Illumination Normalization for Face Recognition and Uneven Background Correction Using Total Variation Based Image Models

Terrence Chen[1] , Wotao Yin[2], Xiang Sean Zhou[3], Dorin Comaniciu[3], and Thomas S. Huang[1]
[1]University of Illinois at Urbana Champaign, [2]Columbia University, [3]Siemens Corporate Research
{[1]tchen5, huang}@ifp.uiuc.edu, [2]wy2002@columbia.edu, {[3]xzhou, comanici}@scr.siemens.com

Abstract

We present a new algorithm for illumination normalization and uneven background correction in images, utilizing the recently proposed TV+L^1 model: minimizing the total variation of the output cartoon while subject to an L^1-norm fidelity term. We give intuitive proofs of its main advantages, including the well-known edge preserving capability, minimal signal distortion, and scale-dependent but intensity-independent foreground extraction. We then propose a novel TV-based quotient image model (TVQI) for illumination normalization, an important preprocessing for face recognition under different lighting conditions. Using this model, we achieve 100% face recognition rate on Yale face database B if the reference images are under good lighting condition and 99.45% if not. These results, compared to the average 65% recognition rate of the quotient image model and the average 95% recognition rate of the more recent self quotient image model, show a clear improvement. In addition, this model requires no training data, no assumption on the light source, and no alignment between different images for illumination normalization. We also present the results of the related applications - uneven background correction for cDNA microarray films and digital microscope images. We believe the proposed works can serve important roles in the related fields.[4]

1. Introduction

Illumination is one of the most significant factors affecting the appearance of an image. It often leads to diminished structures or inhomogeneous intensities of the image due to different albedos (texture) of the object surface and the shadows cast from different light source directions. On the other hand, uneven background, also known as background bias, background intensity inhomogeneity, or nonuniform background, is the problem that an ideal image f is corrupted by an uneven background signal b so that the observed image $I = f + b$. Recovering f from I is not an easy task when b is non-uniform. In essence, both the varying illumination and the uneven background are inhomogeneous intensity patterns that are either multiplicative or additive. In this paper, we propose total variation based image models to solve these two problems. We use face recognition under different lighting conditions to evaluate the effectiveness of our illumination normalization schemes, and use microarray and digital microscope images with apparent background bias to illustrate the ability of our proposed models for scale-dependent additive background signal removal.

1.1. Illumination normalization

Illumination normalization is an important task in the field of computer vision and pattern recognition. In real world, one of the most important problems of illumination normalization is face recognition under varying illumination. Face recognition has many applications, such as public security, identity authentication, etc. It has been proven both experimentally [1] and theoretically [27] that the differences due to varying illumination is more significant than the differences between individuals in face recognition. Various methods have been proposed for face recognition, including Eigenface [23], Fisherface[4], Probablistic and Bayesian Matching [12], subspace LDA [28], Active Shape Model and Active Appearance Model [10], LFA[15], EBGM[25], and SVM[9]. Nevertheless, the performance of most existing algorithms is highly sensitive to the variation of illumination.

To attack the problem of face recognition under illumination variation, several algorithms have also been proposed. The Illumination Cone methods [5][7], spherical harmonic based representations [16] [3] [26], and quotient image based approaches [19] [18] [24] are proposed for this purpose. However, not only the performances of most of the methods are still far from ideal, many of these methods either require assumptions of the light source or need a large

[4]This paper is the result of the internship work of the first two authors at Siemens Corporate Research during the Summer of 2004

number of training sets, which are not considered practical in real applications.

1.2. Uneven background correction

Uneven background is another important problem in the field of image processing. The main difficulty of solving this ill-posed inverse problem is to correct the background without distorting the foreground signals. In this paper, background bias correction for cDNA microarray slides and digital microscope images are our motivating applications. cDNA microarrays consist of tens of thousands of individual DNA sequences printed in parallel on a glass microscope slide. They are designed to detect specific genes and to measure their activities in tissue samples by monitoring the differential hybridization of the two DNA or RNA samples to the sequences on the array. On a microarray slide, the measured fluorescence intensity of a spot is a combination of the image background intensities near the spot and the intensities determined by the hybridization level of the mRNA samples with the spotted DNA. Background correction is to identify the image background intensities near cDNA spots and then quantify the extracted foreground intensities. Other than microarray images, digital microscope images also suffer from uneven background corruption, which leads to the nonuniform intensities in target specimen. Accurate uneven background correction can not only facilitate the observation of the specimen, but also improve the accuracy of further image analysis, such as segmentation, quantification, etc.

