, y)$ and $L^c_{Ga,a=1}(x, y)$, respectively. The illumination component is then compensated to obtain enhanced illumination, denoted as $L^{e,c}_{i=1}(x, y)$, before decomposing to receive the illumination and reflectance component at the 2nd layer, and so on.

To enhance the illumination of the face image, the illumination at the $i^{th}$ layer is compensated by multiplying simultaneously three illumination components of the RGB color channels by adaptive compensation weight coefficients. The compensation operation for the $i^{th}$ layer illumination is briefly described in Algorithm 1. Accordingly, the mean pre-compensation value of each illumination component is first calculated. Then, by using the highest mean value as the reference, the individual compensation weight coefficients for the remaining color channels are obtained adaptively by using their ratios to the reference value. The mean values of each color channel illumination component of the color face image and corresponding illumination component of the Gaussian template are calculated as follows:

$$
\mu_{I_i=1,2,...,n}^{c(r,g,b)} = \frac{1}{M \times N} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} L^c_{i}(x, y),
$$

and

$$
\mu_{Ga,a=1,2,...,n}^{c(r,g,b)} = \frac{1}{M \times N} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} L^c_{Ga,a}(x, y).
$$

Then, the highest mean of three illumination components of the color face image is determined as follows:

$$
M_i=1,2,...,n = \max_{i=1,2,...,n} (\mu_{I_i=1,2,...,n}^{c(r,g,b)}).
$$

**Algorithm 1: $i^{th}$ illumination layer compensation**

**Input:** $i^{th}$ illumination layers $I_i^{c(r,g,b)}$ and $G_a_i^{c(r,g,b)}$.

**Output:** $i^{th}$ compensated illumination layers $L_i^{e,c(r,g,b)}$.

1. $\mu_{I_i=1,2,3,...,n}^{c(r,g,b)} \leftarrow$ calculate means of illumination of $I_i^{c(r,g,b)}$ via Eq. (13);
2. $\mu_{Ga,a=1,2,3,...,n}^{c(r,g,b)} \leftarrow$ calculate mean of illumination of $G_a^{c(r,g,b)}$ via Eq. (14);
3. $M_i=1,2,3,...,n \leftarrow$ obtain the highest mean value of the $i^{th}$ illuminations of color channels via Eq. (15);
4. $\xi_i^c \leftarrow$ calculate the compensation weight coefficient via Eq. (16);
5. $L_i^{e,c} \leftarrow$ compensate the $i^{th}$ illumination layer via Eq. (17);
6. **Return** $L_i^{e,c}$.

Our purpose is to adjust the illumination components automatically to obtain the enhanced color face image with a near normal distribution. Therefore, if the face image is dark, the illumination is compensated with a scalar larger than 1.0. In contrast, if the face image is too bright, the illumination of the image is reduced gradually by multiplying the illumination components by a scalar smaller than 1.0. To do this, we propose an adaptive illumination compensation weight coefficient for each color channel at the $i^{th}$ layer as follows:

$$
\xi_i^c = \frac{M_i}{\mu_{I_i}^{c(r,g,b)}} \sqrt{\frac{\mu_{Ga,a}^{c(r,g,b)}}{\mu_{I_i}^{c(r,g,b)}}}, \ i = 1,2,3, \ldots, n.
$$

The enhanced illumination component at the $i^{th}$ layer is obtained by multiplying by the illumination compensation weight coefficient $\xi_i^c$ as follow:

$$
L_i^{e,c}(x, y) = \xi_i^c \times L_i^{c}(x, y).
$$

In this way, the illumination is enhanced adaptively in each layer decomposition. In addition, the illumination of the face image becomes more uniform since its variation is gradually decreased. To this end, the uniform illumination is obtained and enhanced before composing with the integrated...
reflectance component to produce the enhanced color face image, denoted as \( I^c_\xi \) as follows:

\[
I^c_\xi (x, y) = R^c(x, y) \cdot L^c_{\text{uniform}}(x, y),
\]

(18)

where \( L^c_{\text{uniform}}(x, y) \) is the enhanced uniform illumination component, and \( R^c(x, y) \) denotes the integrated reflectance component, which is determined as follows:

\[
R^c(x, y) = \frac{I^c(x, y)}{L^c_{1,i=1}(x, y)} \cdot \prod_{i=1}^{n} \frac{L^c_{i+1}(x, y)}{L^c_{i,i=1}(x, y)}
\]

(19)

To this end, the enhanced color face image, denoted as \( I^c_\xi(x, y) \), is finally obtained by merging the three color channel images \( I^c_\xi(x, y) \).

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/
proposed method can be automatically applied to a database of any kind to improve the quality of face images for deep learning-based face recognition, as discussed in the next section.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we implemented experiments to demonstrate the performance of our proposed method in color face image enhancement. Firstly, we adopted image quality assessment measures (IQAs), namely PIQUE [32], BRISQUE [33], and NIQE [34], to evaluate the quality image enhancement capability, especially the facial naturalness recovery. Then, to illustrate the effectiveness of the proposed method in improving the face recognition performance, we implemented experiments on the four most famous color face databases with large illumination variations, namely, LFW [28], UB-C [29], CMU Multi-PIE [30], and CMU-PIE [31] datasets. We adopted state-of-art deep learning-based face recognition models FaceNet [36] and ArcFace [37] to evaluate our proposed method's performance comprehensively. The performance of the proposed method is further compared with other effective previous methods, namely, ASVDF [14], ASVDW [15], FIN-GAN [19], GRIR [35], and to demonstrate the high performance of our proposed method for face recognition.

A. THE ASSOCIATIVE FILTER PARAMETERS

The effectiveness of the proposed method firstly relies on the performance of the illumination estimation task of the retinex decomposition and then is based on the computing of compensating coefficients. In this subsection, we discuss the configuration of the associative filter to select the appropriate parameters used for face enhancement in our proposed method.

As discussed in section II.A, the weighted average of the coarse image (maximum local image) and the difference of the pixels between the rough image and the reference image are the key ideas of the associative filter. Therefore, we conducted experiments for choosing the size of the local patch $\Omega$ and the sliding window $\Omega'$. There were three different types of associative filters, denoted as $filter_{53}$, $filter_{97}$, and $filter_{157}$, to be discussed in our experiments. The configuration of each filter is shown in Table I.

The criteria for choosing the suitable filter in our proposed method are based on its performance in removing the fine detail of a color face image and computational time consumption. Therefore, we first evaluate each filter's computational time consumption for each layer decomposition of a color face image of size $128 \times 128$ pixels. The results of three types of filters are shown in Table II. As seen, the first filter, namely $filter_{53}$, only consumes 0.07 seconds, whereas the second and third filters, denoted as $filter_{97}$ and $filter_{157}$, cost about three and six times compared to $filter_{53}$ to decompose a layer, respectively.

These three different types of filters are further evaluated in face enhancement performance. We first collected 500 dark face images and 500 bright face images from the CMU Multi-PIE database to compare their performances. Then, we grouped these images into two categories: dark and bright. Finally, we applied our proposed AMRF method with different filters to enhance the color face images for each category. The averages of the compensation weight coefficients for the face images in multiple layers are used for evaluating the performance of the three types of filters. The experimental results are depicted in Fig. 7. As seen, the compensation weight coefficients of the proposed method using all three filters alternately for both dark, and bright images are generally stable. In particular, the average of the compensation weight coefficients is gradually decreased to close to 1.0 during the increment of layer index for the dark
images. In contrast, the coefficients slowly increase to be near 1.0 when the number of layers increases for the bright images. Therefore, the coefficients of both dark and bright images are converged at the value of 1.0 and reach a stable state of illumination. As seen in Figs. 7(a), (b), and (c), the steady-state of coefficients of all three filters are reached when the number of layers in our proposed method is over 10. In other words, the illumination of facial images is completely enhanced after ten layers. In addition, the coefficients in Figs. 7(b) and (c) at the period from the 5th to 9th layer, the average weight compensation coefficients for filters filter_97 and filter_157, are not stable since the coefficients for dark images are less than 1.0 and bigger than 1.0 for the bright images. However, the coefficients for the filter_53 are more stable, smoother in this period, and converge once the layer index reaches over 10. Therefore, by using a filter with the filter_53 configuration, the color face image is enhanced stably and faster than other filters. After comprehensively comparing performances and the computational time consumption of three filters used for our proposed method, filter_53 outperforms the two remaining filters. Indeed, filter_53 is the smallest size for local patches and sliding windows. Therefore, the computational time consumption used for estimating the illumination is less than others. Besides, its performance in face image enhancement outperforms higher than others. Thus, in this study, we adopted the filter_53 for our proposed method to reduce the computational time consumption and maintain the high performance in enhancing color face images. Furthermore, as discussed above, the number of layers in our proposed method is fixed at ten layers. In addition, the computational time can be further decreased by down-sampling the estimated illumination after each layer decomposition [27]. Therefore, we use the sampling scaled ratio of 1.25 for each layer decomposition. If the sampling rate is more significant, the image quality will be affected. On the other hand, it will cost you time if it is smaller. Besides, we adopt the linear interpolation method to save the computational time while the quality of face images is retained sufficiently.

