Illumination Invariant Feature Extraction Based on Natural Images Statistics – Taking Face Images as An Example

Lu-Hung Chen* Yao-Hsiang Yang* Chu-Song Chen#
Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan
luhung.chen,yhyang@statistics.twbbs.org song@iis.sincia.edu.tw

Ming-Yen Cheng
Department of Mathematics, National Taiwan University, Taipei, Taiwan
cheng@math.ntu.edu.tw

Abstract

Natural images are known to carry several distinct properties which are not shared with randomly generated images. In this article we utilize the scale invariant property of natural images to construct a filter which extracts features invariant to illumination conditions. In contrast to most of the existing methods which assume that such features lie in high frequency part of spectrum, by analyzing the power spectra of natural images we show that some of these features could lie in low frequency part as well. From this fact, we derive a Wiener filter approach to best separate the illumination-invariant features from an image. We also provide a linear time algorithm for our proposed Wiener filter, which only involves solving linear equations with narrowly banded matrix. Our experiments on variable lighting face recognition show that our proposed method does achieve the best recognition rate and is generally faster compared to the state-of-the-art methods.

1. Introduction

1.1. Scale Invariance in Natural Images Statistics

It has been aware in the fields of computer vision and image processing that the classes of natural images carry several distinct properties which are not shared with other classes of general random images (e.g. white noise, wire-frame rendering, etc.). This fact has underlain many heuristics in feature extraction and led to several classes of natural images data (e.g. images in the woods [21], range images [16], illumination maps [8], etc.) and to statistical models of natural images from these data since the '90s. These statistical descriptions serve as an alternative for image analysis when exact physical models failed. Such statistical models stir up many ideas in applications such as image denoising [18] and intrinsic images [29]. Please refer to Srivastava et al. [23] for a more general survey about the models and applications of this topic.

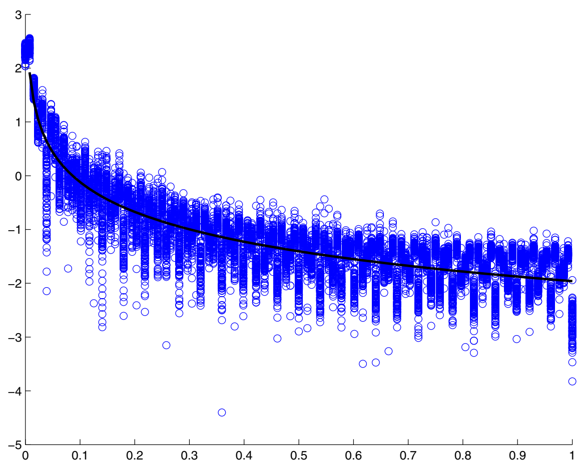

Figure 1. A typical power spectrum of a human face in x-direction (blue) and the fitted power law spectrum (black).

Throughout this paper, we restrict ourselves to the most basic statistical properties of natural images, namely the power law spectrum: $P(\omega) \propto \omega^{-\alpha}$ where ω is the frequency and the exponent α is typically a number close to 2. This property has been discovered repeatedly in several different sources of natural images [21][16][8]. It manifests the most general scale invariance property for many classes of natural images. A typical power spectrum and the fitted power law spectrum of a human face is shown in the Figure 1.

In this work, we wish to demonstrate how to use this simple assumption to construct a filter extracting illumination-

* First authors with equal contribution; # Corresponding author.

invariant features. We also suggest a method to estimate power spectra indirectly from well-designed datasets which provide a set of objects in certain class under a given set of lighting conditions. For this purpose, we use human face databases as an example. However the underlying rationale could be also applied to other classes of objects. We will also show that, while applying to variable lighting face recognition, this filter outperforms the state-of-the-art methods within *one sample per person setting*. We shall give a brief survey of *illumination-invariant feature extraction* for face recognition in the next section.

