Mathematics

Research article

Application of Fisher information to CMOS noise estimation

Mingying Pan and Xiangchu Feng[*]

School of Mathematics and Statistics, Xidian University, Xi'an, China

* **Correspondence:** Email: xcfeng@mail.xidian.edu.cn; Tel: +8602988201432.

Abstract: Analysis of the accuracy of estimated parameters is an important research direction. In the article, the maximum likelihood estimation is used to estimate CMOS image noise parameters and Fisher information is used to analyse their accuracy. The accuracies of the two parameters are different in different situations. Two applications of it are proposed in this paper. The first one is a guide to image representation. The standard pixel image has higher accuracy for signal-dependent noise and higher error for additive noise, in contrast to the normalised pixel image. Therefore, the corresponding image representation is chosen to estimate the noise parameters according to the dominant noise. The second application of the conclusions is a guide to algorithm design. For standard pixel images, the error of additive noise estimation will largely affect the final denoising result if two kinds of noise are removed simultaneously. Therefore, a divide-and-conquer hybrid total least squares algorithm is proposed for CMOS image restoration. After estimating the parameters, the total least square algorithm is first used to remove the signal-dependent noise of the image. Then, the additive noise parameters of the processed image are updated by using the principal component analysis algorithm, and the additive noise in the image is removed by BM3D. Experiments show that this hybrid method can effectively avoid the problems caused by the inconsistent precision of the two kinds of noise parameters. Compared with the state-of-art methods, the new method shows certain advantages in subjective visual quality and objective image restoration indicators.

Keywords: parameter accuracy; noise estimation; CMOS image; Fisher information; hybrid total least square
Mathematics Subject Classification: 62F12, 65Y04, 68U10, 94A08

1. Introduction

Image denoising is a subject of extensive research. There are many types of image noise, mainly including additive noise, multiplicative noise, and mixed noise, etc. For additive noise, the image

degradation process can be formulated as

$$\mathbf{x}(i, j) = \mathbf{s}(i, j) + k_0 \boldsymbol{\delta}(i, j), \quad (i, j) \in \boldsymbol{\Omega}, \tag{1.1}$$

where $\mathbf{x}(i, j)$ is the pixel value of the observed image $\mathbf{x}$ at (i, j), $\mathbf{s}$ is the original image, $k_0 \boldsymbol{\delta}$ is independent and identically distributed (i.i.d.) Gaussian noise, $\delta(i, j) \sim \mathcal{N}(0, 1), \forall (i, j) \in \boldsymbol{\Omega}$, k_0 is a constant and $\boldsymbol{\Omega}$ is a bounded open subset of $\mathbb{R}^2$ with Lipschitz boundary. The general model for multiplicative noise is

$$\mathbf{x}(i, j) = \mathbf{s}(i, j) \mathbf{N}(i, j), \quad (i, j) \in \boldsymbol{\Omega}, \tag{1.2}$$

where $\mathbf{N}$ is i.i.d. noise with mean 1. Many denoising algorithms have been proposed to solve (1.1) and (1.2), which mainly include image filtering algorithms [1], partial differential equations [2–11], wavelet-based methods [12–15], sparse representation methods [16], non-local means (NL-means) based methods [17–23] and neural network-based methods [24–26], etc. Noise generated by common image sensors usually has specific patterns. In CMOS sensors, we see a fixed-pattern of noise that is a mixture of additive and multiplication Gaussian noise [27]

$$\mathbf{x}(i, j) = \mathbf{s}(i, j) + (k_0 + k_1 \mathbf{s}(i, j)) \boldsymbol{\delta}(i, j), \quad (i, j) \in \boldsymbol{\Omega}, \tag{1.3}$$

where $\mathbf{x}(i, j)$ is the pixel value of the observed image $\mathbf{x}$ at (i, j), $\mathbf{s}$ is the original image, k_0 and k_1 are noise parameters and $\delta \sim \mathcal{N}(0, 1)$. There are many methods to remove the noise, which mainly include equation-based methods [28–35], NL-means based methods [36, 37] and neural network-based methods [38], etc. Although learning-based models achieve good results, they have a strong dependence on the dataset. In contrast, traditional methods are more appealing to us because they do not rely on datasets and achieve better results.

Most of the above methods assume that the noise parameters are known. However, it is unrealistic due to the complexity of the degradation process, which greatly reduces the applicability of the algorithm. Although there are many models for noise estimation [39–51], most of them are for additive or multiplicative noise, and there are relatively few studies for CMOS noise estimation.

Investigation in parameter estimation, based on the maximum likelihood equation is an important research direction [48, 52–54]. Liu et al. [48] applies the maximum likelihood theory to the noise estimation of images. It shows that flat image blocks derive their information mainly from noise, so their noisy parameters are estimated from weak textured image blocks by the maximum likelihood estimation. This method has achieved good results.

Analysing the accuracy of the estimated parameters gives a clear picture of how well the parameters match the real data. It reduces the negative impact of parameter errors on experimental results. This is of great importance in practical applications such as image denoising. There is guidance not only for traditional methods but also for emerging neural network denoising methods.

Fisher information, which originated in the 1930s, is an important concept in mathematical statistics. It is a way of measuring the amount of information that an observable random variable X carries about an unknown parameter θ of a distribution that models X. Ronald Fisher introduced the maximum likelihood estimation method in 1912 and emphasized the role of Fisher information in the asymptotic theory of maximum likelihood estimation [55]. He pointed out its mathematical significance in that it could be used to estimate the variance of the maximum likelihood equation and reflect the accuracy of the estimated parameters. Later, Fisher information was used not only in

mathematics but also in physics [56–58] and machine learning [59, 60]. In this paper, the method in [48] is adapted for use in the CMOS mixed noise model (1.3), and the accuracy of the estimated parameters is analysed using Fisher information.

The remainder of the paper is organized as follows. Section 2 introduces related work on noise estimation and section 3 presents a new estimation algorithm for CMOS mixed noise. Section 4 introduces the definition of Fisher information and gives a theoretical analysis of the noise parameter error. In Section 5, based on the theoretical analysis of noise estimation, we present two applications of the theorem. Subsection 5.1 and 5.2 show guidance for image representation and algorithm design respectively. Finally, we provide some discussion and conclude the paper in Section 6.

