國立臺灣大學網媒所

碩士論文

自然影像的機率分析

楊耀翔 撰

101
7

國立臺灣大學電機資訊學院資訊網路與多媒體所
碩士論文

Graduate Institute of Networking and Multimedia

College of Electrical Engineering and Computer Science

National Taiwan University

Master Thesis

自然影像的機率分析
A Probabilistic Analysis of Natural Images

楊耀翔
Yang, Yao-Hsiang

指導教授：陳祝嵩 教授
王傑智 教授

Advisor: Chen, Chu-Song, Ph.D.
Wang, Chieh-Chih, Ph.D.

中華民國一百零一年七月
July, 2012

Abstract

Natural image statistics has long been interested by the researchers of the computer vision community during the last twenty years. There has been reports on many distinct characteristics of image data collecting under various conditions. Several applications based on the assumptions derived from natural image statistics have been proved to be both useful and intelligible in compare to those works which do not employ any explicit assumptions about images.

In this thesis, I wish to summarize those experimental results and explanations in previous studies. Then I wish to investigate the role of these facts in the research. After explicating the theoretical framework, two novel applications incorporating the Fisher's discriminant analysis and Wiener filtering are conduct in order to demostrate the power of this approach. I wish these rudimentary results are shown to be promising enough to make people more confident about the possible future toward a more complete formal theory of image analysis.

*Graduate Institute of Networking and Multimedia, National Taiwan University, Taipei 106, Taiwan.

Email: yhyang@statistics.twbbs.org

中文摘要

自然影像統計，對於計算視覺社群的研究者而言，在過去二十年間早已引發相當的興趣。已有不少研究報告在許多的不同條件下收集到的影像資料之不同特徵。相較於其他不對影像作明確假設的應用，基於自然影像統計假設的許多應用研究也已證明相當有用並易於理解。

在本文中，我希望摘要那些過去研究的實驗結果與解釋。而後，我希望調查這些事實在研究中可以扮演的角色。在闡釋理論的架構之後，兩項新的應用，分別利用費雪判別分析與維納濾波，將被推演用以展示這一方法的效力。我希望這些初步的結果能顯示足夠的前途，讓人們對未來發展更加完備的影像分析之形式理論有足夠信心。

致謝

本篇論文的完成，首先要感謝陳祝嵩老師與王傑智老師提供的各種資源與不厭其煩的指導。在構成本篇論文兩項專題的撰寫與投稿過程中，兩位老師的經驗提供了許多寶貴的意見，讓我對當前學術界的潮流與標準有所領會，並且得以選擇合適的題材進行研究。

其次，我必須感謝陳律閎博士在研究當中提供的協助。我們花費了相當多的時間對研究相關的設計與結果的解釋進行討論，如果沒有他的幫助，這些研究恐怕無法如期完成。自然，文章中錯誤的部分必須由我負責。

此外，尚要感謝鄭明燕老師在統計學方面的指導，在方法論的操作與最新的統計文獻上，鄭老師提供了足夠札實的訓練，這是刺激我產生新的構思的基礎。

最後，我希望將這篇論文獻給我的雙親，他們在生活中提供的關懷、支持與幫助，讓我能在研究中無後顧之憂，這是本篇論文最牢靠的基礎。

Contents

List of Figures

List of Tables

1 Introduction

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, wireframe 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 [37], range images [23], illumination maps [12], 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 [30] and intrinsic images [47]. Please refer to Srivastava et al. [39] for a more general survey about the models and applications of this topic.

Throughout the thesis, 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, and some basic consequences drived from it. This property has been discovered repeatedly in several different sources of natural images [37][23][12]. 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.

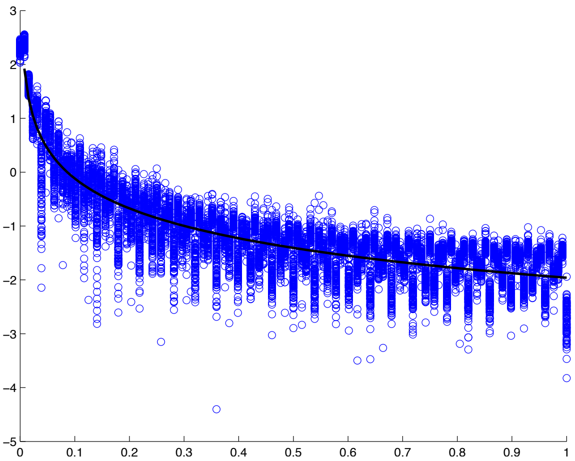


Figure 1: A typical power spectrum of a human face in x-direction (blue) and the fitted power law spectrum (black).

2 Basic Interpretation of Natural Image Statistics

We first distinguish two kind of interpretations of the likelihood functions, namely the **seeker** and the **generator**[*]. In the ordinary Bayesian framework, priors are always interpreted as generators. We instead employ a different interpretation in the following applications. Suppose the resolution is given as $n \times m$:

Definition 1. *An **X-generator** is a probability density function f: $\mathbb{R}^{n\times m} \to \mathbb{R}$ with the interpretation that it generates all the data of X. A dataset containing instances of X with random sampling is an **X-generator** by its nature.*

[*] This distinction is similar to Wittgenstein's "seeing as" v.s. "seeing that" in his famous Philosophical Investigations [48]. However, we employ his term not in the exact sense but only for convienience.

Definition 2. *An **X-seeker** is a computable likelihood function l:* $\mathbb{R}^{n\times m} \to \mathbb{R}$ *with the interpretation that it sees all the data as X across different environments (i.e. with different X-generators). The supervisor labeling X and not-X in the supervised learning is an **X-seeker** by its nature. It is the ideal X-seeker which is the target of approximation.*

Since there is no presupposed “true model”, we shall evaluate an X-seeker with different X-generators and compare its result to the ideal X-seeker. Note there may still be different metrics given an fixed X-generator. Now the classification problem in the supervised learning shall be interpreted as approximating an ideal X-seeker well.

We now give a brief acount for the basic property of natural image statistics. A **wanderer** seeks nothing and sees data **as** image (i.e. giving each data a likelihood of being image), which shall not be equal to the generator which generates data with this given probability. The scale-invariant property can be interpreted as a wanderer gathering images at different scale indifferently. Now the goal of natural image statistics shall be interpreted as constructing a good wanderer which approximates an ideal wanderer well.