1.2. Illumination-Invariant Feature Extraction

There has been an abundant literature on *illumination-invariant feature extraction* in the area of face recognition, since variable lighting face recognition is an important issue for many applications in computer vision. It has been shown that, in face recognition, variations due to different illuminations are more significant than the inherent differences between face identities [1]. To solve the variable lighting face recognition problem, some approaches use 3D human-face models for training [4]. Despite its nice performance, collecting a series of 3D face scans is usually not feasible in practice. Analyses of illumination cone [9] or spherical harmonics [3][19] have shown that variable lighting images of a convex Lambertian object lie in a low-dimensional subspace. This property has been employed by several studies for variable lighting face recognition such as Lee et al. [17] and Chen and Chen [5]. To ensure the performance, this type of approaches requires the collection of multiple images of each subject under fixed or simply different lighting conditions. However, multiple variable-lighting training images of people are difficult to gather for a practical system. For the above reasons, we present an approach that can perform variable lighting face recognition within the *one sample per person setting*. Our approach utilizes the power law spectra of the illuminance and the reflectance. For the *one sample per person setting*, there has already been an abundance of literature, and variation over pose, expression and other issues have been treated as well. Of course, here we restrict ourselves to the case of variation over illumination only. For recent general methods and applicable environments, one may consult the literature survey of Tan et al. [24] for details.

The recent literature of variable lighting face recognition can be actually divided into two broad categories: (1) the *illumination normalization* which tries to estimate light source(s) and reconstructs the person's normal lighting (or uniform lighting) face, e.g. LAP [13] or estimates the generative model [9]; and (2) the *illumination-invariant feature extraction* which only extracts discriminative features that are invariant for the same person under different lighting conditions and discriminative between different people, e.g. LTV [6].

(a) (b) (c)

Figure 2. (a) Human face under serious lighting condition and its corresponding results produced by represented methods of (b) *illumination normalization* obtained from spherical harmonics and (c) *illumination-invariant feature extraction* from the proposed method.

Illumination normalization is more difficult than *illumination-invariant feature extraction* in the sense that it needs to reconstruct exactly the normal lighting (or uniform lighting) images from data and this is hardly resolved within *one sample per person setting*. It is hard to estimate the normal lighting (or uniform lighting) face given a single image because the problem is highly ill-posed. However, extracting features by reducing the lighting variance without losing too much the discriminative power is much easier. Typical results produced by representative methods in these two categories are shown in Figure 2.

We now turn to describe the idea development of the *illumination-invariant feature extraction* briefly. To the best of our knowledge, nearly every method in variable lighting face recognition, following the initiative work of Horn [15], utilizes the fact that the illuminance of an image (as a two-dimensional spatial function) is typically smoother than the reflectance. Thus they assume that most illumination variation lies in the low frequency part of the spectrum. Some of these methods formulate illumination variation in different bases (such as principal components [19], spherical harmonics [17][5], wavelets [31], and cosine functions [7]). Others derive some low-pass filter fulfilling certain constrains directly [11][28]. The most recent methods perform cartoon decomposition and extract the large scale part [6][30] in a variational scheme. They take different approaches to eliminate the illumination variation, but they do not examine in details the relationship between the illuminance and the reflectance *across the whole frequency spectrum*. In short, most methods actually truncate the low frequency part of the spectrum under certain kind of decomposition.

Although it has long been known that low frequency part contains most variations of illumination [3], only until recently, Xie et al. [30] started to investigate seriously the possibility that part of the relevant features may lie in the low frequency part[1] and cannot be simply discarded. However this approach only divides the frequency spectrum into the low and high frequency parts, extracts features accordingly, and combines the results from the two groups. Two limitations could be: (1) their method decomposes the image at an arbitrary frequency and does not use all combinations of different frequency grouping schemes; and (2) they do not utilize the relationship across the whole spectrum and thus the feature extraction strategies they emploied cannot be grounded theoretically.

In this work, we wish to provide a discriminative method which best separates the spectra of illuminance and reflectance across the whole frequency spectrum. For this purpose, we suggest a simple Wiener filter derived from power law spectra assumption of natural images and examine the statistical behaviour of face images under different lighting conditions. Wiener filter is a Bayesian optimal filter for separating two spectra of sources given the observed composite signal. To amazing, it has not been well employed for computer vision before. It decomposes the sum of two signals drawn from two given stationary processes with different autocorrelation functions, where the Fourier transform of the autocorrelation function is the power spectrum density in the frequency domain. Thus our method retains features at every frequency while previous ones preserve features only at the high frequency part. The underlying difference between our method and previous direct thresholding methods across the entire spectrum is shown in the Figure 3. The rest of the paper is organized as follows: In Section 2 we begin with the reflectance model, followed by the derivation of our *natural delighting filtering* and discuss the choice of their parameters in different image sizes. We then give the procedure of indirect measurement of the power law spectra we need in Section 3, and describe the experimental evaluation of our assumption in Section 4. Finally, conclusions and future directions are given in Section 5.

