

# Total Variation Models for Variable Lighting Face Recognition

Terrence Chen<sup>†</sup> Wotao Yin<sup>\*</sup> Xiang Sean Zhou<sup>‡</sup> Dorin Comaniciu<sup>‡</sup> Thomas S. Huang<sup>†</sup>

<sup>†</sup>University of Illinois at Urbana Champaign  
405 N. Mathews Ave., Urbana, IL 61801

<sup>\*</sup>Columbia University  
New York City, NY 10027

<sup>‡</sup>Siemens Corporate Research  
755 College Road East, Princeton, NJ 08540

## Abstract

In this paper, we present the logarithmic total variation (LTV) model for face recognition under varying illumination, including natural lighting condition, where we can hardly know the strength, the directions, and the number of light sources. The proposed LTV model has the capability to factorize a single face image and obtain the illumination invariant facial structure, which is then used for face recognition. The merit of this model is that neither does it require any lighting assumption nor does it need any training process. Besides, there is only one parameter which could be easily set. The LTV model is able to reach very high recognition rates on both Yale and CMU PIE face databases as well as on a face database containing 765 subjects under outdoor lighting conditions.

**Keywords:** I.5.4.d Face and gesture recognition ; I.5.4.m Signal processing; I.4 Image Processing and Computer Vision; I.5.2.c Pattern analysis;

## 1 Introduction

Illumination normalization is an important task in the field of computer vision and pattern recognition. One of its most important applications is face recognition under varying illumination. It has been proven both experimentally [1] and theoretically [47] that in face recognition the differences caused by varying illumination are more significant than the inherent differences between individuals. Various methods have been proposed for face recognition, including Eigenface [43], Fisherface[5], Probabilistic and Bayesian Matching [26], subspace LDA [48], Active Shape Model and Active Appearance Model [23], LFA[28], EBGM[45], and SVM[17]. Nevertheless, the performances of most existing algorithms are highly sensitive to the illumination variation. To attack

the problem of face recognition under varying illumination, several methods have been proposed. The predominant ones include the Illumination Cone methods [6][14], spherical harmonic based representations [31] [4] [46], quotient image based approaches [36] [35] [44], and correlation filter based method [34]. However, not only are the performances of most of them still far from ideal, many of these methods require either knowledge of the light source or a large number of training data, which are not practical for most real world scenarios. Let's take some of the most recent methods as examples: Lee et al.'s *nine points of light* [24] method needs perfect alignment between different images, Savvides et al.'s *Corefaces* [34] needs several training images to reach perfect results, and the recognition rate of Wang et al.'s *self quotient image* [44] still has room for improvement.

In addition to methods designed for face recognition, there have been methods developed to remove lighting effects on general images. Most generally, an image  $I(x, y)$  is regarded as a product of reflectance  $R$  and the illuminance effect  $L$  [19]. Getting  $R$  from an input image  $I$  is regarded as an ill-posed problem [32]. Assuming  $L$  changes slowly compared to  $R$ , homomorphic filtering [39] separates slow and fast changes by applying high-pass filter on the log of the image. Horn et al. [18] took Laplacian of the log of the image to remove  $L$ . However, the assumption is not true for images under natural lighting conditions, where shadow boundaries may create abrupt changes in  $L$ , and hence these methods would create "halo" artifacts.

Similarly, Land's "Retinex" model [22] estimated the reflectance  $R$  as the ratio of the image  $I$  using the low pass estimator  $L$ . Jobson et al. [20] reduced the "halo" artifacts by combining several low-pass copies as the estimation of  $L$ . To reduce the "halo" artifacts, discontinuity preserving filtering can be used to estimate  $L$ , such as anisotropic diffusion [30], bilateral filtering[41], or mean shift filtering [11]. Relevant works include LCIS by Tumblin et al. [42] using anisotropic diffusion, Durand et al. [12] using bilateral filtering, and perceptually adjusted weighted least squares within a variational framework by Brajovic [7, 8]. Kimmel et al. [21] and Elad et al. [13] provided good reviews of Retinex and related illumination compensation methods. Some face recognition results of relevant works were reported in [16, 38]. These works have reduced the halo artifacts a lot though not entirely. However, the parameter selection of these models are mostly empirical and complicated, and/or the number of parameters can be as large as around eight [42].

In this paper, we propose a novel model utilizing the  $TV+L^1$  model [9]to factorize an image,

which has some unique advantages compared to existing solutions, especially the simple parameter selection. The advantages can be seen from our original analysis of the  $TV+L^1$  model in section 2.3 and the experimental evaluation in section 3.

## 2 Methodology

In this section, we introduce the Logarithmic Total Variation (LTV) model and explain the way it removes varying illumination for face images. We begin with the discussion about the reflectance model, followed by the analysis of the  $TV+L^1$  model, which serves as the basis of the proposed model. Finally, we propose the LTV model and discuss the choice of its parameter.

