Illumination Invariant Face Recognition:
A Survey of Passive Methods

Ramji M. Makwana*
Department of Computer Engineering, A. D. Patel Institute of Technology, Anand, 388121, India

Abstract

Face recognition under varying illumination is one of the challenging problems in real-time applications. Numerous methods have been developed by the research community to handle the problem. Existing surveys of methods are either too old or do not cover performance analysis of illumination invariant methods. This paper is more extensive than previous surveys and covers recently developed methods. The paper focuses on passive methods which solve the illumination problem by investigating the visible light images in which the face appearance has been altered by varying illumination. The methods are classified into four broad categories, namely (1) subspace-based statistical methods (2) illumination invariant representation methods, (3) model based methods, (4) other illumination handling methods. The other illumination handling category includes the methods which do not fall under first three categories. Performance analysis and discussion of methods and an evaluation of results is presented to determine the suitability and applicability of the method(s) for specific applications. It is observed from the survey of methods that illumination invariant representation based methods are better in terms of the number of training images required, the simplicity, computational complexity and robustness.

Keywords—Face Recognition, Illumination Invariance, Illumination Model, Retinex Theory, Self Quotient Image.

1. Introduction

Face Recognition is one of the important biometric research problems that use automated methods to recognize the identity of a person based on one's facial characteristics. Face Recognition became more important and competent with other biometric applications due to rapid advances in technologies such as digital cameras, portable digital computing devices, internet and wireless communication. The performance of any Face Recognition system is adversely affected by changes in the facial appearance caused by variation in lighting and pose [2][20][35][40]. Existing survey of methods for illumination invariant recognition can be found in [11][39][40]. However these surveys are either too older or do not cover fundamental theory insight and analysis of illumination invariant methods. This paper is extensive than previous surveys and covers recently developed methods. Existing methods addressing the illumination variation problem, falls into two main classes: (1) Active methods and (2) Passive methods. Passive methods attempt to overcome this problem by studying the visible spectrum images in which face appearance has been changed by illumination variations. The paper focuses only on passive methods.

The paper is organized as follows. Section 2 presents subspace based statistical methods. Section 3 provides details of illumination invariant representation methods. Section 4 is regarding model based methods. Section 5 covers the fourth category of methods. Section 6 presents a discussion on performance of methods.
2. Subspace Based Statistical Method

Numerous statistical methods [4][9][13][14][18] have been used by researchers for Face Recognition. Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Linear Discriminant Analysis (LDA) are most widely used approaches. Independent component analysis (ICA) is a computational method for separating a multivariate signal into additive subcomponents assuming the mutual statistical independence of the non-Gaussian source signals. ICA approach used in [18] to extract global features seems to be an adequate method due to its simplicity and speed. The independent components obtained by ICA algorithm are used as feature vectors for classification and these features are invariant to illumination.

Existing methods dealing with just one of these variations are often unable to cope with the other variations. The problem is even more difficult in applications where only one gallery image per person is available. Chen, Lovell and Shan [8] describe a recognition method, Adapted Principal Component Analysis (APCA), which can simultaneously deal with large variations in both illumination and facial expressions using only a single gallery image per person. Experimental results show that APCA performs much better than other recognition methods including PCA and LDA. Gudur and Asari [13] presented Gabor Wavelet based Modular PCA (GW-MPCA) to deal with illumination and facial expression. The method divides face image into sub-images and then applies Gabor wavelet on each sub-image with different scale and orientation. Every image in the database is divided into \( N \) smaller sub-images. The size of each sub-image is \( L^2/N \) (size of each image being \( L \times L \)). The sub-images are represented as:

\[
\Gamma_{ij}(m,n) = \Gamma_{ij}(\frac{L}{\sqrt{N}}(j-1)+m, \frac{L}{\sqrt{N}}(i-1)+n)\forall i, j
\]

where \( i \) vary from 1 to \( Z \), \( Z \) is the number of images in the database, \( j \) varies from 1 to \( N \), where \( N \) is the number of sub-images and \( m \) and \( n \) vary from 1 to \( L/\sqrt{N} \). They found that performance of GW-MPCA is better than PCA and Modular PCA. Chen et. al. [9] addresses nonlinear feature extraction and Small Sample Size (S3) problems in face recognition. In sample feature space, the distribution of face images is nonlinear because of complex variations in illumination. The performance of classical linear method, such as Fisher Discriminant Analysis (FDA), will degrade. To overcome pose and illumination problems, Shannon wavelet kernel is constructed and utilized for nonlinear feature extraction. Based on a modified Fisher criterion, simultaneous diagonalization method is exploited to deal with S3 problem, which often occurs in FDA based methods. Shannon Wavelet Kernel based Subspace Fisher Discriminant (SWK-SFD) method has been developed. The proposed method not only overcomes some drawbacks of existing FDA based algorithms, but is also computationally efficient. The proposed method gives superior results compared to existing FDA based methods.