B. COLOR FACE IMAGE QUALITY ASSESSMENT

Besides facial illumination enhancement capability, the proposed method enables the naturalness and smoothness of enhanced images. To demonstrate the performance of the proposed method in the naturalness enhancement for the color face images objectively, we assessed the quality of result enhanced images with three state-of-the-art image quality assessment metrics (IQA), namely perception-based image quality evaluator (PIQUE) [32], blind/referenceless image spatial quality evaluator (BRISQUE) [33], and naturalness image quality evaluator (NIQE) [34]. The assessment was carried out on 1,000 cropped 128×128-pixel color face images under various illumination conditions from the CMU-PIE, CMU Multi-PIE, LWF, and IJB-C datasets.

PIQUE [32] measure is a no-reference image quality metric used for blind image quality assessment through block-wise distortion estimation. The PIQUE value is determined as follows:

$$PIQUE = \frac{\left(\sum_{k=1}^{N_{SA}} D_{sk}\right) + C_1}{N_{SA} + C_1},$$

(20)

$$D_{sk} = \begin{cases} 1, & \text{if } (\sigma_{pq} < T_{STD}) \text{ and } (\sigma_{blk} > 2 \times \beta), \\ (1 - v_{blk}), & \text{if } (\sigma_{blk} > 2 \times \beta), \\ v_{blk}, & \text{if } (\sigma_{pq} < T_{STD}), \end{cases}$$

(21)

where, $N_{SA}$ is the number of spatially active blocks, $C_1$ and $\beta$ are a positive constant and a parameter set empirically, $v_{blk}$ is the variance feature of the block, $T_{STD}$ is the set threshold, and $\sigma_{blk}$, $\sigma_{pq}$ are the block and segment standard deviations, respectively. A lower value of $PIQUE$ indicates high perceptual quality and vice versa.

Like the PIQUE measure, BRISQUE [33] is a blind/no-reference image quality assessment model using scene statistics of locally normalized luminance coefficients to quantify losses of naturalness of a given image. The BRISQUE measure is determined based on three main steps. Firstly, the BRISQUE model extracts the natural scene statistics of a given image $I$ by computing locally normalized luminances via local mean subtraction and divisive normalization as follows:

$$\tilde{I}(x, y) = \frac{I(x, y) - \mu(x, y)}{\sigma(x, y) + C},$$

(22)

where, $C$ is the constant to prevent instabilities, $\mu(x, y)$ and $\sigma(x, y)$ are the local mean and local variance fields, respectively. The normalized luminance image is then used to generate four pair-wise products, horizontal, vertical, left-diagonal, and right-diagonal, to capture pixels’ neighborhood relationships before calculating feature vectors by generalized Gaussian distribution (GGD) for predicting an image quality score.

To comprehensively demonstrate our proposed method’s facial naturalness enhancement performance, we further adopted the NIQE measure [34] to evaluate the color face image quality along with PIQUE and BRISQUE measures. NIQE is a completely blind image quality assessment evaluator based on the construction of a quality aware collection of statistical features, which are derived from a corpus of natural images. The quality of a given image is expressed as the distance between the quality aware feature model and the multivariate Gaussian model (MVG) as follows:

$$D(v_1, v_2, \Sigma_1, \Sigma_2) = \sqrt{(v_1 - v_2)\left(\frac{\Sigma_1 - \Sigma_2}{2}\right)^{-1}(v_1 - v_2)},$$

(23)

where, $v_1, v_2$ are the mean vectors, and $\Sigma_1, \Sigma_2$ are the covariance matrices of the natural MVG model and the given image’s MVG model. A smaller score of the NIQE measure indicates better perceptual quality and vice versa.
Our evaluation experiments alternately used the mentioned IQA metrics to compute average image quality scores of 1,000 original color face images and their enhanced images by ASVDW, ASVDF, GRIR, FIN-GAN and proposed AMRF methods. It is noted that all color face images used for the quality evaluation are uncompressed bitmap image files (BMP) to preserve the highest quality of the images.

| Methods     | Original | ASVDW | ASVDF | GRIR | FIN-GAN | AMRF  |
|-------------|----------|-------|-------|------|---------|-------|
| PIQUE       | 31.461   | 28.384| 28.942| 31.024| 28.443  | 27.801|
| BRISQUE     | 29.258   | 27.113| 27.475| 28.932| 27.522  | 27.028|
| NIQE        | 18.874   | 18.589| 18.751| 18.876| 18.723  | 18.312|

The experimental results of the image quality assessment are shown in Table III. As it can be seen, all enhancement methods used in the evaluation can enhance the quality of the color face images. However, there is a large range of variations between the IQA scores of these methods. In particular, while the GRIR method can only slightly improve the quality of the color face images, other methods, such as ASVDW, ASVDF, and FIN-GAN can considerably improve the color face image quality. However, our proposed AMRF method even performs better in enhancing the quality of the color face images since its IQA scores are significantly lower than the scores for the original and other methods’ images.

C. RESULTS OF THE LFW DATASET

The Labeled Faces in the Wild (LFW) [28] is a public benchmark dataset compiled by the Computer Vision Laboratory of Massachusetts State University Amherst for face recognition. The LFW dataset contains 13,233 face images of 5,749 subjects collected from the web. Each face image was labeled with the name of the subject. The face images in the LFW dataset are captured under unconstrained environmental conditions. Thus, it provides a large range of variations seen in daily life, such as lighting, expressions, poses, ages, gender, clothing, background, camera quality, and more. In this study, the LFW dataset is utilized to evaluate our proposed method’s performance in complex scenarios.

The LFW dataset involves four different sets of images, including the original and three different types of “aligned” images named “funneled images” (ICCV 2007), LFW-a, and “deep funneled” images (NIPS 2012). This study adopted the “deep funneled” images (NIPS 2012) set for evaluating the proposed AMRF method performance. Firstly, we extracted the face images in the NIPS 2012 set of the LFW dataset using a deep learning-based face detection model, namely, multi-task cascaded convolutional neural network (MTCNN) [38]. MTCNN is a framework developed for both face detection and face alignment. The process of MTCNN involves three stages of convolutional networks that can recognize faces and landmark locations. All extracted face images from the LFW dataset are then resized to 160×160-pixel images.

![Fig. 8](image-url)  
**Fig. 8.** Twelve Bill Clinton’s color face images under different illumination conditions in the LFW dataset. (a) Original images. (b) ASVDW images. (c) ASVDF images. (d) GRIR images. (e) AMRF images.

Fig. 8 depicts twelve extracted Bill Clinton’s color face images from the images in the LFW dataset with variant illumination conditions of different lighting environments. These images are taken with a large range of variations of many factors, such as pose, background, camera quality, and expression, that are the leading causes affecting human face recognition systems. Thus, to achieve a higher performance of a face recognition system, the quality of the color face images needs to be enhanced by using face image enhancement methods. Figs. 8(b), (c), and (d) show the ASVDW, ASVDF, and GRIR images, respectively, corresponding to original images in Fig. 8(a). As noticed, the ASVDF images have problems with insufficient illumination enhancement, while the ASVDW images are over-enhanced to produce many over-lighting areas on the face images. Although the GRIR can slightly enhance the illumination, it cannot recover the facial skin color and the naturalness of the
face images. In contrast, the proposed AMRF method can efficiently enhance the face images under various illumination conditions, as depicted in Fig. 8(e). The AMRF images can achieve a balance in illumination enhancement to make the face images more natural and smoother than the ASVDW, ASVDF, and GRIR images.