In the following chapters, we demonstrate two applications within this framework. Each application contains two different but correlated parts: (1) the identification/ construction of X-seekers; (2) the Bayesian framework links X-seekers to classifiers.

3 Application: Fisher's Discriminant Analysis

3.1 Introduction

Nowadays, dimension reduction techniques have become standard tool-kits in the fields of pattern recognition and computer vision, which stand either directly for the purpose of recognition or indirectly for pre-processing. Two of the most popular and fundamental methods among this ever-growing family, namely Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), both stem directly from the classical works in the field of multivariate statistical analysis. However, the standard assumptions which make the asymptotic behaviours hold often are severely violated in most situations, such as sufficiently large sample size or normality.

One of the most important violation is due to the size of dimensionality (i.e. the input data space) and that of the 'intrinsic' dimensionality (i.e. the dimension of the feature space) are too large in comparison to the size of training data. Recently, Hall et al. [17] coined the term **High Dimension Low Sample Size (HDLSS)** in order to describe this situation, which replace the "large p, small n", an old-fashioned statistical colloquialism (and somewhat a misnomer, as Hall et al. [17] indicated). Here we only mention two specific problems within this setting. The additional assumptions of these two problems are respectively: (1) the data are assumed predominantly lie within a nonlinear manifold, or (2) the data are first drawn from an infnite-dimensional space, say a functional space, with a smoothness assumption and then down-sampled. As in the case of dimension reduction, two different approaches respectively for tackling the two problems aforementioned have been taken independently. One is to change the linear subspace projections toward nonlinear one, and the other approach is to think of the high-dimensional data as a function, in order to exploit its spatial relationship. Later in this chapter, we refer the first approach as nonlinear extension and the second one as functional extension of dimension reduction respectively.

In this chapter, we shall argue that the functional extension may be interpreted and explicitly derived in Bayesian approach when data are modeled as images or sequences of images. We shall first establish a theoretical framework separating the functional extension form the nonlinear extension. The more reasonable and realistic priors can be derived from the natural image statistics which will be referred as **Natural Image Priors** later on. And then we will show that previous method such as Penalized Discriminant Analysis has corresponding prior of intra-class variation. Our proposed method assume the Natural Image Prior for both intra-class variation and inter-class variation. Finally, we demonstrate that functional extension incorporating Natural Image Priors outperforms previous methods in a number of experimental settings. The most important finding is that: (1) the result will be better when prior comforms to the natural image statistics more, and (2) the accuracy of estimation of the inter-class variation may become an overwhelming issue in some datasets.

3.1.1 Related Works

The statistics community have been generally aware of the fact that HDLSS problem arisen everywhere since 1980s. The idea of nonlinear dimension reduction is first introduced as Principal Curves and Surfaces in Hastie [19], though the Multidimensional Scaling (MDS) [52] may be viewed as a forerunner of it and similar ideas emerged from several other areas, including Self-Organizing Map (SOM) [24]. The functional extension has been done in Functional PCA of Ramsay [32], Functional Canonical Correlation Analysis (Functional CCA) of Leurgans et al. [27], and Penalized Discriminant Analysis (PDA) of Hastie et al. [20]. The nonlinear extension and functional extension are by no mean restricted within the dimension reduction setting. For example, Support Vector Machine and in general the kernel learning utilize the fact that the 'intrinsic' dimensionality of data is much lower than the 'extrinsic' one to conquer the "curse of dimensionality". On the other hand, Functional Data Analysis (FDA), first coined by Ramsay and Dalzell [33], covers several other areas such as functional regression. Please see Ramsay and Silverman [34] for detailed studies. The authors have noticed that French statisticians, like Deville [11], had pioneered this area long before the whole statistics community become widely interested in this topic.

Similar considerations arise naturally in computer vision, especially in face recognition and detection. The introduction of Eigenface [44], Fisherface [4], and several

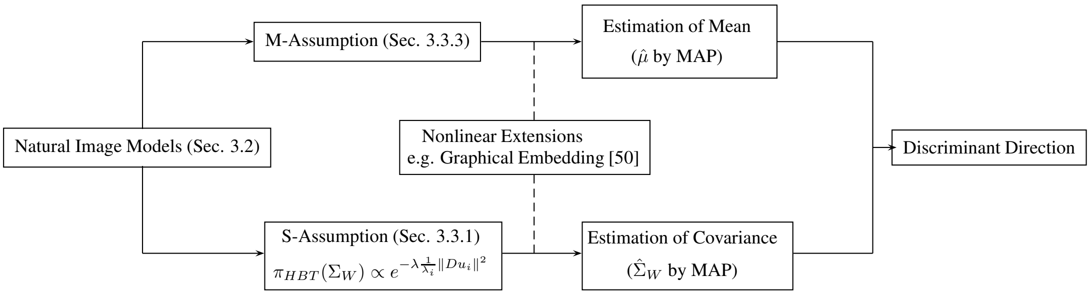

M-Assumption (Sec. 3.3.3)
Estimation of Mean
($\hat{\mu}$ by MAP)
Natural Image Models (Sec. 3.2)
Nonlinear Extensions
e.g. Graphical Embedding [50]
Discriminant Direction
S-Assumption (Sec. 3.3.1)
$\pi_{HBT}(\Sigma_W) \propto e^{-\lambda \frac{1}{\lambda_i} \|Du_i\|^2}$
Estimation of Covariance
($\hat{\Sigma}_W$ by MAP)

Figure 2: Components of Proposed Method

others has made dimension reduction an important role. Both of the same difficulties mentioned above have arisen here, and several ingenious methods have been suggested in response to this situation. The nonlinear extension was first introduced into computer vision from machine learning community. ISOMAP [42], Locally Linear Embedding (LLE) [35], Laplacian Eigenmap [5], Marginal Fisher Analysis (MFA) [50] are some novel and effective methods proposed along this line among several others. On the other hand, the tensor-based approach was employed in order to resolve the second difficulty. An ever-increasing literature contains methods such as 2DPCA [51], 2DLDA, Uncorrelated Multilinear PCA (UMPCA) [29], Uncorrelated Multilinear LDA (UMLDA) [28]. Yen et al. [50] have done a through study and unified all these methods under the Graph Embedding scheme. Since our main focus is on the functional extension, we restrict ourself within the Graph Embedding scheme for the nonlinear extension. We shall address here that this restriction is set up for a basis of comparison and we believe that cooperating functional extension with other method is potentially possible.