### 2.1 Reflectance models

From our observation, the shapes, the contours, and the relative positions of small-scale facial objects (e.g., eyes, noses, mouths, eyebrows) can be key features for face recognition. The surface albedos on or in the boundaries of lower nose, mouth, eyes, eyebrows, and chins are often different from the albedos of the large-scale skin areas and background due to different textures (lips, eyeballs, eyebrow hair) and geometries (nares). Hence, similar to  $I = RL$  and based on the Lambertian model, to obtain these albedos, we propose to solve the following equation for surface under any lighting conditions, including the natural ones:

$$I(x, y) = \rho(x, y)S(x, y) \quad (1)$$

where  $I(x, y)$  is the intensity of a 2D surface image at location  $(x, y)$ ,  $\rho$  is the albedo and  $S$  is the final light **received** at location  $(x, y)$  that generates the observed intensity. Compared to the Lambertian surface,  $S(x, y)$  equals  $A \cos \theta$  in the Lambertian model, where  $A$  is the strength of the light source, and  $\theta$  is the angle between the light source direction and the surface normal. That is, no matter what kinds of and how many light sources there are, the intensity of each location reflects the strength of the light it receives and is with a multiplicative relationship. Obtaining  $\rho$  and  $S$  by solving (1) does not give complete surface information, but for illumination invariant face recognition, we are only interested in the variation pattern of albedos  $\rho$  in an input face image  $I$ . Hence, the problem is simplified to how we can retrieve the variation pattern of  $\rho$  from a given

surface  $I$  with possibly varying intensity. Next, we analyze the  $\text{TV}+L^1$  model used to obtain the  $\rho$  of small-scale facial features in our LTV model.

## 2.2 The $\text{TV}+L^1$ model

In the  $\text{TV}+L^1$  model [3, 27, 9], an input image  $f$  is decomposed into large-scale output  $u$  and small-scale output  $v$ , where  $f$ ,  $u$  and  $v$  are functions in  $\mathbb{R}^2$ .  $u$  contains background hues and important boundaries as sharp edges.  $v$ , which is the rest of the image, is characterized by small-scale patterns. Since the level sets of large-scale signal  $u$  have simpler boundaries than those of small-scale signal  $v$ , we can obtain  $u$  from  $f$  by solving a variational regularization problem:

$$\min \int |\nabla u| + \lambda \|f - u\|_{L^1}, \quad (2)$$

where regularizing term  $\int |\nabla u|$  is the total variation (TV) of  $u$  and  $\lambda$  is a scalar threshold on scale. Let  $u_\lambda$  denote the optimal solution of (2). There are two approaches used for solving (2). PDE [9] solves for  $u_\lambda$  as a solution of the Euler-Lagrange equation of (2), which is essentially its first-order optimality condition:

$$\nabla \cdot \left( \frac{\nabla u}{|\nabla u|} \right) + \lambda \frac{f - u}{|f - u|} = 0. \quad (3)$$

Artificial time evolution iterations can approximately reach the steady state of the above heat PDE. This approach is easy to implement and requires little amount of memory. However, since the second term  $(f - u)/|f - u|$  is non-smooth, time step  $\delta t$  must be very small when the system is near its steady state. This causes numerical difficulties. This problem can be avoided by using a more direct approach [15] which casts (2) as a second-order cone program (SOCP) and solves it using modern interior-point methods [2]. The SOCP algorithm achieves better accuracy but requires more memory.

## 2.3 Edge-preserving and scale-dependent additive signal decomposition

In this section, we analyze the properties of the  $\text{TV}+L^1$  model for the purpose of edge-preserving and scale-dependent additive signal decomposition and provide theoretical justification for our proposed application. Just like many other TV-based models (e.g., the Rudin-Osher-Fatemi model [33]), the  $\text{TV}+L^1$  model keeps the sharp object edges in  $u$  [40]. This property is very important in illumination normalization as you can see in section 3 that the sharp boundaries of the shadows

cast on faces are entirely kept in  $u$  and therefore, their appearance does not affect the recognition process, which is based on  $v$ . What distinguishes the  $\text{TV}+L^1$  model from other TV-based models is its unique property of scale-dependent, but intensity-independent, decomposition. Using different  $\lambda$ s when applying  $\text{TV}+L^1$  to input signals  $f$ , we get  $u$ 's that only contain signals whose scales are larger than  $2/\lambda$ . In addition, all the jumps of the kept signals are exactly preserved in  $u$ . Next, we illustrate the scale-dependency and intensity-independency properties using simple 2D examples.