FaceNet [36] is a deep learning-based face recognition method that directly learns a mapping of faces images to a compact Euclidean space. The distances between the face images in the Euclidean space are used to measure the face images' similarity. By using FaceNet, face recognition, verification and clustering can be implemented simply by using FaceNet embeddings as feature vectors to represent face images in Euclidean space. However, to achieve high accuracy, the FaceNet embeddings need to be optimized by training a deep convolutional network with a triplet loss function, which determines the relative difference of the distances between the matching pairs and non-matching pairs. In our study, we implemented two different scenarios to evaluate the performance of FaceNet on the LFW dataset.

In the first scenario, we use the pre-trained FaceNet model as a classifier to predict and generate feature vectors for face images. Then these feature vectors are used as input vectors to predict the identity of given face images. In our experiment, we adopted two pre-trained FaceNet models provided by David [39]. Both pre-trained FaceNet models adopted the Inception ResNet V1 network [40] using softmax loss for training on the CASIA-WebFace [41], and VGGFace2 [42] datasets. The pre-trained FaceNet models were trained on the datasets including cropped 160×160-pixel face images with 1% of the training images were used for validation. The models were trained by 275,000 steps with a batch size of 90 and the dataset is evaluated once every 5,000 steps. We use the pre-trained FaceNet models to extract FaceNet embeddings (128-dimensional feature vector) for each face image and compare its embeddings with the embeddings of another face image by using the similarity measures, namely, Euclidean or cosine distances. Then, the face match is determined by comparing the distance to a predefined threshold. We alternately evaluate the face recognition accuracy of the pre-trained FaceNet models with a batch size of 1 on the testing sets from the original, ASVDW, ASVDF, GRIR, and AMRF images. In addition, to make the model able to distinguish whether images belong to the same person, we randomly generate a pair character list based on the validating set in advance that contains two categories: the pairwise combination images of the same person labeled as “1” and the pairwise combination images of different people labeled as “0”. In our study, we generated a number of pairs for validating set equal to the number of subjects in the LFW dataset for each pair category. Thus, there were 5,749 pairs created for the same person and 5,749 pairs for different people. The average recognition rates of the pre-trained FaceNet models for the LFW images are shown in Table IV. As seen, the average recognition rates of the AMRF images are higher than ASVDW, ASVDF, and GRIR images in all evaluations on the LFW dataset. In particular, the recognition rate of the FaceNet model pre-trained by CASIA-WebFace for the AMRF images is higher than the original, ASVDW, ASVDF, and GRIR images by 0.81%, 0.46%, 0.52%, and 0.57%, respectively, when using cosine distance metric and by 0.55%, 0.25%, 0.21%, and 0.29% with Euclidean distance metric. By using the FaceNet model pre-trained by VGGFace2, the average recognition rate of the AMRF images is greater than the original, ASVDW, ASVDF, and GRIR images by 1.44%, 0.27%, 0.42%, and 0.31% when using cosine distance metric and by 1.96%, 0.15%, 0.33%, and 0.36% for Euclidean distance metric.

In the second scenario, we implement training the FaceNet model from scratch without pre-trained weights on the LFW training set and evaluate the recognition accuracy by using the trained FaceNet model. In this study, we only use one image for training, one image for validating, and the remaining images are used for testing for each person. For the person with only one image, we use that image for the training, validating, and testing sets. Therefore, the training and the validating sets contain 5,749 face images corresponding to 5,749 people for each set, while the testing set involves the remaining 10,652 images. While training the model with triplet loss function is heavy and slow, training the Inception ResNet V1 model as a classifier (not using triplet loss) is significantly easier and faster. Therefore, we implement training the FaceNet model using the LFW dataset and the softmax loss in this evaluation scenario. We train the model on the training set in 100,000 steps corresponding to 100 epochs of size 1,000 and evaluate the validation set's accuracy once every 5,000 steps. The batch size for each training step is 128 in our experiment. The learning rate is scheduled to be set to 5×10−2 for epochs in the range of 1-60, 5×10−3 for epochs in the range of 61-80, and 5×10−4 until the number of epochs reaches 100. The data augmentation is also applied during the training process to improve the training performance. Accordingly, we only apply the “random flip” function for the data augmentation task. The RMSProp optimizer is adopted for the training model. In addition, the size of the generated embedding is set to 512 dimensions. The trained FaceNet model is then used for evaluating the recognition accuracy on the testing set of the LFW dataset. As the first evaluation scenario, we evaluate the recognition rate using two distance metrics: cosine and Euclidean distances. The average recognition rates of the FaceNet model trained by the training
set of the LFW dataset are shown in Table V. As noticed, although the model is only trained in 100,000 steps, the average recognition rate of the FaceNet model evaluated on the original images is relatively high to achieve 93.81% and 93.26%, corresponding to the use of cosine Euclidean distance metrics, respectively. However, the recognition results can still be significantly improved by using our proposed method to enhance the face images. In particular, the average recognition rates of the AMRF images are higher than the original, ASVDW, ASVDF, and GRIR images by 1.44%, 0.09%, 0.2%, and 0.24% for the cosine distance metric and by 0.96%, 0.39%, 0.41%, and 0.48% for the Euclidean distance metric.

### Table V

| Methods          | Recognition rate (%) |
|------------------|----------------------|
|                 | Original | ASVDW | ASVDF | GRIR | AMRF |
| FaceNet (LFW) Cosine | 93.81±0.052 | 95.16±0.032 | 95.05±0.027 | 95.01±0.031 | 95.25±0.041 |
| Euclidean           | 93.26±0.045 | 94.83±0.021 | 94.81±0.052 | 94.74±0.023 | 95.22±0.034 |

To show more efficiency of the proposed AMRF method used in the pre-processing step for human face recognition systems, we do more experiments by using another state-of-art deep learning-based face recognition model, named ArcFace to evaluate the recognition accuracy LFW dataset. To enhance the discriminative power for large-scale face recognition, ArcFace adopts an additive angular margin loss to obtain discriminative features effectively for face recognition. Using ArcFace, the embedding feature is optimized to enforce higher similarity for samples in a class and diversity for samples in different classes by distributing them around the feature center on a hypersphere with a radius. Accordingly, the angle between the current feature and the ground truth weight is calculated using the arccos function. Then, the target angle is added by an angular margin penalty to enhance the intra-class compactness and inter-class discrepancy simultaneously. Next, the target logit is obtained using the cosine function and rescaling by a feature norm. Finally, the subsequent steps are implemented as same as in the softmax loss.

In our experiment, we used the ArcFace model implemented by using TensorFlow [43] to evaluate the recognition accuracy for the LFW dataset. The LFW dataset is first separated into training and validating sets. Then, we select 75% images for each person for the training set while using the remaining 25% images for the validating set. For the person with only one image, we use this image for both training and validating sets. Since the input image size of the ArcFace is 112×112 pixels, we compile the images in both training and validating sets to 112×112-pixel images. Therefore, the training set contains 10,279 squared face images of size 112×112 pixels with labeled classes. The validating set involves 7,023 squared face images used for evaluating the recognition accuracy of the model. In addition, similar to the FaceNet model, we randomly prepare the pair character list containing pairwise combinations for the ArcFace model to evaluate the recognition accuracy. In our experiment, we generated 5,000 pairs for matching combinations and 5,000 pairs for non-matching combinations. For a person containing k images in validating set, we permute and combine to generate k(k − 1)/2 pairs. For each person, pairs for matching combinations were chosen randomly from other images of that person, and pairs for non-matching combinations were randomly selected from a list of other person’s images. In this study, we implement the evaluation using the ArcFace model for the LFW dataset in three different scenarios.