In this paper, we propose to attack the two relevant problems by the total variation (TV) based image models. We utilize the properties of the TV+L^1 model [6] and show why and how it can be adapted for uneven background correction. Starting from the TV+L^1 model we propose the total variation based quotient image (TVQI) model for illumination normalization. Extensive experimental evaluations are conducted to illustrate the effectiveness and advantages of our approach.

2. Methodology

We first introduce the total variation modes, especially the TV+L^1 model, followed by its property analysis. Based on its properties and the needs in illumination normalization, we propose the novel TVQI model later in this section.

In the TV-based framework, an image f is modelled as the sum of image cartoon u and texture v, where f, u and v are defined as functions (or flow fields) in appropriate spaces. Cartoon contains background hues and important boundaries as sharp edges. The rest of the image, which is texture, is characterized by small-scale patterns. Since cartoon u is more regular than texture v, we can obtain u from image f by solving a variational problem:

$$\min \int_\Omega |\nabla u| + \lambda \|t(u,f)\|_B, \tag{1}$$

where $\int_\Omega |\nabla u|$ is the total variation of u over its support Ω, $\|t(u,f)\|_B$ is some measure of the closeness between u and f, and λ is a scalar weight parameter. The choice of the measure $\|\cdot\|_B$ depends on applications.

The first use of this model was due to Rudin, Osher, and Fatemi (ROF) [17] for image denoising where they use $\|t(u,f)\|_B = \|f-u\|_{L^2}$. The essential merit of total variation based image model is the edge-preserving property [22]. A simple way to understand this property is to notice the following. First, minimizing the regularization measure $\int |\nabla u(x)| \mathrm{d}x$ only tends to reduce the total variation of u over its support, a value that is independent of edge smoothness. Second, unless $\|t(u,f)\|_B$ specifically penalizes sharp edges, minimizing a fidelity term $\|t(u,f)\|_B$ (e.g., L^1 or L^2-norm of $f-u$) generally tends to keep u close to f, and thus, also keeps edges of f in u. Finally, minimizing $\int |\nabla u| + \lambda \|t(u,f)\|_B$, with λ sufficiently big, will keep sharp edges. ROF uses the L^2-norm, which penalizes big point-wise differences between f and u, so it aims at removing small point-wise differences (like noise) from f. Mainly due to this good edge-keeping property, the ROF model has been generalized and modified in many ways in the research communities. One of them uses the L^1-norm as the fidelity term [2, 14, 6].

2.1. A closer look at TV+L^1 for additive signal decomposition

Formally, the TV+L^1 model is formulated as:

$$\min_u \int_\Omega |\nabla u(x)| + \lambda |f(x) - u(x)| \, dx. \tag{2}$$

To solve (2), Goldfarb and Yin [8] casts (2) as a second-order cone program and solves it using the modern interior-point method. We analyze the properties of the TV+L^1 model for the purpose of additive signal decomposition, providing theoretical justification for our proposed application of background correction in images. Intuitive proofs of key properties are provided in the Appendix.

Just like the L^2-norm, the L^1-norm keeps u close to f (but under a different measure), so the edge-preserving property can be easily seen by following the similar argument for the ROF model . We give the TV+L^1 analytical results of some easy problems in $\mathbb{R}^2$. First, we would mention that Chan and Esedoglu [6] proved that solving equation (2) is equivalent to solving the following level-set-based geo-

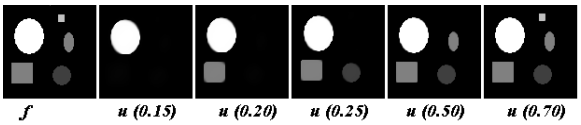

Figure 1. Original image f and different level of u when applying different (λ).

metrical problem:

$$\min_u \quad \int_{-\infty}^{+\infty} Per\left(\{x : u(x) > \mu\}\right) + \lambda Vol(\{x : u(x) > \mu\} \oplus \{x : f(x) > \mu\})\, d\mu \tag{3}$$

where $Per(\cdot)$ is the perimeter function, $Vol(\cdot)$ is the volume function, and $S_1 \oplus S_2 := (S_1 \setminus S_2) \cup (S_2 \setminus S_1)$, for any sets S_1 and S_2. Using equation (3), we can prove the following geometric properties of the solution $v(\lambda) = f - u(\lambda)$ in (2):

- Suppose $f = c_1 1_{B_r(y)}(x)$, a function with the intensity c_1 in the disk centered at y and with radius r, and the intensity 0 anywhere else. Then

$$v(\lambda) = \begin{cases} c_1 1_{B_r(y)}(x) & 0 \le \lambda < 2/r, \\ \{s 1_{B_r(y)}(x) : 0 \le s \le c_1\} & \lambda = 2/r, \\ 0 & \lambda > 2/r. \end{cases} \tag{4}$$