3. Illumination Invariant Representation

This section presents methods for finding illumination invariant representation for illumination invariant face recognition. These methods are divided in main three sub categories named Retinex theory based methods, Gradient based method, and Local binary pattern based method.

3.1. Retinex Theory Based Methods

Majority of existing illumination invariant representation methods are related to the Retinex theory developed and presented by Land and McCann in [19][31]. The theory tries to explain the basic principles governing the process of image formation and/or scene perception and states that an image \( I(x, y) \) can be modeled as the product of the reflectance \( R(x, y) \) and luminance \( L(x, y) \) functions as follow:

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

Here, the reflectance \( R(x, y) \) relates to the characteristics of the objects comprising the scene of an image and is dependant on the reflectivity (or albedo) of the scene’s surfaces, while the luminance \( L(x, y) \) is determined by the illumination source and relates to the amount of illumination falling on the observed scene. Since the reflectance \( R(x, y) \) relates solely to the objects in an image, it is obvious that (when successfully estimated) it acts as an illumination invariant representation of the input image. To determine the reflectance of an image, and thus, to obtain an illumination invariant image representation, the luminance \( L(x, y) \) of an image is commonly estimated first. This estimate \( L(x, y) \) is then exploited to compute the reflectance via the manipulation of the image model given by the expression (2) as follow:

\[
\ln R(x, y) = \ln I(x, y) - \ln L(x, y)
\]

\[
R(x, y) = I(x, y)/L(x, y)
\]

where, the Eq. (4) denotes an element-wise division of the input image \( I(x, y) \) with the estimated luminance \( L(x, y) \). The reflectance computed with (3) is referred to as the logarithmic reflectance and the reflectance computed with (4) is referred to as the quotient reflectance. As already emphasized, the luminance is considered to vary slowly with the spatial position [28] and can, therefore, be estimated as a smoothen version of the original image \( I(x, y) \). Various smoothing filters and smoothing methods have been proposed in the literature resulting in different illumination normalization procedures that were successfully applied to the problem of face recognition under severe illumination changes.

The single scale retinex algorithm [16], for example, computes the estimate of the luminance \( L(x, y) \) by simply smoothing the
input image $I(x, y)$ with a Gaussian smoothing filter. The illumination invariant image representation is then computed using the expression for the logarithmic reflectance. While such an approach generally produces good results with a properly selected Gaussian, its broader use in robust face recognition systems is still limited by an important weakness: at large illumination discontinuities caused by strong shadows are cast over the face and halo effects are often visible in the computed reflectance [28].

To avoid this problem the authors of the algorithm extended their normalization method to a multi scale form, i.e. Multi-Scale Retinex (MSR) method [17], where Gaussians with different widths are used and basically outputs of different implementations of the single scale retinex algorithm are combined to compute the final illumination invariant face representation.

Shashua, et al. [29] uses a new approach called Quotient Image (QI) with class based re-rendering (synthesis) to handle varying lighting conditions. Advantage of the approach is that it requires very small set of example faces. But it requires to register (align) test image with images in database, which adds to computational complexity. They achieve two major goals: First, they do not make a reconstruction assumption and thereby tolerate small databases without pixel-wise alignment. Second, they solve problem for a system of $N = 3$ parameters (instead of 3N). As a byproduct of the method of optimization, an intermediate image is obtained, an illumination invariant signature image which can also be used for purposes of visual recognition.

Wang et al. [33][35] improved the QI method by introducing Self Quotient Image (SQI) based method, which requires only one face image and no image registration (alignment) required. They solve the problem of luminance estimation by introducing an anisotropic smoothing filter. Once the anisotropic smoothing operation produces an estimate of the luminance $L(x, y)$, the quotient reflectance $R(x, y)$ is computed in accordance with (4). However, due to the anisotropic nature of the employed smoothing filter flat zones in the images are not smoothed properly. It also has additional advantage of handling images with shadow.

Srisuk and Petpon [30] extended the Self Quotient Image (SQI) to Gabor based SQI where the 2D Gabor filter is applied instead of weighted Gaussian filter in order to increase more efficiency of the face recognition. Experimental result on Yale B face database has shown that the method reached a very high recognition rate even in the case of extremely varying illumination.

