An Efficient Statistical Method for Image Noise Level Estimation

Guangyong Chen[1], Fengyuan Zhu[1], and Pheng Ann Heng[1,2]

[1] Department of Computer Science and Engineering, The Chinese University of Hong Kong
[2]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences

Abstract

In this paper, we address the problem of estimating noise level from a single image contaminated by additive zero-mean Gaussian noise. We first provide rigorous analysis on the statistical relationship between the noise variance and the eigenvalues of the covariance matrix of patches within an image, which shows that many state-of-the-art noise estimation methods underestimate the noise level of an image. To this end, we derive a new nonparametric algorithm for efficient noise level estimation based on the observation that patches decomposed from a clean image often lie around a low-dimensional subspace. The performance of our method has been guaranteed both theoretically and empirically. Specifically, our method outperforms existing state-of-the-art algorithms on estimating noise level with the least executing time in our experiments. We further demonstrate that the denoising algorithm BM3D algorithm achieves optimal performance using noise variance estimated by our algorithm.

1. Introduction

Noise level is an important parameter to many algorithms in different areas of computer vision, including image denoising [3, 5, 6, 7, 16], optical flow [12, 26], image segmentation [1, 4] and super resolution [9]. However, in real world situations the noise level of particular images can be unavailable and is required to be estimated. So far, it still remains to be a challenge to accurately estimate the noise level for different noisy images, especially for those with rich textures. Therefore, a robust noise level estimation method is highly demanded.

One important noise model widely used in different computer vision problems, including image denoising, is the additive, independent and homogeneous Gaussian noise model, where "homogeneous" means that the noise variance is a constant for all pixels within an image and does not change over the position or color intensity of a pixel. The goal of noise level estimation is to estimate the unknown standard deviation σ of the Gaussian noise with a single observed noisy image.

The problem of estimating noise level from a single image is fundamentally ill-posed. During last decades, numerous noise estimating methods [2, 17, 13, 20, 24] have been proposed. However, all of these methods are based on the assumption that the processed image contains a sufficient amount of flat areas, which is not always the case for natural image processing. Recently, new algorithms have been proposed in [19, 23] with state-of-the-art performance. The authors of [19, 23] claim that these methods can accurately estimate the noise level of images without homogeneous areas. However, these methods suffer from the following two weaknesses. Firstly, as concluded in [19], the convergence and performance of selecting low-rank patches is not theoretically guaranteed and does not have high accuracy empirically. Secondly, as theoretically explained in Sec. 2.2 of this paper, both [19] and [23] underestimate the noise level for processed images, since they take the smallest eigenvalue of the covariance of selected low-rank patches as their noise estimation result.

To tackle these problems, we propose a new algorithm for noise level estimation. Our work is based on the observation that patches taken from the noiseless image often lie in a low-dimensional subspace, instead of being uniformly distributed across the ambient space. This property has been widely used in subspace clustering methods [8, 29]. The low-dimensional subspace can be learned by the method of Principal Component Analysis (PCA) [14]. As analyzed in 2.1, the noise variance can be estimated from the eigenvalues of redundant dimensions. In this way, the problem of noise level estimation is reformulated to the issue of selecting redundant dimensions for PCA. This problem has been investigated as a model selection problem in the fields of statistics and signal processing, including [10, 15, 21, 28]. However, these methods focus on using less latent components to represent observed signals. As a result, their meth-

Figure 1. An example to illustrate that patches of a clean image lie in a low-dimension subspace. (a) The clean image. (b) The noisy image with noise $\sigma = 50$. (c) The eigenvalues of clean image and noisy image, where the red curve represents the eigenvalues of clean image and the blue curve represents the eigenvalues of noisy image.

ods always consider signals components as noises resulting in the overestimation of image noise. In this paper, we propose an effective method to solve this issue with the statistical property that the eigenvalues of redundant dimensions are random variables following a same distribution, which is demonstrated in Sec. 2.2. As proved in Sec. 2.3, the proposed method is expected to achieve the accurate noise estimation when the number of principal dimensions is smaller than a threshold.

Thus, the contributions of this paper can be summarized as below:

- The statistical relationship between the noise level σ^2 and the eigenvalues of covariance matrix of patches is firstly estimated in this paper.

- A nonparametric algorithm is proposed to estimate the noise level σ^2 from the eigenvalues in polynomial time, whose performance is theoretically guaranteed. As demonstrated empirically, our method is the most robust and can achieve best performance for noise level estimation in most cases. Moreover, our method consumes least executing time and is nearly 8 times faster than [19, 23].

The rest of the article is organized as following. Based on the patch-based model, we propose our method in Sec. 2. In Sec. 3, we present the comparison of the experimental results with discussions. Finally, the conclusion and some future works are presented in Sec. 4.