By this property, when applying different values of λ, objects of different scales can be kept in u or v (c.f. figure 1). Furthermore, we can extend this property to the following:

- Suppose $f = c_1 1_{B_{r_1}(y)}(x) + c_2 1_{B_{r_2}(y)}(x)$, where $0 < r_2 < r_1$ and $c_1, c_2 > 0$.

$$v(\lambda) = \begin{cases} (c_1 1_{B_{r_1}(y)} + c_2 1_{B_{r_2}(y)})(x) & \lambda < 2/r_1, \\ c_2 1_{B_{r_2}(y)}(x) & 2/r_1 < \lambda < 2/r_2, \\ 0 & \lambda > 2/r_2. \end{cases} \tag{5}$$

We give the proof of these two properties in the appendix. We note that these properties can be further expanded to $c_1, c_2, \ldots$ with radius $r_1, r_2, r_3, \ldots$, and also to any objects with C^2 boundaries. Figure 2 illustrates these properties.

Another property of the TV+L^1 model we would mention here is its minimal signal distortion. From (5), we know that the signal of the disk is either kept in u or in v unless the value of λ is exactly $2/r$ (this is negligible after discretization). This means that the TV+L^1 model is not likely to separate an entire signal into both u and v, but only in one of them. This is one of the intrinsic differences between TV+L^1 and TV+L^2, and is also the reason why TV+L^1 is more geometrically interesting to our applications.

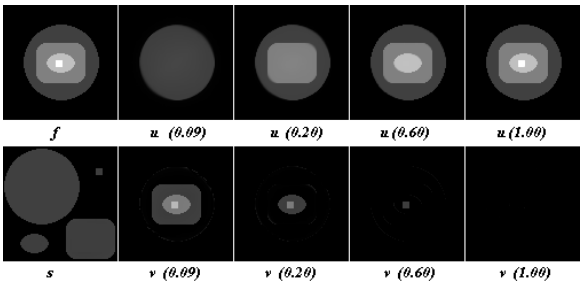

Figure 2. Additive signals with one included in the other can be extracted one by one using increasing values of (λ). s shows the intensities of the four shapes before addition.

To summarize, assuming a background bias of larger scale than the foreground (E.g., uneven illumination field, sensor bias, process distortion, etc.), a TV+L^1-based signal decomposition algorithm with an appropriate λ can remove the background bias from the foreground signals, with *tunable control based on scale*, *preservation of edges*, and *minimal signal distortion*. Next, we show that these properties can also be utilized for illumination normalization for face recognition. Our proposed algorithm is an integration of the quotient image concept [19] and the TV+L^1 decomposition scheme, and we call it the TVQI algorithm.

2.2. TVQI for illumination normalization

Although it has been shown that reflectance of an image can be helpful for identification [13], it is less useful for face recognition, where geometric features (i.e. shape, size of eyes, nose, etc.) are more important to distinguish different people. Hence, our goal is to acquire an illumination invariant face feature image for a group of images of the same subject. Based on the previous discussion, the cartoon u in the TV+L^1 model keeps the large-scale background intensities and important boundaries of the original image f and leaves, in v, the small scale signals, edges and textures, which are the most important features. However, the TV+L^1 model has certain limitation if it is directly applied for normalizing illumination. That is, the signals of intrinsic structures in the image are not normalized (cf. figure 3, (v)). For this reason, we propose the total variation based quotient image (TVQI) model.

The intuition of the TVQI model is the Lambertian surface model. According to the Lambertian model, the intensity at the (x, y) position of an image f is defined as:

$$f_{x,y} = A_{x,y} \rho_{x,y} \cos\theta_{x,y} \tag{6}$$

where A is the strength of the light source, ρ, the albedo (texture) of the image, is the surface reflectance associated

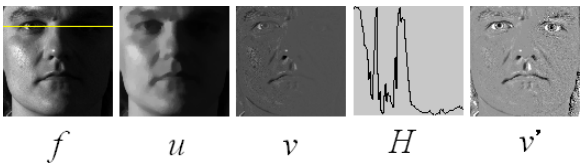

Figure 3. f, u, and v of TV+L^1 model. Weak signal in v (the right hand side) is hardly observed. H shows the intensity profile of the horizontal line in f. $v\prime$ is the TVQI result.