Classified Appearance-based Quotient Image (CAQI) [26] is proposed by Nishiyama, Kozakaya and Yamaguchi. They classified facial appearances caused by illumination into four main components: diffuse reflection, specular reflection, attached shadow and cast shadow. Diffuse reflection occurs when the incident light is scattered by the object. The pixel value of diffuse reflection is determined by albedo that is invariance. Specular reflection occurs when the incident light is cleanly reflected by the object. Attached shadow occurs when the object itself blocks the incident light. Cast shadow occurs when a different object blocks the incident light.

The non-local means (NLM) algorithm [7] is a recently proposed image de-noising method, which, unlike existing de-noising methods, considers pixel values from the entire image for the task of noise reduction. The algorithm is based on the fact that, for every small window of the image several similar windows can be found in the image as well, and above all windows can be exploited to de-noise the image. Let us denote an image contaminated with noise as $I(x) \in R^{a\times b}$, where $a$ and $b$ are image dimensions in pixels, and let $z$ stand for an arbitrary pixel location $x = (x, y)$ within the noisy image. The NL means algorithm constructs the de-noised image $I_{d}(x)$ by computing each pixel value of $I_{d}(x)$ as a weighted average of pixels comprising $I_{d}(x)$:

$$I_{d}(x) = \sum_{z \in \Omega(z)} w(z, x) I_{d}(x),$$

where, $w(z, x)$ represents the weighting function that measures the similarity between local neighborhoods of the pixel at the spatial locations $z$ and $x$. Here, the weighting function is defined as follows:

$$w(z, x) = \frac{1}{Z(z)} e^{-\frac{\sum_{\Omega(z)} \sum_{\Omega(z)} |I_{d}(x) - I_{d}(z)|^{2}}{\sigma^{2}}},$$

and $Z(z) = \sum_{z \in \Omega(z)} e^{-\frac{\sum_{\Omega(z)} \sum_{\Omega(z)} |I_{d}(x) - I_{d}(z)|^{2}}{\sigma^{2}}}$

In eq. (6) $G_{\sigma}$ denotes a Gaussian Kernel with the standard deviation $\sigma$, $\Omega(z)$ and $\Omega(z)$ denote the local neighborhoods of the pixels at the locations $x$ and $z$, respectively, $h$ stands for the parameter that controls the decay of the exponential function, and $Z(z)$ represents a normalizing factor [7]. From the presented equations it is clear that if the local neighborhoods of a given pair of pixel locations $z$ and $x$ display a high degree of similarity, the pixels at $z$ and $x$ will be assigned relatively large weights when computing their de-noised estimates.

Rather than using the original NL means algorithm for estimation of the luminance of an image, an improved method [31] is propose to exploit an adaptive version of the algorithm, where the decay parameter $h$ is a function of local contrast and not a fixed and pre-selected value. At regions of low contrast, which represent homogeneous areas, more smoothing should be done while in regions of high contrast the less smoothing should be done.

3.2. Gradient Based Method

The gradient domain is very important to image processing. The pixel points are not completely independent of each other,
there is some relationship between neighboring pixel points. The gradient domain explicitly considers such relationships between neighboring pixel points such that it is able to reveal underlying inherent structure of image data.

Zhang et al. [35] proposed a novel method to extract illumination insensitive features for face recognition under varying lighting called the Gradientfaces. Theoretical analysis shows Gradientface is an illumination insensitive measure, and robust to different illumination, including uncontrolled, natural lighting. In addition, Gradientface is derived from the image gradient domain such that it can discover underlying inherent structure of face images since the gradient domain explicitly considers the relationships between neighboring pixel points. Therefore, Gradientface has more discriminating power than the illumination insensitive measure extracted from the pixel domain. The author has evaluated the performance of the method using recognition rate performance parameter. Evaluation of the method on CMU PIE and Yale B database shows that Gradientface method is an effective method for face recognition under varying illumination. The proposed method is compared with other three methods named multi-scale retinex method, self quotient image method and local total variance method. The proposed method performs better than other methods. Though performance of method is evaluated on different dataset, evaluation of methods using other parameters like false acceptance rate, false rejection rate, equal error rate etc. is lacking.