2. Wiener Filtering for Natural Illumination

According to Barrow and Tenenbaum's intrinsic image model [2], an image $I(x,y)$ can be represented as

$$I(x,y) = R(x,y)\, L(x,y)\,,$$

where $I(x,y)$ is the intensity, $R(x,y)$ is the reflectance, and $L(x,y)$ is the illuminance of the pixel location (x,y). For the face recognition problem, R contains important features such as the shapes and locations of eyes and noses. Hence our main interest is to retrieve the variation pattern of R from a given image I. Horn [15] suggested to take the logarithm of I and transforms the model into an additive

[1] Actually the decomposition they use is the cartoon decomposition, but we identify the large scale part to the low frequency part for clarity of argument.

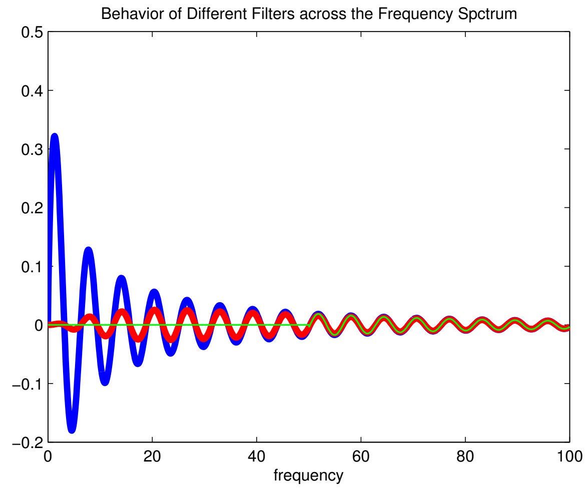

Figure 3. The effects of Wiener filtering (red) and directly thresholding method with the threshold at frequency 50 (green) in comparison with the original spectrum (blue). It can be shown that both directly thresholding method and Wiener filtering preserve features at the high frequency part, while Wiener filtering retains features at the low frequency part as well. The original spectrum is generated from $\exp(-x^{0.35}) \cdot \sin(x)$ as an example.

one:

$$\begin{aligned} f(x,y) = \log I(x,y) &= \log R(x,y) + \log L(x,y) \\ &= \nu(x,y) + \mu(x,y)\,. \end{aligned}$$

Since μ is smoother than ν, we can first estimate μ by some smoothing filter. Denote the estimator by $\hat{\mu}$. Let $\hat{\nu} = f - \hat{\mu}$, we then estimate R by $\hat{R} = \exp(\hat{\nu})$.

Let f, ν, μ be drawn from three wide-sense stationary processes and the latter two are uncorrelated, which follows the natural image statistics of power laws of the single scan-line data. Restricted within the *one sample per person setting*, the estimation problem of μ is highly ill-posed. However, we can still apply Wiener filter to get the Bayesian optimal estimate of μ from a single image since the filter's behaviour is completely determined by the autocorrelations of ν and μ. From this observation, we do not need to extract the exact μ and ν, but can still separate them with the optimal filtering setting. Here the filter does not even depend on the mean of each process, while *illumination normalization* methods target at estimating the normal lighting face thus depend on the exact distribution of ν.

Since the stationary condition of natural images is only satisfied in the one-dimensional case but severely violated in the two-dimensional case, and x-, y-directions are two dominant directions in a two dimensional power spectrum [20], we restrict ourselves to the one-dimensional power law spectrum. We shall filter the image in the x-direction and the y-direction consecutively. So we study only one-dimensional signal in the rest of this section. The 1-D setting also makes the derived filter highly efficient to compute. We assume further that both μ and ν follow power law spectrum:

$$\begin{aligned} P_\mu(\omega) &\propto \omega^{-\alpha_\mu}, & (1)\\ P_\nu(\omega) &\propto \omega^{-\alpha_\nu}, & (2)\end{aligned}$$

where P_μ, P_ν are power spectrum densities of μ and ν, respectively, and α_μ, α_ν are some positive real numbers. Since most of the natural images statistics were estimating on the logarithmic domain, we could still apply the power law spectrum assumption in our case.

First, let us consider the Wiener filter in the frequency domain:

$$\begin{aligned} \mathcal{F}\{l\}(\omega) &= \frac{P_\mu(\omega)}{P_\mu(\omega) + P_\nu(\omega)} & (3)\\ &= \frac{\lambda}{\lambda + \omega^\delta}, & (4)\end{aligned}$$

where l is the Wiener filter in the spatial domain, $\lambda > 0$ is the ratio of power spectra P_μ and P_ν at the frequency $\omega = 1$, and $\delta = \alpha_\mu - \alpha_\nu$.