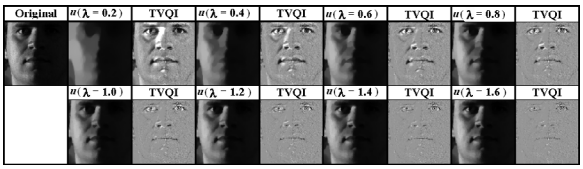

Figure 4. Effect of different λ in TVQI methodology. 1st column: original image f; even columns: u using different values of λ, odd columns: TVQI obtained by f/u.

with each point in the image, and θ is the angle between the surface normal and the light source. (6) tells us that the intensity of image point (x, y) is proportional to the strength of the received light at this point. Therefore, the intensity and the variance of the small-scale signals are proportional to the intensity of the background in their vicinities. Figure 3 (H) illustrates this property (i.e. the signal variance under strong light is much larger than that under weak light). However, for recognition purpose, the small-scale signals in the dark area must be amplified and all important small-scale signals should have close amplification levels after illumination normalization. For this purpose, we approximate the normalized image by $v'_{x,y} = f_{x,y}/u_{x,y}$ for every point (x, y) using the output u of the TV+L^1 model. Since u is large-scale background intensity of the input f, dividing f by u gives an illumination normalized image, in which small-scale signals are uniformly highlighted. Therefore, we propose the TVQI model as:

$$u = \arg\min_u \int_\Omega |\nabla u(x)| + \lambda |f(x) - u(x)|\, dx,$$

$$TVQI = v' = \frac{f}{u}, \tag{7}$$

where f is the original face image. Figure 3 (v') shows the result obtained by applying this model to f. The remaining task is to choose an appropriate λ.

1. $f_1 \leftarrow$ input image

2. Remove noise in dark region: (+TVQI only)
   $f_1 \leftarrow \arg\min_u \int_\Omega |\nabla u(x)| + \lambda_{L_2} ||f_1(x) - u(x)||^2_{L_2}\, dx$

3. Calculate the denominator :
   $u = \arg\min_u \int_\Omega |\nabla u(x)| + \lambda_{L_1} |f_1(x) - u(x)|\, dx$

4. Obtain the illumination normalized image:
   $(+)TVQI = v'_1 = \frac{f_1}{u}$

5. Repeat 1 - 4 for other images $f_2, ..., f_m$, where m is the total number of images.

6. Apply face recognition on v'_is of the original image f_is.

Table 1. The (+)TVQI model.

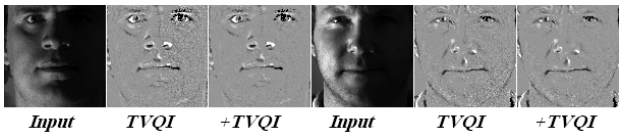

Figure 5. The preprocessing effect of +TVQI

2.3. Selecting λ for face recognition

The choice of lambda is straightforward - it is inverse proportional to signal scale. To extract circular pattern, lambda is inverse proportional to the circle radius (4). In the same way, we recommend the following lambda values for different face sizes: 100x100 (pixels): 0.8; 200x200: 0.4; 400x400: 0.2. We want to emphasize that a single lambda should work for all faces in the same size. All of these mean an appropriate lambda can easily be selected by using the recommended value, by a simple training process, or even by investigation (cf. figure 4). This is because we only need to find lambda for one face size.

2.4. +TVQI

In Figure 3, grainy noise can be seen on the right half of $v\prime$. This is due to the division operation in the TVQI model, which propagates small variations and noise in the dark region of the image. One possible remedy is to use the aforementioned TV+L^2 model [17] with a larger λ ($\geq$ 1) as a preprocessing step for noise removal before applying TVQI. We call the resulting combination the *+TVQI algorithm*. Table 1 shows this algorithm. (TVQI is the same except without step 2). Figure 5 illustrates the effect of +TVQI. One can see clear quality improvement for a human observer. However, interestingly, after running our face recognition algorithm, +TVQI has only similar performance as TVQI ($\pm 0.05\%$). One explanation is that the TV+L^2 step destroys small signals while denosing. Another possibility is that on our test dataset the TVQI algorithm has already achieved a very high recognition rate ($> 99\%$),

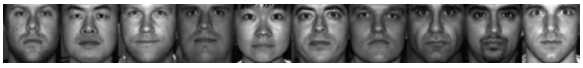

Figure 6. 10 subjects in Yale face database B.

leaving little room for improvement. Further investigation of +TVQI on more difficult data is among our future efforts. In the following sections, we only present evaluation results for the TVQI algorithm.