For the first scenario, we adopt a pre-train ArcFace model in 100,600 steps on the MS-Celeb-1M dataset [44] to evaluate the entire LFW dataset. The backbone network used for the pre-trained ArcFace model was ResNet50 V2 [45]. The feature embedding size is set to 512, while the logit margin and logit scale are set to 0.5 and 64.0, respectively. The results of the first evaluation scenario using the pre-trained ArcFace model are shown in Table VI as “ArcFace_1”. As can be seen, the average recognition rate by using the pre-trained ArcFace model for the LFW validating set of the AMRF images is higher than the results of the original, ASVDW, ASVDF, and GRIR images by 1.23%, 0.48%, 0.6%, and 0.81%, respectively.

### Table VI

| Methods                      | Recognition rate (%) |
|------------------------------|----------------------|
|                              | Original | ASVDW | ASVDF | GRIR | AMRF |
| ArcFace_1                    | 93.10±0.021 | 93.85±0.033 | 93.73±0.027 | 93.52±0.024 | 94.33±0.022 |
| ArcFace_2                    | 93.26±0.041 | 93.71±0.052 | 93.52±0.033 | 93.44±0.031 | 93.97±0.030 |
| ArcFace_3                    | 97.71±0.009 | 98.44±0.006 | 98.35±0.004 | 98.16±0.029 | 98.98±0.008 |

The second evaluation scenario applies the ArcFace model from scratch without pre-trained weights to the LFW dataset. We also split the LFW dataset into a training set and validated set same as the second scenario using FaceNet. In training processes, the number of training epochs is set to 100, the epoch size is set to 1,000 steps, and the batch size is 32. The backbone network used in these experiments is ResNet50 V2, where the dimension of the final output feature embedding vector is 512. The logits scale and logit margin are set to 64.0 and 0.5, respectively. The learning rate of the training process is scheduled as 4×10−3 for epochs in the range of 1-40, 2×10−3 for epochs in the range of 41-60, 1.2×10−3 for epochs in the range of 61-80, and 4×10−4 the number of epochs reach 100. The results for the second evaluation scenario are shown in Table V as “ArcFace_2”. As seen, the result of the ArcFace model without pre-trained weights on the AMRF images is higher than the original, ASVDW, ASVDF, and GRIR images by 0.71%, 0.26%, 0.45%, and 0.53%.

For the third evaluation scenario, we train the ArcFace model on the training set with pre-trained weights generated by training on the MS-Celeb-1M dataset in 100,600 steps and...
then evaluate the validation set's recognition accuracy. The results for the third evaluation scenario are shown in Table V as “ArcFace_3”. As seen, the result evaluated on the AMRF images is greater than the original, ASVDW, ASVDF, and GRIR images by 1.27%, 0.54%, 0.63%, and 0.82%, respectively.

In general, the FaceNet and ArcFace models can achieve relatively high performance in the face recognition rate for the LFW dataset. However, the experimental results demonstrated that the performance of these face recognition models can be significantly improved by using pre-processing methods to enhance the quality of face images. Furthermore, using our proposed AMRF method, we even achieve better results in improving the face recognition accuracy evaluated on the LFW dataset.

D. RESULTS OF THE IJB-C DATASET

IARPA Janus Benchmark C (IJB-C) [29] is a dataset including video still-frames and photos for face recognition benchmarking compiled in cooperation with the Intelligence Advanced Research Projects Activity (IARPA) and National Institute of Standards and Technology (NIST) in 2017. Compared to the LFW and other datasets in the public domain, the IJB-C dataset is more challenging and will advance the state of the art in unconstrained face recognition. The IJB-C dataset contains 138,000 face images, 11,000 face videos, and 10,000 non-face images. In this paper, we evaluate the performance of our proposed AMRF method on two sub-datasets, namely “frames” and “img”, in the IJB-C dataset. In particular, the “frames” sub-dataset contains 118,966 face images, in which there are 118,483 labeled face images from 3,530 subjects and 483 unlabeled face images. The “img” sub-dataset involves 22,366 face images, in which there are 22,257 labeled from 3,531 subjects and 109 unlabeled face images. Our study only uses 118,483 face images of 3,530 subjects for the “frames” sub-dataset and 22,249 face images of 3,531 subjects for the “img” sub-dataset discarding the remaining unlabeled face images. Similar to the LFW dataset, we also adopt the MTCNN model to extract face images from the photos in both “frames” and “img” sub-datasets. The extracted face images are then resized to squared face images of size 160×160 pixels.

Fig. 9(a) shows eighteen Hillary Clinton’s color face images extracted from photos taken under variant illumination conditions of the IJB-C dataset. As seen, the face images are highly affected by different lighting of the environment to produce an extensive range of variations in illumination on the face images. Consequently, it leads to inefficiencies in the performance of human face recognition systems. To improve the recognition performance, the face images are required to enhance the quality before operating recognizing tasks. The GRIR method can only enhance the illumination of the color face images, while the facial skin color is not recovered and the images do not look natural, as displayed in Fig. 9(d). Adopting ASVDW and ASVDF methods makes the face images more precise, and the environmental lighting effect is relatively illuminated, as shown in Fig. 9(b) and (c), respectively. However, the ASVDW images are over-enhanced to produce over-lighting areas on the face images. On the other hand, the illumination on the ASVDF face images is insufficiently enhanced. The balance can be achieved when we use the proposed AMRF method to enhance the quality of the face images. As depicted in Fig. 9(e), the illumination of the face images is not only enhanced sufficiently, but the effect of the environmental lighting is also illuminated efficiently.
Consequently, the face images are more natural, clearer and the facial skin color tone is also recovered. Therefore, by using the proposed enhancement method, the face image quality can be improved to increase the recognition rate of the human face recognition system. To demonstrate the efficiency of the proposed method on the IJB-C dataset, we also use the FaceNet and ArcFace models with the same scenarios used for the LFW dataset, as discussed in sub-section IV.C. First, we use the FaceNet model to evaluate the recognition accuracy on the IJB-C dataset with two different scenarios: evaluate the recognition rates by using a pre-trained FaceNet model and train the FaceNet model from scratch.

**TABLE VII**

**AVERAGE RECOGNITION RATES OF THE ORIGINAL, ASVDW, ASVDF, GRIR, AND AMRF IMAGES BY USING THE PRE-TRAINED FACENETS FOR THE IJB-C DATASET**

| Methods                      | Sub-dataset | Metrics | Original | FaceNet (ASVDF) | ASVDW | ASVDF | GRIR | AMRF |
|------------------------------|-------------|---------|----------|-----------------|-------|-------|------|------|
| FaceNet (CASIA-WebFace)      | frames      | Cosine  | 88.05 ± 0.023 | 88.19 ± 0.032 | 88.12 ± 0.043 | 88.09 ± 0.037 | 88.65 ± 0.039 |
|                              | img         | Cosine  | 87.55 ± 0.043 | 87.71 ± 0.022 | 87.73 ± 0.048 | 87.63 ± 0.040 | 88.06 ± 0.036 |
|                              |             | Euclidean| 91.05 ± 0.012 | 91.38 ± 0.045 | 91.45 ± 0.042 | 91.38 ± 0.032 | 91.74 ± 0.031 |
| FaceNet (VGGFace2)           | frames      | Cosine  | 92.30 ± 0.017 | 93.13 ± 0.027 | 93.07 ± 0.025 | 92.95 ± 0.022 | 93.40 ± 0.035 |
|                              | img         | Cosine  | 92.06 ± 0.028 | 92.66 ± 0.014 | 92.70 ± 0.031 | 92.51 ± 0.017 | 93.01 ± 0.028 |
|                              |             | Euclidean| 96.00 ± 0.015 | 96.50 ± 0.024 | 96.48 ± 0.043 | 96.29 ± 0.018 | 96.52 ± 0.015 |