3.3. Local Binary Pattern Based Method

Multi-orientation information in face images can be captured by calculating derivatives in different directions. Each face image can be described by a subset of band filtered images containing Steerable Pyramid (SP) coefficients which characterize the face textures, followed by the local binary pattern (LBP) operator. The combination of SP and LBP enhances the representation power of the spatial histogram greatly. Arousii et al. [3] introduced an efficient local appearance feature extraction method based on steerable pyramid for face recognition. Local information is extracted from SP sub-bands using LBP. The underlying statistics allows reducing the required amount of data to be stored. SP decomposition is used to split the features in a face image into different sub-bands at different levels, with approximations and details. Based on the theorem of steerable filter, the derivatives of an image in any direction can be interpolated by several basis derivative functions. The SP representation of a face image is derived by convolving the face image with the SP filters. Let \( f(x, y) \), be the face image and its convolution with a SP filter \( \psi_{\mu,\nu}(z) \), is defined as follows:

\[
G_{\mu,\nu}(x, y, \mu, \nu) = f(x, y) * \psi_{\mu,\nu}(z)
\]  

where '*' denotes the convolution operator, \( \nu \) and \( \mu \) is the number of scales and orientations respectively. Convolving the image with each of the \( \nu \) times \( \mu \) SP filters can then generate the SP features. In this approach, a face image is modeled as a histogram sequence by the following procedure: (1) An input face image is normalized and transformed to obtain multiple SP sub-bands by applying multi-scale and multi-orientation SP filters; (2) Each sub-band is converted to Local SP Binary Pattern (SPBP) map; (3) Each SPBP map is further divided into non-overlapping rectangle regions with specific size, and histogram is computed for each region; (4) The SPBP histograms of all the SPBP maps are concatenated to form the final histogram sequence as the model of the face SPBPS (Steerable Pyramid Binary Pattern Sequence).

4. Model Based Methods

4.1. Lambertian Reflectance Model

The Lambertian model is based on Lambert’s law [14]. According to Lambert’s law, if a light ray of intensity \( I \) coming from the direction \( u_i \) reaches a surface point with albedo \( \rho \) and normal direction \( v_r \), the intensity \( i \) reflected by the point due to this light is given by

\[
i = I(u_i) \rho \max(v_r \cdot u_i, 0)
\]  

A 2-D face image based approach, which combines Eigen light field and Lambertian reflectance model, is presented by Zhou and Chellappa [14]. The generalized photometric stereo algorithm presented, combines the identity subspace with the illumination model and provides an illumination-invariant description. But assumes a fixed pose and cannot easily handle pose variations, i.e., its illumination-invariant identity description is not invariant to variations in pose. An important advantage of the approach is that it can generalize to novel illuminations. To make the recognition more robust under any situation, they assumed that the lighting conditions for the training, gallery and probe sets are completely unknown when recovering the identity signatures.
Basri and Jacobs [5] has proved that the set of all Lambertian reflectance functions obtained with arbitrary distant light sources lies close to a 9D linear subspace. In general, the set of images of a convex Lambertian object obtained under a wide variety of lighting conditions can be approximated accurately by a low-dimensional linear subspace. They also provide a simple analytic characterization of this linear space. They obtain these results by representing lighting using spherical harmonics and describing the effects of Lambertian materials as the analog of a convolution. Both the lighting function, \( l \), and the Lambertian Kernel, \( k \), can be written as sum of spherical harmonics. The harmonic expansion of \( l \) is defined by

\[
l = \sum_{n=1}^{\infty} \sum_{|m| \leq n} l_{nm} Y_{nm}(\alpha, \beta)
\]

where \( l_{nm} \) is amplitude of light at order \( n \) and \( Y_{nm} \) is an \( n \)th order harmonic. Kernel \( k \) is defined by

\[
k(u) = \sum_{n=0}^{\infty} \sum_{|m| \leq n} k_{nm} Y_{nm}(\alpha, \beta)
\]

An image of an object under certain illumination conditions can be constructed from the respective reflectance function in a simple way: each point of the object inherits its intensity from the point on the sphere whose normal is the same. This intensity is further scaled by its albedo. One can write this explicitly, as follows: Let \( p_i \) denote the \( i \)th object point. The \( n \) denotes the surface normal at \( p_i \), and \( \rho \) denote the albedo of \( p_i \). Function \( r(n) \) denotes reflectance function. The illumination can be expanded with the coefficients \( l_{nm} \) (9). The image, \( I_i \) of \( p_i \), is defined as:

\[
I_i = \rho_i r(n_i), \quad \text{where} \quad r(n_i) = \sum_{m=-n}^{n} \sum_{k} l_{kn} T_{kn}(n_i)
\]

Then, any image is a linear combination of harmonic images, \( b_{nm} \), of the form

\[
b_{nm}(p_i) = \rho_i r_{nm}(n_i) \quad \text{with} \quad I_i = \sum_{n=0}^{\infty} \sum_{|m| \leq n} b_{nm} Y_{nm}(\alpha, \beta)
\]

The method is an efficient recognition algorithm in which they are using an accurate approximation to the model’s images. One can compare models to images in a 9D space that captures at least 98 percent of the energy of all the model’s images. Most of the approaches to deal with variation in illumination make single light source assumption which does not hold in most real conditions. Aggarwal and Chellappa [1] proposed an approach to Face Recognition in the presence of multiple illumination sources. They improved earlier approaches using Lambertian reflectance by taking into account the inherent hard non-linearity present in the Lambert’s law. They deal with multiple illumination sources by using an objective function, which can be minimized repeatedly for different illumination sources and the one with minimum error can be taken as the correct hypothesis.