2.5. L^1 norm versus L^2 norm

One may doubt that the TV+L^2 model can also be combined with the quotient image model in (7). We explain here why the L^1 norm is better for this task. Generally speaking, the TV+L^2 model is more suitable in denoising and the TV+L^1 model works much better in scale-based image decomposition. The L^2 term $\| f-u \|^2$ penalizes big $f-u$ values much more than small $f-u$ values, so TV+L^2 allows most small point-wise values (like most noises) 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 total variation, it makes $f-u$ contain nearly all signals with scale $\leq 1/\lambda$ w.r.t. the G-norm and with their original amplification. In TVQI, the main story is the objects in v are highlighted by the process $f/u = 1 + v/u$. TV+L^1 only includes small scale facial components like mouth, nose, eyes, eyebrows, and wrinkles in v. Therefore, the final comparison between two faces is really made between the positions of the small-scale facial parts in the two faces. This is a result by combining quotient image and TV+L^1. We cannot replace TV+L^1 by TV+L^2 because the v of TV+L^2 does contain a small part of large-scale image objects. It is feasible to first apply TV+L^2 to remove tiny noise before applying TV+L^1 (c.f. +TVQI). Although this makes figure 5 look more appealing, this can hardly improve the recognition rate simply because TV+L^2 always takes away some useful signal together with noise.

3. Experimental Results

In this section, we evaluate the proposed algorithms by different experiments. Yale face database B is used for comparing the performance of illumination normalization capability of the TVQI model with other existing solutions. Real cDNA microarray and digital microscope images are used to validate the TV+L^1 model for background correction.

3.1. TVQI illumination normalization results

The performance of illumination normalization is evaluated according to the face recognition rate by the normalized correlation similarity measurement. There are many other similarity measurements, including those for comparing images subject to different scales, rotation, and other transformations. However, we do not focus on these measurements in this paper. Recognition is defined as matching a query image y to a set of *reference images* $\mathbf{T}$. y belongs to one of the subjects in $\mathbf{T}$ but under unknown illumination. We name an image of subject x the *ideal* image if the angle of the light source direction is 0. In all experiments, the TVQI model is compared with three other existing solutions, quotient image (QI) [19], quotient image relighting (QIR) [18], and self-quotient image (SQI) [24]. Figure 7 shows some illumination normalization results by different methods on the same input images.

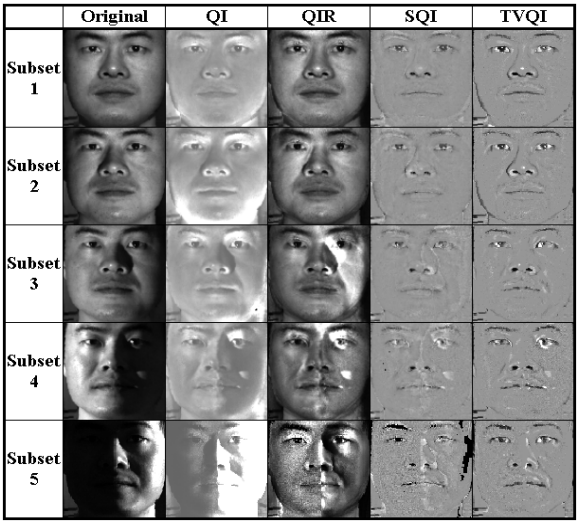

Figure 7. Illumination normalization effect comparison.

3.1.1 Data preparation

The frontal face images of the 10 subjects in the Yale face database B, each with 64 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 database are divided into 5 subsets based on the angle of the the light source directions. The 5 subsets are: subset 1 (0° to 12°), subset 2 (13° to 25°), subset 3 (26° to 50°), subset 4 (51° to 77°), subset 5 (above 78°) [7]. Out of the 640 images (10 x 64), 7 corrupted images are discarded. As a result, there are total 633 images: 70, 118, 118, 138, 189 images in subset 1 to 5, respectively. Figure 6 shows the *ideal* images (angle of light source = 0°) of the 10 subjects.

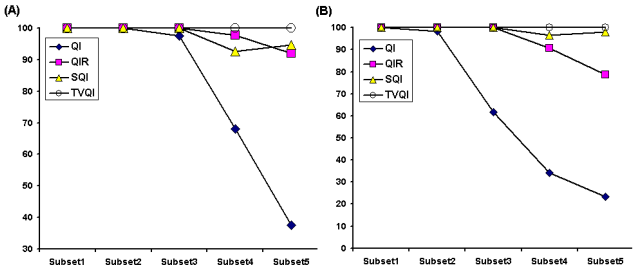

Figure 8. Recognition Rate (%) comparison using (A) subset 1, and (B) only the *ideal* image, as the *reference images*.

3.1.2 Using subset 1 as *reference images*

In the first experiment, 70 images I_j^k for subject k, $k \in K = \{1, \ldots, 10\}$ under illumination j in subset 1 is used as the *reference images*. The recognition process is to find the nearest neighbor $I_{j'}^{k'}$ of a given query image y^l, where $l \in K$. If l is equal to $k\prime$, the recognition is successful; otherwise, it is failed. In the experiment, we use all the images in subset 2 to 5 as the query images and evaluate the recognition rate. Figure 8(A) compares the recognition rates of different solutions. Recognition rate of the proposed TVQI method achieves 100% in all cases. This validates the better illumination normalization effect of TVQI.