- Suppose  $f = c_0 + c_1 1_{B_r(y)}(x)$  (i.e.,  $f$  is a function which equals  $c_0 + c_1$  in the disk centered at  $y$  and with radius  $r$  and equals  $c_0$  anywhere else).

$$u_\lambda = \begin{cases} c_0 & 0 \leq \lambda < \frac{2}{r}, \\ \{c_0 + s 1_{B_r(y)}(x) : 0 \leq s \leq c_1\} & \lambda = \frac{2}{r}, \\ c_0 + c_1 1_{B_r(y)}(x) \equiv f & \lambda > \frac{2}{r}. \end{cases} \quad (4)$$

A proof can be found in [9, 10]. We note that  $\lambda$ , the parameter which determines whether  $u_\lambda$  contains  $c_1 1_{B_r(y)}(x)$  or not, depends only on the disk radius  $r$  but not on the values  $c_0$  and  $c_1$  and the disk center  $y$ . When  $\lambda = 2/r$ ,  $\text{TV}+L^1$  has multiple optimal solutions, but in general, the solutions are not unique for at most countably many  $\lambda$ s [9] (i.e., taking 0 Lebesgue measure). Therefore, we omit these values in the forthcoming analysis and in the numerical tests. This property can be extended to:

- Suppose  $f = c_0 + c_1 1_{B_{r_1}(y_1)}(x) + c_2 1_{B_{r_2}(y_2)}(x)$ , where  $0 < r_2 < r_1$  and  $c_1, c_2 > 0$  and  $B_{r_2}(y_2) \subset B_{r_1}(y_1)$ , then

$$u_\lambda = \begin{cases} c_0 & 0 < \lambda < \frac{2}{r_1}, \\ c_0 + c_1 1_{B_{r_1}(y_1)}(x) & \frac{2}{r_1} < \lambda < \frac{2}{r_2}, \\ f & \lambda > \frac{2}{r_2}. \end{cases} \quad (5)$$

Figure 1 (A) illustrates this property and is proved by us in [10]. In 2D images, the scale of a simple disk signal is defined as its radius  $r$  divided by 2. To extend the property to the facial feature signals with non-regular shapes and varying intensity, we would expect a generalized scale measure that is consistent with the  $1/\lambda$  scale threshold of  $\text{TV}+L^1$ . The  $G$ -norm of the  $G$  space is the answer.

**Definition 2.1** [25] *Let space  $G$  denote the Banach space consisting of all generalized functions  $v(x)$  defined on  $\mathbb{R}^n$  which can be written as*

$$v = \text{div}(\vec{g}), \quad \vec{g} = (g_1, \dots, g_n) \in (C_0^1)^n. \quad (6)$$

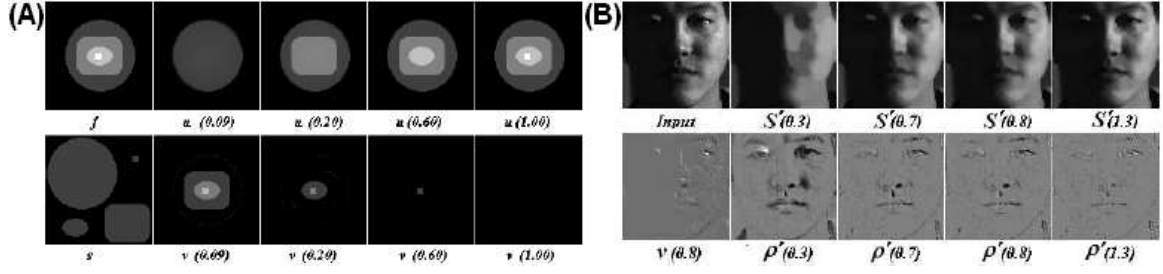


Figure 1: (A) Additive signals with one included in the other can be extracted one by one using increasing  $\lambda$ .  $s$  separately shows the four components piled in  $f$ . (B) Input image and  $v$ : the result of the  $\text{TV}+L^1$  model.  $\rho'$  and  $S'$ : the results of the LTV model with the  $\lambda$  used in parentheses.

Its norm  $\|v\|_G$  is defined as the infimum of all  $L^\infty$  norms of  $\|\vec{g}(x)\|_{l^2}$  and the infimum is computed over all decompositions (6) of  $v$ . In short, supposing the infimum is attained at  $\vec{g}^*$ , we can write

$$\vec{g}^* := \operatorname{arginf}\{\|\vec{g}(x)\|_{l^2} \|_{L^\infty} : v = \operatorname{div}(\vec{g})\},$$

$$\|v\|_G = \|\vec{g}^*(x)\|_{l^2} \|_{L^\infty}.$$

In the previous examples, the  $G$ -norms of 1D signal  $f = 1_{[0,r]}$  and 2D disk signal  $f = 1_{B_r(0)}$  are both  $r/2$  with

$$\text{1D: } g^*(x) = \int_0^x f(t) dt - r/2,$$

and [25]

$$\text{2D: } \vec{g}^* = (x_1 \omega(|x|), x_2 \omega(|x|)),$$

where  $\omega(t) = 1/2$  if  $0 \leq t \leq r$  and  $\omega(t) = r^2/(2t^2)$  if  $t > r$ . The  $G$ -norms of the other signals in the previous examples can be easily derived from these two.