3.2 The Models of the Natural Images

We now distinguish two tasks within the HDLSS setting, namely: the identification of the **geometric model** and the **probabilistic model**. By geometric model, we mean that the manifold embedded in the full space on which the data lie and there may also be some re-parametrization through an atlas or a global map. By probabilistic model, we mean the prior probability from which the data were drawn. The geometric model determine **what the subspaces are**, while the probabilistic model determine **how the subspaces correlate**.

We may consider a rigid object under a single light source for simplicity. The geometric model maps $S^2 \times S^1 \times \mathbb{R}^3 \backslash \{0\} \times S^2 \times S^1$ (perspective × rigid motion, where S^n is the n-sphere lying in the $n+1$-dimensional space) into image space of the $n \times m$ resolution. The probabilistic model may set-up the independence between $X_1 \in S^2 \times S^1$ (perspective) and $X_2 \in \mathbb{R}^3 \backslash \{0\} \times S^2 \times S^1$ (rigid motion). It may be plausible to assume further that both X_1 and X_2 are uniform priors which may be improper for the latter one in this case. Both the manifold structure and the likelihood principle are invariant under diffeomorphisms, and thus we may assume everything is fine with re-parametrization, i.e. intrinsic representation and extrinsic representation are equivalent. Since we may constrain the likelihood on any manifold and the likelihood is arbitrary given a manifold, the identifications of these two models function quite independently. Note the estimation of a specific manifold a posteriori is different with the geometric model constraint given a priori. A close parallel distinction for probabilistic model is a well-known fact in the Bayesian approach. For our special purpose, Cai et al. [7] have successfully shown that the Penalized Discriminant Analysis (we will show that it can be derived from a probabilistic model in the next section) is compatible with Graph Embedding framework (a general geometric model). This can be easily extended to our proposed method.

In this chapter, we restrict ourselves to basic statistical properties of natural images, especially the fact that the distribution of derivative is white noise. This property can be derived from power law spectra with order 2 which has already been introduced in Section 1. The power law spectra manifest the most general scale invariance property for many classes of natural images. We first restrict ourselves in one coordinate, says x (the corresponding frequency domain is ω). For the power law, we have in general:

$$P(\omega) \propto \omega^{-\alpha}.$$

Since in most scenarios, α is approximately 2, and thus we have:

$$E\left[\frac{\mathcal{F}\{f\}(\omega)\mathcal{F}\{f\}^*(\omega)}{2\pi}\right] = P(\omega) \propto \omega^{-2},$$

where f is the signal, and which implies:

$$E\left[\frac{df}{dx} * \frac{df}{dx}\right] \propto \delta(x),$$

which means the $\frac{df}{dx}$ is a white noise and we may assume the derivatives with different x are i.i.d..

Above derivation is based upon stationarity. Since the power spectrum density (PSD) calculates the average spectrum, we do not need to worry about the non-stationary processes. Combining the likelihoods of x-coordinate and y-coordinate, we may have the following likelihood assumption:

$$\pi_G(f) \propto \sqrt{\lambda} e^{-\frac{\lambda}{2}\|\nabla f\|^2}, \tag{3.1}$$

where f is a 2-dimensional function. Note that the empirical distribution in the literature fits the generalized Gaussian distribution more. Here we assume the Gaussian distribution only for simplification of computational issues.

3.3 Fisher's Discriminant with Natural Image Priors

Hastie et al. [20] have already suggested the Penalized Discriminant Analysis for adapting to smooth signals, and Cai et al. [7] have successfully shown that their formulation combined with Graph Embedding framework do significantly surpass the tensor-based method on several real datasets. However, they did not justify their formulation within the Bayesian frameworks and explicit probabilistic assumption about the images, and thus missed the link between the image model and the penalty. On the contrary, we find out that the penalty suggested by them deviates significantly from the natural image model. This finding can only be explained in terms of strict derivation with explicit smoothness assumption, and thus we now proceed to introduce the **Fisher's Discriminant with Natural Image Priors**. By Natural Image

Priors, we simply mean the following distribution derived in the last section:

$$\pi_G\left(f\right) \propto \sqrt{\lambda} e^{-\frac{\lambda}{2}\|\nabla f\|^2}, \tag{3.2}$$

where f is a 2-dimensional function. Some modifications will be made when it apply to different parameters, and we will state explicitly when we meet the issue.

Recall that the Fisher's discriminant is based on the 2-class special case in which the discriminant direction is:

$$\Sigma_W^{-1}\left(\mu_1 - \mu_2\right), \tag{3.3}$$

where Σ_W is the common covariance matrix for these two class, μ_1 and μ_1 are the means of the respective classes. Since the parameters of the ordinary Fisher's discriminant are derived based on maximum likelihood estimation, we further have:

$$\ln L\left(\mu_1, \mu_2, \Sigma_W \middle| \left\{y_{i,j}\right\}\right) = -\frac{1}{2} \sum_{i=1,2;j=1,\ldots,n_i} \left(y_{i,j} - \mu_i\right)^T \Sigma_W^{-1}\left(y_{i,j} - \mu_i\right) - \frac{n}{2}\ln|\Sigma_W|, \tag{3.4}$$

$$\hat{\mu}_i = \frac{1}{n_i} \sum_{j=1,\ldots,n_i} y_{i,j}, i = 1, 2, \tag{3.5}$$

$$\hat{\Sigma}_W = \frac{1}{n} \sum_{i=1,2;j=1,\ldots,n_j} \left(y_{i,j} - \hat{\mu}_i\right)^T \left(y_{i,j} - \hat{\mu}_i\right), \tag{3.6}$$

where the log-likelihood is up to a constant and $\left\{y_{i,j}\right\}$ is the sample of size $n = n_1 + n_2$.