2. Our Method: Theory and Implementation

An observable image I can be decomposed into a number of patches $X_s = \{x_t\}_{t=1}^s \in \mathbb{R}^{r \times s}$. Given a multi-channel image I with size $M \times N \times c$, X_s contains $s = (M-d+1)(N-d+1)$ patches of size $d \times d \times c$, whose left-top corner positions are taken from set $\{1, \ldots, M-d+1\} \times \{1, \ldots, N-d+1\}$. To simplify the following calculations, all patches are further rearranged into vectors with

$r = cd^2$ elements in this paper. For any arbitrary vector x_t in the observable set X_s, it can be decomposed as:

$$x_t = \hat{x}_t + e_t, \tag{1}$$

where $\hat{x}_t \in \mathbb{R}^{r \times 1}$ is the corresponding noise-free image patch lying in the low-dimensional subspace, $e_t \in \mathbb{R}^{r \times 1}$ denotes the additive noise and $\mathbb{E}(\hat{x}_t^T e_t) = 0$. As I is contaminated by Gaussian noise $N(0, \sigma^2)$ with zero-mean and variance σ^2, e_t follows a multivariate Gaussian distribution $N_r(0, \sigma^2 \mathbf{I})$ with mean 0 and covariance matrix $\sigma^2 \mathbf{I}$. With such setting, *estimating noise level of an image with the set of patches X_s is equivalent to estimating the noise level σ^2 of the dataset X_s.*

2.1. Eigenvalues

As illustrated in Fig.1 (c), most eigenvalues of the clean image are 0, which confirms our previous discussion: *"patches taken from clean image often lie in a single low-dimensional subspace"*. However, it can be observed that most eigenvalues of the noisy image surround the true noise variance 50 instead of being 50 exactly. Thus, it is still difficult to obtain an accurate noise estimation from eigenvalues directly. To investigate the relationship between eigenvalues and noise level comprehensively, we first formulate the calculation of eigenvalues from another perspective.

Assume that clean patches lie in m-dimensional linear subspace, where m is a predefined positive integer with $m \ll r$, we can formulate equation (1) as follows:

$$x_t = Ay_t + e_t. \tag{2}$$

$A \in \mathbb{R}^{r \times m}$ denotes the dictionary matrix spanning the m-dimension subspace with constraint $A^T A = \mathbf{I}$ and $y_t \in \mathbb{R}^{m \times 1}$ denotes the projection point of x_t on the subspace spanned by A. PCA has been widely used to infer the linear model described in equation (2), where A consists of the m eigenvectors with the m largest eigenvalues of the covariance matrix $\Sigma_x = \frac{1}{s} \sum_{t=1}^s (x_t - \mu)(x_t - \mu)^T$ with $\mu = \frac{1}{s} \sum_{t=1}^s x_t$.

Given an additional matrix $U \in \mathbb{R}^{(r-m) \times r}$, we define a rotation matrix $R = [A, U]$ which satisfies $R^T R = \mathbf{I}$. R can be effectively solved by the eigen-decomposition of the covariance matrix Σ_x, just as PCA. Thus, equation (2) can be rewritten as:

$$x_t = R \begin{bmatrix} y_t, \\ \mathbf{0} \end{bmatrix} + e_t, \qquad (3)$$

where $\mathbf{0}$ means the column vector with size $(r - m) \times 1$ and each element as 0. Multiplying R^T on both size of equation (3) leads to the following equation:

$$R^T x_t = \begin{bmatrix} y_t, \\ \mathbf{0} \end{bmatrix} + R^T e_t = \begin{bmatrix} y_t + A^T e_t, \\ U^T e_t \end{bmatrix}. \qquad (4)$$

It can be observed that the noise component e_t has been separated from the observable x_t by the rotation matrix R, and concentrates on $(m + 1)$-th column to r-th column in the vector $R^T x_t$. Based on the Gaussian properties, $n_t = U^T e_t$ is an random variable following Gaussian distribution $N_{r-m}(0, \sigma^2 \mathbf{I})$

As R consists of the eigenvectors of Σ_x, it satisfies

$$\Sigma_x R = R\Phi, \qquad (5)$$

where $\Phi \in \mathbb{R}^{r \times r}$ is a diagonal matrix with the eigenvalues as its diagonal elements. Thus, the covariance matrix of $R^T x_t$ can be calculated as:

$$\begin{aligned} \frac{1}{s} \sum_{t=1}^{s} (R^T x_t)(R^T x_t)^T &= \frac{1}{s} \sum_{t=1}^{s} R^T x_t x_t^T R \\ &= R^T \Sigma_x R \\ &= \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_r \end{bmatrix}, \end{aligned} \qquad (6)$$

which states that i-th eigenvalue λ_i is also the variance calculated on i-th dimension of vector $R^T x_t$. From equation (4), we can find that the variances on the principal dimensions are expected to be larger than σ, and the variances calculated on the redundant dimensions are expected to be equal to σ. Given $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r$, we can represent $\mathcal{S} = \{\lambda_i\}_{i=1}^{r}$ with $\mathcal{S}_1 \bigcup \mathcal{S}_2$, where $\mathcal{S}_1 = \{\lambda_i\}_{i=1}^{m}$ denotes the variance on principal dimensions and $\mathcal{S}_2 = \{\lambda_i\}_{i=m+1}^{r}$ denotes the variance on redundant dimensions.

Let $n_t[i]$ denote the i-th element of the vector $n_t = U^T e_t$, the eigenvalues in the subset $\mathcal{S}_2$ is denoted as

$$\lambda_i = \frac{1}{s} \sum_{t=1}^{s} n_t[i]^2, \forall i \in \{m+1, m+2, \ldots, r\}, \qquad (7)$$

where $n_t[i]$ is a random variable following Gaussian distribution $N(0, \sigma^2)$. As a function of random variables, each eigenvalue $\lambda_i \in \mathcal{S}_2$ is itself a random value and can be viewed as an estimation of the noise variance σ^2.

2.2. Variance of Finite Gaussian Variables

Both [19] and [23] assumed the minimum eigenvalue λ_r to be the estimation of noise variance σ^2. However, as shown in their experiments, their noise estimation results are consistently smaller than the true noise variance σ^2. In this section, we will give a brief theoretical analysis of this phenomena.