2. Related works on noise estimation

Noise estimation methods are mainly divided into two categories: filter-based and block-based.

The key to the filter-based noise estimation method is to extract the differencing image by convolving the noisy image with a specially designed filter. After revealing the image structure, the filtered image is used as noise to estimate the noise level. In 1995, Donoho et al. proposed a classic and commonly used image estimation algorithm in [39], which involves transforming the image to the wavelet domain for estimation. The following year, Immerkaer et al. achieved their goal by designing appropriate filters and operating them on the image, as described in [40]. Zoran and Weiss [41] described the related work of Zoran et al. using the discrete cosine transform. In recent years, filtering-based noise estimation algorithms are evolving and more and more related studies are proposed [42–44]. Although this kind of method can achieve certain performances, the accuracy of the algorithm's estimation is greatly reduced when dealing with images with many textures.

The block-based noise estimation algorithm estimates the noise related information from the appropriate image blocks after division. Research shows that the information of weak textured patches mainly comes from noise, and its noise parameters can be calculated by appropriate processing. In [45], Shin et al. proposed a block-based algorithm in which an image is split into numerous patches. Then smooth patches are selected and the noise level is computed from them. The main issue of the method is how to identify the weak textured or smooth blocks for various scenes in the presence of noise.

In [46], Danielyan et al. estimated image noise by stacking similar blocks in an image together and using a 3D transform, and retaining some of the transformed information as input samples. The essence of the algorithm is to achieve the separation of signal and noise by using the non-local self-similarity of the image. Relatively good results are achieved. However, because the algorithm requires block matching and related transformations, it greatly increases the running time of the algorithm and reduces its efficiency.

In [47], Liu et al. exploited this conclusion and proposed a noise estimation algorithm for Gaussian additive noise. The article considered that the minimum eigenvalue of the noiseless weak texture block is 0. The noise level of the image is the minimum eigenvalue of the covariance matrix of the selected weak texture block obtained by principal component analysis (PCA).

Liu et al. [48] proposed a parameter estimation method for a general signal-dependent noise model that can represent different types of noise. The observed noisy pixel value can be expressed by

$$\mathbf{x}(i, j) = \mathbf{s}(i, j) + k_0 \boldsymbol{\delta}_0(i, j) + k_1 \mathbf{s}(i, j)^\gamma \boldsymbol{\delta}_1(i, j), \quad (i, j) \in \boldsymbol{\Omega} \tag{2.1}$$

where (i, j) is the coordinate position, $\mathbf{x}$ is the observed image, $\mathbf{s}$ is the original image, γ is the exponential parameter, k_0 and k_1 are noise parameters, $\delta_0, \delta_1 \sim \mathcal{N}(0, 1)$ and they are independent. This is a generalized noise model that can represent various types of noise by changing the values of k_0, k_1, and γ. Liu et al. [48] uses the maximum likelihood estimator to estimate noise parameters. Firstly, based on the noise model and the independence of δ_0 and δ_1, the theoretical variance function is determined as

$$\sigma^2 = k_0^2 + k_1^2 s^{2\gamma}, \tag{2.2}$$

where s is the pixel value at any position in the image. Then we select M weak textured blocks and estimate the mean $\hat{s}_i$ and variance $\hat{\sigma}_i^2$ according to the i-th observed weak textured block. The likelihood with selected weak textured patches is

$$\mathcal{L} = \prod_{i=1}^{M} \frac{1}{\sqrt{2\pi(k_0^2 + k_1^2 \hat{s}_i^{2\gamma})}} \exp\left\{-\frac{\hat{\sigma}_i^2}{2(k_0^2 + k_1^2 \hat{s}_i^{2\gamma})}\right\}. \tag{2.3}$$

The energy functional can be derived from negative log-likelihood function as

$$E(\gamma, k_0^2, k_1^2) = \sum_{i=1}^{M} \left[\log(k_0^2 + k_1^2 \hat{s}_i^{2\gamma}) + \frac{\hat{\sigma}_i^2}{k_0^2 + k_1^2 \hat{s}_i^{2\gamma}}\right]. \tag{2.4}$$

The gradient-descent algorithm is applied here to solve the problem, and we get the estimated noise parameters. This method also achieved good results, but the error analysis of the estimated parameters is not given in [48]. In addition, there are many related studies [49, 50].

It is worth noting that (1.3) studied in this paper and (2.1) are different, even though they look similar. More importantly than γ equals 1, the two types of noise in (1.3) are correlated, i.e. $\delta_0 = \delta_1$. In the next section, the noise estimation method in [48] will be modified to fit our model and theoretical proof of the method will be given.

3. Noise estimation in the CMOS image

In (1.3), the noise image $\mathbf{x}$ is known and $\mathbf{s}, k_0, k_1$ are unknown. We need to estimate k_0, k_1 from $\mathbf{x}$. Similar to [48], the noise estimation algorithm proposed in this paper is mainly divided into four steps: 1. Determine the expression of its variance according to the noise model; 2. Select weak textured blocks, and use their information to estimate the mean and variance; 3. Determine likelihood function; 4. Find the energy functional and the values of these parameters when the functional reaches the minimum.

First, the image variance expression is determined. According to the image noise model (1.3), we have

$$\sigma^2 = (k_0 + k_1 s)^2. \tag{3.1}$$

where s is the pixel value at any position in the image. Note the difference between (2.2) and (3.1).

In the second step, according to the gradient information of the noise image blocks, M weak textured blocks of $\sqrt{N} \times \sqrt{N}$ are selected and noted as $\{\mathbf{w}_i\}$, $i = 1, 2, \cdots, M$. For the i-th block $\mathbf{w}_i$, we use the

mean values to approximate the noise-free pixel values and estimate the noise variance of the patches with the power of the noisy block along to the eigenvector associated to the minimum eigenvalue

$$
\begin{aligned}
\bar{s}_i &= \frac{1}{N} \sum_{j=1}^{N} x_j, \\
\hat{\sigma}_i^2 &= \|\mathbf{u}_{\min}^T \cdot \mathbf{x}\|^2,
\end{aligned}
\tag{3.2}
$$

where $\bar{s}_i$ is the noise-free signal estimate of $\mathbf{w}_i$, x_j is the j-th pixel value in the observed block, $\hat{\sigma}_i^2$ is the estimation of the noise variance, $\mathbf{x}$ is the vector representation of $\mathbf{w}_i$, and $\mathbf{u}_{\min}$ is the minimum eigenvector of the covariance matrix of the weak textured blocks.