We refer to the filters derived with respect to (4) the *natural delighting filters (NDFs)*, which are varying with δ. In particular, we shall show later in the Section 3 that δ is around 2 for the face images case. Approximating δ by 2 the result could be easily computed from spatial domain without Fourier transform since solving linear equations with narrowly banded matrix only requires $O(n)$ time. The resulted filter in the spatial domain is determined by:

$$\lambda l[f] + \frac{\partial^2}{\partial x^2}(l[f]) = \lambda f,$$

which can be expressed in the following narrowly banded matrix equation after discretization:

$$(\lambda I + D^T D)\boldsymbol{\mu} = \lambda \boldsymbol{f}, \qquad (5)$$

where $\boldsymbol{f} = (f_1, \cdots, f_n)^T$ is a row or a column of the input image, $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_n)^T$, and D is a $(n-1)\times n$ difference matrix: $D_{i,j} = -1$ if $i = j$, $D_{i,j} = 1$ if $i = j - 1$, and $D_{i,j} = 0$ otherwise. It can be easily verified that the solution of (5) is equivalent to a linear smoothing spline [27] that is easy to compute. We can boost the computation speed by using the analytic solutions introduced in [10] and reduce the computation complexity to exactly $4n$ flops. Solving equation (5) along x- and y-directions consecutively is referred to as the NDF of $\delta = 2$ in this paper. Note that as we approximate δ by 2 for computational issues, λ could no longer be the ratio of power spectra and need to be chosen empirically.

Before we go into the details of experimental evaluation, we shall briefly mention the procedure of parameter selections. Assume that a suitable parameter is found for some scale (image size), we can apply the following derivation to find the suitable parameters for any other scales, which is referred to as the scale-transform property for parameter selection in this paper:

$$\begin{aligned} P_{\tilde\mu}(\omega) &= \mathcal{F}\{\tilde\mu\}\overline{\mathcal{F}\{\tilde\mu\}}\\ &= \frac{1}{\sigma^2}\mathcal{F}\{\mu\}\left(\frac{\omega}{\sigma}\right)\overline{\mathcal{F}\{\mu\}\left(\frac{\omega}{\sigma}\right)}\\ &= \frac{1}{\sigma^2}P_\mu\left(\frac{\omega}{\sigma}\right),\\ \frac{P_{\tilde\mu}(\omega)}{P_{\tilde\mu}(\omega)+P_{\tilde\nu}(\omega)} &= \frac{\frac{1}{\sigma^2}P_\mu\left(\frac{\omega}{\sigma}\right)}{\frac{1}{\sigma^2}P_\mu\left(\frac{\omega}{\sigma}\right)+\frac{1}{\sigma^2}P_\nu\left(\frac{\omega}{\sigma}\right)}\\ &= \frac{\lambda}{\lambda + \frac{1}{\sigma^\delta}\omega^\delta} = \frac{\sigma^\delta\lambda}{\sigma^\delta\lambda + \omega^\delta},\end{aligned}$$

where $\tilde\mu$ and $\tilde\nu$ are the scaled signals of μ and ν with $\tilde\mu(x) = \mu(\sigma x)$, $\tilde\nu(x) = \nu(\sigma x)$, respectively. Hence the multiplicative rescaling factor of λ is σ^δ. Based on this property, we can vary the parameters with respect to different image sizes easily. We shall follow this parameter selection procedure in the next section which allows us to transform parameters across different databases. Note that there are in fact two parameters in the x- and y-directions respectively.

3. Verification of δ

We estimate the power spectrum density of a stationary sequence $x(1), \cdots, x(n)$ by periodogram

$$\hat P_X(\omega) = \frac{1}{n}|\sum_{t=1}^{n} x(t)e^{-j\omega n}|^2 = \frac{1}{n}|DFT(x)|^2,$$

while other methods such as Welch's method can be used as well. For more details on the estimation of power spectrum density, we refer to [14]. Note that from the Wiener-Khinchin theorem, the power spectrum density of a stationary signal x is the fourier transform of its autocorrelation

$$\rho_x(\tau) = \frac{E[(x(t)-\bar x)(x(t+\tau)-\bar x)]}{\sigma^2},$$

where $\bar x$ is the mean and σ^2 is the variance of x. Due to the violation of the constancy of first moment in our case, we extend the definition of autocorrelation function to cover this non-stationary case:

$$\bar\rho_y(\tau) = \rho_{y-\bar y}(\tau),$$

which simply normalizes the signal by subtracting the mean signal before the calculation of second moment. Note here we assume the constancy of second moment after such normalization. Since $\bar\rho_y = \bar\rho_{y+c}$ for any fixed function c, it is easy to show that the power spectrum density remains the same when a fixed signal is added.