They assume that an image of an arbitrarily illuminated face can be approximated by a linear combination of the images of the same face in the same pose, illuminated by nine different light sources placed at pre-selected positions. Hence, a face image can be defined as

\[
h = \sum_{i=1}^{n} \alpha_i h_i
\]

where, \( h_{i, \text{max}}(\rho n^T S_i, 0) \) and \( \rho \) is face albedo, \( n^T \) is face surface normal and \( S_i = \{s_1, s_2, ..., s_9\} \) are the predefined illumination directions. The recovery of the identity vector \( f \) and illumination \( s \) can be posed as an optimization problem as follows:

\[
\varepsilon(f, s) = \| h - h_{\text{rec}} \|^2 + (1/f^2 - 1)^2 \text{ where } h_{\text{rec}} = \sum_{i=1}^{n} f \sum_{j=1}^{9} \alpha_j \max(T_is_j, 0)
\]

Using the function (14), it is possible to potentially recover the illumination-free identity vector \( f \) without any prior knowledge of the number of light sources or any need to check different hypotheses for the same. The objective function is minimized with respect to \( f = \{f_1, f_2, ..., f_n\} \) and \( \alpha = \{\alpha_1, \alpha_2, ..., \alpha_9\} \). This gives us the illumination free identity vector \( f \) which is used for recognition. The optimization is done in an iterative fashion by fixing one parameter and estimating the other.

Georgiades et. al. [12] present a generative appearance based method for recognizing human faces under variation in lighting and viewpoint. The method exploits the fact that the set of images of an object in a fixed pose, but under all possible illumination conditions, is a convex cone in the space of images. Using a small number of training images of each face taken with different lighting directions, the shape and albedo of the face can be reconstructed. In turn, this reconstruction serves as a generative model that can be used to render or synthesize images of the face under novel poses and illumination conditions. The pose space is then sampled and, for each pose, the corresponding illumination cone is approximated by a low-dimensional linear subspace whose basis vectors are estimated using the generative model. They tested the method on 4,050 images from the Yale B Face Database; these images contain 405 viewing conditions (9 poses and 45 illumination conditions) for 10 individuals. The method performs almost without error, except on the most extreme lighting directions, and significantly outperforms popular recognition methods that do not use a generative model.

Qing et. al. [27] presents a model-based approach to identify faces robustly under generic illumination with the face relighting technology. Relighting of face images to a canonical illumination is based on the harmonic image model. Benefiting from the observations that human faces share similar shape and the albedos of the face surfaces are quasi-constant, first estimate the nine low-frequency components of the illumination from the input facial image. The facial image is then normalized to the canonical illumination by re-rendering it using the illumination ratio image method. For the purpose of face recognition, two kinds of canonical illuminations, the uniform illumination and a frontal flash with the ambient lights, are considered, among which the
former encodes merely the texture information, while the latter encodes both the texture and shading information. Experiments on the PIE and the Yale B database shows that the proposed relighting normalization can significantly improve the performance of a face recognition system when the probes are collected under varying lighting conditions.

4.2. Active Appearance Model

In AAM the 2D shape is represented by a triangulated 2D mesh with \( h \) vertices, which correspond to the salient points of the object. Mathematically, the shape vector \( s \) consists of the 2D coordinates of the vertices that make up the mesh as \( s = (x_1, y_1, ..., x_h, y_h)^T \) and shape variation is expressed by a linear combination of a mean shape \( s_0 \) and \( p \) shape basis vectors \( s_t \) as [14],

\[
s = s_0 + \sum_{t=1}^{p} \beta_t s_t
\]

where, \( \{\beta_t\}_{t=1}^{p} \) are shape parameters, and \( t \)th shape basis vector \( s_t \) is the \( t \)th eigenvector that corresponds to the \( t \)th largest eigenvalue. A standard approach to compute the linear shape model is to apply PCA to a set of shape vectors gathered from manually landmarked training images and aligned using the Procrustes analysis. Once a mean shape \( s_0 \) is obtained, the training images are warped to the mean shape using the piece-wise affine warps that is defined between the corresponding triangles in the landmarked shape of the training images and the mean shape. Then, it is possible to define the appearance as the shape normalized image over the region covered by the mean shape \( s_0 \). The appearance variation is expressed by a linear combination of a mean appearance \( A_0 \) and \( q \) appearance basis images \( A_i \) as

\[
A = A_0 + \sum_{i=1}^{q} \alpha_i A_i
\]

where, \( \alpha_i \) is the appearance parameter, and the \( i \)th appearance basis image \( A_i \) is the \( i \)th eigenface image that corresponds to the \( i \)th largest eigenvalue. Like the shape model, the appearance model is computed from manually landmark training images by collecting the shape normalized images and applying PCA.