For the first scenario, we use two FaceNet models pre-trained alternately by the CASIA-WebFace and VGGFace2 to classify the face images in the “frames” and “img” sub-datasets of the IJB-C dataset. The results of the evaluations on both “frames” and “img” sub-datasets are shown in Table VII. As seen, the average recognition rates of the AMRF images in “frames” sub-dataset by using FaceNet model pre-trained by CASIA-WebFace are greater than the original, ASVDW, ASVDF, and GRIR images by 0.6%, 0.46%, 0.53%, and 0.56% for the cosine distance metric, and by 0.51%, 0.35%, 0.33%, and 0.43% for the Euclidean distance metric. By using the FaceNet pre-trained by the VGGFace2, the recognition rates of the AMRF images in the “frames” sub-dataset are higher than the original, ASVDW, ASVDF, and GRIR images by 1.1%, 0.27%, 0.33%, and 0.45% for the cosine distance metric, and by 0.95%, 0.35%, 0.31%, and 0.5% for the Euclidean distance metric. The average recognition rates of the AMRF in “img” sub-dataset by using FaceNet model pre-trained by the CASIA-WebFace dataset are greater than the original, ASVDW, ASVDF, and GRIR images by 0.69%, 0.16%, 0.29%, and 0.36% for the cosine distance metric, and by 1.29%, 0.08%, 0.2%, and 0.21% for the Euclidean distance metric. Also, the recognition rate improvements compared to the original, ASVDW, ASVDF, and GRIR images when using the FaceNet model pre-trained by the VGGFace2 dataset to evaluate the testing images of the AMRF in the “img” sub-dataset are 0.52%, 0.02%, 0.04%, and 0.23% for the cosine distance metric, and it is 0.35%, 0.09%, 0.14%, and 0.22% for the Euclidean distance metric.

**TABLE VIII**

**AVERAGE RECOGNITION RATES OF THE ORIGINAL, ASVDW, ASVDF, GRIR, AND AMRF IMAGES BY USING THE FACENETS FOR THE IJB-C DATASET**

| Methods | Sub-dataset | Metrics | Original | FaceNet (ASVDW) | ASVDF | GRIR | AMRF |
|---------|-------------|---------|----------|-----------------|-------|------|------|
| FaceNet | frames      | Cosine  | 79.98 ± 0.023 | 80.31 ± 0.028 | 80.22 ± 0.187 | 80.11 ± 0.177 | 80.39 ± 0.104 |
|         | img         | Cosine  | 80.57 ± 0.052 | 81.32 ± 0.155 | 80.97 ± 0.148 | 80.72 ± 0.131 | 81.66 ± 0.111 |
| FaceNet | frames      | Euclidean| 75.61 ± 0.140 | 77.18 ± 0.201 | 77.01 ± 0.152 | 76.82 ± 0.154 | 77.22 ± 0.145 |
|         | img         | Euclidean| 75.70 ± 0.188 | 76.55 ± 0.178 | 76.32 ± 0.184 | 76.17 ± 0.176 | 76.74 ± 0.144 |

In the second scenario, we train the FaceNet model from scratch without pre-trained weights on the training sets of the “frames” and “img” sub-datasets and then evaluate the recognition rates using this trained FaceNet model testing sets. The “frames” sub-dataset is split into training, validating, and testing sets in the experiment. Each training and validating set contains 3,530 and 3,531 non- matching pairs of 160×160 pixel images corresponding to 3,530 subjects, and the testing set includes the remaining 111,870 images. With the same splitting protocol, each training and validate set of the “img” sub-dataset contains 3,531 face images corresponding to 3,531 subjects, while the validating set includes the remaining 15,954 face images. We use the same training configuration used for the LFW dataset for this experiment. However, for evaluating the testing sets of the “frames” and “img” sub-dataset, we create two different pair character lists to contain pairwise combination images of matching and non-matching people. As a result, we generate 3,530 matching pairs and 3,530 non-matching pairs corresponding to the number of subjects in the “frames” sub-dataset. Similarly, the pair character list of the “img” sub-dataset contains 3,531 matching pairs and 3,531 non-matching pairs. Although the IJB-C is more challenging than the LFW and other datasets for face recognition, the results are significantly improved in terms of average recognition rate by using our proposed method. As shown in Table VIII, the average recognition rates of the FaceNet model for the AMRF images in the “frames” sub-dataset are higher than original, ASVDW, ASVDF, and GRIR images by 0.41%, 0.08%, 0.17%, and 0.28% for the cosine distance metric, and by 1.09%, 0.34%, 0.69%, and 0.94% for the Euclidean distance metric. For the “img” sub-dataset, the average recognition rates for the AMRF images by the FaceNet model are also greater than the original, ASVDW, ASVDF, GRIR images by 0.71%, 0.04%, 0.21%, and 0.4% for the cosine distance metric, and by 1.04%, 0.19%, 0.42%, and 0.57% for the Euclidean distance metric.

The performance of our proposed method is further demonstrated by using the ArcFace model to evaluate the recognition accuracy on the IJB-C dataset through three scenarios as used for the LFW dataset. The configurations for all scenarios are set as same as used for the LFW dataset. Firstly, we split the “frames” and “img” into their training and validating sets, in which we use 75% images for training and the remaining 25% images for validation. The face images in
all sets are compiled to 112×112 pixels as the input image size requirement of the ArcFace model. Therefore, the training set of the “frames” sub-dataset contains 87,632 squared face images, while the validating set involves the remaining 31,021 face images. For the “img” sub-dataset, its training set contains 15,329 face images and the validating set contains 6,929 face images.

### TABLE IX

**AVERAGE RECOGNITION RATES OF THE ORIGINAL, ASVDW, ASVDF, GRIR, AND AMRF IMAGES BY USING THE ARCFACE MODELS FOR THE IJB-C DATASET**

| Methods      | Recognition rate (%) | Sub-dataset | Original | ASVDW | ASVDF | GRIR | AMRF |
|--------------|----------------------|-------------|----------|-------|-------|------|------|
| ArcFace_1    |                      | frames      | 79.51±0.042 80.35±0.043 80.24±0.058 | 80.15±0.044 | 80.88±0.055 |
|              |                      | img         | 80.12±0.059 81.23±0.047 81.11±0.052 | 81.02±0.035 | 81.51±0.058 |
| ArcFace_2    |                      | frames      | 71.10±0.063 72.40±0.018 72.13±0.052 | 71.94±0.039 | 72.81±0.031 |
|              |                      | img         | 72.31±0.052 73.71±0.028 73.51±0.042 | 73.34±0.042 | 73.93±0.052 |
| ArcFace_3    |                      | frames      | 89.54±0.055 90.98±0.052 90.87±0.046 | 90.52±0.027 | 91.08±0.033 |
|              |                      | img         | 90.54±0.057 92.11±0.051 92.18±0.035 | 92.03±0.037 | 92.30±0.048 |

In the first scenario using the ArcFace model, we use the ArcFace model pre-trained in 100,600 steps on the MS-Celeb-1M dataset to evaluate the recognition accuracy on the validating sets of both “frames” and “img” sub-datasets. The results of this evaluation scenario are shown in Table IX as “ArcFace_1”. The average recognition rate of the ArcFace model for the AMRF images in the “frames” sub-dataset is better than the original, ASVDW, ASVDF, and GRIR images by 1.37%, 0.53%, 0.64%, and 0.75%. For the “img” sub-dataset, the result for the AMRF images is higher by 1.39%, 0.28%, 0.4%, and 0.49% for the original, ASVDW, ASVDF, and GRIR images, respectively.

In the second and third evaluation scenarios on the IJB-C dataset by the ArcFace model by training the training sets of the “frames” and “img” sub-datasets without pre-trained weights and with pre-trained weights in 100,000 steps. The best trained weights of the model are used to evaluate the recognition rates on the validating sets. The results of these evaluation scenarios are shown in Table IX as “ArcFace_2” for the training ArcFace model without pre-trained weights and “ArcFace_3” for training ArcFace model with pre-trained weights. As seen, by using our proposed method, the recognition rates are better than using the original images and applying ASVDW, ASVDF, and GRIR methods for the “frames” sub-dataset by 1.71%, 0.41%, 0.68%, and 0.87%, respectively, for the trained ArcFace model without pre-trained weights. For the images in the “img” sub-dataset, the average recognition rate by using the AMRF method is higher than the original, ASVDW, ASVDF, and GRIR images by 1.62%, 0.22%, 0.38%, and 0.59%, respectively. In the third scenario, for the “frames” sub-dataset, the result of the AMRF images is better than the original, ASVDW, ASVDF, GRIR images by 1.54%, 0.1%, 0.21%, and 0.56%. For the “img” sub-dataset, this improvement is 1.76%, 0.19%, 0.12%, and 0.27%, respectively, compared to the original, ASVDW, ASVDF, and GRIR images. These results demonstrate that our proposed method helps improve significantly the performance of the human face recognition system, even outperforming the ASVDW, ASVDF, and GRIR methods.