The connection between the  $\text{TV}+L^1$  model and  $G$ -norm is given in the following theorem [10, 29]:

**Theorem 2.2** Let  $f \in L^1(\Omega)$ , where  $\Omega$  is bounded and contains the support of  $f$ , be the input and  $u_{\lambda,\epsilon}$  ( $v_{\lambda,\epsilon} = f - u_{\lambda,\epsilon}$ ) be the unique solution of the approximate  $\text{TV}+L^1$  model

$$\min \int_{\Omega} |\nabla u| + \lambda \sqrt{(f-u)^2 + \epsilon} dx.$$

According, let  $\operatorname{sign}_\epsilon(v)(x) := v(x)/\sqrt{|v(x)|^2 + \epsilon}$  denote the approximate signum function. Then,

$$u_{\lambda,\epsilon} \equiv 0 \text{ if } \|\operatorname{sign}_\epsilon(f)\|_G \leq 1/\lambda,$$

and

$$\|\text{sign}_\epsilon(v_{\lambda,\epsilon})\|_G = 1/\lambda \text{ if } \|\text{sign}_\epsilon(f)\|_G > 1/\lambda.$$

Let  $u_\lambda$  denote the unique solution of the  $TV+L^1$  model. Then

$$\lim_{\epsilon \downarrow 0+} \|u_{\lambda,\epsilon} - u_\lambda\|_{L^1} = 0, \quad \lim_{\epsilon \downarrow 0+} \|v_{\lambda,\epsilon} - v_\lambda\|_{L^1} = 0.$$

The reason that we use the solutions of approximate  $TV+L^1$  to derive the  $G$ -norm based properties and let the solution sequence converge to the  $TV+L^1$  solution is that the  $L^1$ -norm functional is non-differentiable and its derivative is a set-valued function. The main point of this theorem is that  $TV+L^1$  decomposes  $f$  into  $u$  and  $v$  using  $1/\lambda$  as a threshold on  $\|\text{sign}_\epsilon(v)\|_G$ . The previous examples on simple signals demonstrate this. To see that smaller scale (or stronger oscillation) functions have smaller  $G$ -norms, consider  $v(x) = \cos(nx)$  and  $g^* = \text{sign}(nx)/n$ .  $\|v\|_G = 1/n$  decreases when the oscillation of  $v$  increases (i.e.,  $n$  increases). Another example [25] is, for any  $f \in L^\infty(\mathbb{R}^2)$ ,  $\|\exp(i\omega x)f(x)\|_G \leq C/|\omega|$ , where  $C$  only depends on  $f$ . This explains the scale-dependency of  $TV+L^1$  decomposition. The intensity-independency property simply follows from the use of the signum function in  $G$ -norm.

To determine an appropriate  $\lambda$  for the small-scale facial features is straightforward. It is clear from the above analysis that a single  $\lambda$  is good for all the faces with the same size. Alternatively, we can pick a  $\lambda$  that is slightly smaller than  $1/\|m\|_G$ , where  $m$  is the facial feature mask function which equals 1 over the small-scale facial objects and 0 anywhere else. Today for the large numbers of functions for which we have not known an analytic way to derive the  $\bar{g}^*$ 's (hence the  $G$ -norms), Goldfarb and Yin [15] introduced a second-order cone programming based method to compute  $G$ -norm numerically.

Besides, readers may know the famous  $TV+L^2$  model (the Rudin-Osher-Fatemi model [33]). It is a model more suitable in denoising, but the  $TV+L^1$  model works much better in scale-based image decomposition. The  $L^2$  term  $\|f - u\|_{L^2}$  penalizes big  $f(x) - u(x)$  values much more than small  $f(x) - u(x)$  values, so the  $TV+L^2$  model allows most small point-wise values (like most noise) in  $f - u$ . The  $L^1$  term  $|f - u|$ , however, penalizes the difference between  $f$  and  $u$  in a linear way. The  $L^1$  term does not favor noises, but when used with  $TV(u)$ , it lets  $f - u$  contain nearly all the signals with scale  $\leq 1/\lambda$  w.r.t.  $G$ -norm and with their original amplification. This is also one main difference between our work and Brajovic's work [8]. The  $TV+L^1$  model can successfully

extract small-scale signals like the edges, lines, and corners of facial features form a face image with an appropriate  $\lambda$ , but being an additive model, it cannot reflect the multiplicative illumination effect in equation (1). Figure 1(B) Column 1 depicts the  $v$  output of the  $TV+L^1$  model applied to the input image which is under an extreme lighting condition. Just like the input, the left half of  $v$  barely contains any perceivable signal. To overcome this limitation, we propose the LTV model.