3.1.3 Using only the *ideal* images as *reference images*

In the second experiment, instead of all the 70 images in subset 1, we only use the 10 *ideal* images, one for each subject, as the *reference images*. Figure 8(B) shows the results. The performance of both QI and QIR degrade heavily in this experiment. The recognition rates of the TVQI model remain 100%. This shows the robustness and consistency of the TVQI model.

3.1.4 Using other subsets as *reference images*

In the last experiment, we tried a more challenging but meaningful experiment which is rarely done in the literature. Instead of using the images in subset 1, we choose images in all the other subsets (2 to 5) as *reference images*. In real world, perfect *reference images* are not always available. A robust algorithm should perform well even with the *reference images* under varying lighting conditions. Figure 9 compares the recognition results and table 2 shows the average recognition rate (%). In figure 9, the label T$\mathbf{X}$-Subset$\mathbf{Y}$ in the horizontal axis denotes the query image is from subset $\mathbf{Y}$ and the *reference images* is from subset $\mathbf{X}$. For each test T$\mathbf{X}_i$-Subset$\mathbf{Y}_i$, we use 10 images (one for each subject) in subset $\mathbf{X}_i$ as the *reference images* for each round

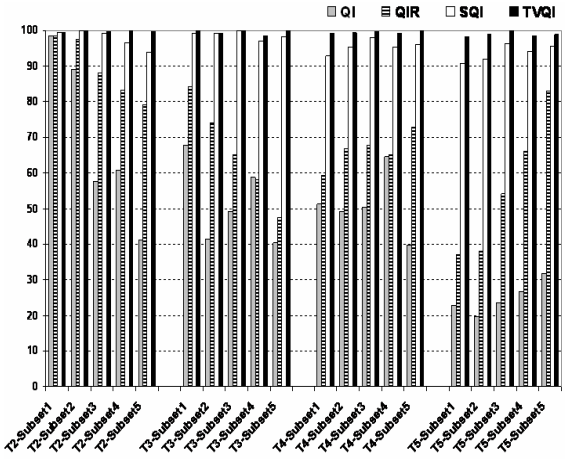

Figure 9. Recognition Rate (%) comparison when all different subsets are used as the reference images.

Methodolgy	QI	QIR	SQI	$TVQI$
average	47.27	69.44	94.95	99.45

Table 2. Average recognition Rate (%) comparison using images in all different subsets as the *reference images*.

i, we then change another 10 images in subset $\mathbf{X}_{i+1}$ as the *reference images* for round $i+1$, and so on. For each round i, all images in subset $\mathbf{Y}_i$ are used as the query image one by one to evaluate the recognition rate R_i. The final recognition rate $R = \frac{1}{N}\sum_{i=1}^{N} R_i$, where N is the round number. The performance of all the other methods degrades when the bad illuminated images (in subset 4 and subset 5) are used as the *reference images*. In contrast, the recognition rate of the TVQI method are always above 98.51% and averaged at 99.45%. Hence, TVQI is proved experimentally to be a very robust and consistent algorithm for illumination invariant face recognition.

3.2. Background correction results

We have explained the reasons why TV+L^1 can be used for uneven background correction (2.1). In this section, we use real cDNA microarray and digital microscope images to show the capability of uneven background correction of the TV+L^1 model. Figure 10 shows a real microarray image. For comparison, we use both the TV+L^1 method and the state-of-the-art method, morphological opening (MO) [21], in the field of bioinformatics to estimate the background of the image. Specifically, morphological opening applies a local minimum filter, which is an erosion process,

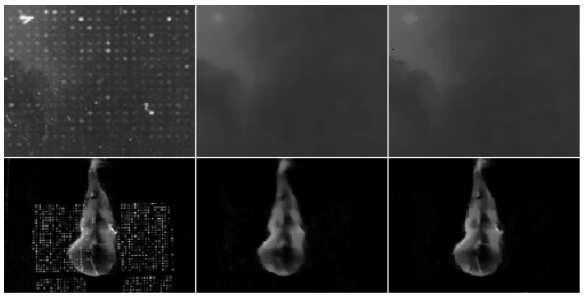

Figure 10. Background estimation on cDNA microarray image, Left: original images. Middle: estimated backgrounds using MO. Right: estimated backgrounds using TV+L^1.

followed by a local maximum filter, which is a dilation process, to estimate the background of the image. From the first row of figure 10, the water stain at upper left corner in the first image was estimated by morphological opening, but the shape of the stain is smoothed. Instead, $TV + L^1$ not only estimates the correct background but also keeps the shape of the stain. This property can also be seen from the second row of figure 10. Furthermore, we also apply it to digital microscope images with apparent nonuniform background. Figure 11 shows some results of the microscope images and the background corrected results by the TV+L^1 model. The MO method used for comparison shows that removing background bias without destroying the foreground structures is not an easy task. The images are downloaded from the Olympus web site [20].