Then the likelihood of the weak textured blocks requires knowledge of statistics. We think of $\mathbf{w}_i$ as a set in which any element obeys a normal distribution with $\bar{s}_i$ as the mean and $(k_0 + k_1 \bar{s}_i)^2$ as the variance. The square of the difference between the element in $\mathbf{w}_i$ and $\bar{s}_i$ is $\hat{\sigma}_i^2$. The probability function of $\mathbf{w}_i$ is

$$
P(\mathbf{w}_i) = \frac{1}{\sqrt{2\pi(k_0 + k_1 \bar{s}_i)^2}} \exp\left\{-\frac{\hat{\sigma}_i^2}{2(k_0 + k_1 \bar{s}_i)^2}\right\}.
\tag{3.3}
$$

Since k_0 and k_1 are unknown, (3.3) should actually be written as $P(\mathbf{w}_i|k_0, k_1)$, which is the likelihood function of k_0 and k_1. There is

$$
(k_0, k_1) = \arg\max_{k_0, k_1} P(\mathbf{w}_i|k_0, k_1).
\tag{3.4}
$$

For $\{\mathbf{w}_i\}$, $i = 1, 2, \cdots, M$, the likelihood function is

$$
\mathcal{L} = \prod_{i=1}^{M} \frac{1}{\sqrt{2\pi(k_0 + k_1 \bar{s}_i)^2}} \exp\left\{-\frac{\hat{\sigma}_i^2}{2(k_0 + k_1 \bar{s}_i)^2}\right\}.
\tag{3.5}
$$

Finally, for ease of solution, we have subjected (3.5) to negative logarithmic operations and noted it as the energy functional $E(k_0, k_1)$, which is

$$
\begin{aligned}
E(k_0, k_1) &= -\log\left\{\prod_{i=1}^{M} \frac{1}{\sqrt{2\pi(k_0 + k_1 \bar{s}_i)^2}} \exp\left\{-\frac{\hat{\sigma}_i^2}{2(k_0 + k_1 \bar{s}_i)^2}\right\}\right\} \\
&= \sum_{i=1}^{M} \left[\frac{1}{2}\log 2\pi(k_0 + k_1 \bar{s}_i)^2 + \frac{\hat{\sigma}_i^2}{2(k_0 + k_1 \bar{s}_i)^2}\right]
\end{aligned}
\tag{3.6}
$$

The objective becomes to find the values of the parameters when $E(k_0, k_1)$ is minimized. Since the constant coefficients in (3.6) have no effect on the final result, $E(k_0, k_1)$ is updated to

$$
E(k_0, k_1) = \sum_{i=1}^{M} \left[\log(k_0 + k_1 \bar{s}_i)^2 + \frac{\hat{\sigma}_i^2}{(k_0 + k_1 \bar{s}_i)^2}\right].
\tag{3.7}
$$

We use the gradient-descent algorithm to solve the energy functional, from which the estimated results of the two parameters $\hat{k}_0$ and $\hat{k}_1$ are obtained.

The estimation of noise parameters is an iterative process. Because of the presence of noise, the selection of weak texture blocks is inaccurate and multiple iterations can lead to more accurate estimates. The flowchart of the algorithm is shown in Figure 1.
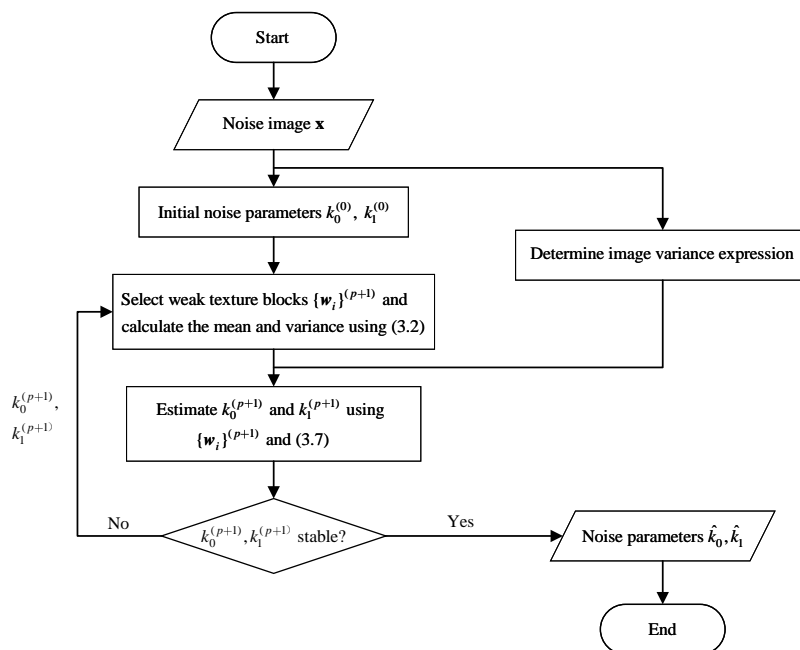
Figure 1. Flowchart of the iterative noise level estimation algorithm.

4. Error analysis of noise parameters

The error analysis is given below using the Fisher information.