To estimate the value of δ, we need to approximate P_μ and P_ν, the respective power spectra of μ and ν, respectively. Let $f_{s,t}$ denote the logarithm of the image indexed by subject s and illumination condition t. Remember that μ and ν are the illumination and reflectance parts in the log domain, respectively. First, let us consider the case where the subject is fixed and the illumination varies. By the assumption that μ and ν are independent, the autocorrelation of the sequence $f_{s,1}, \cdots, f_{s,T}$, where T is the number of illumination conditions, can be approximated by the autocorrelation of the sequence $\mu_{s,1}, \cdots, \mu_{s,T}$ since $\nu_{s,t_1} \approx \nu_{s,t_2}$ for all pairs of t_1, t_2 (the same subject has the same reflectance), and thus the autocorrelation of the sequence $\nu_{i,1}, \cdots, \nu_{i,J}$ should be close to 0. On the other hand, let us consider the case where the illumination is fixed and the subject varies. In this case, the autocorrelation of the sequence $f_{1,t}, \cdots, f_{S,t}$, where S is the number of subjects, can be approximated by the autocorrelation of the sequence $\nu_{1,t}, \cdots, \nu_{S,t}$ since $\mu_{s_1,t} \approx \mu_{s_2,t}$ (the same illumination), and thus the autocorrelation of the sequence $\mu_{1,t}, \cdots, \mu_{S,t}$ should be close to 0. Hence, we approximate P_μ by estimating the power spectrum density of illumination changes for each subject, and approximate P_ν by estimating the power spectrum density of different subjects for each illumination condition.

Since we assume that P_μ and P_ν follow the power law as in 1, by taking logarithm on P_μ and P_ν we have

$$\begin{aligned} \log P_\mu(\omega) &= \gamma_\mu - \alpha_\mu \log \omega, \\ \log P_\nu(\omega) &= \gamma_\nu - \alpha_\nu \log \omega, \end{aligned}$$

where γ_μ and γ_ν are some constants. Hence, we can estimate α_μ and α_ν by regressing the logarithm of P_μ and P_μ on the logarithm of frequency ω, respectively. When α_μ and α_ν are estimated, δ can be estimated by $\hat{\alpha}_\mu - \hat{\alpha}_\nu$ as well. We carry out the estimation procedure on Yale B face database since it includes more illumination conditions than the other datasets. The estimated values of α_μ are 3.97 in the x-direction and 4.04 in the y-direction, and the estimated values of α_ν are 2.03 in the x-direction and 2.09 in the y-direction. Thus the estimated values of δ are 1.94 and 1.95 in the x-direction and y-direction, respectively, which are indeed both around 2; thus we can approximate δ by 2 to reduce the computational complexity. Figure 4 shows the log-log plot of the regressed P_μ and P_ν in the x-direction and y-direction, where α_μ and α_ν are the slopes of the corresponding regression lines. As we have estimated the power law parameters from only the Yale B database, the one-sample-per-person face recognition experiments are done for other different or broader databases in the following.

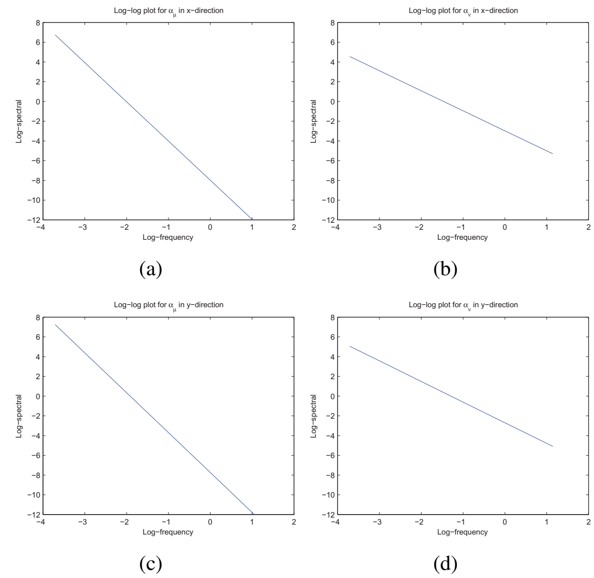

Figure 4. The log-log plot of the regressed (a) P_μ, (b) P_ν in the x-direction, (c) P_μ, (d) P_ν in the y-direction.