### E. RESULTS OF THE CMU MULTI-PIE DATABASE

The CMU Multi-PIE [30] is a face image database established by researchers at Carnegie Mellon University to address shortcomings, such as the limited number of subjects, single recording session, and captured expression, from the previous famous CMU PIE database [31]. On the contrary to the LFW and IJB-C datasets, the face images in the CMU Multi-PIE database were taken under constrain lighting, expression, and pose conditions. This database contains 755,370 images captured from 337 subjects under 15 viewpoints and 19 illumination conditions in four recording sessions. With many challenges, the CMU Multi-PIE can be considered a standard benchmark face database in the face recognition domain. In this paper, we use session 4 of the CMU Multi-PIE due to its large number of subjects and range of variations in capturing conditions to demonstrate the performance of our proposed method in improving the recognition rate of human face recognition systems.

Session 4 of the CMU Multi-PIE database contains 215,100 photos taken under 19 illumination in 15 viewpoints with three face expressions of 239 subjects. In this study, we adopt the images taken by 13 cameras while discarding the photos taken by “08_1” and “19_1”, which are upside-down imaging cameras. Then, we extract the face images by using the MTCNN model and compile the extracted face images to 160×160 pixels. Finally, the CMU Multi-PIE used in our experiment includes 186,420 squared face images taken from 13 viewpoints from 239 people.

Fig. 10 displays eighteen color face images in the CMU Multi-PIE datasets. As seen, these face images are taken under different illumination conditions. Therefore, the illumination of the face images is hugely affected by the environmental lighting conditions. Using the ASVDW and ASVDF methods, the quality of the face images can be improved, as shown in Fig. 10(b) and (c). The GRIR method can partly enhance the illumination of the face images, but the naturalness is not recovered, as depicted in Fig. 10(d). Although the FIN-GAN color face images shown in Fig. 10(e) can resolve the challenge with the illumination variations, the facial skin color is not recovered. In addition, since the FIN-GAN method requires a target image as the reference face image during the training process, the result face images are not natural, smooth and the structure of the face image is also partly affected due to the inconsistent of the poses between the given face images and its reference face images. On the other hand, the face images enhanced by our proposed AMRF methods are even clearer, smoother than the result images enhanced by other methods, as depicted in Fig. 10(f). In addition, the skin color tone of the images is recovered effectively to be more natural on the AMRF face images rather than other enhancement methods.
The performance of the proposed method is demonstrated by using the FaceNet and ArcFace models with the same evaluation scenarios as used in sub-section IV.C and IV.D. The results of the first scenario by using the pre-trained FaceNet models to evaluate the recognition accuracy of the CMU Multi-PIE are shown in Table X. As seen, the average recognition rate of the pre-trained FaceNet model by using CASIA-WebFace for the AMRF images is higher than the original, ASVDW, ASVDF, GRIR, and FIN-GAN by 1.78%, 0.74%, 0.73%, 0.98%, and 0.24% by using the cosine distance metric and by 0.9%, 0.4%, 0.49%, 0.64%, and 0.16% by using the Euclidean distance metric. By using the FaceNet model pre-trained by VGGFace2, the result for the AMRF images is greater than the original, ASVDW, ASVDF, GRIR, and FIN-GAN images by 0.94%, 0.22%, 0.29%, 0.36%, and 0.13% by using the cosine distance metric and by 1.07%, 0.16%, 0.21%, 0.42%, and 0.25% with the Euclidean distance metric.

![Figure 10. Eighteen color face images under different illumination conditions of a subject in the CMU Multi-PIE dataset.](image)

In another scenario, the FaceNet model is trained by testing and validating sets for 100,000 steps with the same configuration used in sub-section IV.C. We prepare the training set by selecting one face image per the viewpoint of every expression type for each subject in the dataset. Therefore, there are 78 face images of each subject used for training. In total, the training set contains 9,321 images. We also take the same number of images as the training set for the validating set. Then, the testing set contains the remaining 167,778 face images. The recognition accuracy is evaluated on the testing set of the CMU Multi-PIE dataset and shown in Table XI. As seen, the average recognition rate of the FaceNet model for the AMRF images is higher than the original, ASVDW, ASVDF, GRIR, and FIN-GAN images by 1.31%, 0.64%, 0.81%, 0.93%, and 0.34% by using the cosine distance metric and by 0.99%, 0.39%, 0.62%, 0.69%, and 0.3% by using the Euclidean distance metric.

**TABLE X**

| Methods          | Metrics | FaceNet (CASIA-WebFace) | FaceNet (VGGFace2) |
|------------------|---------|-------------------------|--------------------|
|                  | Cosine  | Euclidean               | Cosine             |
| FaceNet          | 90.65   | 90.77                   | 90.17              |
| (CASIA-WebFace)  | ±0.017  | ±0.016                  | ±0.014             |
| ASVDW            | 91.69   | 91.27                   | 91.13              |
|                  | ±0.013  | ±0.026                  | ±0.012             |
| ASVDF            | 91.70   | 91.18                   | 91.01              |
|                  | ±0.005  | ±0.012                  | ±0.011             |
| GRIR             | 91.45   | 91.03                   | 91.36              |
|                  | ±0.018  | ±0.012                  | ±0.011             |
| FIN-GAN          | 92.19   | 91.51                   | 91.67              |
|                  | ±0.014  | ±0.022                  | ±0.013             |
| AMRF             | 92.43   | 91.67                   | 91.71              |
|                  | ±0.011  | ±0.013                  | ±0.011             |

**TABLE XI**

| Methods             | Metrics | Recognition rate (%) |
|---------------------|---------|----------------------|
| FaceNet             | Cosine  | Original ASVDW ASVDF ASVDF GRIR FIN-GAN AMRF |
| (CMU Multi-PIE)     | Euclidean | 90.53 91.20 91.03 90.91 91.50 91.84 |
| AMRF                | ±0.008  | ±0.009 ±0.007 ±0.004 ±0.009 ±0.005 |
| GRIR                | ±0.006  | ±0.004 ±0.005 ±0.011 ±0.006 |
| FIN-GAN             | 91.02   | 91.62 91.39 91.32 91.71 92.07 |
| AMRF                | ±0.005  | ±0.004 ±0.005 ±0.011 ±0.006 |

For the experiments with the ArcFace model, we establish the training set by selecting 75% images of the images taken by a camera for the expression type of each subject. Therefore, the training set of the CMU Multi-PIE dataset in this experiment contains 139,815 face images. Then, the validating set is built to contain the remaining 46,605 face images. The first scenario shows the results in Table XII as “ArcFace_1”. As seen, the average recognition rate of the pre-trained
ArcFace model for the AMRF images is higher than the original, ASVDW, ASVDF, GRIR, and FIN-GAN images by 1.22%, 0.72%, 0.84%, 0.99%, and 0.12%, respectively.

In the other scenarios, the ArcFace models are trained using the training set from scratch (without pre-trained weights) and the pre-trained weights (on the MS-Celeb-1M dataset). The results of these two evaluation scenarios are alternately shown in Table XII as “ArcFace_2” and “ArcFace_3”. In particular, the average recognition rate of the ArcFace model trained without pre-trained weights for the AMRF images is higher than the original, ASVDW, ASVDF, GRIR, and FIN-GAN images by 0.64%, 0.24%, 0.31%, 0.46%, and 0.23%, respectively. In the third scenario, this improvement achieves 0.94%, 0.47%, 0.4%, 0.67%, and 0.29%, which correspond to the greater average recognition rate of the AMRF images compared to the original, ASVDW, ASVDF, GRIR, and FIN-GAN images.