Let $X = (X_1, X_2, \ldots, X_n)^T \sim f(x, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p$ be a Cramer-Rao regularity distribution family, $X_1, X_2, \ldots, X_n$ be i.i.d, and $X_1 \sim f(x_1, \boldsymbol{\theta})$. The logarithm of the density function is $l(\boldsymbol{\theta}, x_1) = \log f(x_1, \boldsymbol{\theta})$, and its derivative with respect to the variable $\boldsymbol{\theta}$ is $S(x_1, \boldsymbol{\theta}) = \dot{l}(\boldsymbol{\theta}, x_1) = \frac{\partial \log f(x_1, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$. Then the expectation and variance of $\boldsymbol{\theta}$ are

$$\mathrm{E}_{\boldsymbol{\theta}}[\dot{l}(\boldsymbol{\theta}, x_1)] = 0,$$
$$\mathrm{Var}_{\boldsymbol{\theta}}[\dot{l}(\boldsymbol{\theta}, x_1)] = \mathrm{E}_{\boldsymbol{\theta}}[-\ddot{l}(\boldsymbol{\theta}, x_1)] = i(\boldsymbol{\theta}).$$

Further, since $X \sim f(x, \boldsymbol{\theta}) = \prod\limits_{i=1}^{n} f(x_i, \boldsymbol{\theta})$, we have

$$L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}, x) = \log f(x, \boldsymbol{\theta}) = \sum_{i=1}^{n} l(\boldsymbol{\theta}, x_i).$$

The derivative of the above equation to $\boldsymbol{\theta}$ has the expectation and variance

$$\mathrm{E}_{\boldsymbol{\theta}}[\dot{L}(\boldsymbol{\theta})] = \mathrm{E}_{\boldsymbol{\theta}}[S(X, \boldsymbol{\theta})] = 0,$$
$$\mathrm{Var}_{\boldsymbol{\theta}}[\dot{L}(\boldsymbol{\theta})] = \mathrm{E}_{\boldsymbol{\theta}}[-\ddot{L}(\boldsymbol{\theta})] = I(\boldsymbol{\theta}) = ni(\boldsymbol{\theta}).$$

Definition 1 (Fisher Information)**.** *$i(\boldsymbol{\theta})$ is the Fisher information of X_1, and $I(\boldsymbol{\theta})$ is the Fisher information of X.*

We can derive the relevant properties of the maximum likelihood estimator from Fisher information.

Proposition 1 (Strong Consistency). *Let* $X = (X_1, X_2, \ldots, X_n)^T \sim f(x, \theta)$, $\theta \in \boldsymbol{\Theta} \subset \mathbb{R}^p$ *be a Cramer-Rao regularity distribution family,* $X_1, X_2, \ldots, X_n$ *be i.i.d, and* $\boldsymbol{\Theta} \subset \mathbb{R}^p$ *be an open set. Then the likelihood function* $\dot{L}(\boldsymbol{\theta}) = 0$ *must have a strongly consistent solution* $\hat{\theta}_n(X) = \hat{\theta}(X_1, X_2, \ldots, X_n)$ *when* $n \to +\infty$. *That is, for* $\theta_0 \in \boldsymbol{\Theta}$ *there is*

$$P_{\theta_0}\{X : \lim_{n \to +\infty} \hat{\theta}_n(X) = \theta_0\} = 1, \quad \theta_0 \in \boldsymbol{\Theta}.$$

Proposition 2 (asymptotic normality). *Let* $X = (X_1, X_2, \ldots, X_n)^T \sim f(x, \theta)$, $\theta \in \boldsymbol{\Theta} \subset \mathbb{R}^p$ *be a Cramer-Rao regularity distribution family,* $X_1, X_2, \ldots, X_n$ *be i.i.d, and* $\boldsymbol{\Theta} \subset \mathbb{R}^p$ *be an open set. Assume that the likelihood function* $\dot{L}(\boldsymbol{\theta}) = 0$ *has a consistent solution* $\hat{\theta}_n(X) = \hat{\theta}(X_1, X_2, \ldots, X_n)$ *when* $n \to +\infty$, *and* $L^{(3)}(\boldsymbol{\theta})$ *exists and is continuous in* $\boldsymbol{\Theta}$. *Then* $\hat{\theta}_n(X)$ *is the best asymptotic normal estimate of* θ, *and*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{L} \mathcal{N}(0, i^{-1}(\theta_0)).$$

Using the above propositions, we can derive the variance of the maximum likelihood estimator.

Corollary 1. *The asymptotic normality holds for* $\forall \theta \in \boldsymbol{\Theta}$, *i.e.* $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{L} \mathcal{N}(0, i^{-1}(\theta))$, $\sqrt{n}(\hat{\theta}_n - \theta) = O_p(1)$ *and* $\mathrm{Var}_\theta[\sqrt{n}\hat{\theta}_n(X)] \to i^{-1}(\theta)$, $\forall \theta \in \boldsymbol{\Theta}$.

It is worth noting that $\forall \theta \in \boldsymbol{\Theta}$, $\mathrm{Var}_\theta[\sqrt{n}\hat{\theta}_n(X)] \to i^{-1}(\theta)$, and $I(\theta) = ni(\theta)$, so $\mathrm{Var}_\theta[\hat{\theta}_n(X)] \to I^{-1}(\theta)$.

The parameters obtained by the maximum likelihood estimation all satisfy asymptotic normality, which means that the parameters converge uniformly to normal distributions. The error increases as the standard deviation increases. According to the above propositions and corollary, we can obtain that the variances of $\hat{k}_0$ and $\hat{k}_1$ estimated with (3.7) satisfy the following asymptotic convergence.

Theorem 3. *Let* $a_i = 2(\frac{1}{(k_0 + k_1 \bar{s}_i)^2} - \frac{3\hat{\sigma}_i^2}{(k_0 + k_1 \bar{s}_i)^4})$. *When the number of samples* $n \to \infty$, k_0 *and* k_1 *estimated using* (3.7) *satisfy*

$$\mathrm{Var}[k_0] \quad \to \quad \frac{\sum\limits_{i=1}^{M} a_i \bar{s}_i^2}{(\sum\limits_{i=1}^{M} a_i)(\sum\limits_{i=1}^{M} a_i \bar{s}_i^2) - (\sum\limits_{i=1}^{M} a_i \bar{s}_i)^2},$$

$$\mathrm{Var}[k_1] \quad \to \quad \frac{\sum\limits_{i=1}^{M} a_i}{(\sum\limits_{i=1}^{M} a_i)(\sum\limits_{i=1}^{M} a_i \bar{s}_i^2) - (\sum\limits_{i=1}^{M} a_i \bar{s}_i)^2}.$$

Proof. The Fisher information of the sample is $I(\boldsymbol{k}) = \mathrm{E}_{\boldsymbol{k}}[-\ddot{E}(k_0, k_1)]$, and

$$\ddot{E}(k_0, k_1) = \begin{bmatrix} \ddot{E}_{k_0 k_0} & \ddot{E}_{k_0 k_1} \\ \ddot{E}_{k_1 k_0} & \ddot{E}_{k_1 k_1} \end{bmatrix}.$$