The conventional AAM often diverge, when the input image has pose, expression and illumination variations which are not included in the training set. Although variation specific models like view-based AAM or person-specific AAM perform robustly for variations of pose or subject, they require off-line training using a variety of AAM for different subjects and poses.

Lee and Kim [20] overcome the limitation of AMM by tensor-based AAM, which applies multi-linear algebra to the shape and appearance models of the conventional AAM to improve the fitting performance. Tensor-based AAM consists of an image tensor and a model tensor. The image tensor estimates the pose, expression, and illumination of the input image. The model tensor generates variation-specific AAM basis vectors using a single trained model tensor and the estimated variations in an on-line manner. To estimate the variations from the image tensor, they proposed two different approaches: discrete variation estimation and continuous variation estimation. The conventional AAM often diverge, when the input image has pose, expression and illumination variations which are not included in the training set. Although variation specific models like view-based AAM or person-specific AAM perform robustly for variations of pose or subject, they require off-line training using a variety of AAM for different subjects and poses.

Bilinear Illumination Model [21] is shown to be superior to Lambertian [14] and Phong model. Phong model [14] is a 2-D Models, which deals with illumination by approximating the diffuse and specular reflection on a surface. Its parameters are the intensity of ambient light, the intensity of directed light, its direction, the specular reflection of human skin and angular distribution of specular reflections on human skin. It easily surpasses the limitations of a Lambertian reflectance model. But disadvantage is that, all the extracted textures are registered in a shape-free vector space, so they lose all shape information for matching.

Lee, Moghaddam, Pfister, and Machiraju [21] present a method to generate an illumination subspace for arbitrary 3D faces based on the statistics of measured illuminations under variable lighting conditions from many subjects. A bilinear model based on the higher-order singular value decomposition is used to create a compact illumination subspace given arbitrary shape parameters from a parametric 3D face model. First 3D point-to-point correspondences across different 3D faces are obtained. Illumination samples (intensities) from each reflectance image are projected from the 3D points on the face, yielding registered 2D samples which are thereby aligned with the 3D shape, all in one common vector space. They also compute a diffuse texture from all illuminated images for each subject. Assuming that facial texture is not coupled with the shape and reflectance, they factor out diffuse texture from the illumination samples as follow:

\[
w_k = \frac{t_k}{t_k}, \quad k = 1...N,
\]

where, an illumination sample, \( t_k \) is diffuse texture at a 3D point \( p_k \) with \( N \) as the number of 3D mesh points. Call \( w \) as a texture-free illumination component, which differs from reflectance since it includes cast shadows. Using a fitting procedure based on minimizing the distance of the input image to the dynamically changing illumination subspace, it is possible to reconstruct a shape-specific illumination subspace from a single photograph. The reconstructed illumination subspaces is used in several face recognition experiments with variable lighting conditions and obtain accuracies that are very competitive with previous methods that require training sessions or multiple images of the subject. The advantage of the method is that only one image is required to construct 3D shape. With the Yale B face database, the method was comparable to the prior state-of-the-art methods despite the
much simpler computation for obtaining an illumination-invariant face representation from a single image.

4.3. Three-D Morphable Model

Face recognition across variations in a wide range of illuminations, including cast shadows and specular reflections is presented by Blanz and Vetter [6]. To account for these variations, the algorithm simulates the process of image formation in 3D space and it estimates 3D shape and texture of faces from single images. The estimate is achieved by fitting a statistical, Morphable Model of 3D faces to images. The model is learned from a set of textured 3D scans of heads. They describe the construction of the morphable model, an algorithm to fit the model to images, and a framework for face identification. In this framework, faces are represented by model parameters for 3D shape and texture. The morphable face model is based on a vector space representation of faces that is constructed such that any convex combination of shape and texture vectors $S_i$ and $T_i$ of a set of examples describes a realistic human face:

$$S = \sum_{mm} a_i S_i \quad \text{and} \quad T = \sum_{mm} b_i T_i$$

(18)

Continuous changes in the model parameters $a_i$ generate a smooth transition such that each point of the initial surface moves toward a point on the final surface. Just as in morphing, artifacts in intermediate states of the morph are avoided only if the initial and final points are corresponding structures in the face, such as the tip of the nose. They tested performance of their algorithm using publicly available CMU-PIE database. The recognition rate varies from 89.0% to 95% for front, side and profile view.