Recall that the eigenvalues of redundant dimensions is the variance of finite Gaussian variables, which is demonstrated in Sec. 2.1, we first prove the following lemma that the eigenvalues of redundant dimensions can be viewed as the realizations of a special Gaussian distribution under certain conditions.

Lemma 1. *Given a set of random variable $\{n_t[i]\}_{t=1}^{s}$ with each element following Gaussian distribution $N(0, \sigma^2)$ independently, the distribution of the noise estimation $\hat{\sigma}_i^2 = \frac{1}{s} \sum_{t=1}^{s} n_t[i]^2$ converges to the distribution of $N(\sigma^2, \frac{2\sigma^4}{s})$ when s is large.*

Proof. Since $n_t[i]$ is generated from the Gaussian distribution $N(0, \sigma^2)$, it can be proved that the random variable $\frac{n_t[i]}{\sigma}$ follows the standard normal distribution $N(0, 1)$. By the definition of Chi-squared distribution, we get

$$\sum_{t=1}^{s} (\frac{n_t[i]}{\sigma})^2 \sim \chi_s^2.$$

Combined with equation (7), we can obtain

$$\frac{s}{\sigma^2} \hat{\sigma}_i^2 \sim \chi_s^2.$$

From the properties of Chi-square distribution,

$$\frac{\chi_s^2 - s}{\sqrt{2s}} \xrightarrow{d} N(0, 1) \qquad (8)$$

as $s \to \infty$. Then we have

$$\sqrt{s} \frac{\hat{\sigma}_i^2 - \sigma^2}{\sqrt{2}\sigma^2} \xrightarrow{d} N(0, 1),$$

which implies that the noise estimation $\hat{\sigma}_i^2$ also converges to a Gaussian distribution when s is large based on the property of Gaussian distribution. Let $w = \sqrt{s} \frac{\hat{\sigma}_i^2 - \sigma^2}{\sqrt{2}\sigma^2}$, then $\hat{\sigma}_i^2 = \sigma^2 + \frac{\sqrt{2}\sigma^2}{\sqrt{s}} w$. Consequently, we can write the probability distribution function (pdf) of $\hat{\sigma}_i^2$ as below

$$\hat{\sigma}_i^2 \sim N(\sigma^2, \frac{2\sigma^4}{s}). \qquad (9)$$

$\square$

Lemma 1 can be further demonstrated empirically by Monte Carlo simulation, where the sample variance is estimated 10^6 times repeatedly. In each trial, $s = 10^3$ samples are drawn from a normal distribution $N(0, \sigma^2)$ with
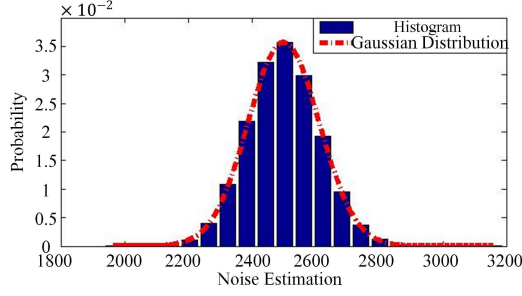
Figure 2. The simulation results for the variance estimation, which shows that the distribution of the variance can be well approximated by a Gaussian distribution.
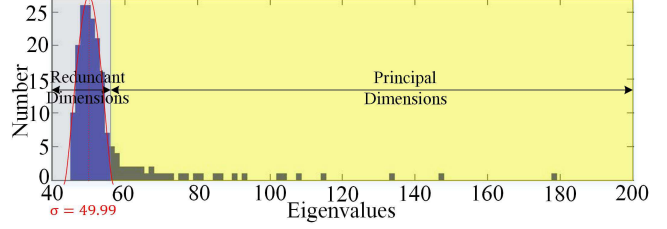

Figure 3. The histogram of eigenvalues of Fig.1 (b). It can be observed that the eigenvalues follow a Gaussian distribution, despite several large outliers.

$\sigma^2 = 50^2$. The histogram of the sample variance estimations of all trials is drawn in Fig.2, whose result is coincided with Lemma 1 and states that the histogram of noise variance $\hat{\sigma}_i^2$ can be effectively approximated by a Gaussian distribution.

For the issue of noise estimation, the number of decomposed patches $s = (M - d + 1)(N - d + 1)$ is typically large enough. Let $d = 8$ by default, an observed image with resolution 512×512 has $s = (512 - 8 + 1)^2 = 255025$ decomposed patches, which means that there are 255025 samples for the estimation of the eigenvalue of each redundant dimension. As illustrated in Fig.2, 1000 samples are large enough for the satisfaction of equation (8). Thus, from the equation (7) with lemma 1, we can derive that the eigenvalues of redundant dimensions $\{\lambda_i\}_{i=m+1}^r$ are the realizations of Gaussian distribution $N(\sigma^2, \frac{2\sigma^4}{s})$.

Let $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$ denote the cumulative distribution function of a standard Gaussian distribution, Blom [25, 27] has proven the following theorem.