![Figures 11(a)-(f)](image)

**FIGURE 11.** Ten color face images under different illumination conditions of a subject in the CMU-PIE dataset. (a) Original images. (b) ASVDW images. (c) ASVDF images. (d) GRIR images. (e) FIN-GAN. (f) AMRF images

### F. RESULTS OF THE CMU-PIE DATABASE

The CMU-PIE human face database [31] comprises 41,368 facial images of 68 people at the CMU-3D studio with three types of variations: pose, illumination, and expression. In this database, each person is imaged across 13 different poses, under 43 different illumination conditions, and with 4 different facial expressions. From the CMU-PIE database, we collected 45 images with different illumination variations for each of 68 people for experiments. These facial images are captured by the central camera (c27) and include two lighting conditions: “With Room Lights” and “Without Room Lights”. For the “With Room Lights” lighting condition, there are 24 types of illumination variations (from f00 to f23), whereas the “Without Room Lights” lighting condition contains facial images with 21 types of illumination variations (from f00 to f21). Finally, 3,060 cropped 128×128-pixel facial images with different lighting conditions in the CMU-PIE database are chosen for our experiments in this study. In this experiment, we conducted experiments to compare AMRF images to the original, ASVDW, ASVDF, GRIR, and FIN-GAN images with the evaluation scenarios as same as experiments for the CMU Multi-PIE in sub-section VI.E.

Fig. 11 displays ten color face images of a subject under various illumination conditions in the CMU-PIE dataset. The environmental lighting conditions affect the face image quality, as shown in Fig. 11(a). The face image quality can be enhanced by using the enhancement methods, as depicted in Figs. 11(b)-(f). As it can be seen, the proposed AMRF method outperforms other methods in terms of illumination enhancement and facial skin color recovery.

| Methods | Metrics | Recognition rate (%) |
|---------|---------|----------------------|
|         | Original | ASVDW | ASVDF | GRIR | FIN-GAN | AMRF |
| FaceNet (CASIA-WebFace) | Cosine | 98.55 ± 0.02 | 99.23 ± 0.00 | 99.21 ± 0.00 | 98.82 ± 0.00 | 99.30 ± 0.00 | 99.41 ± 0.00 |
| FaceNet (VGGFace2) | Cosine | 98.64 ± 0.00 | 99.12 ± 0.00 | 99.01 ± 0.00 | 98.91 ± 0.00 | 99.28 ± 0.00 | 99.37 ± 0.00 |
| Euclidean | ±0.005 | ±0.005 | ±0.005 | ±0.004 | ±0.004 | ±0.004 | ±0.004 |
| Euclidean | ±0.005 | ±0.005 | ±0.005 | ±0.004 | ±0.004 | ±0.004 | ±0.004 |

The experiments to demonstrate the performance of our proposed method in face recognition accuracy improvement are conducted by using the FaceNet and ArcFace models with the same scenarios and configurations as used in subsection IV. E. The results of the first scenario using the pre-trained FaceNet models to evaluate the recognition accuracy of the CMU-PIE are shown in Table XIII. As seen, the average recognition rate of the pre-trained FaceNet model using CASIA-WebFace dataset for the AMRF images is greater than the original, ASVDW, ASVDF, GRIR, and FIN-GAN images by 0.86%, 0.18%, 0.2%, 0.59%, and 0.11% by using the cosine distance metric and by 1.6%, 0.21%, 0.3%, 1.19%, and 0.15% by using the Euclidean distance metric, respectively. When adopting the FaceNet model pre-trained by using VGGFace2 dataset, the recognition rate for the AMRF images is higher than the original, ASVDW,
ASVDF, GRIR, and FIN-GAN images by 0.73%, 0.25%, 0.36%, 0.46%, and 0.09% for the cosine distance metric and by 1.02%, 0.14%, 0.27%, 0.46%, and 0.1% for the Euclidean distance metric.

In the second scenario, we train the FaceNet model without pre-trained weights on the CMU-PIE training set and then evaluate the recognition accuracy on the testing set by using the trained FaceNet model. In this experiment, for each subject in the CMU-PIE dataset, we use one image for training, one image for validating, and the remains for testing. Therefore, the training and validating sets contain 68 face images corresponding to 68 subjects, and the testing set involves 2,924 face images. The average recognition rates of the FaceNet model trained by the CMU-PIE training set are shown in Table XIV. As noticed, the result of the proposed AMRF method is higher than the original images and the ASVDW, ASVDF, GRIR, and FIN-GAN methods by 2.98%, 0.16%, 0.56%, 1.72%, and 0.09% by using cosine distance metric and by 1.45%, 0.28%, 0.5%, 0.69%, and 0.19% for using Euclidean distance metric.

| Methods            | Metrics | Recognition rate (%) |
|--------------------|---------|----------------------|
| FaceNet (CMU Multi- PIE) | Cosine | Original ASVDF ASVDF GRIR FIN-GAN AMRF |
|                    | Euclidean | 93.12 ±0.03 | 95.94 ±0.04 | 95.54 ±0.05 | 94.38 ±0.02 | 96.01 ±0.05 | 96.10 ±0.04 |
|                    | GRIR | | | | | |
|                    | FIN-GAN | 92.85 ±0.04 | 94.02 ±0.03 | 93.80 ±0.04 | 93.61 ±0.05 | 94.11 ±0.05 | 94.30 ±0.04 |

In the third scenario, we train the ArcFace model using the pre-trained weights from the MS-Celeb-1M dataset. The result of the recognition rate by using the ArcFace model pre-trained weights from the MS-Celeb-1M dataset is shown in Table XV as “ArcFace_2”. As seen, the result for the AMRF images is higher than the original, ASVDW, ASVDF, GRIR, and FIN-GAN images by 1.01%, 0.22%, 0.31%, 0.56%, and 0.17%, respectively. In the last evaluation, we train the ArcFace model using the pre-trained weights from the MS-Celeb-1M dataset. The result of the AMRF images is also better than the original, ASVDW, ASVDF, GRIR, and FIN-GAN images by 1.43%, 0.46%, 0.72%, 1.12%, and 0.15%, respectively, as shown in Table XV as “ArcFace_3”.

**G. COMPUTATIONAL COMPLEXITY**

In this sub-section, the efficiency of the proposed method is presented by measuring the CPU running time on different images sizes. The computational complexity of the AMRF method is O(n²). A longer computation time is required for a larger image. Due to the small size of the face images usually used in face recognition, the proposed method can be used in a human face recognition system to improve performance. Compared to previously proposed methods, our proposed method costs less computational complexity and time consumption for face image enhancement. Table XVI summarizes the computation time results of the ASVDW, ASVDF, GRIR, and our proposed methods by using different face image sizes. The evaluations were conducted in Microsoft Visual C++ 2010 on a workstation with an Intel Core i7-9700, RAM 32GB on Windows 10.

**TABLE XV**

| Methods            | Recognition rate (%) |
|--------------------|----------------------|
| ArcFace_1          | 80.42 ±0.33 | 81.77 ±0.30 | 81.58 ±0.34 | 81.34 ±0.40 | 82.22 ±0.38 | 82.27 ±0.30 |
| ArcFace_2          | 82.68 ±0.42 | 83.47 ±0.31 | 83.38 ±0.48 | 83.13 ±0.50 | 83.52 ±0.48 | 83.69 ±0.47 |
| ArcFace_3          | 83.92 ±0.23 | 84.89 ±0.22 | 84.63 ±0.20 | 84.23 ±0.31 | 85.20 ±0.20 | 85.35 ±0.23 |

| Methods            | CPU time (Second/image) |
|--------------------|-------------------------|
| ASVDW              | 0.28 ±0.05 | 1.13 ±0.07 | 2.52 ±0.11 |
| ASVDF              | 0.16 ±0.03 | 0.58 ±0.04 | 6.23 ±0.08 |
| GRIR               | 0.30 ±0.05 | 1.21 ±0.06 | 3.08 ±0.08 |
| AMRF               | 0.05 ±0.03 | 0.11 ±0.03 | 0.41 ±0.04 |

**V. CONCLUSION**

This paper presents a novel image enhancement method called AMRF for color face recognition under varying lighting conditions. The proposed method used the associative filter-based retinex model to represent effectively color face images into multiple layers of illumination and reflectance. Then, the face image can be enhanced by compensating the multiple layers of illumination by adaptive compensation coefficients that are first determined by using a corresponding Gaussian template as the reference. The enhanced face image is obtained by composing the enhanced illumination components with the integrated reflectance component. The experiments were performed on four of the most famous color face datasets named LFW, IJB-C, CMU Multi-PIE, and CMU-PIE to demonstrate the performance of our proposed method in improving the recognition accuracy of deep learning-based face recognition models, namely, FaceNet, and ArcFace. The results reveal that the AMRF method does not only effectively enhances the quality of face images to become more precise, more natural, and smoother but can also notably improve the recognition rate, especially...
for the face dataset with significant illumination variations and computational time for human face recognition systems.