5. Other Illumination Handling Methods

The basic methods for illumination invariant Face Recognition are Histogram Equalization, Gamma Intensity Correction, Logarithmic Transform, Edge feature based method, etc. They are already discussed in [11] or [40]. Here, few more methods are covered briefly. Most of the methods do not classify face image into small-scale feature image and large scale feature image. But, Xie et. al. [36] presented a new framework, in which a face image is first decomposed into a large-scale feature image and a small-scale feature image. Normalization is then mainly performed on the large-scale feature image, and meanwhile a smooth operator is applied on the small-scale feature image. Finally, the processed large and small scale feature images are combined together to generate an illumination normalized face image.

Liau and Isa [23] propose a method for implementing illumination invariant face recognition based on discrete cosine transform (DCT). This is done to address the effect of varying illumination on the performance of appearance based face recognition systems. The proposed method aims to correct the illumination variation rather than to simply discard it. In the proposed method, illumination variation, which lies mainly in the low frequency band, is normalized in the DCT domain. Other effects of illumination variation which manifest themselves by the formation of shadows and specular defects are corrected by manipulating the properties of the odd and even components of the DCT. The proposed method possesses several advantages. First, it does not require the use of training images or an illumination model and can be applied directly to the test image. Second is the simplicity of the system. It needs only one parameter to be determined, making it simple to implement. Experimental results on Yale face database B using PCA and support vector machines (SVM)-based face recognition algorithms showed that the proposed method gives good enough performance compared to other well-known but more complicated methods.

A new approach for handling illumination and pose variations is proposed by C. Choi and S. Choi, [10] which is based on two-dimensional face images. Face images in different poses under various illumination conditions are classified into several pose classes using the geometrical relations between facial feature points. Then, the direction of light is determined in each pose class using the binary images obtained by simple mathematical calculation. Let $a_{pc}$ denote the average value of the gray-level intensity of all the pixels in a face image $I_p$ with pose $p$ and illumination class $c$. Then generate a modified new image $I'_{pc}$ from $I_p$ as follow:

$$I'_{pc}(i,j) = \begin{cases} 0 & I_p(i,j) < a_{pc} \\ \frac{255}{2} & I_p(i,j) \geq a_{pc} \end{cases}$$

(19)

This binary image, $I'_{pc}$, reflects the effect of a light source in different directions on a face image and can therefore be used to estimate the illumination category. They divide the category $c$ into right (r), front (f) and left (l). The limitation of the system is that it considers only three illuminations directions.
Table 1. Method with their Error Rate and Recognition Rate on Yale B and CMU PIE Face Database

<table>
<thead>
<tr>
<th>CT</th>
<th>METHODOLOGY</th>
<th>Yale B (%)</th>
<th>CMU PIE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Subset 2</td>
<td>Subset 3</td>
</tr>
<tr>
<td>A</td>
<td>Gabor KPCA [24]</td>
<td>RR varies from 86.0 to 99.0 (Fig. 10 [21]) as features set size varies from 50 to 300</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>APCA [8]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>SQI [35]</td>
<td>98.0</td>
<td>99.0</td>
</tr>
<tr>
<td></td>
<td>GQI [30]</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>CSQI [26]</td>
<td>95.0</td>
<td>88.0</td>
</tr>
<tr>
<td></td>
<td>LSPPS [3]</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>NLM [31]</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Wiener [15]</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Gradient-face [38]</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>B</td>
<td>3DMM [6]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>S. Harmonics [40]</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Illumination Cone [12]</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>MLS Method [1]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>LJP [22]</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>S. Harmonics [27]</td>
<td>0.0</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>Face Lighting [34]</td>
<td>99.0</td>
<td>85.0</td>
</tr>
<tr>
<td>C</td>
<td>BIM [21]</td>
<td>-</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Back propagation Network (E) [25]</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Counter-propagation network (E) [25]</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>RLS(LOO-DCT) [36]</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>DCT-SVM [23]</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