Theorem 1. *Given n independent random variables $x_1, x_2, \ldots, x_n$ generated from the normal distribution $N(\sigma^2, \nu^2)$ with order $x_1 \geq x_2 \geq \ldots \geq x_n$, then the expected value of x_i can be approximated by $\mathbb{E}(x_i) \approx \sigma^2 + \Phi^{-1}(\frac{n-\alpha+1-i}{n-2\alpha+1})\nu$ with $\alpha = 0.375$.*

In both [19] and [23], researchers chose the minimum eigenvalues as their noise estimation. However, the eigenvalues in $\mathcal{S}_2$ follow the Gaussian distribution. Based on theorem 1, we can obtain that the expectation of the minimum eigenvalue is $\sigma^2 + \nu\Phi^{-1}(\frac{1-\alpha}{m-r-2\alpha+1})$ with $\nu = \frac{2\sigma^4}{s}$. Since points in the decomposed set X_s lie around a low-dimension subspace, the number of redundant dimensions $m - r$ is larger than 1 naturally, we obtain $m - r - \alpha > 1 - \alpha$ and $\frac{1-\alpha}{m-r-2\alpha+1} < 0.5$. Finally, we obtain $\sigma^2 + \nu\Phi^{-1}(\frac{1-\alpha}{m-r-2\alpha+1}) < \sigma^2$. It can be seen that the smallest eigenvalue is essentially smaller than the noise level when the number of redundant dimension is larger than 1. The results given by [19, 23] tend to be more inaccurate as the number of redundant dimensions increases.

2.3. Dimension Selection

As discussed earlier, more accurate estimation can be obtained by taking all elements in the $\mathcal{S}_2$ into consideration and the problem of underestimating noise variance can be avoided. Fig.3 draws the histogram of eigenvalues calculated from the noisy image, Fig.1 (b), which consists of the eigenvalues of redundant dimensions and some upper outliers. However, the number of redundant dimensions is not known previously. In this section, we propose a new algorithm for dimension selection which is illustrated in Alg.1. The number of principal dimensions is estimated by indicating whether there are upper outliers in the $\mathcal{S}_2$ or not. Starting with the initial status with $\mathcal{S}_1 = \emptyset$ and $\mathcal{S}_2 = \mathcal{S}$, we remove the largest value in $\mathcal{S}_2$ until there are not any outliers in $\mathcal{S}_2$ based on our criterion. A theoretical bound is provided to guarantee its performance, as stated in the following theorem.

Theorem 2. *If $\forall x_t \in X_s$ lie in a m-dimension subspace, and m satisfies the following condition*

$$m < r - \frac{(1 - 2\alpha)\delta(\beta) + \alpha}{1 - \delta(\beta)}, \qquad (10)$$

where $\delta(\beta) = \Phi\left((1 + \frac{1}{\beta})\Phi^{-1}(0.5 + \frac{\beta}{2})\right)$ is a function with respect to $\beta = \frac{m}{r}$. Then, for the mean vector μ of the dataset $\mathcal{S}$, the following two statements hold.

1. *μ is expected to be larger than the median value of $\mathcal{S}$ when there are any upper outliers in the dataset $\mathcal{S}$ or $m > 0$.*

2. *μ is expected to be equal to the median value of $\mathcal{S}$ when there are no upper outliers in the dataset $\mathcal{S}$ or $m = 0$.*

Proof.
For the 1st Statement
If $\forall x_t \in X_s$ lie in a m-dimension subspace, then there are m outliers in the set of eigenvalues $\mathcal{S}$. Thus, the expected

mean μ of the set $\mathcal{S}$ can be given as

$$
\begin{aligned}
\mathbb{E}\mu &= \sigma^2 + \frac{1}{r}\sum_{i=1}^{m}(\mathbb{E}\lambda_i - \sigma^2) \\
&\geq \sigma^2 + \beta(\mathbb{E}\lambda_m - \sigma^2) \\
&= \sigma^2 + \beta\nu\Phi^{-1}(\frac{r-m-\alpha}{r-m-2\alpha+1}),
\end{aligned}
\tag{11}
$$

where $\nu = \frac{\sqrt{2}\sigma^2}{\sqrt{s}}$.

Let $k = r - m$ denote the number of eigenvalues of redundant dimensions, it can be indicated that there are k variables following the Gaussian distribution $N(\sigma^2, \nu^2)$ in dataset $\mathcal{S}$, as proved in lemma 1. For an arbitrary sample drawn from Gaussian distribution $N(\sigma^2, \nu^2)$, it is smaller than the value $\mathbb{E}\mu$ with the probability $\Phi(\frac{\mathbb{E}\mu-\sigma^2}{\nu})$. Thus, in dataset $\mathcal{S}$, the expected number of samples that is smaller than $\mathbb{E}\mu$ can be denoted as

$$
\begin{aligned}
&k\Phi(\frac{\mathbb{E}\mu-\sigma^2}{\nu}) \\
&\geq k\Phi\left(\beta\Phi^{-1}(\frac{k-\alpha}{k-2\alpha+1})\right)
\end{aligned}
\tag{12}
$$

Based on the definition of $\Phi(x)$, we can get $1-\delta(\beta) > 0$. Thus, the given condition (10) is equivalent to

$$
\frac{k-\alpha}{k-2\alpha+1} > \delta(\beta)
$$
$$
\Updownarrow
$$
$$
\frac{k-\alpha}{k-2\alpha+1} > \Phi\left((1+\frac{1}{\beta})\Phi^{-1}(0.5+\frac{\beta}{2})\right)
$$
$$
\Updownarrow
$$
$$
\Phi\left(\frac{\beta}{1+\beta}\Phi^{-1}(\frac{k-\alpha}{k-2\alpha+1})\right) > \frac{1+\beta}{2}
$$
$$
\Updownarrow
$$
$$
k\Phi\left(\beta\Phi^{-1}(\frac{k-\alpha}{k-2\alpha+1})\right) > \frac{r}{2}.
\tag{13}
$$

As $\Phi(x)$ is an increasing function, $\mathbb{E}\mu$ is larger than the median value of the set $\mathcal{S}$ when $\beta > 0$.