When we generalize this procedure for the multi-class case, it is conventional to include a new parameter Σ_{bet} which characterizes the variance of the class means:

$$\Sigma_{Bet} \equiv \sum_{i=1}^{c} \left(\mu_i - \bar{\mu}\right)\left(\mu_i - \bar{\mu}\right)^T, \tag{3.7}$$

where $\bar{\mu} = \frac{1}{c}\sum_{i=1}^{c}\mu_i$. And now the discriminant directions become the spectrum of a generalized eigenproblem:

$$\hat{\beta} \equiv \arg\max_{\beta} \frac{\beta^T \hat{\Sigma}_{bet} \beta}{\beta^T \hat{\Sigma}_W \beta}. \tag{3.8}$$

The above estimation is based on Fisherian paradigm. Given Natural Image Priors, we can estimate parameters in a Bayesian way. Now we restricted ourselves for the MAP estimator for simplicity. The estimation can be done by simply adding log priors to the objective function. Since the discriminant directions are solely determined by decomposition of total variance, i.e. $\hat{\Sigma}_{bet}$ and $\hat{\Sigma}_W$, and $\hat{\Sigma}_{bet}$ is in turn determined by the estimated class means $\hat{\mu}_i$, the Bayesian assumptions of μ_i and Σ_W can be separated from each other. From now on, we shall call the Bayesian assumption of μ_i as **M-Assumption** and that of Σ_W as **S-Assumption**. We will treat them separately and derive the estimator from each assumption. Finally we will evaluate the effect of each of them on several different tasks in the experimental section. Now we first show that the Penalized Discriminant Analysis [20] can be derived from a special S-assumption.

Here we introduce two elemenatray results of matrix calculus as a lemma which will be used in the later derivation:

Lemma 1 (Matrix Calculus)**.**

$$\frac{d(Tr(A\mathbf{X}))}{d\mathbf{X}} = A.$$

$$\frac{d(\ln|\mathbf{X}|)}{d\mathbf{X}} = \mathbf{X}^{-1}.$$

3.3.1 Penalized Discriminant Analysis: a special S-Assumption

We define the Hastie-Buja-Tibshirani prior (abbv. the HBT prior) as follow:

$$\pi_{HBT}(\Sigma_W) \propto e^{-\lambda \sum_i \frac{1}{\lambda_i} \|du_i\|^2}, \tag{3.9}$$

where u_i and λ_i are eigenfuntions and eigenvalues of Σ_W respectively and d is a differential operator which is Δ, the Laplacian operator, in Cai et al. [7]. However, there is no sufficient evidence or rationale why this operator is an appropriate choice.

In our experiment setting, we shall only examine the case for $d = \nabla$, the gradient operator, which better conforms to natural image statistics in comparison to the Laplacian operator proposed in Cai et al. [7]. The following derivation is still valid for the general case. The corresponding maximum a posteriori estimators must satisfy:

$$\min_{\mu_1,\mu_2,\Sigma_W} \sum_{i=1,2} \sum_{j=1,\ldots,n_i} \frac{1}{n} \left(y_{i,j} - \mu_i\right)^T \Sigma_W^{-1} \left(y_{i,j} - \mu_i\right) + \ln|\Sigma_W| + \mathrm{Tr}\left(\lambda \Omega \Sigma_W^{-1}\right), \tag{3.10}$$

where $\Omega = D^T D$ and D is the discretized version of the operator $d = \nabla$.

And hence, by first-order condition, we have:

$$\begin{cases} \sum_{j=1,\ldots,n_i} \Sigma_W^{-1} \left(y_{i,j} - \mu_i\right) = 0, \\ -\Sigma_W^{-1} \left(\sum_{i=1,2;j=1,\ldots,n_i} \frac{1}{n} \left(y_{i,j} - \mu_i\right) \left(y_{i,j} - \mu_i\right)^T \right) \Sigma_W^{-1} + \Sigma_W^{-1} - \Sigma_W^{-1} \lambda \Omega \Sigma_W^{-1} = 0, \end{cases} \tag{3.11}$$

$$\Rightarrow \begin{cases} \mu_i = \sum_{j=1,\ldots,n_i} y_{i,j} = \bar{y}_i, \\ \Sigma_W = \left(\sum_{i=1,2;j=1,\ldots,n_i} \frac{1}{n} \left(y_{i,j} - \bar{y}_i\right) \left(y_{i,j} - \bar{y}_i\right)^T \right) + \lambda \Omega = S_{yy} + \lambda \Omega, \end{cases} \tag{3.12}$$

where $\bar{y}_i$ are the sample class means and S_{yy} is the sample covariance. The above equation implies the discriminant direction is:

$$\hat{\Sigma}_W^{-1}\left(\hat{\mu}_1 - \hat{\mu}_2\right) = \left(S_{yy} + \lambda\Omega\right)^{-1}\left(\bar{y}_1 - \bar{y}_2\right), \tag{3.13}$$

which in turn can be generalized to the multi-class case:

$$\hat{\beta} \equiv \arg\max_{\beta} \frac{\beta^T \hat{\Sigma}_{Bet}\beta}{\beta^T\left(S_{yy} + \lambda\Omega\right)\beta}, \tag{3.14}$$

which is exactly the criterion of Penalized Discriminant Analysis [20].

However, we have found certain evidence in the empirical data standing against the HBT prior assumption. The more relevant principal component deviates less from the typical images, while the HBT prior assumes it is more probable that the less relevant principal component deviates less from the typical images in comparison to the more relevant one.

In fact, spectral decomposition of the within-class variance shall generate certain low-rank results, since the relevant dimensions shall be low in comparison to the very large size of the resolution of images. The HBT prior implicitly assumes the within-class variance is full-rank, and this can only be justified for the sake of regularization, i.e. for the sake of solving the generalized eigenproblem. The problem discussed here is actually beyond the scope of this chapter, we wish to explore the full range of intricacies in the future works.

3.3.2 M-Assumption

The prior of μ with M-Assumption:

$$\pi_M(\mu_i) \propto e^{-\lambda\|\nabla\mu_i\|^2}, \tag{3.15}$$

and the objective function becomes:

$$\min_{\mu_1,\mu_2,\Sigma_W} \sum_{i=1,2} \left[\sum_{j=1,\dots,n_i} \frac{1}{n} \left(y_{i,j} - \mu_i\right)^T \Sigma_W^{-1} \left(y_{i,j} - \mu_i\right) + \frac{\lambda}{n} \|D\mu_i\|^2 \right] + \ln|\Sigma_W|, \tag{3.16}$$

where D is the discretized version of the operator ∇.