It is necessary to find second order partial derivatives of $E(k_0, k_1)$

$$\ddot{E}_{k_0 k_0} = 2 \sum_{i=1}^{M} (\frac{-1}{(k_0 + k_1 \bar{s}_i)^2} + \frac{3\hat{\sigma}_i^2}{(k_0 + k_1 \bar{s}_i)^4}),$$

$$\ddot{E}_{k_0 k_1} = \ddot{E}_{k_1 k_0} = 2 \sum_{i=1}^{M} (\frac{-\bar{s}_i}{(k_0 + k_1 \bar{s}_i)^2} + \frac{3\hat{\sigma}_i^2 \bar{s}_i}{(k_0 + k_1 \bar{s}_i)^4}),$$

$$\ddot{E}_{k_1 k_1} = 2 \sum_{i=1}^{M} (\frac{-\bar{s}_i^2}{(k_0 + k_1 \bar{s}_i)^2} + \frac{3\hat{\sigma}_i^2 \bar{s}_i^2}{(k_0 + k_1 \bar{s}_i)^4}).$$

So the Hessian matrix is

$$\ddot{E}(k_0, k_1) = \begin{bmatrix} \ddot{E}_{k_0 k_0} & \ddot{E}_{k_0 k_1} \\ \ddot{E}_{k_1 k_0} & \ddot{E}_{k_1 k_1} \end{bmatrix} = \begin{bmatrix} 2\sum\limits_{i=1}^{M}(\frac{-1}{(k_0+k_1\bar{s}_i)^2} + \frac{3\hat{\sigma}_i^2}{(k_0+k_1\bar{s}_i)^4}) & 2\sum\limits_{i=1}^{M}(\frac{-\bar{s}_i}{(k_0+k_1\bar{s}_i)^2} + \frac{3\hat{\sigma}_i^2\bar{s}_i}{(k_0+k_1\bar{s}_i)^4}) \\ 2\sum\limits_{i=1}^{M}(\frac{-\bar{s}_i}{(k_0+k_1\bar{s}_i)^2} + \frac{3\hat{\sigma}_i^2\bar{s}_i}{(k_0+k_1\bar{s}_i)^4}) & 2\sum\limits_{i=1}^{M}(\frac{-\bar{s}_i^2}{(k_0+k_1\bar{s}_i)^2} + \frac{3\hat{\sigma}_i^2\bar{s}_i^2}{(k_0+k_1\bar{s}_i)^4}) \end{bmatrix}.$$

We can get the Fisher information of samples as

$$I(\boldsymbol{k}) = \begin{bmatrix} 2\sum\limits_{i=1}^{M}(\frac{1}{(k_0+k_1\bar{s}_i)^2} - \frac{3\hat{\sigma}_i^2}{(k_0+k_1\bar{s}_i)^4}) & 2\sum\limits_{i=1}^{M}(\frac{\bar{s}_i}{(k_0+k_1\bar{s}_i)^2} - \frac{3\hat{\sigma}_i^2\bar{s}_i}{(k_0+k_1\bar{s}_i)^4}) \\ 2\sum\limits_{i=1}^{M}(\frac{\bar{s}_i}{(k_0+k_1\bar{s}_i)^2} - \frac{3\hat{\sigma}_i^2\bar{s}_i}{(k_0+k_1\bar{s}_i)^4}) & 2\sum\limits_{i=1}^{M}(\frac{\bar{s}_i^2}{(k_0+k_1\bar{s}_i)^2} - \frac{3\hat{\sigma}_i^2\bar{s}_i^2}{(k_0+k_1\bar{s}_i)^4}) \end{bmatrix}.$$

Because $a_i = 2(\frac{1}{(k_0+k_1\bar{s}_i)^2} - \frac{3\hat{\sigma}_i^2}{(k_0+k_1\bar{s}_i)^4})$,

$$I(\boldsymbol{k}) = \begin{bmatrix} \sum\limits_{i=1}^{M} a_i & \sum\limits_{i=1}^{M} a_i\bar{s}_i \\ \sum\limits_{i=1}^{M} a_i\bar{s}_i & \sum\limits_{i=1}^{M} a_i\bar{s}_i^2 \end{bmatrix}.$$

The inverse of $I(\boldsymbol{k})$ is

$$I^{-1}(\boldsymbol{k}) = \frac{\begin{bmatrix} \sum\limits_{i=1}^{M} a_i\bar{s}_i^2 & -\sum\limits_{i=1}^{M} a_i\bar{s}_i \\ -\sum\limits_{i=1}^{M} a_i\bar{s}_i & \sum\limits_{i=1}^{M} a_i \end{bmatrix}}{(\sum\limits_{i=1}^{M} a_i)(\sum\limits_{i=1}^{M} a_i\bar{s}_i^2) - (\sum\limits_{i=1}^{M} a_i\bar{s}_i)^2}.$$

When $n \to \infty$, there is

$$\text{Var}[k_0] \rightarrow \frac{\sum\limits_{i=1}^{M} a_i\bar{s}_i^2}{(\sum\limits_{i=1}^{M} a_i)(\sum\limits_{i=1}^{M} a_i\bar{s}_i^2) - (\sum\limits_{i=1}^{M} a_i\bar{s}_i)^2},$$

$$\text{Var}[k_1] \rightarrow \frac{\sum\limits_{i=1}^{M} a_i}{(\sum\limits_{i=1}^{M} a_i)(\sum\limits_{i=1}^{M} a_i\bar{s}_i^2) - (\sum\limits_{i=1}^{M} a_i\bar{s}_i)^2}.$$

$\square$

The difference between the above two is $\bar{s}_i^2$ on the numerator. Theorem 3 has very important implications for our study.

5. Application of Theorem 3

Theorem 3 shows that the precision of the parameters obtained differs for different estimation objects and has different effects on the results. Therefore the precision of the parameters obtained from Theorem 3 is an important guide to our work. This section focuses on the application of the Theorem 3.