For the 2nd Statement

When $\beta = 0$, there exist no outliers in the dataset $\mathcal{S}$ since $N(\sigma^2, \nu^2)$ is a symmetric function and the mean value is expectedly to be the median value of the dataset $\mathcal{S}$ absolutely.

In summary, we can conclude that we can decide whether there exists no outliers in $\mathcal{S}$ by checking if μ is its median value, when there are more than $\frac{(1-2\alpha)\delta(\beta)+\alpha}{1-\delta(\beta)}$ redundant dimensions in the dataset $\mathcal{S}$, whose eigenvalues follow Gaussian distribution $N(\sigma^2, \nu^2)$.
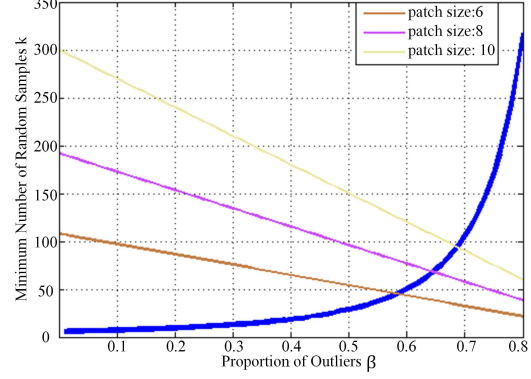
$\square$

Figure 4. The blue line plots the function $k = \frac{(1-2\alpha)\delta(\beta)+\alpha}{1-\delta(\beta)}$ and the brown line plots the function $k = 192(1-\beta)$. It indicates the relationship between k and β when the patch size is 8. The intersection point of these two curves is 65%, which means that we only require $192(1-\beta) = 67$ elements following Gaussian distribution $N(\sigma^2, \nu)$ for accurate estimation of noise variance σ^2

As illustrated in Fig.4, the dimension of subspace must be smaller than a threshold when estimating its mean value. In this paper, the patch size is set as 8 by default, then the dimension r of $X = \{x_t\}_{t=1}^{s}$ is 192. As shown in Fig.4, our method is expected to achieve the accurate mean value of Gaussian distribution when the dimension of subspace is smaller than $m = 192 \times (1-\beta) = 67$. During the procedure, the statements described in theorem 2 always hold as the number of random variables k don't change during the procedure, while $\beta \to 0$ along with the decreasing of the number of outliers m, which leads to a lower requirement of the number k.

2.4. Implementation and Time Complexity

The core idea of our method is to obtain the accurate noise variance by removing the eigenvalues of principal dimensions $\mathcal{S}_1$ from $\mathcal{S}$. As shown in Alg.1, we first decompose the observed image I into the set of patches with patch size d, and calculate the eigenvalues $\mathcal{S} = \{\lambda_i\}_{i=1}^{r}$ of the decomposed dataset X_s. Initialized with $\mathcal{S}_1 = \emptyset$ and $\mathcal{S}_2 = \mathcal{S}$, our method use the difference between the mean value μ and the median value φ of the subset $\mathcal{S}_2$ to indicate whether there are outliers in the subset $\mathcal{S}_2$ or not. If $\mu \neq \varphi$, the largest value in $\mathcal{S}_2$ is taken out and put into the subset $\mathcal{S}_1$. This procedure will stop until the condition $\mu = \varphi$ is achieved. The MATLAB version of the source code is included in the supplementary file.

According to the steps described in Alg.1, we can analyze the time complexity of our method step by step. The complexity of generating sets X with s $r-$dimensional samples from the observed image I is $\mathcal{O}(sr)$. Calculating the mean vector μ and covariance matrix Σ of the dataset X are $\mathcal{O}(sr)$ and $\mathcal{O}(sr^2)$ respectively. The eigen-decomposition of Σ is $\mathcal{O}(r^3)$. Finally, the sorting process in

the 4th step consumes $\mathcal{O}(r^2)$ and the checking procedure in the $5 - 9$the steps take $\mathcal{O}(r^2)$. Thus, the time complexity of our method is $\mathcal{O}(sr^2 + r^3)$, which means that our algorithm is very fast and can be solved in polynomial time.

Algorithm 1 Estimating Image Noise Level

Require: Observed Image $I \in \mathbb{R}^{M \times N \times c}$, Patch Size d.

1: Generating dataset $X = \{x_t\}_{t=1}^s$, which contains $s = (M - d + 1)(N - d + 1)$ patches with size $r = cd^2$ from the image I.
2: $\mu = \sum_{t=1}^s x_t$
3: $\Sigma = \frac{1}{s} \sum_{t=1}^s (x_t - \mu)(x_t - \mu)^T$
4: Calculating the eigenvalues $\{\lambda_i\}_{i=1}^r$ of the covariance matrix Σ with $r = d^2$ and order $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r$
5: **for** i = 1:r **do**
6:    $\tau = \frac{1}{r-i+1} \sum_{j=i}^r \lambda_j$
7:    **if** τ is the median of the set $\{\lambda_j\}_{j=i}^r$ **then**
8:       $\sigma = \sqrt{\tau}$ and **break**
9:    **end if**
10: **end for**
11: **return** noise level estimation σ

---

3. Results and Discussions

To evaluate the performance of the proposed method, we apply it on two benchmark datasets respectively: TID2008 [22] and BSDS500 test set [1]. To further evaluate its performance on real noisy images, we apply our method on 100 noisy images of a static scene captured by a commercial digital camera under low-light condition. We compare its performance with two state-of-the-art methods [19, 23], whose source codes can be downloaded from their homepage [1] [2]. For a fair comparison, all the methods are implemented in the environment of Matlab R2013a (Intel Core(TM) i7 CPU 920 2.67GHz$\times$ 4). For [19] and [23], we use the default parameters reported in their papers. As [19] estimates the noise level for each channel separately when handling colorful image, the final noise level estimation is obtained by averaging the noise level estimation of each channel. The effectiveness of our approach is finally evaluated by using the automatically estimated noise level as input parameter to the denoising algorithm of BM3D on both synthetic and real noisy images.