Now the solution must satisfy the following first-order conditions:

$$\begin{cases} \sum_{j=1,\dots,n_i} \Sigma_W^{-1} \left(\mu_i - y_{i,j}\right) + \lambda\Omega\mu_i = 0, \\ -\Sigma_W^{-1} \left(\frac{1}{n} \sum_{i=1,2;j=1,\dots,n_i} \left(y_{i,j} - \mu_i\right)\left(y_{i,j} - \mu_i\right)^T \right) \Sigma_W^{-1} + \Sigma_W^{-1} = 0, \end{cases} \tag{3.17}$$

$$\Rightarrow \begin{cases} \left(\Sigma_W^{-1} + \frac{\lambda}{n_i}\Omega\right)\mu_i = \Sigma_W^{-1}\left(\frac{1}{n_i}\sum_{j=1,\dots,n_i} y_{i,j}\right) = \Sigma_W^{-1}\bar{y}_i, \\ \Sigma_W = \frac{1}{n}\sum_{i=1,2;j=1,\dots,n_i}\left(y_{i,j} - \mu_i\right)\left(y_{i,j} - \mu_i\right)^T = S_{yy} + \sum_{i=1,2}\frac{n_i}{n}\left(\bar{y}_i - \mu_i\right)\left(\bar{y}_i - \mu_i\right)^T, \end{cases} \tag{3.18}$$

where $\Omega = D^T D$, $\bar{y}_i$ are the sample class means and S_{yy} is the sample covariance.

3.3.3 Proposed method: approximated M-Assumption + S-Assumption

We now combine the results of the previous subsections in order to suggest our method. Since solving the system (3.18) requires iteration, we approximate the result with a single iteration for reducing computation time. By the S-assumption, we have:

$$\hat{\Sigma}_W = S_{yy} + \lambda_1\Omega. \tag{3.19}$$

After substituting the first order conditions (3.18) based on M-Assumption, we have:

$$\hat{\mu}_i = \left(I + (S_{yy} + \lambda_1 \Omega) \frac{\lambda_2}{n_i} \Omega \right)^{-1} \bar{y}_i. \tag{3.20}$$

Finally, we estimate Σ_{Bet} as in ordinary Fisher's discriminant:

$$\hat{\Sigma}_{Bet} \equiv \sum_{i=1}^{c} \left(\hat{\mu}_i - \bar{\hat{\mu}}\right) \left(\hat{\mu}_i - \bar{\hat{\mu}}\right)^T, \tag{3.21}$$

and the discriminant directions are the spectra the generalized eigenproblem of $\hat{\Sigma}_{Bet}$ and $\hat{\Sigma}_W$.

3.4 Experimental Results

In this section, we demonstrate the performances of the dimension reduction techniques on two different tasks, namely: face recognition and object categorization. The Yale database [25] is used for the face recognition task, which contains 165 gray scale images of 15 individuals, each of 11 images with different lighting conditions and facial expressions (normal, happy, sad, sleepy, surprised, and wink). The ETH-80 [26] dataset is used for the object categorization, which contains 3280 images of 8 categories. Each category contains 10 different objects with 41 views per object. All the images in both datasets are well aligned and cropped. Each cropped image is resized to 32 × 32 pixels, with 256 gray levels per pixel. We rescale the pixel values to $[0, 1]$ (divided by 255). Sample images of Yale and ETH-80 database are shown in Fig. 3 and 4.

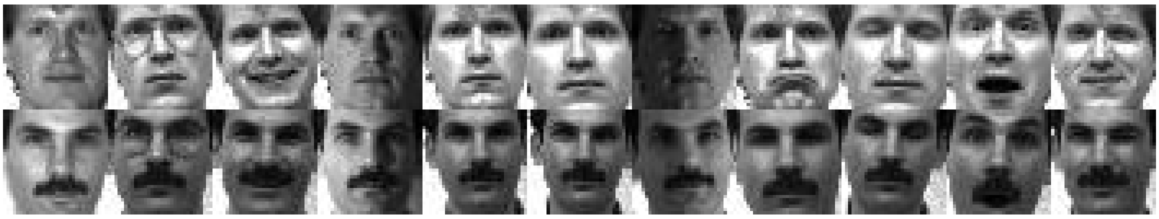


Figure 3: Some sample images from Yale database

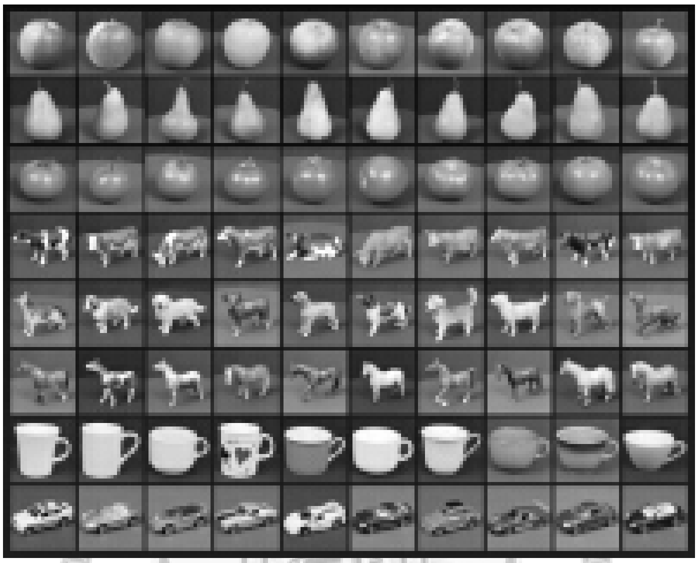


Figure 4: Some sample images from ETH-80 data set

Each image set is then partitioned into the gallery (training) and probe (testing) set with different numbers. For the Yale database, m images per person are randomly selected for training, and the remaining images are for testing. For the ETH-80 dataset, m images per *category* are randomly selected for training, and the remaining images are for testing. Notice that in our setting the training set may not contain all the 10 objects for each category.

We compare our MS-LDA method (Section 3.3.3) with the following algorithms:

Fisherface [4] as the baseline, UMLDA [28], and Cai's Spatially Smooth LDA [7] which uses Laplacian smoothing term (S-LDA-L hereafter). We also replace the Laplacian smoothing term with the Natural Image Prior (S-LDA-N hereafter) to see if the Natural Image Prior (1) is more appropriate. The codes for Fisherface and Cai's S-LDA-L are download from Cai's website. The parameters are selected by cross-validation for all the algorithms.