4. Experiments

We compare our proposed NDF method with several methods that can perform lighting invariant feature extraction based on only a single image without pre-training. The methods compared include the SQI [28], LTV [6], RLS [30], and TT [25], which achieve the highest recognition rates on the datasets we use as far as we know. We use the codes provided by the authors of LTV [6] and TT [25] for the implementation, and use the same parameters as reported in [6] and [25]. The SQI method [28] has no available codes, and so we implement it on our own. The RLS method [30] has also no available codes, and we directly use their recognition rates reported in [30] for comparison. As mentioned in Section 2, λ can no longer be derived directly from the ratio of power spectra since we set δ to be 2 to boost the computational speed. Therefore we choose our parameter λ empirically by the average recognition rate on Yale B face database, and use the same optimal λ for the other databases. The empirically optimal λ is 0.5 for the x-direction and 0.35 in the y-direction for image size 100×100.

We use CMU PIE [22], Extended Yale B [17], and CMU Multi-PIE [12] face databases for evaluation. Although there are other datasets containing faces under different lighting conditions, their conditions are either few or lacking of sufficient serious-lighting situations. The above datasets are systematically designed for lighting (with poses and expressions fixed), where CMU PIE contains 68 humans, each with 21 different illuminations. Extended Yale B includes the original Yale B face database with images of

	SQI [28]	LTV [6]	RLS [30]	TT [25]	NDF
Ideal	97.19	100	99.9	100	100
Average	95.19	99.35	–	100	100

Table 1. Recognition rate (%) of CMU PIE face database

10 humans in 64 different lighting conditions (divided into 5 subsets according to [17]), and the extended part with 16128 images of 28 humans captured under the same conditions as in Yale B. CMU Multi-PIE, which is a recently published face database, contains images of 337 subjects each with 20 different illumination conditions. To concentrate on the performances due to illumination changes, only frontal faces are used with all images simply aligned by eyes, and the image size is 100×100. Thus our result on CMU Multi-PIE differs from the one recently reported by [13], which follows the standard testing procedure described by [12] with multiple poses and expressions.

Since approaches of this type serve as preprocessing methods for lighting normalization, unlike a general face recognition system, the recognition method used is simply *cross correlation* (CC) as also done in many previous related approaches [28][5][6][30], so that lighting will be the only factor affecting the recognition results and the influences of other factors are removed. Since a general face recognition system is presented in [25] which uses methods such as distance transformation and kernel LDA for recognition, we employ the preprocessing and feature extraction steps of the TT method [25], and only change the recognition step to CC in our implementation for fair comparison.

[**CMU PIE Database**]: This is a simple dataset that can demonstrate the basic performance of our method. We first use the ideal images (i.e., frontal lighting images) as the reference images (called the *ideal case*). The recognition rates of SQI and RLS are 97.19% and 99.9% respectively and all of LTV, TT, and our NDF methods reach the 100% rate. Then we use images under all 21 lighting conditions as the reference images in turn and average the results (called the *average case*). In this case, the recognition rates of SQI and LTV are 95.19% and 99.65%, respectively, while both TT and our NDF model reach the 99.94% rate. The results for the RLS method in the average case have not been reported in [30]. Note that the recognition rates obtained for LTV and SQI of our implementation are consistent to those reported in [6] and [28] for this dataset.

[**Extended Yale B Dataset**]: Extended Yale B is the most challenging dataset for illumination-robust recognition as the lighting directions vary from left 130° degrees to right 130°. We report the results for all the 5 subsets. Table 2 shows the recognition rates of the ideal case when the ideal images (frontal lighting images) are used as the reference images, and table 3 shows the average recognition rates when images under all 64 lighting conditions are used as the reference images. Note that for the RLS method, only the recognition rates of subsets 2–5 were reported for the ideal case in [30]. Our proposed model achieves much higher recognition rates than all the other methods under all lighting conditions for this more challenging dataset.

Subset	SQI [28]	LTV [6]	RLS [30]	TT [25]	NDF
1	89.87	88.72	–	93.98	**100**
2	**100**	**100**	**100**	**100**	**100**
3	80.55	80.14	87.1	93.61	**100**
4	80.78	78.11	87.6	98.25	**100**
5	77.11	80.3	84.8	99.03	**100**
Average	84.23	84.48	–	97.33	**100**

Table 2. Recognition rate (%) of Extended Yale B face database when the ideal images are used as reference images.