## 2.4 The LTV model

To factorize a surface under multiplicative models (e.g. the Lambertian model), we take the approach by extracting the small intrinsic facial structures, where the albedos vary a lot. We observe that one of the differences between the intrinsic structure and the illumination pattern of a face image is the scale difference. The former is mostly composed of lines, edges, and small-scale objects. The later, which is consisted of direct light illumination and/or shadows cast by bigger objects (e.g., noses), is often of a larger scale. From the analysis in Section 2.3, it is clear that we can utilize the edge-preserving and scale-dependent decomposition capacity of the  $TV+L^1$  model.

To apply the additive  $TV+L^1$  model to the multiplicative model, we take the logarithm of the input image, followed by applying the  $TV+L^1$  model:

$$\begin{aligned}
 I(x, y) &= \rho(x, y) \cdot S = (\rho/\rho_l) \cdot (S\rho_l) = \rho' \cdot S' \\
 u &= \operatorname{argmin}_u \int_{\Omega} |\nabla u| + \lambda |\log(I) - u| dx, \\
 v &= \log(I) - u \\
 u &\approx \log(S'), \text{ and } v \approx \log(\rho')
 \end{aligned} \tag{7}$$

where  $\rho_l$  denotes the albedos of large scale skin areas and background. Since logarithm preserves structures and  $TV+L^1$  decomposes images by scales, the albedos  $\rho_l$  of large-scale areas are kept along with the  $S$  in  $u$  (or  $S'$ ). Nevertheless,  $\rho' = \rho/\rho_l$  is sufficient for face recognition as it promotes the variation patterns of the albedos of the small-scale facial features.

It is worth pointing out that for face recognition purposes, the intrinsic structures in  $\rho'$ , which are in general of a smaller scale than extrinsic illumination artifacts and shadows, contribute as discriminants. As long as the illumination fields or shadows are of a larger scale than the intrinsic structures, the LTV model can remove them by keeping them in  $S'$ . This assumption is validated by the experimental results in Section 3. Before presenting these results, we first illustrate how to



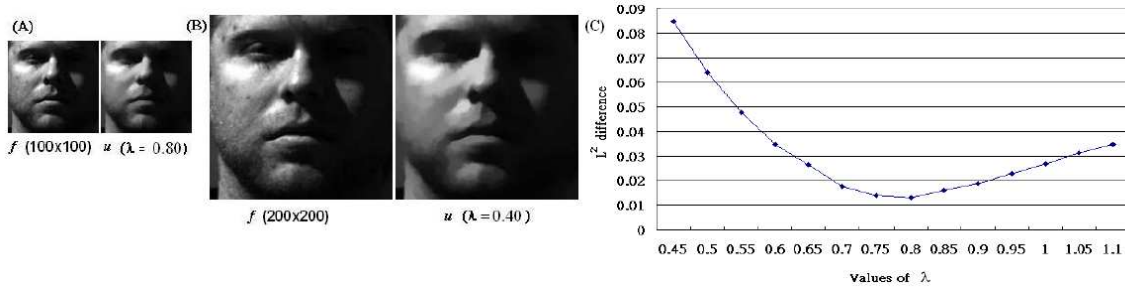


Figure 2: (A) Input face  $f$  with size  $100 \times 100$  and  $u$  obtained by  $\lambda = 0.80$ . (B) Input face  $f$  with size  $200 \times 200$  and  $u$  obtained by  $\lambda = 0.40$ . (C) The least square difference between  $u$  ( $\lambda = 0.40$ ) in figure 2 (B) and the  $u$  with different values of  $\lambda$  (horizontal axis) in figure 2 (A).

select an appropriate  $\lambda$ , the only parameter in the LTV model.

## 2.5 Parameter selection

One of the main advantages of our model over existing solutions is that there is only one parameter need to be set, and the choice of it is only dependent on the scales but independent of the intensities of the image, hence independent of the illumination. In other words, one parameter can serve for all face images with similar size. Finding an appropriate  $\lambda$  is straightforward simply because the features of different faces have similar scales and  $\lambda$  is in inverse proportion to the scale of the face (4)(5). Following the guidelines of  $\lambda$  selection in Sec. 2.3, we recommend the following  $\lambda$  values for different face sizes: 100x100 (pixels):  $\lambda = 0.7$  to 0.8, 200x200:  $\lambda = 0.35$  to 0.4, 400x400:  $\lambda = 0.175$  to 0.2. Figures 2 illustrates and presents statistically that the LTV output with  $\lambda = 0.4$  of a face in a 200x200 size is almost the same as the LTV output with  $\lambda = 0.8$  of the same face in a down sampled 100x100 size. Therefore, a single  $\lambda(0.75)$  is consistently used throughout our experiments in section 3. Figure 1 (B) illustrates the effects of different  $\lambda$  and figure 3 (A) illustrates the LTV algorithm.