3.4.1 Face Recognition

The recognition accuracy of different algorithms on Yale database are listed in Table 1. Notice that the performance for Cai's S-LDA-L reported here is better than the result reported in the original paper [7]. From Table 1 we can observe that both S-LDA-L and the S-LDA-N significantly outperform Fisherface, which indicates that S-Assumption (Section 3.3.1) is important. The accuracy of S-LDA-N is better than S-LDA-L method, which shows that proper regularization is also important. Finally, the proposed MS-LDA method generally outperforms both S-LDA-L and S-LDA-N, which reveals that not only the S-assumption but also the M-assumption are important.

Gallery Size	2	3	4	5	6	7	8
Fisherface	0.5556	0.6618	0.7225	0.7620	0.7939	0.8027	0.8169
UMLDA	0.5778	0.6083	0.6762	0.7000	0.7667	0.7867	0.8000
S-LDA-L	0.5951	0.7178	0.7718	0.8238	0.8597	0.8710	0.8942
S-LDA-N	0.6116	0.7378	0.7898	0.8389	0.8657	0.8890	0.9004
MS-LDA	**0.6511**	**0.7663**	**0.8177**	**0.8576**	**0.8845**	**0.8957**	**0.9196**

Table 1: Average recognition rates (%) of Yale database.

3.4.2 Object Categorization

The recognition accuracy of different algorithms on ETH-80 dataset are listed in Table 2. Both S-LDA-L and S-LDA-N significantly outperform Fisherface. Although UMLDA performs better than S-LDA-L and S-LDA-N on this dataset, the proposed MS-LDA method still significantly outperforms all the other methods. Despite the fact that both S-LDA-L and S-LDA-N utilize the general prior structure while UMLDA constraining itself on tensor structure which seems unrealistic, UMLDA does account some effects of image structure for the inter-class variation (i.e. something parallel the M-assumption). The results demonstrate that for some datasets the effect of the inter-class variation Σ_{Bet} may overwhelm the intra-class variation Σ_W, and the proposed MS-LDA method still works very well in such case.

Gallery Size	Fisherface	UMLDA	S-LDA-L	S-LDA-N	MS-LDA
20	0.5192	0.6081	0.5630	0.6111	**0.6968**
100	0.5875	0.7405	0.7144	0.7187	**0.8039**

Table 2: Average recognition rates (%) of ETH-80 dataset.

3.5 Concluding Remarks

In this chapter, we introduce a Bayesian framework for Fisher's discriminant, and propose the Natural Image Priors derived form natural image statistics. We apply the Natural Image Priors to image class mean (M-assumption) and the within-class variation (S-assumption), and derive the novel MS-LDA method which is a maximum a posteriori probability (MAP) estimate. Previous work that utilizes the spatial smoothness property of images can be shown as special cases of this framework, but with only the S-assumption and with different prior. Experimental results on the Yale face database and the ETH-80 object categorization dataset show that the proposed natural image prior is indeed superior to the prior used in previous method, and the M-assumption are shown to be especially important in ETH-80 object categorization dataset. Finally, our Bayesian framework can be easily combined with the extensions of Fisher's discriminant as well, such as multilinear LDA, MFA, etc. Due to the algorithmic bottleneck, several approximations are made during the derivation in order to save the computation time. We wish to investigate new algorithms to overcome the time issue without approximations in the future.

4 Application: Illumination Invariant Feature Extraction

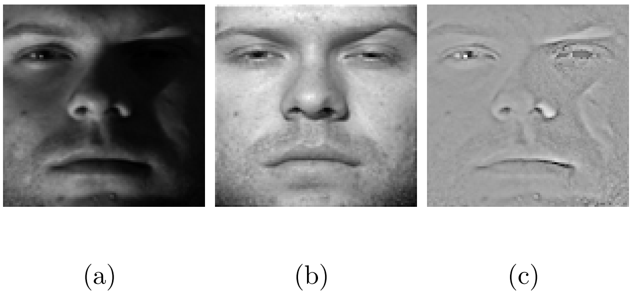
4.1 Introduction

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

4.1.1 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 [6]. Despite its nice performance, collecting a series of 3D face scans is usually not feasible in practice. Analyses of illumination cone [13] or spherical harmonics [3][31] 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. [25] and Chen and Chen [8]. 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. [40] 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 [18] or estimates the generative model [13]; and (2) the *illumination-*



(a) (b) (c)

Figure 5: (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.

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 [9].

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 5.

We now turn to describe the idea development of the *illumination-invariant fea-*

ture extraction briefly. To the best of our knowledge, nearly every method in variable lighting face recognition, following the initiative work of Horn [22], 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 [31], spherical harmonics [25][8], wavelets [53], and cosine functions [10]). Others derive some low-pass filter fulfilling certain constrains directly [16][46]. The most recent methods perform cartoon decomposition and extract the large scale part [9][49] 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. [49] started to investigate seriously the possibility that part of the relevant features may lie in the low frequency part[†] 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

[†] 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.

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 6. The rest of the paper is organized as follows: In Section 4.2 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 4.3, and describe the experimental evaluation of our assumption

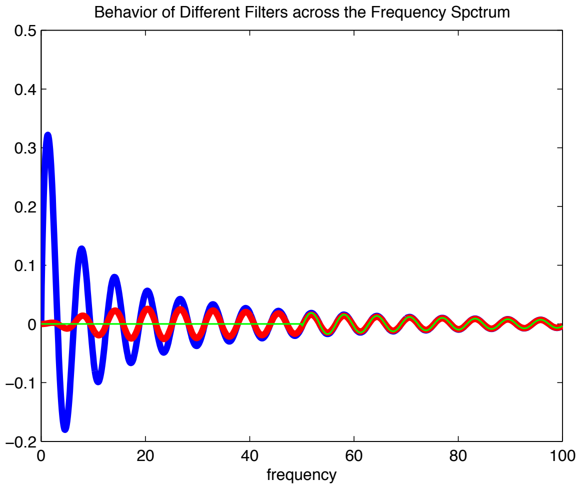


Figure 6: The effects of Wiener filtering (red) and directly thresholding method with the threshold at frequency 50 (green) in comparison with the original spectrum (blue). It can be shown that both directly thresholding method and Wiener filtering preserve features at the high frequency part, while 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.

in Section 4.4. Finally, conclusions and future directions are given in Section 4.5.