3.1. Parameter Configurations

As described in Alg. 1, the patch size d is the only free parameter required to be pre-specified. In this section, we mainly discuss the influence of d on the performance of our method.

---

[1] http://www.ok.ctrl.titech.ac.jp/res/NLE/AWGNestimation.html
[2] http://physics.medma.uni-heidelberg.de/cms/projects/132-pcanle

For patch size defined as $r = cd^2$, there is a trade-off between the statistical significance of the result and the practical executing time. As there exist correlation between pixels for images with texture, the patch size should be large enough to represent the texture patterns. Moreover, larger patch size d will lead to a higher β value, which can be observed in Fig.4. It means that larger patch size can handle higher proportion of outliers. However, the patch size d cannot be arbitrary large, since larger patch size d will lead to smaller sample size s of the generated dataset $X = \{x_i\}_{t=1}^s$, which will be contradict with the assumption made in Lemma 1 that s should be large enough. Moreover, as described previously, the time complexity of our method is related to the patch size $r = cd^2$. A larger patch size will lead to longer running time. In summary, the suitable patch size should be chosen with the consideration of the running speed and the assumption used in Lemma 1. In this paper, the patch size d is chosen to be 8 by default for all experiments.

3.2. Noise Level Estimation Results

We add synthetic white noise with different known variance to each testing image, and estimate the noise level in the modified noisy image with different algorithms. As discussed below, three measurements are used for the evaluation of our performance.

3.2.1 Performance Evaluation Criterion

Given the observed image $\mathcal{I}$ with noise variance σ, each method can be considered as a function $\hat{\sigma} = f(\mathcal{I})$ and the optimal estimator should minimize the expected square error (MSE) which is defined as following:

$$\mathbb{E}(f(\mathcal{I}) - \sigma)^2. \tag{14}$$

No matter which method (function f) we use, we can decompose the mean square error (MSE) (14) as below:

$$MSE = \text{Bias}^2(f(\mathcal{I})) + \text{Std}^2(f(\mathcal{I})), \tag{15}$$

where $\text{Std}(f(\mathcal{I})) = \sqrt{\mathbb{E}[f(\mathcal{I}) - \mathbb{E}(f(\mathcal{I}))]^2}$ denotes the robustness of an estimator $f(\mathcal{I})$ and $\text{Bias}(f(\mathcal{I})) = \mathbb{E}|\sigma - \mathbb{E}(f(\mathcal{I}))|$ denotes the accuracy of an estimator. Thus, in this paper, we utilize the following three measurements to evaluate the performance of different estimators,

- Bias: evaluate the accuracy of an estimator.

- Std: evaluate the robustness of an estimator.

- $\sqrt{MSE}$: evaluate the overall performance of an estimator.

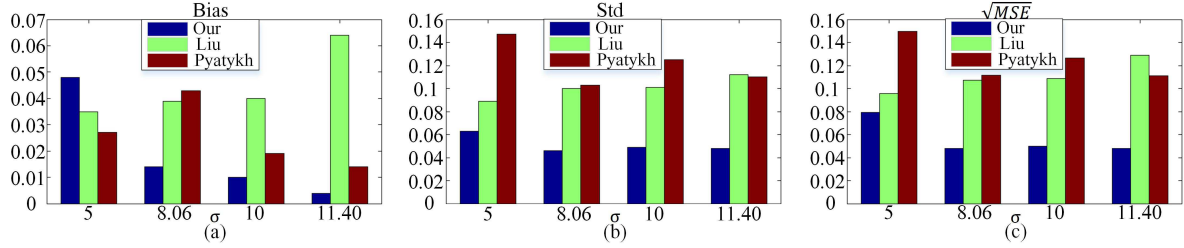Note that smaller Bias, Std or $\sqrt{MSE}$ value means better performance.

Figure 5. The performance of different methods on TID2008. Our method outperforms other two algorithms from the perspective of both accuracy and robustness.

Method	$\min t(s)$	$\bar{t}(s)$	$\max t(s)$
Proposed	**0.5409**	**0.5785**	**0.6435**
Liu et al.	3.5541	4.1901	4.9197
Pyatykh et al.	3.3829	3.4462	5.0448

Table 1. Execution Time of the different methods on TID2008. The proposed algorithm is the fastest.

3.2.2 Performance on TID2008

The TID2008 dataset has been widely used for the evaluation of full-reference image quality assessment metrics, which contains 25 reference images without any compression. As described in [19, 23], the reference images from TID2008 still contain a small level of noise. Thus, the same experimental settings are applied in our experiments, where each estimation result is corrected by the equation $\sigma_{corr}^2 = \sigma_{est}^2 - \sigma_{ref}^2$, with σ_{est}^2 being the estimation result given by each method and σ_{ref}^2 denoting the inherent noise contained in each reference image. To make a fair comparison, we take the σ_{ref}^2 calculated by [23] for each reference image.