5.1. Guidance for image representation

Popular image processing software usually represents a pixel in 8 bits, so this pixel depth allows 256 different intensities to be recorded. In a greyscale or colour image, a pixel can take on any value between 0 and 255, with each value representing a different brightness. This is known as the standard representation of an image. To facilitate the representation, there is a normalisation of the image pixels. The intensity of a pixel is expressed within a given range between a minimum and a maximum, inclusive. This range is represented in an abstract way as a range from 0 and 1, with any fractional values in between. We use each of these two images as an example to analyse the effect of the image pixel interval on the noise estimation results.

For a standard pixel image, most of the pixel values are greater than 1. This means that the mean $\bar{s}_i$ of any image block in the image is greater than 1. There must be $\bar{s}_i^2 > 1$, so the variance of k_0 is greater than the variance of k_1, i.e. the accuracy of the multiplicative noise parameter is greater than that of the additive noise at this point.

For normalized images with pixel values less than or equal to 1, the mean value of most of the image blocks in the image $\bar{s}_i < 1$. There must be $\bar{s}_i^2 < 1$, so the variance of k_0 is less than the variance of k_1, and the corresponding parameter accuracy is the opposite.

In order to verify the theoretical results, this paper uses different image patterns to represent the images in the standard test set separately. The noise with different parameters is also added and estimated with (3.7).

For images at standard pixels, several sets of noise with different parameters are selected in this paper to analyse the estimation accuracy of the parameters k_0 and k_1 when the image pixel values are between 0 and 255. Figure 2 shows the results. It can be seen that the error of the estimation result of additive noise gets larger as the noise level increases. In other words, the estimated signal-dependent noise parameter k_1 is more accurate, while the additive noise parameter k_0 has a larger error.
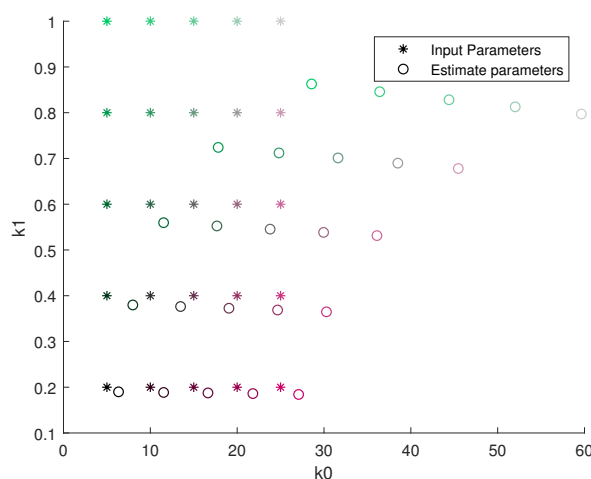

Figure 2. Standard image noise estimation results. $*$ denotes the actual input parameters, $\circ$ denotes the estimated parameters, and the color indicates the correspondence between input and output.

For the images after normalisation, the ratio of the parameter to 255 in the above experiments is

used as the noise parameter in this paper to analyse the estimation accuracy of the parameters k_0 and k_1 when the image pixel values are between 0 and 1, respectively. The results are shown in Figure 3. As can be seen from the graph, the accuracy of the two noise parameters of the normalised image is the opposite of that of the standard image.
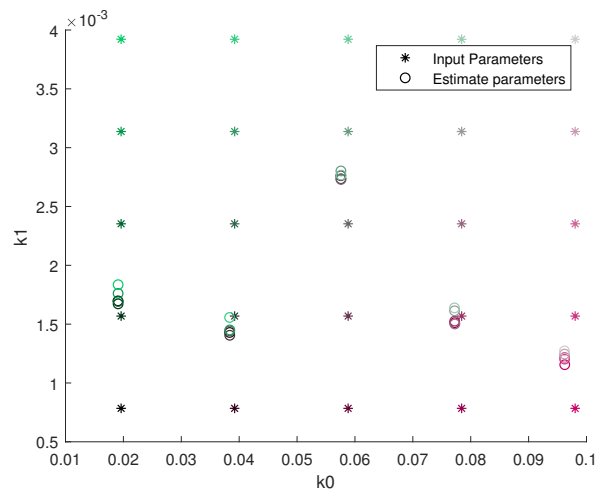

Figure 3. Normalised image noise estimation results. ∗ denotes the actual input parameters, ∘ denotes the estimated parameters, and the color indicates the correspondence between input and output.

The above results enlighten us that different representations are taken to estimate image parameters for different images. For example, the additive noise in the image dominates in the case of relatively large additive noise parameters. In order to make the denoising better, we want it to be estimated more accurately. In this case, the normalised image should be selected to estimate the noise parameters. Similarly, when the multiplicative noise parameter is relatively large, the standard image is chosen to estimate the noise parameter.

In this paper, two types of experiments are conducted on randomly selected images from the standard test image set. In one category, the additive noise level is relatively small while the multiplicative noise is large; in the other category, appropriate parameters are set to make the additive noise level relatively large while the multiplicative noise is small. The denoising experiments were carried out using the method in [36], and the PSNR values were used to measure the effectiveness of the denoising. The experimental results are shown in Table 1.

Table 1. Comparison of PSNR with different noise parameters and different image representations.

Parameter setting	Noisy	Standard	Normalised
$k_0 = 5, k_1 = 1$	5.556	**20.621**	14.961
$k_0 = 5, k_1 = 0.5$	11.560	**24.822**	24.652
$k_0 = 25, k_1 = 0.05$	19.903	28.299	**28.332**
$k_0 = 25, k_1 = 0.1$	19.107	27.924	**28.331**

The above experimental results are in line with our expectations and prove our conclusions correct.

5.2. Guidance for algorithm design

The two parameters of the standard image have different accuracies. Removing both types of noise at the same time may have a negative impact on the results of the image. Therefore, we propose a divide-and-conquer hybrid total least squares algorithm (HTLS). Since image noise is mixed, we suggest to remove the signal-dependent noise first and then the additive noise.

5.2.1. HTLS

1. Signal-Dependent Noise Removal

Let an intermediate variable $\mathbf{y}$ be

$$\mathbf{y}(i, j) = \mathbf{s}(i, j) + k_0 \boldsymbol{\delta}(i, j), \quad (i, j) \in \boldsymbol{\Omega}. \tag{5.1}$$

We estimate the image $\mathbf{y}$ containing only additive noise from the noisy image $\mathbf{x}$ by TLS.