4.2 Wiener Filtering for Natural Illumination

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

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

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 [22] suggested to take the logarithm of I and transforms the model into an additive 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 auto-correlations 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 [36], 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 chapter. 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:

$$\begin{aligned} P_{\mu}(\omega) &\propto \omega^{-\alpha_{\mu}}, \quad (4.1) \\ P_{\nu}(\omega) &\propto \omega^{-\alpha_{\nu}}, \quad (4.2) \end{aligned}$$

where P_{μ}, P_{ν} are power spectrum densities of μ and ν, respectively, and $\alpha_{\mu}, \alpha_{\nu}$ 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:

$$\begin{aligned} \mathcal{F}\{l\}(\omega) &= \frac{P_{\mu}(\omega)}{P_{\mu}(\omega) + P_{\nu}(\omega)} \quad (4.3) \\ &= \frac{\lambda}{\lambda + \omega^{\delta}}, \quad (4.4) \end{aligned}$$

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 (4.4) the *natural delighting filters (NDFs)*, which are varying with δ. In particular, we shall show later in the Section 4.3

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:

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

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

$$(\lambda I + D^T D)\boldsymbol{\mu} = \lambda \boldsymbol{f}, \tag{4.5}$$

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 (4.5) is equivalent to a linear smoothing spline [45] that is easy to compute. We can boost the computation speed by using the analytic solutions introduced in [14] and reduce the computation complexity to exactly $4n$ flops. Solving equation (4.5) along x- and y-directions consecutively is referred to as the NDF of $\delta = 2$ in this chapter. 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 chapter:

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

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.

4.3 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 [21]. 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.

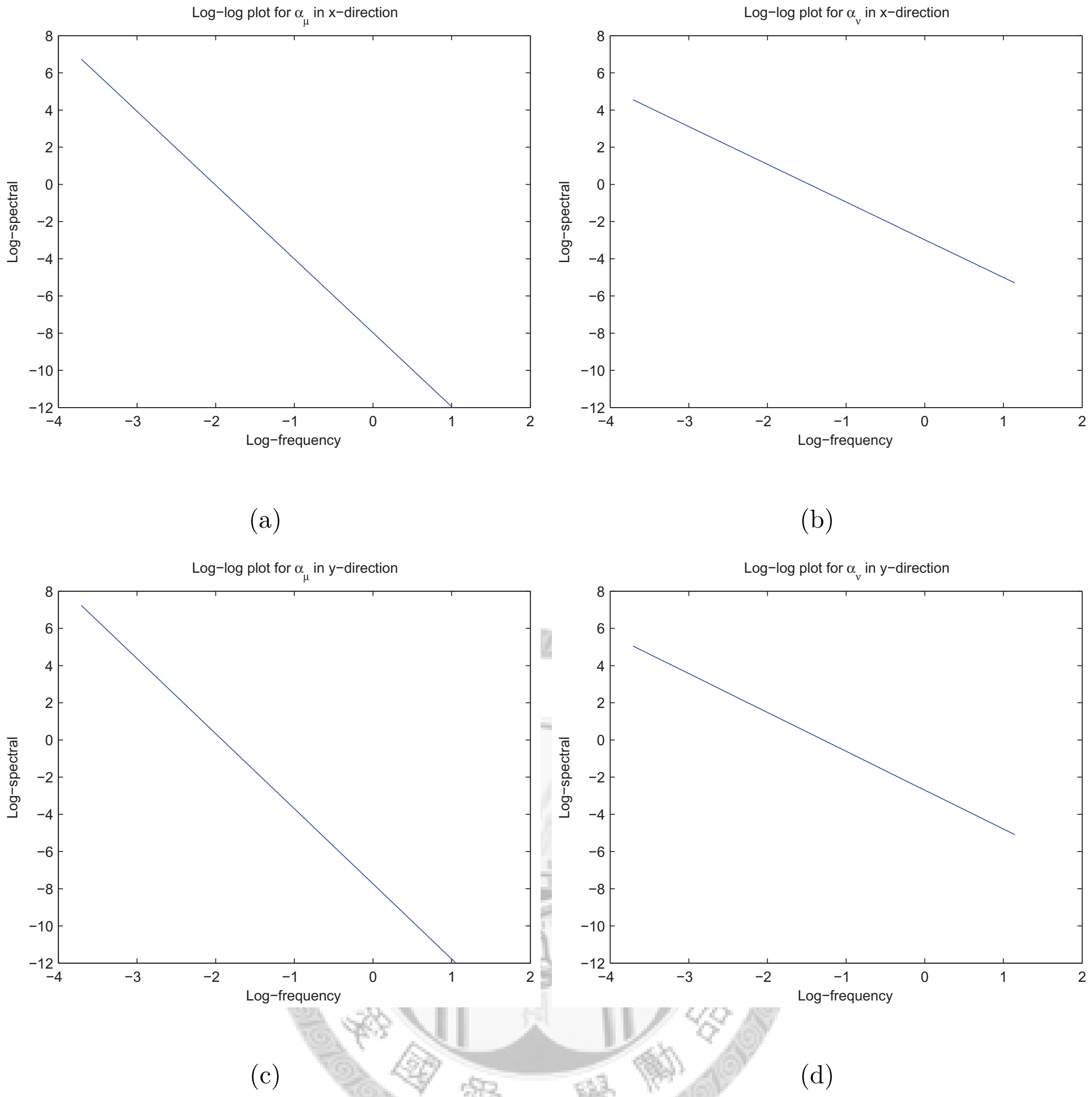
Since we assume that P_μ and P_ν follow the power law as in 4.1, by taking logarithm on P_μ and P_ν we have

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

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 7 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.

4.4 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 [46], LTV [9], RLS [49], and TT [41], 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 [9] and TT [41] for the implementation, and use the same parameters as reported in [9] and [41]. The SQI method [46] has no available codes, and so we implement it on our own. The RLS method [49] has also no available codes, and we directly use their recognition rates reported in [49] for comparison. As mentioned in Section 4.2, 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.



(a)

(b)

(c)

(d)

Figure 7: 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.