## 2.6 TVQI

One variation of the LTV model, which parallels the idea of Brajovic [7] and Wang [44] is that we can divide the original image  $I$  by  $S$  to recover  $\rho$ :  $\rho = \frac{I}{S}$ . Using the  $TV+L^1$  model with similar parameter  $\lambda$  as the LTV model, we can get an estimation of  $S$  using  $u$  in (2). We call this model the total variation quotient image model (TVQI)[10]. The LTV model and the TVQI model are

essentially the same. The only difference is that the log operator removes the noise of the image which makes it possible to promote the useful signals more and improve the performance.

### 3 Experimental Results

In this section, we evaluate the proposed algorithm by several experiments. We first compare our models with 3 other methods (QI[36], QIR[35], SQI[44]) on Yale face database B and CMU PIE database. Then an outdoor database is used for evaluating the performance of natural lighting condition. We evaluate the face recognition by two different methods, template matching and PCA. The former uses a very simple similarity metric, the normalized correlation, to match images. Recognition is defined as matching, a query image  $y$  to a set of *reference images*  $\mathbf{T}$ . We name an image of subject  $x$  the *ideal* image if the angle of the light source direction is 0.

#### 3.1 Data preparation

We use both Yale face database B [14] and CMU PIE database [37] as our testbed to compare different algorithms. The frontal face images of the 10 subjects in the Yale face database B, each with 64 different illumination, and the frontal face images of the 68 subjects in CMU PIE database, each with 21 different illumination, are used for evaluation. All images are roughly aligned between different subjects and resized to 100 x 100. Images are cropped so that only the face region of each image is used. Images in the Yale face database B are divided into 5 subsets based on [14]. Other than these two benchmark databases, we also evaluate our methods on an outdoor database containing 765 subjects, each with 2 to 5 different illumination.

#### 3.2 Results on Yale face database B

In the first experiment, we use each image (except the ideal cases for each person) from subset 1 to subset 5 as a query image and match it with the 10 *ideal* images, which serves as the *reference images*. Table 3 (B) shows the results. The recognition rates of both the LTV model and its variation (TVQI) reach 100%.

After simple template matching, we also conducted PCA recognition on Yale database, the

(A)	1. $I \leftarrow$ input image ( $I = \rho' \cdot S'$ )	(B)	Subset	1	2	3	4	5
	2. Take the logarithm: $f \leftarrow \log(I), f = u + v, u = \log(S'), v = \log(\rho')$		QI	100	98.3	61.86	34.06	23.28
	3. Solve for $u$ and $v$ in the TV+ $L^1$ model: $u = \arg \min_u \int_{\Omega}  \nabla u(x)  + \lambda  f(x) - u(x)  dx$		QIR	100	100	100	90.58	78.84
	4. Obtain the intrinsic part: $v = f - u = \log(\rho')$		SQI	100	100	100	96.37	97.88
	5. Calculate the real intrinsic structure represented by $\rho'$		TVQI	100	100	100	100	100
	6. Illumination normalized result $\leftarrow \rho'$		LTV	100	100	100	100	100
(C)			Method	QI	QIR	SQI	TVQI	LTV
			Recog. Rate	75.66	88.65	97.91	100	100

Figure 3: (A) The LTV algorithm. (B) Recognition Rate (%) on Yale database B. (C) Average recognition Rate (%) on CMU PIE database.



Figure 4: Left: The 21 different lighting conditions for each single subject in CMU PIE database. Right: The illumination normalized images by the LTV model.

results show that by using 2 images per subject as the training set, the LTV model reaches 99.25% recognition rate in average. By using 3 images per subject as the training set, the LTV model reaches 99.99% recognition rate in average. By average, we tried all the combinations of 2 and 3 from the 64 illumination as training set and get the average recognition results.

### 3.3 Results on CMU PIE database

Since CMU PIE face database has much less illumination on each subject (21), we do not classify the images into different subsets according to the angle of the light source directions. In this experiment, we also use the *ideal images* as the *reference images*. Figure 3 (C) shows the recognition rates of the experiment. Figure 4 shows the illumination normalization images by the LTV model of the input images. We also conducted PCA recognition on CMU PIE database. The results show that by using 2 images per subject as the training set, the LTV model reaches 99.79% recognition rate in average. By using 3 images per subject as the training set, the LTV model reaches 99.99% recognition rate in average. The results of PCA recognition are at least as good as most recently published works, e.g. Corefaces [34]. As shown in the results, the LTV and TVQI models reach great recognition rates in the experiments on both Yale and the CMU PIE databases, which demon-



Figure 5: (A) Recognition rate(%) on outdoor face database.  $K$  is the number of nearest neighbors retrieved. (B) The *LTV* results (even rows) on images under natural lighting condition(odd rows).

strates that the proposed *LTV* model can obtain excellent illumination invariant recognition results. Since PCA selects the most determinant features first, the good results by PCA recognition can imply that our assumption that small facial structures contribute as the key factor for face recognition may be true.