Subset	SQI [28]	LTV [6]	TT [25]	NDF
1	79.70	74.97	91.92	**97.15**
2	77.59	77.01	89.02	**94.96**
3	67.20	71.59	86.25	**94.42**
4	72.68	78.96	90.89	**96.47**
5	72.96	86.93	93.80	**98.54**
Average	73.25	78.91	90.41	**96.43**

Table 3. Average recognition rates (%) of Extended Yale B face database.

Note that we employed the pre-processing and feature extraction of the TT method [25] but changed the recognition step to CC. Compared to our implementation, the recognition rate reported in [25] on this dataset of their original method, which uses distance transformation in the recognition step, is 99.0% for the ideal case (they have not shown the result of the average case). This indicates that the TT method [25] does not necessarily perform better than our method even when more complicated similarity metrics are used.

[**CMU Multi-PIE Dataset**]: CMU Multi-PIE is a recently reported larger dataset which contains images from much more subjects than CMU PIE and Extended Yale B datasets. Since the illumination conditions are relatively simple for this dataset, we only report the average recognition rates. The average recognition rates for SQI and LTV are 98.85% and 99.56%, and both of TT and our NDF achieve 100%, respectively. Note that to concentrate on the influences of illumination changes, only the frontal images with the same expressions are used for each subject in our experiments. Thus our recognition rates on this dataset are higher than those recently reported by [13], which includes all the

SQI [28]	LTV [6]	TT [25]	NDF
98.85	99.56	100	100

Table 4. Average recognition rate (%) of CMU Multi-PIE face database

images under different viewpoints and expressions.

[**Computational Speed**]: In comparison of the computational speeds, since the RLS method uses LTV as a sub-step, it cannot be faster than LTV. The SQI method requires several iterations for multiple smoothing, making its computation time unstable. On the other hand, the TT method only needs a single run which shall be faster than SQI. Hence, we only report the computation time of LTV, TT, and our NDF for comparisons. We run all these implementations on a 2.8 GHz Core 2 Duo. The computation time of LTV is 11470 milliseconds (ms) per image for image size 100×100, and 47060 ms for 192×168. TT requires 10.5 ms and 32.4 ms for image sizes 100×100 and 192×168, respectively. As to our method, we need only 0.9 ms and 3.2 ms for image sizes 100×100 and 192×168, respectively.

5. Conclusion and Future Works

This paper explores a new direction to extract lighting invariant features. Although natural image statistics such as the power laws have been studied for many years, their application to anti-lighting feature discovery has not been explored. In this paper, we show how to extract these salient features in an optimal way by using Wiener filtering. Our derived method is very simple but effective.

We summarize our findings as following:

1. As to extract lighting invariant features, previous studies typically focus on modeling the lighting (spectrum) only, but overlooking to model the reflectance. We show that, as the low frequency bands also contain part of the reflectance information in terms of the power law in the Fourier domain, we shall consider both spectra instead of only the spectrum of lighting.

2. Our results show that, to separate the spectra from two different power laws, it is not enough to simply cut out the low frequencies (for delighting) or projecting the signal to the low-frequency space (for retrieving lighting), as several of the previous studies did (e.g. logDCT [7] and Intrinsic Subspace [5]). Instead, we shall consider the whole spectra. The optimal filter (obtained by Wiener filtering) appears generally as a curve (but not a straight line) in the spectral domain.

In short, this paper investigates to what extent the scale-invariance property can be used to separate the illumination from an image. We illustrate that, within the face category, this kind of separation can outperform state-of-the-art *illumination-invariant feature extraction* methods according to the recognition rates. The only assumptions of our work are power law spectrum and stationary signal from the natural image statistics. Based on these properties, We derive our simple and computationally highly efficient NDF as a special case of the Wiener filtering. Thus the proposed models can be directly applied to a single image without any prior information about the 3D geometry or the position of light sources. Although we restricted ourselves to the category of faces due to the data availability and evaluation consideration, our method can be extended to extract illumination-invariant features from any object category.

Two other directions for further study can be drawn from our conclusions. First, the constants of power-law spectra may vary among different datasets, especially among different settings of object classes (such as artificial versus natural scenes shown in Torralba and Oliva [26]), and result in different Wiener filters with different values of δ. In the future, we wish to evaluate on other datasets to explore the power of current method under different conditions. Second, other kind of natural image statistics may be derived for intrinsic images. In this paper, we employ the power laws which can be derived indirectly from the well-designed existing face datasets. In the future, we can also consider the natural statistics of intrinsic images in other transform domains (such as wavelets). New dataset may be needed for directly estimating natural image statistics from a large set of intrinsic images like illumination or reflectance.