6. Discussion and Concluding Remarks

The paper has classified various methods for illumination invariant Face Recognition into four category. It is observed that, Model based and Illumination invariant representation based methods have been used most widely. Performance of these methods is mostly evaluated on main two face databases i.e. CMU PIE [4] and Yale B face database [22]. TABLE-I and TABLE-II presents an amalgamation of results from different research and therefore methods may vary and direct comparison is not likely to be reliable. TABLE I gives details of methods used in various research papers with their Error Rate (ER) and Recognition Rate (RR) on Yale B and PIE database. TABLE II gives details of methods with their ER and RR on various face databases mention in the table. The methods are presented category (CT) wise. CT takes value A, B, C and D, which refers to Subspace based statistical methods, illumination invariant representation methods, model based methods and other methods respectively. In TABLE I, subset-2 to subset-5 represents image subsets of Yale B face database given in respective papers. First category (A) of methods is simple, straight forward and computationally efficient. But recognition accuracy is less compared to methods given in category ‘B’, and ‘D’ (TABLE I). Model based methods can give good accuracy but it requires more number of images for developing a model. Illumination invariant representation methods may requires maximum three and minimum one training image. There is no direct way of comparing computational complexity of second and third category of methods but based on experience and intuition it is possible to say that model generation, estimation of illumination or synthesizing of illumination of images may require more computation time compared to getting invariant representation using retinex based methods (B). Also, TABLE-I and TABLE-II shows that RR of second category of methods is better compared to third category of methods. Very few papers evaluate their methods using error rate and ROC curve. The fourth category of methods is used by few researchers and their performance is not better than second and third categories of methods (TABLE-1) if evaluation result on all subsets of Yale B database is compared. Yet there is a need to evaluate the methods on common database with same evaluation parameters.
Table 2. Methods with their Error Rate and Recognition Rate on Different Databases

<table>
<thead>
<tr>
<th>Database</th>
<th>CT</th>
<th>Methods</th>
<th>ER (%)</th>
<th>RR (%)</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asian Face</td>
<td>A</td>
<td>ICA [18]</td>
<td>-</td>
<td>86.00 to 100.00</td>
<td>The database contains the true-color face images of 103 people, 53 men and 50 women with 4 illumination variations per person.</td>
</tr>
<tr>
<td>ORL</td>
<td></td>
<td>GW-MPCA [13]</td>
<td>-</td>
<td>82.30 to 88.90</td>
<td>Database contains ten different images of 40 distinct subjects in up-right, frontal position.</td>
</tr>
<tr>
<td>AR Face</td>
<td></td>
<td>GW-MPCA [13]</td>
<td>-</td>
<td>86.00 to 89.95</td>
<td>-</td>
</tr>
<tr>
<td>Vetter</td>
<td>B</td>
<td>QI with Rendering [29]</td>
<td>0.33</td>
<td>-</td>
<td>Vetter's database contains 200 faces each under nine lighting conditions, making a total of 1,800 images.</td>
</tr>
<tr>
<td>ORL</td>
<td></td>
<td>LSPBPS [3]</td>
<td>-</td>
<td>96.0</td>
<td>Database contains ten different images of 40 distinct subjects in up-right, frontal position. 5 face images per person are chosen randomly as training images while the remaining 5 images are set as test images.</td>
</tr>
<tr>
<td>FERET</td>
<td></td>
<td>Gabor KPCA [24]</td>
<td>-</td>
<td>99.5</td>
<td>The data set consists of 600 FERET frontal face images corresponding to 200 subjects, such that each subject has three images of size 256x384 with 256 gray-scale levels.</td>
</tr>
<tr>
<td>Vetter</td>
<td></td>
<td>QI with Rendering [29]</td>
<td>0.33</td>
<td>-</td>
<td>Vetter's database contains 200 faces each under nine lighting conditions, making a total of 1,800 images.</td>
</tr>
<tr>
<td>FERET</td>
<td>C</td>
<td>3DMM [6]</td>
<td>1.0 to 9.8</td>
<td>87.9 to 95.9</td>
<td>-</td>
</tr>
</tbody>
</table>

Three dimensional methods may give better performance and comes out to be very useful if it is combined with two dimensional methods. Soft computing based methods may further improve recognition accuracy and decrease error rate. A method can be used for surveillance application if error rate is lower and recognition rate is high. Low false acceptance rate and moderate false rejection rate is desirable for banking related applications. But there is need to harvest good features of two or more methods for improving error rate and recognition rate. It is observed from evaluation results (TABLE I) that illumination invariant representation based methods gives better performance and is more suitable in terms of number of training images, simplicity, computational complexity, and robustness. Although the Yale and CMU PIE datasets are well chosen for investigation into illumination effects, they do not reflect real-world data (i.e. they are very constrained data sets) and it is hard to tell how these algorithms would work in the real world.

Acknowledgements
A. D. Patel Institute of Technology and Sophisticated Instrumentation Centre for Applied Research and Testing (SICART), Anand, Gujarat, India have provided support to accomplish the work. Author is thankful for the support.

References