4. Conclusion

In this paper, we propose the total variation quotient image (TVQI) model for face recognition under varying illumination and illustrate its effectiveness for illumination normalization. The advantage of this method is that it requires no training images, no assumption of the light source, and no alignment between images for illumination normalization. In addition, the recognition results of TVQI outperforms the existing solutions in a significant manner. The experiment shows that our method can be used for identifying unknown faces even if the available *reference face images* are not obtained under good lighting condition. Furthermore, we also propose the use of the TV+L^1 model as a suitable tool for correcting intensity inhomogeneities in image background. We demonstrate it using experiments on real cDNA microarray and digital microscope images. We believe the proposed works have significant contribution to computer vision, image processing, public security, and other related fields.

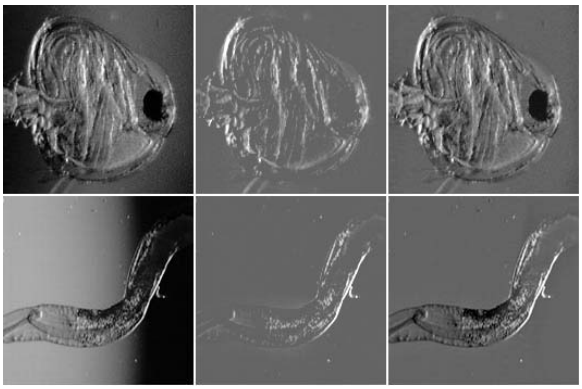

Figure 11. Uneven background correction on digital microscope images, 1st column: original images. 2nd column: background corrected images using MO. 3rd column: background corrected images using TV+L^1.

Appendix

Proof of property (4):

Proof By assumption, $f = c_1 1_{B_r(y)}(x)$. Without loss of generality, we assume $c_1 > 0$. Clearly, solution $u(x)$ of (2) is bounded between 0 and c_1, for all almost all $x \in \Omega$. It follows that (3) is simplified to:

$$\begin{aligned}\min_u \quad & \int_0^{c_1} Per\left(\{x : u(x) > \mu\}\right) \\ & +\lambda Vol\left(\{x : u(x) > \mu\} \oplus \{x : f(x) > \mu\}\right)\, d\mu.\end{aligned} \tag{8}$$

Since $\{x : f(x) > \mu\} \equiv B_r(y)$ for $\mu \in (0, c_1)$, $S(\mu) := \{x : u(x) > \mu\}$ must solve the following geometry problem:

$$\min_S Per\left(S(\mu)\right) + \lambda Vol\left(S(\mu) \oplus B_r(y)\right), \tag{9}$$

for almost all $\mu \in (0, c_1)$. First, $S(\mu) \subseteq B_r(y)$ holds because, otherwise, $\bar{S}(\mu) := S(\mu) \cap B_r(y)$ achieves lower objective value than $S(\mu)$. Then, it follows that

$$Vol\left(S(\mu) \oplus B_r(y)\right) = Vol\left(B_r(y) \setminus S(\mu)\right). \tag{10}$$

Therefore, to minimize (9) is to minimize the perimeter of S while maximize its volume. By Isoperimetric Theorem, $S(\mu)$ must be either empty or a disk. Let r_S denote the radius of S, it follows that $r_S = r$ if $\lambda > 2/r$, $r_S = 0$ if $0 \le \lambda < 2/r$, and $r_S \in \{0, r\}$ if $\lambda = 2/r$. Property (4) follows from relationship $v = f - u$. ■

Proof of property (5):

Proof Clearly, (3) can be simplified to:

$$\begin{aligned}\min_{u \in BV(\Omega)} \quad & (\int_0^{c_1} + \int_{c_1}^{c_1+c_2}) Per\left(\{x : u(x) > \mu\}\right) \\ & +\lambda Vol\left(\{x : u(x) > \mu\} \oplus \{x : f(x) > \mu\}\right)\, \mathrm{d}\mu.\end{aligned} \tag{11}$$

Since, for $\mu \in (0, c_1)$, $\{x : f(x) > \mu\} \equiv B_{r_1}(y)$, and for $\mu \in (c_1, c_1 + c_2)$, $\{x : f(x) > \mu\} \equiv B_{r_2}(y)$, this problem can be simplified as:

$$\begin{array}{rl} \min_{u \in BV(\Omega)} \int_0^{c_1} & Per\left(\{x : u(x) > \mu\}\right) \\ & +\lambda Vol\left(\{x : u(x) > \mu\} \oplus B_{r_1}(y)\right) \mathrm{d}\mu \\ + \int_{c_1}^{c_1+c_2} & Per\left(\{x : u(x) > \mu\}\right) \\ & +\lambda Vol\left(\{x : u(x) > \mu\} \oplus B_{r_2}(y)\right) \mathrm{d}\mu. \end{array} \tag{12}$$