### 3.4 Outdoor database

In this section, we conduct experiments on a database with 2662 frontal face images under outdoor natural lighting conditions, including 395 females and 370 males. There are 2 to 5 images per person under various illumination. We evaluate our face recognition algorithm by the following scenario, which is designed for real applications. For each image  $I_q$ ,  $q \in \{1, \dots, 2662\}$ , we search for the  $K$  nearest neighbors of  $I_q$  in the remaining 2661 images. If an image with the same subject is retrieved, it is a successful recognition. Since there are at least 2 images per person, for each query image  $I_q$ , there is at least one image in the remaining 2661 images matches  $I_q$ . This scenario is one possible way for a police officer to identify a criminal or look for a suspect in the database or in similar situations. We compare our methods with *SQI* and a simple histogram equalization (*HE*) algorithm. Figure 5 (A) shows the result. From figure 5 (A), the proposed *LTV* algorithm can achieve nearly perfect result (99+%) even when  $K$  is 1 and can achieve 100% when  $K$  is only 4. Although images in the evaluated outdoor database are not allowed to be published, figure 5 (B) shows some *LTV* results on images under outdoor lighting condition, which are very similar to the images in the original database and were added into the database during evaluation. Since PCA recognition is not suitable for this scenario and database, the results are obtained using simple template matching algorithm.

## 4 Discussion and Conclusion

In this paper, we propose the LTV model as a preprocessing technique for face recognition under varying illumination. This method works on any single image without knowledge of 3D face models or light sources. It minimizes the notorious halo artifacts and leaves only illumination invariant small scale facial structures with only one simple parameter to set. One assumption of our work is that small scale facial structures may be the key for frontal face recognition. Since our model reaches very high recognition rate using PCA recognition, this assumption can be true. The proposed approach has strong potential to be applied to relevant applications, such as face alignment, face tracking, where the results are easily affected by illumination variation.

## References

- [1] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: The problem of compensating for changes in illumination direction" *IEEE Tran. PAMI*, 19(7), pp. 721-732, 1997.
- [2] F. Alizadeh, and D. Goldfarb. "Second-order cone programming" *Math. Prog.* 95(1), pp. 3-51, 2003.
- [3] S. Alliney "Digital filters as absolute norm regularizers" *IEEE Tran. on Signal Processing*, 40, pp. 1548-1562, 1992.
- [4] R. Basri, and D. Jacobs, "Lambertian reflectance and linear subspaces" *NEC research institute technical report*, 2000-172R.
- [5] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, "Eigenfaces vs Fisherfaces: recognition using class specific linear projection" *IEEE Tran. PAMI*, 20(7), 1997.
- [6] P. N. Belhumeur, D. J. Kriegman, "What is the set of images of an object under all possible lighting conditions?" *CVPR*, 1996.
- [7] V. Brajovic, "A Model for Reflectance Perception in Vision" *SPIE, EMT*, 2003.
- [8] V. Brajovic, "Brightness Perception, Dynamic Range and Noise: A Unifying Model for Adaptive Image Sensors" *CVPR*, 2004.
- [9] T. F. Chan, and S. Esedoglu, "Aspects of Total Variation Regularized  $L^1$  Function Approximation" *UCLA CAM Report 04-07*, Feb, 2004.
- [10] T. Chen, W. Yin, X.-S. Zhou, D. Comaniciu, T. S. Huang, "Illumination Normalization for Face Recognition and Uneven Background Correction Using Total Variation Based Image Models" *CVPR*, 2005.
- [11] D. Comaniciu, P. Meer, "Mean Shift: A Robust Approach toward Feature Space Analysis", *IEEE Tran. PAMI*, 24(5), 2002.