According to (5.1), we know $\mathbf{s}(i, j) = \mathbf{y}(i, j) - k_0 \boldsymbol{\delta}(i, j)$, then (1.3) can be rewritten as

$$\mathbf{x}(i, j) = \mathbf{y}(i, j) + k_1(\mathbf{y}(i, j) - k_0 \boldsymbol{\delta}(i, j))\boldsymbol{\delta}(i, j), \quad (i, j) \in \boldsymbol{\Omega}. \tag{5.2}$$

We need to estimate $\mathbf{y}$ from $\mathbf{x}$.

Select the top n image blocks that are most similar to the target blocks in the noisy image, pull these matrices into column vectors, and reconstitute the matrix X with $X = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]$, where $\boldsymbol{x}_i \in \mathbb{R}^m, i = 1, 2, \ldots, n$ are the column vectors of the image blocks. The total least square problem is

$$\begin{aligned} \min_{\boldsymbol{\alpha}} \quad & \|E, \boldsymbol{e}_0\|_F^2, \\ s.t. \quad & \boldsymbol{y}_0 + \boldsymbol{e}_0 = (X + E)\boldsymbol{\alpha}, \end{aligned} \tag{5.3}$$

where $\boldsymbol{e}_0$ and E are the perturbations of $\boldsymbol{y}_0$ and X, respectively, and $\boldsymbol{y}_0$ is a vector arranged according to the corresponding block of the target block $\boldsymbol{x}_0$ in $\mathbf{y}$.

For solving the optimization problem (5.3), we need to know the singular value decomposition of $[X, \boldsymbol{y}_0]$, i.e. $[X, \boldsymbol{y}_0] = U\boldsymbol{\Sigma}V^T$. But $\boldsymbol{y}_0$ is unknown. Since $\mathbf{x}$ and $\mathbf{y}$ have the same expectation, it is necessary to use their statistical information. We define $\varepsilon\{\cdot\}$ as the expectation operator and define

$$P = [X, \boldsymbol{y}_0]^T [X, \boldsymbol{y}_0] = (U\boldsymbol{\Sigma}V^T)^T (U\boldsymbol{\Sigma}V^T) = V\boldsymbol{\Sigma}^2 V^T.$$

When $m \gg n + 1$, $P \approx \varepsilon\{P\}$, so

$$\begin{aligned} P \ &= \ \varepsilon\{[X, \boldsymbol{y}_0]^T [X, \boldsymbol{y}_0]\} \\ &= \ \begin{bmatrix} \varepsilon\{X^T X\} & \varepsilon\{X^T \boldsymbol{y}_0\} \\ \varepsilon\{\boldsymbol{y}_0^T X\} & \varepsilon\{\boldsymbol{y}_0^T \boldsymbol{y}_0\} \end{bmatrix}. \end{aligned}$$

Let $\boldsymbol{y}_i$ be a column vector arranged by the image block in $\mathbf{y}$ at the same position as $\boldsymbol{x}_i$, and $Y = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n]$, where $\boldsymbol{y}_i \in \mathbb{R}^m, i = 1, 2, \ldots, n$, and let $\boldsymbol{P}_{XX} = \varepsilon\{X^T X\}$, so

$$P = \begin{bmatrix} \boldsymbol{P}_{XX} & Y^T \boldsymbol{y}_0 \\ \boldsymbol{y}_0^T Y & \boldsymbol{y}_0^T \boldsymbol{y}_0 \end{bmatrix}, \tag{5.4}$$

and

$$P_{XX} = Y^T Y - \{k_0 k_1 \sum_{a=1}^{m} (y_{a,i} + y_{a,j})\}_{i,j} + k_1^2 \sum_{a=1}^{m} \text{diag}(y_{a,1}^2, \ldots, y_{a,n}^2) + 2mk_0^2 k_1^2 I + mk_0^2 k_1^2 \mathbf{1}_n, \qquad (5.5)$$

where $\mathbf{1}_n$ is a matrix of ones.

When $m \gg n + 1$, we approximate $\sum_a y_{a,i}$ as $\sum_a x_{a,i}$, which can be computable. It is a fact that the i-th diagonal entry of $Y^T Y$ is $\sum_a y_{a,i}^2$. According to (5.5), $Y^T Y$ can be estimated using the following procedure.

- Compute $P_{XX} = X^T X$;
- Compute $P_{XX} + \{k_0 k_1 \sum_{a=1}^{m} (y_{a,i} + y_{a,j})\}_{i,j} - 2mk_0^2 k_1^2 I - mk_0^2 k_1^2$;
- Divide diagonal elements of the above matrix by $(1 + k_1^2)$.

$Y^T y_0, y_0^T Y$ and $y_0^T y_0$ in (5.4) can be estimated from $Y^T Y$. The new α_{TLS} is computed from $\alpha_{\text{TLS}} = -[v_{1,n+1}, \ldots, v_{n,n+1}]^T v_{n+1,n+1}^{-1}$, where $[v_{1,n+1}, \ldots, v_{n+1,n+1}]^T$ is the right singular vector corresponding to the minimum singular value σ_{n+1} of $[X, y_0]$ and given by the eigen decomposition of P in (5.4).

2. Additive Noise Removal

Additive noise estimation algorithms usually achieve better results. The above process removes the signal-dependent noise from the image to get $\mathbf{y}$, and

$$\mathbf{y}(i, j) = \mathbf{s}(i, j) + k_0' \delta'(i, j), \quad (i, j) \in \Omega,$$

where k_0' is the noise parameter of the new image, which can be estimated directly with [47].

We can use the existing algorithms to remove the additive noise. BM3D, which is more classical and has a better denoising effect, is one of them and can be used in the second step of HTLS.

3. HTLS

The proceeds of HTLS are summarized as follows:

Step 1. Signal-dependent noise removal.

a) Signal-dependent noise estimation. We use the noise estimation algorithm based on weak textured blocks to estimate the noise parameters. We need to determine the noise variance expression of the image and ML estimator, and then find k_0 and k_1.

b) Signal-dependent noise removal. The original image and the additive noise are viewed as a unit $\mathbf{y}$, and the noise parameters estimated earlier are used to solve $\mathbf{y}$.