As shown in Fig.5, the proposed algorithm outperforms other two algorithms in most cases by removing the upper outliers in the set of eigenvalues $\mathcal{S}$. It can be observed that [23] outperforms our method when $\sigma^2 = 25$, but [23] is less robust in this situation, as shown in Fig.5 (b). Note that the results of [19, 23] shown in Fig.5 are coincident with the results reported in their original papers.

Furthermore, the average execution time of either [23] or [19] is nearly 4 seconds per image as both of them require to select low-rank patches during their procedure. In comparison, the proposed method takes only 0.5 seconds per image, which is almost 8 times faster than the previous methods. The specific executing time of each algorithm can be found in Tab. 1.

3.2.3 Performance on BSDS500

It has been argued that all images in the TID2008 dataset contain small or large homogeneous areas [23]. To further demonstrate the robustness of our method on images full of

textures, we compare the performance of different methods on a more challenging dataset, 200 test images of BSDS500 [1]. Images in BSDS are achieved in different environments and contain rich features. Fig.1 shows the image ♯77062 in the test set of BSDS500, which contains rich texture information and is hard to classify the homogeneous field for noise estimation. To evaluate the performance of different methods for a larger range of noise level, we synthesize 5 noisy images with different noise levels from $\sigma_n = 10$ to 50 for each image. As illustrated in Fig.6, our method outperforms other two algorithms under all three measurements significantly. Moreover, we have compared our method with Gavish's work [10], our method achieves better performance, e.g. on BSDS500 with $\sigma_n = 50$, the $\sqrt{MSE}$ of Gavish's method is 1.612, while ours is 0.183.

3.2.4 Performance on the Real-world Dataset

Figure 7. (a) An example of the 100 images captured by Nikon D5200. (b) An enlarged patch of this image. (c) The clean version of this patch.

We further evaluate the performance of our method on real-world noisy images. Following the experimental setting in [18], we capture 100 images of a static scene under low-light condition with Nikon D5200 (ISO 6400, exposure time 1/20s and aperture f/5) and the mean of the 100 images is served as the clean image. With the clean image and 100 noisy images, the noise variance of each image can be easily estimated as the ground truth variance, e.g. the ground truth σ of Fig.7 (a) is 5.9549. Then we apply both our method and competitive ones on each of the 100 images
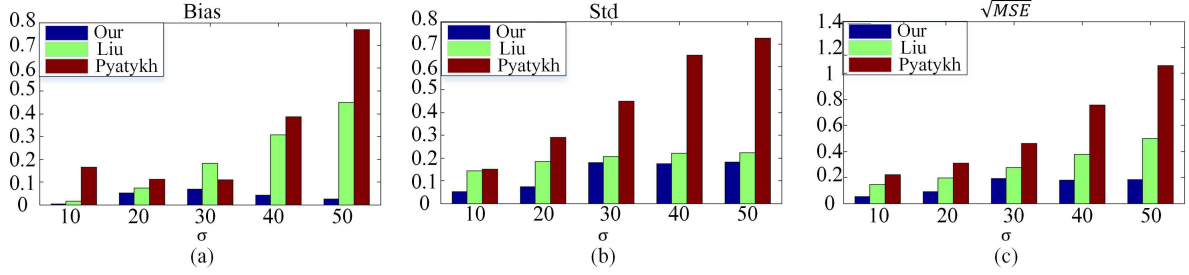
Figure 6. The performance of different methods on the test dataset of BSDS500. Our method outperforms other two algorithms from the perspective of both accuracy and robustness.

Method	Bias	Std	$\sqrt{\text{MSE}}$
Proposed	**0.1198**	0.2234	**0.2535**
Liu et al.	3.3463	0.0957	3.3476
Pyatykh et al.	3.3919	0.0957	3.3932

Table 2. Performance of the different methods on the real-world dataset. The proposed algorithm significantly outperforms others.

individually for noise estimation. Table 2 demonstrates the experimental result and the proposed one significantly outperforms the other two. It shows that our method is more suitable for practical noise estimation applications.

3.3. Image Denoising Results

Image denoising is a very important preliminary step for many computer vision methods. During last decades, impressive improvements have been made in this area. However, the noise level σ is regarded as a known parameter in many algorithms with good performance. BM3D [6] is one of such algorithms which requires noise level as an input parameter. In this section, we apply our noise level estimation algorithms in the application of BM3D image denoising.

All images in TID2008 and BSDS500 are used here to evaluate the effectiveness of using our noise level estimation results as the input parameter of BM3D. As confirmed by Fig.8 (a), the performance of BM3D with our estimated noise level is almost the same as that with true noise level. Due to the article limitation, we only use PSNR [11] to evaluate the denoising performance. A real-world denoising examples using BM3D with our estimation results is demonstrated in Fig.8 (b) and (c).