- [12] F. Durand, and J. Dorsey, "Fast Bilateral Filtering for the Display of High-Dynammic-Range Images" *ACM Tran. on Graphics*, 21(3), 2002.
- [13] M. Elad, R. Kimmel, D. Shaked, and R. Keshet, "Reduced Complexity Retinex Algorithm Via the Variational Approach" *J. on Visual Communication and Image Representation*, 14(4), 369-388, 2003.
- [14] 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 PAMI*, 23(6), pp. 643-660, 2001.
- [15] D. Goldfarb, and W. Yin, "Second-order cone programming methods for total variation-based image restoration" *Columbia CORC TR-2004-05*, May, 2004.
- [16] R. Gross, and V. Brajovic, "An Image Preprocessing Algorithm for Illumination Invariant Face Recognition" *4th Int. Con. on Audio and Video Based Biometric Person Authentication*, 10-18, 2003.
- [17] G. Guo, S. Z. Li, and K. Chan, "Face recognition by support vector machines" *FG*, pp. 196-201, 2000.
- [18] B. K. P. Horn, "Determining Lightness from an Image" *Computer Graphics and Image Processing*, 3(1), 277-299, 1986.
- [19] B. K. P. Horn, "Robot Vision" *MIT Press*, 1986.
- [20] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multi-scale Retinex for Bridging the Gap between Color Images and the Human Observation of Scenes" *IEEE Tran. on Imag. Proc.*, 6(7), 965-976, 1997.
- [21] R. Kimmel, M. Elad, D. Shaked, R. Keshet, I. Sobel, "A Variational Framework to Retinex" *IJCV*, 52(1), 7-23, 2003.
- [22] E. H. Land, and J. J. McCann, "Lightness and Retinex Theory" *J. Opt. Soc. Am.*, 61(1), 1971.
- [23] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Automatic interpretation and coding of face images using flexible models" *IEEE Tran. PAMI* 19(7), pp. 743-756, 1997.
- [24] K.-C. Lee, J. Ho, and D. Kriegman, "Nine points of light: Acquiring subspaces for face recognition under variable lighting" *CVPR*, 2001.
- [25] Y. Meyer, "Oscillating patterns in image processing" *University Lecture Series 22*, AMS, 2000.
- [26] B. Moghaddam, T. Jebara, and A. Pentland, "Bayesian face recognition" *Pattern Recognition*, 33, pp. 1771-1782, 2000.
- [27] M. Nikolova, "Minimizers of cost-functions involving nonsmooth data-fidelity terms" *SIAM Journal on Numerical Analysis*, vol. 40:3, pp. 965-994, 2002.
- [28] P. Oenev, and J. Atick, "Local feature analysis: A general statistical theory for object representation" *Network: Computation in Neural System*, 7, pp. 477-500, 1996.
- [29] S. Osher, and O. Scherzer, "G-norm properties of bounded variation regularization" *UCLA CAM Report* 04-35, 2004.
- [30] P. Perona, and J. Malik, "Scale-space and edge detection using anisotropic diffusion" *IEEE Tran. PAMI*, 12(7), 629-639, 1990.

- [31] R. Ramamoorthi, and P. Hanrahan, "On the relationship between radiance and irradiance: determining the illumination from images of a convex Lambertian object" *J. Opt. Soc. Am.*, 18(10), 2001.
- [32] R. Rammamorthi, and P. Hanrahan, "A Signal-Processing Framework for Inverse Rendering" *ACM SIGGRAPH*, 2001.
- [33] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms" *Physica D*, 60, pp. 259-268, 1992.
- [34] M. Savvides, B. V. K. Vijaya Kumar, and P. L. Khosla, "Corefaces - Robust Shift Invariant PCA based Correlation Filter for Illumination Tolerant Face Recognition" *CVPR*, 2004.
- [35] S. Shan, W. Gao, B. Cao, and D. Zhao, "Illumination normalization for robust face recognition against varying lighting conditions" *Int'l Workshop on Analysis and Modeling of Faces and Gestures*, 2003.
- [36] A. Shashua, and T. Riklin-Raviv, "The quotient image: Class-based re-rendering and recognition with varying illuminations" *IEEE Tran. PAMI*, Vol. 23(2), pp. 129-139, 2001.
- [37] T. Sim, S. Baker, and M. Bsat, "The CMU Pose, Illumination, and Expression(PIE) Database" *IEEE International Conference on Automatic Face and Gesture Recognition*, May, 2002.
- [38] J. Short, J. Kittler, and K. Messer, "A Comparison of Photometric Normalisation Algorithms for Face Verification", *FGR*, 2004.
- [39] T. G. Stockham Jr., "Image Processing in the Context of a Visual Model" *Proc. of IEEE*, 60, 828-842, 1972.
- [40] D. Strong, and T. Chan, "Edge-preserving and scale-dependent properties of total variation regularization" *Inverse problems* 19, pp.S165-S187, 2003.
- [41] C. Tomasi, and R. Manduchi, "Bilateral filtering for gray and color images" *ICCV*, 1998.
- [42] J. Tumblin, and G. Turk "LCIS:A Boundary Hierarchy for Detail-preserving Contrast Reduction" *ACM SIGGraph*, 1999.
- [43] M. Turk, and A. Pentland, "Eigenfaces for Recognition" *Journal of cognitive neuroscience*, 3(1), pp. 71-86, 1991.1.
- [44] H. Wang, S. Z. Li, and Y. Wang, "Generalized Quotient Image" *CVPR*, 2004.
- [45] L. Wiskott, J. M. Fellous, N. Kruger, and C. V. D. Malsburg, "Face recognition by elastic bunch graph matching" *IEEE Tran. PAMI*, 19(7), pp. 775-779, 1997.
- [46] L. Zhang, D. Samaras, "Face recognition under variable lighting using harmonic image exemplars" *CVPR*, 2003.
- [47] W. Zhao, and R. Chellappa, "Robust face recognition using symmetric shape-from-shading" *Technical report, Center for Automation Research, University of Maryland*, 1999.
- [48] W. Zhao, and R. Chellappa, "Robust image-based 3D face recognition" *CAR-TR-932, N00014-95-1-0521, CS-TR-4091, Center for auto research, UMD,2000.1*.