It follows from Property (4) that

1. when $\lambda < 2/r_1$, $\{x : u(x) > \mu\} = \emptyset$, for $\mu \in (0, c_1 + c_2)$, minimizes both parts of (12);
2. when $2/r_1 < \lambda < 2/r_2$, $\{x : u(x) > \mu\} = B_{r_1}(y)$, for $\mu \in (0, c_1)$, minimizes the first integral and $\{x : u(x) > \mu\} = \emptyset$, for $\mu \in (c1, c_1 + c_2)$, minimizes the second integral;
3. when $\lambda > 2/r_2$, $\{x : u(x) > \mu\} = B_{r_1}(y)$, for $\mu \in (0, c_1)$, minimizes the first integral and $\{x : u(x) > \mu\} = B_{r_2}(y)$, for $\mu \in (c_1, c_1 + c_2)$, minimizes the second integral.

Noting that $v = f - u$, the solutions given in Property (5) imply the optimal $\{x : u(x) > \mu\}$'s listed above. ■

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] S. Alliney. Digital filters as absolute norm regularizers. *IEEE Trans. on Signal Processing*, 40:1548–1562, 1992.

[3] R. Basri and D. Jacobs. Lambertian reflectance and linear subspaces. *NEC research institute technical report*, 2000.

[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*, 20(7), 1997.

[5] P. N. Belhumeur and D. J. Kriegman. hat is the set of images of an object under all possible lighting conditions? *IEEE International Conference on Computer Vision and Pattern Recognition*, 1996.

[6] T. F. Chan and S. Esedoglu. Aspects of total variation regularized l^1 function approximation. *CLA CAM Report 04-07*, 2004.

[7] 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.

[8] D. Goldfarb and W. Yin. Second-order cone programming methods for total variation-based image restoration. *Columbia CORC TR-2004-05*, 2004.

[9] G. Guo, S. Z. Li, and K. Chan. Face recognition by support vector machines. *FG*, pages 196–201, 2000.

[10] A. Lanitis, C. J. Taylor, and T. F. Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):743–756, 1997.

[11] Y. Meyer. Oscillating patterns in image processing. *University Lecture Series 22, AMS*, 2000.

[12] B. Moghaddam, T. Jebara, and A. Pentland. Bayesian face recogntion. *Pattern Recognition*, 33:1771–1782, 2000.

[13] S. K. Nayar and R. M. Bolle. Reflectance based object recognition. *International Journal of Computer Vision*, 17(3), 1996.

[14] M. Nikolova. Minimizers of cost-functions involving nonsmooth data-fidelity terms. *SIAM Journal on Numerical Analysis*, 40(3):965–994, 2002.

[15] P. Oenev and J. Atick. Local feature analysis: A general statistical theory for object reperesentation. *Network: Computation in Neural System*, 7:477–500, 1996.

[16] 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.

[17] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[18] S. Shan, W. Gao, B. Cao, and D. Zhao. llumination normalization for robust face recognition against varying lighting conditions. *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, 2003.

[19] A. Shashua and T. Riklin-Raviv. The quotient image: Class-based re-rendering and recognition with varying illuminations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):129–139, 2001.

[20] O. D. M. G. W. Site. www.mic-d.com/java/digitalimaging/backgroundsubtraction/.

[21] P. Soille. *Morphological Image Analysis: Principles and Applications*. New York, 1999.

[22] D. Strong and T. Chan. Edge-preserving and scale-dependent properties of total variation regularization. *Inverse problems*, 19:S165–S187, 2003.

[23] M. Turk and A. Pentland. Eigenfaces for recogntion. *Journal of cognitive neuroscience*, 3(1):71–86, 1991.

[24] H. Wang, S. Z. Li, and Y. Wang. Generalized quotient image. *IEEE International Conference on Computer Vision and Pattern Recognition*, 2004.

[25] L. Wiskott, J. M. Fellous, N. Kruger, and C. V. D. Malsburg. Face recogntion by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997.

[26] Y. Zhang, M. Brady, and S. Smith. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Trans. Medical Imaging*, 20(1):45–57, 2001.

[27] W. Zhao and R. Chellappa. Robust face recognition using symmetric shape-from-shading. *Technical report, Center for Automation Research, University of Maryland*, 1999.

[28] 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.