Acknowledgement

This paper was supported in part under the grants 99-EC-17-A-02-S1-032, NSC98-2221-E-001-012-MY3, and NSC97-2118-M-002-001-MY3.

References

[1] Y. Adini, Y. Moses, and S.Ullman. Face recognition: The problem of compensating for changes in illumination direction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):721–732, 1997.

[2] H. Barrow and J. Tenenbaum. Recovering intrinsic scene characteristics from images. *Computer Vision System*, 1978.

[3] R. Basri and D. Jacobs. Lambertian reflectance and linear subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):218–233, 2003.

[4] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1063–1074, 2003.

[5] C. P. Chen and C. S. Chen. Lighting normalization with generic intrinsic illumination subspace for face recognition. In *ICCV*, 2005.

[6] T. Chen, X. S. Zhou, and T. S. Huang. Total variation models for variable lighting face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1519–1524, 2006.

[7] W. Chen, M. J. Er, and S. Wu. Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 36(2):458 –466, apr. 2006.

[8] R. Dror, T. Leung, E. Adelson, and A. Willsky. Statistics of real-world illumination. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 2001.

[9] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under differing pose and lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):643–660, 2001.

[10] G. H. Golub and C. F. V. Loan. *Matrix Computations, 3rd Edition*. The Johns Hopkins University Press, 1996.

[11] R. Gross and V. Brajovic. An image preprocessing algorithm for illumination invariant face recognition. In *4th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*. Springer, June 2003.

[12] R. Gross, I. Matthews, J. F. Cohn, T. Kanade, and S. Baker. Multi-pie. In *IEEE ICAFGR*, 2008.

[13] H. Han, S. Shan, L. Qing, X. Chen, and W. Gao. Lighting aware preprocessing for face recognition across varying illumination. In *ECCV*, 2010.

[14] M. H. Hayes. *Statistical Digital Signal Processing And Modeling*. Wiley, 1996.

[15] B. K. Horn. Determining lightness from an image. *Computer Graphics and Image Processing*, 3(4):277 – 299, 1974.

[16] J. Huang, A. Lee, and D. Mumford. Statistics of range images. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, 2000.

[17] K. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):684–698, 2005.

[18] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *Image Processing, IEEE Transactions on*, 12(11):1338 – 1351, 2003.

[19] R. Ramamoorthi. Analytic pca construction for theoretical analysis of lighting variability in images of a lambertian object. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(10):1322–1333, 2002.

[20] D. L. Ruderman. The statistics of natural images. *Network: Computation in Neural Systems*, 5(4):517–548, 1994.

[21] D. L. Ruderman and W. Bialek. Statistics of natural images: Scaling in the woods. *Phys. Rev. Lett.*, 73(6):814–817, Aug 1994.

[22] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression (pie) database. *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, 2002.

[23] A. Srivastava, A. Lee, E. Simoncelli, and S.-C. Zhu. On advances in statistical modeling of natural images. *Journal of Mathematical Imaging and Vision*, 18:17–33, 2003. 10.1023/A:1021889010444.

[24] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang. Face recognition from a single image per person: A survey. *Pattern Recognition*, 39(9):1725 – 1745, 2006.

[25] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE TIP*, 19(2):1635–1650, 2010.

[26] A. Torralba and A. Oliva. Statistics of natural image categories. *Network: Computation in Neural Systems*, 14(3):391–412, 2003.

[27] G. Wahba. *Spline models for observational data*. SIAM: Society for Industrial and Applied Mathematics, 1990.

[28] H. Wang, S. Li, and Y. Wang. Face recognition under varying lighting conditions using self quotient image. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings.*, pages 819–824, May 2004.

[29] Y. Weiss. Deriving intrinsic images from image sequences. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, 2001.

[30] X. Xie, W. Zheng, J.-H. Lai, and P. C. Yuen. Face illumination normalization on large and small scale features. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1 –8, jun. 2008.

[31] T. Zhang, B. Fang, Y. Yuan, Y. Yan Tang, Z. Shang, D. Li, and F. Lang. Multiscale facial structure representation for face recognition under varying illumination. *Pattern Recognition*, 42(2):251–258, 2009.