4. Conclusion and Future Works

In this paper, we propose an efficient method to estimate noise variance from a single noisy image automatically, which is very important for different computer vision algorithms. The performance of this method has been theoretically guaranteed. Furthermore, by comparing our approach with two state-of-the-art ones, we show that the accuracy of the proposed method is the best in most cases with
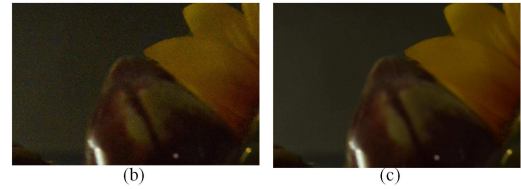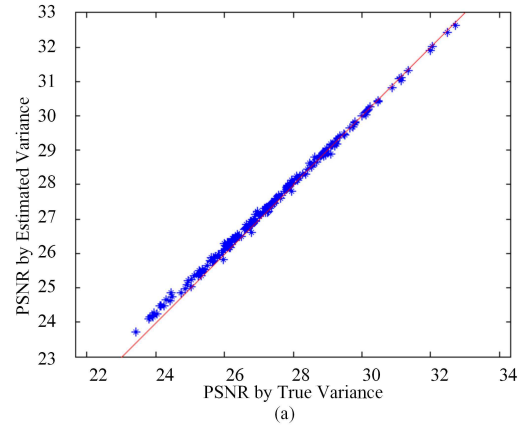

Figure 8. (a) compares the performance of BM3D between our estimated noise level and the true noise level. (b) and (c) show a real noisy image and the denoising image with BM3D+our method.

the least execution time.

There still exist several future works that we are interested in. In this paper, the noise of an image is assumed to be zero-mean additive Gaussian distributed. However, the noise can be more complicated in real world applications which is required to be investigated. Meanwhile, we are interested in developing efficient blind image denoising algorithm in the future.

References

[1] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(5):898–916, May 2011.

[2] R. Bracho and A. C. Sanderson. Segmentation of images based on intensity gradient information. In *Computer Vision and Pattern Recognition, 1985 IEEE Computer Society Conference on*, page 19–23. IEEE, 1985.

[3] A. Buades, B. Coll, and J.-M. Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation*, 4(2):490–530, 2005.

[4] G. Chen, P.-A. Heng, and L. Xu. Projection-embedded byy learning algorithm for gaussian mixture-based clustering. In *Applied Informatics*, volume 1, pages 1–20. Springer, 2014.

[5] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising with block-matching and 3d filtering. In *Electronic Imaging 2006*, pages 606414–606414. International Society for Optics and Photonics, 2006.

[6] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *Image Processing, IEEE Transactions on*, 16(8):2080–2095, 2007.

[7] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *Image Processing, IEEE Transactions on*, 15(12):3736–3745, 2006.

[8] E. Elhamifar and R. Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(11):2765–2781, 2013.

[9] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael. Learning low-level vision. *International journal of computer vision*, 40(1):25–47, 2000.

[10] M. Gavish and D. L. Donoho. The optimal hard threshold for singular values is. *Information Theory, IEEE Transactions on*, 60(8):5040–5053, 2014.

[11] A. Hore and D. Ziou. Image quality metrics: Psnr vs. ssim. In *ICPR*, volume 34, pages 2366–2369, 2010.

[12] B. K. Horn and B. G. Schunck. Determining optical flow. In *1981 Technical Symposium East*, pages 319–331. International Society for Optics and Photonics, 1981.

[13] J. Immerkaer. Fast noise variance estimation. *Computer vision and image understanding*, 64(2):300–302, 1996.

[14] I. Jolliffe. *Principal component analysis*. Wiley Online Library, 2002.

[15] S. Kritchman and B. Nadler. Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory. *Signal Processing, IEEE Transactions on*, 57(10):3930–3941, 2009.

[16] M. Lebrun, A. Buades, and J.-M. Morel. A nonlocal bayesian image denoising algorithm. *SIAM Journal on Imaging Sciences*, 6(3):1665–1688, 2013.

[17] J.-S. Lee. Refined filtering of image noise using local statistics. *Computer graphics and image processing*, 15(4):380–389, 1981.

[18] C. Liu, W. T. Freeman, R. Szeliski, and S. B. Kang. Noise estimation from a single image. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 901–908. IEEE, 2006.

[19] X. Liu, M. Tanaka, and M. Okutomi. Single-image noise level estimation for blind denoising. *Image Processing, IEEE Transactions on*, 22(12):5226–5237, 2013.

[20] P. Meer, J. Jolion, and A. Rosenfeld. A fast parallel algorithm for blind estimation of noise variance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(2):216–223, 1990.

[21] S. Nakajima, R. Tomioka, M. Sugiyama, and S. D. Babacan. Perfect dimensionality recovery by variational bayesian pca. In *Advances in Neural Information Processing Systems*, pages 971–979, 2012.

[22] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti. Tid2008 - a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics*, 10:30–45, 2009.

[23] S. Pyatykh, J. Hesser, and L. Zheng. Image noise level estimation by principal component analysis. *Image Processing, IEEE Transactions on*, 22(2):687–699, 2013.

[24] K. Rank, M. Lendl, and R. Unbehauen. Estimation of image noise variance. *IEE Proceedings-Vision, Image and Signal Processing*, 146(2):80–84, 1999.

[25] J. Royston. Algorithm as 177: Expected normal order statistics (exact and approximate). *Applied Statistics*, pages 161–165, 1982.

[26] H. Scharr and H. Spies. Accurate optical flow in noisy image sequences using flow adapted anisotropic diffusion. *Signal Processing: Image Communication*, 20(6):537–553, 2005.

[27] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, pages 591–611, 1965.

[28] M. O. Ulfarsson and V. Solo. Dimension estimation in noisy pca with sure and random matrix theory. *Signal Processing, IEEE Transactions on*, 56(12):5804–5816, 2008.

[29] R. Vidal. A tutorial on subspace clustering. *IEEE Signal Processing Magazine*, 28(2):52–68, 2010.