We use CMU PIE [38], Extended Yale B [25], and CMU Multi-PIE [15] 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 10 humans in 64 different lighting conditions (divided into 5 subsets according to [25]), 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 [18], which follows the standard testing procedure described by [15] 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 [46] [8][9][49], 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 [41] which uses methods such as distance transformation and kernel LDA for recognition, we employ the preprocessing and feature extraction steps of the TT method [41], and only change the recognition step to CC in our implementation for fair comparison.

	SQI [46]	LTV [9]	RLS [49]	TT [41]	NDF
Ideal	97.19	100	99.9	100	100
Average	95.19	99.35	–	100	100

Table 3: Recognition rate (%) of CMU PIE face database

[**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 [49]. Note that the recognition rates obtained for LTV and SQI of our implementation are consistent to those reported in [9] and [46] 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 4 shows the recognition rates of the ideal case when the ideal images (frontal lighting images) are used as the reference images, and table 5 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 [49]. Our proposed model achieves much higher recognition rates than all the other methods under all lighting conditions for this more challenging dataset.

Note that we employed the pre-processing and feature extraction of the TT method [41] but changed the recognition step to CC. Compared to our implementation, the recognition rate reported in [41] 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 [41] 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

Subset	SQI [46]	LTV [9]	RLS [49]	TT [41]	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 4: Recognition rate (%) of Extended Yale B face database when the ideal images are used as reference images.

Subset	SQI [46]	LTV [9]	TT [41]	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 5: Average recognition rates (%) of Extended Yale B face database.

SQI [46]	LTV [9]	TT [41]	NDF
98.85	99.56	100	100

Table 6: Average recognition rate (%) of CMU Multi-PIE face database [18], which includes all the 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.

4.5 Conclusion and Future Works

This chapter 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 chapter, 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 [10] and Intrinsic Subspace [8]). 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 chapter 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 [43]), 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 chapter, 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.

5 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*, 3:3–26, 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] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711 –720, 1997.

[5] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2002.

[6] 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.

[7] D. Cai, X. He, Y. Hu, J. Han, and T. Huang. Learning a spatially smooth subspace for face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

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

[9] 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.

[10] W. Chen, Meng Joo Er, and Shiqian 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.

[11] J.-C. Deville. Méthodes statistiques et numériques de l'analyse harmonique. *Annales de l'INSEE*, 15:3–101, 1974.

[12] R. O. Dror, T. K. Leung, E. H. Adelson, and A. S. Willsky. Statistics of real-world illumination. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001.

[13] A.S. Georghiades, P.N. Belhumeur, and D.J. 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.

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

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

[16] Ralph Gross and Vladimir 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.

[17] P. Hall, J. S. Marron, and A. Neeman. Geometric representation of high dimension, low sample size data. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 67(3):427–444, 2005.

[18] Hu Han, Shiguang Shan, Laiyun Qing, Xilin Chen, and Wen Gao. Lighting aware preprocessing for face recognition across varying illumination. In *ECCV*, 2010.

[19] T. Hastie. *Principal curves and surfaces.* PhD thesis, Stanford University, 1984.

[20] T. Hastie, A. Buja, and R. Tibshirani. Penalized discriminant analysis. *Annals of Statistics*, 23(1):73–102, 1995.

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

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

[23] J. Huang, A.B. Lee, and D. Mumford. Statistics of range images. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2000.

[24] T. Kohonen. *Self-organization and associative memory*, volume 8 of *Springer Series in Information Sciences.* Springer-Verlag, Berlin, third edition, 1989.

[25] K.C. 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.

[26] B. Leibe and B. Schiele. Analyzing appearance and contour based methods for object categorization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2003.

[27] S. E. Leurgans, R. A. Moyeed, and B. W. Silverman. Canonical correlation analysis when the data are curves. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 55(3):725–740, 1993.

[28] H. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos. Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition. *IEEE Transactions on Neural Networks*, 20(1):103 –123, 2009.

[29] H. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos. Uncorrelated multilinear principal component analysis for unsupervised multilinear subspace learning. *IEEE Transactions on Neural Networks*, 20(11):1820 –1836, 2009.

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

[31] 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.

[32] J. O. Ramsay. When the data are functions. *Psychometrika*, 47(4):379–396, 1982.

[33] J. O. Ramsay and C. J. Dalzell. Some tools for functional data analysis. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 53(3):539–572, 1991. With discussion and a reply by the authors.

[34] J. O. Ramsay and B. W. Silverman. *Functional data analysis.* Springer Series in Statistics. Springer, New York, second edition, 2005.

[35] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

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

[37] D. L. Ruderman and W. Bialek. Statistics of natural images: Scaling in the woods. *Physical Review Letters*, 73(6):814–817, 1994.

[38] 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.

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

[40] Xiaoyang Tan, Songcan Chen, Zhi-Hua Zhou, and Fuyan Zhang. Face recognition from a single image per person: A survey. *Pattern Recognition*, 39(9):1725 – 1745, 2006.

[41] Xiaoyang Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *Image Processing, IEEE Transactions on*, 19(6): 1635 –1650, 2010.

[42] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

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

[44] M.A. Turk and A.P. Pentland. Face recognition using eigenfaces. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1991.

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

[46] Haitao Wang, S.Z. Li, and Yangsheng 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.

[47] Y. Weiss. Deriving intrinsic images from image sequences. In *IEEE International Conference on Computer Vision (ICCV)*, 2001.

[48] L. Wittgenstein. *Philosophical Investigations.* Oxford: Blackwell, 1953.

[49] X. Xie, W. S. Zheng, J. Lai, and P. C. Yuen. Face illumination normalization on large and small scale features. In *CVPR*, 2008.

[50] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1):40–51, 2007.

[51] J. Yang, D. Zhang, A.F. Frangi, and J.-Y. Yang. Two-dimensional pca: a new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):131 –137, 2004.

[52] G. Young and A. Householder. A note on multidimensional psychophysical analysis. *Psychometrika*, 6(5):331–333, 1941.

[53] Taiping Zhang, Bin Fang, Yuan Yuan, Yuan Yan Tang, Zhaowei Shang, Donghui Li, and Fangnian Lang. Multiscale facial structure representation for face recognition under varying illumination. *Pattern Recognition*, 42(2):251–258, 2009.