ACKNOWLEDGMENT

This work was supported in part by the Ministry of Science and Technology, Taiwan, under Grant MOST 110-2221-E-992-076.

REFERENCES

[1] R. Chellappa, C. L. Wilson, and S. Sirohey, “Human and machine recognition of faces: a survey,” in Proc. the IEEE, 1995, pp. 705-741.
[2] Y. Adini, Y. Moses, and S. Ullman, “Face recognition: The problem of compensating for changes in illumination direction,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 7, pp. 721-732, Jul. 1997.
[3] H. Han, S. Shan, X. Chen, and W. Gao, “A comparative study on illumination preprocessing in face recognition,” Pattern Recognit., vol. 46, pp. 1691-1699, 2013.
[4] H. Liu, M. Yang, Y. Gao, and C. Cui, “Local histogram specification for face recognition under varying lighting conditions,” Image Vis. Comput., vol. 32, pp. 335-347, Mar. 2014.
[5] S. Shan, W. Gao, B. Cao and D. Zhao, “Illumination normalization for robust face recognition against varying lighting conditions,” in Proc. IEEE International SIS Conference, 2003, pp. 157-164.
[6] H. Shim, J. Luo, and T. Chen, “A subspace model-based approach to face relighting under unknown lighting and poses,” IEEE Trans. Image Process., vol. 17, no. 8, pp. 1331-1341, Aug. 2008.
[7] H. Dai and X. Gu, “A new method based on bilateral-filtering for illumination invariant face recognition,” in 3rd International Conference on Information and Computing, 2010, pp. 242-245.
[8] L. Torres, J.Y. Reutter, and L. Lorente, “The importance of the color information in face recognition,” in Int. Conf. ICIP, vol. 3, pp. 627-631, 1999.
[9] M. Villegas, R. Paredes, A. Juan, and E. Vidal, “Face verification on color images using local features,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops, 2008, pp. 1-6.
[10] J. Choi, Y. Ro, and K. Platanitis, “Color face recognition for degraded face images,” IEEE Trans. Syst., Man, Cybern. B: Cybern., vol. 39, no.5, pp. 1217-1230, Oct. 2009.
[11] C. Wang, B. Yin, X. Bai, and Y. Sun, “Color face recognition based on 2DPCA,” in 19th Int. Conf. Pattern Recognit., 2008, pp. 1-4.
[12] Y. Z. Goh, A. B. J. Teoh, and M. K. O. Goh, “Wavelet local binary patterns fusion as illuminated facial image preprocessing for face verification,” Expert Syst. Appl., vol. 38, pp. 3959-3972, 2011.
[13] W. Chen, M. J. Er, and S. Wu, “Illusion compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain,” IEEE Trans. Syst., Man, Cybern., Part B (Cybernetics), vol. 36, no. 2, pp. 458-466, Apr. 2006.
[14] J.-W. Wang, N. T. Le, J.-S. Lee, and C.-C. Wang, “Color face image enhancement using adaptive singular value decomposition in Fourier domain for face recognition,” Pattern Recognit. vol. 57, pp. 31-49, Sept. 2016.
[15] J.-W. Wang, N. T. Le, J.-S. Lee, and C.-C. Wang, “Illusion compensation for face recognition using adaptive singular value decomposition in the wavelet domain,” Inf. Sci., vol. 435, pp. 69-93, Apr. 2018.
[16] I. J. Goodfellow et al., “Generative adversarial nets,” in Proc. 27th International Conference on Neural Information Processing Systems, pp. 2672-2680, Dec. 2014.
[17] X. Han, H. Yang, G. Xing, and Y. Liu, “Asymmetric joint GANs for normalizing face illumination from a single image,” IEEE Trans. Multimedia, vol. 22, no. 6, pp. 1619-1633, Jun. 2020.
[18] Y. Zhang, H. Chang-Hui, and X. Lu, “IL-GAN: Illumination-invariant representation learning for single sample face recognition,” J. Vis. Commun. vol. 59, pp. 501-513, Feb. 2019.
[41] D. Yi, Z. Lei, S. Liao, and S. Z. Li, “Learning face representation from scratch,” 2014, arXiv:1801.07698v3. [Online] Available: https://arxiv.org/abs/1411.7923.

[42] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, “VGGFace2: A dataset for recognising faces across pose and age,” 2017, arXiv:1801.07698v3. [Online] Available: https://arxiv.org/abs/1710.08092.

[43] J. Guo and J. Deng, “InsightFace: 2D and 3D face analysis project,” [Online] Available: https://github.com/deepinsight/insightface.

[44] Y. Gou, Lei. Zhang, Y. Hu, X. He, and J. Gao, “MS-Celeb-1M: A dataset and benchmark for large-scale face recognition,” in Computer Vision – ECCV, 2016, pp. 87-102.

[45] K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks,” in Computer Vision – ECCV, 2016, pp. 630-645.

Duong Binh Giap received his B.S. degree in Electronics and Communications Engineering from Posts and Telecommunications Institute of Technology, Hanoi, Vietnam, in 2016. He received his M.S. degree in Electronics Engineering from Le Quy Don Technical University, Hanoi, Vietnam, in 2018. He is currently pursuing a Ph.D. degree from the College of Electrical Engineering and Computer Science, National Kaohsiung University of Science and Technology, Taiwan. His current research interests include digital image processing, pattern recognition, and face recognition.

Tuyen Ngoc Le received his B.S. degree in mathematics and information from Hanoi National University of Education, Hanoi, Vietnam, in 1996. He received his M.S. degree in Computer Science from Le Quy Don Technical University, Hanoi, Vietnam, in 2006, and the Ph.D. degree in Electronic Engineering from National Kaohsiung University of Science and Technology, Taiwan, in 2015. He is currently an Assistant Professor with the Department of Electronic Engineering, Ming-Chi University of Technology, New Taipei City, Taiwan. His research interests include digital image processing; Biometric technologies (face recognition, fingerprint classification, hand gesture), and artificial intelligence-enabled automated optical inspection systems.

Jing-Wein Wang received B.S. and M.S. degrees in Electrical Engineering from the National Taiwan University of Science and Technology, in 1986 and 1988, respectively. He received a Ph.D. degree in Electrical Engineering from National Cheng Kung University, Taiwan, in 1998. He was a chief project leader at the Equipment Design Center of PHILIPS, Taiwan, from 1992 to 2000. In 2000, he joined the faculty of the National Kaohsiung University of Science and Technology and served as the Dean of the College of Electrical Engineering and Information Science from Aug. 2017 to Jul. 2019. Currently, he is a distinguished professor at the Institute of Photonics Engineering. His current research interests include combinatorial optimization, pattern recognition, and wavelets and their applications.

Chia-Nan Wang received the Ph.D. degree from the Department of Industrial Engineering and Management, National Chiao Tung University, Taiwan, in 2004. He is currently a Professor with the Department of Industrial Engineering and Management, National Kaohsiung University of Science and Technology, Taiwan. He has published 80 journal articles and 40 patents. His research interests include manufacturing automation and industrial applications. He achieves around 25 projects of research and academy and has some successful cases of technology transfer. He also achieves awards from many international invention competitions (two Gold Medal Awards, five Silver Medal Awards, and several Bronze Medal Awards). In addition, he has plenty of practical experience in many industries and provides consultant services for several stock companies in Taiwan.