Step 2. Additive noise removal.

a) Additive noise estimation. k_0 is directly estimated by PCA using selected weak textured blocks.

b) Additive noise removal. BM3D is used to remove the additive noise and the noise-free image $\mathbf{s}$ is obtained.

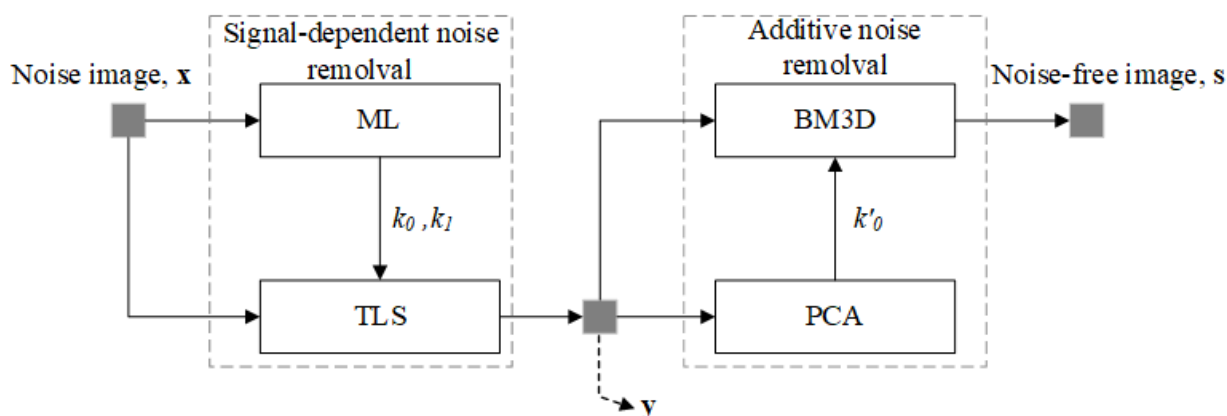The algorithm is illustrated in Figure 4.

Figure 4. Folwchart of HTLS.

5.2.2. Experimental results

We show the experimental results of HTLS. All experiments were implemented on a 64-bit Windows 10 operating system. We measure the quality of the restored images by computing the peak signal-to-noise ratio (PSNR) value.

We randomly select four images from the standard test image set for simulation experiments. They are shown in Figure 5. The first two images are 256×256 and the next two images are 512×512.

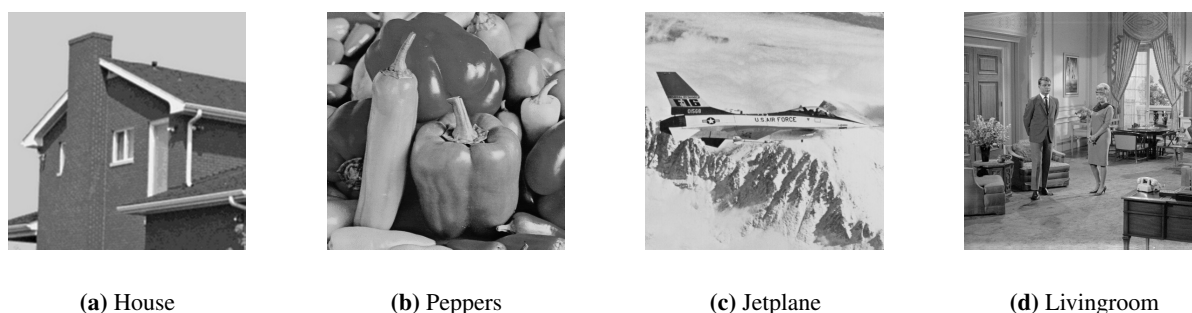**(a)** House   **(b)** Peppers   **(c)** Jetplane   **(d)** Livingroom

Figure 5. Test images.

Adding mixed noise with different noise levels, we can obtain 20 noise images. In this paper, the proposed algorithm is compared with two different algorithms: TLS [36] and Y. Qiu's algorithm [37]. The noise parameters required by the algorithms in [36] are estimated based on the ML estimator. The noise levels set in this paper are $(k_0, k_1) = (5, 0.05), (15, 0.05), (25, 0.05), (15, 0.01), (15, 0.1)$, and the experimental results are shown in Table 2.

From Table 2, we can obtain that for different types of images, the denoising effect of HTLS proposed in this paper is improved compared to TLS. It is worth noting that the adaptive image denoising model proposed by Y. Qiu et al. in [37] has a better denoising effect when the image has obvious bright and dark divisions, but the denoising effect becomes worse when the image brightness division is not obvious. We conducted a large number of experiments and the results are consistent with the above findings.

Figure 6 gives an example denoising results for image jetplane corrupted by signal-dependent noise

Table 2. Denoising performance (PSNR) of different algorithms with different noise parameters.

	(5,0.05)			(15,0.05)			(25,0.05)			(15,0.01)			(15,0.1)		
	[36]	[37]	HTLS	[36]	[37]	HTLS	[36]	[37]	HTLS	[36]	[37]	HTLS	[36]	[37]	HTLS
house	36.17	**36.49**	36.39	32.99	**33.92**	33.84	31.07	**32.22**	32.09	34.33	35.30	**35.32**	31.64	**32.35**	32.13
peppers	34.23	**34.66**	34.44	30.88	**31.55**	31.49	28.91	**29.56**	29.53	32.20	33.08	**33.13**	29.61	**29.96**	29.80
jetplane	34.49	34.61	**34.67**	31.60	31.95	**31.97**	29.83	30.24	**30.25**	33.18	33.60	**33.65**	30.18	30.34	**30.39**
livingroom	32.82	**33.24**	33.17	29.67	30.37	**30.39**	27.93	28.60	**28.63**	30.91	31.60	**31.69**	28.50	29.12	**29.41**

with $k_0 = 25$ and $k_1 = 0.05$. It shows that the proposed method is visually better than other methods used for comparison: the ability of detail representation is better, the edges of the image are clearer, and there are fewer artifacts in the flat regions.

(a) noise image   (b) TLS [36]   (c) Y. Qiu [37]   (d) the proposed HTLS
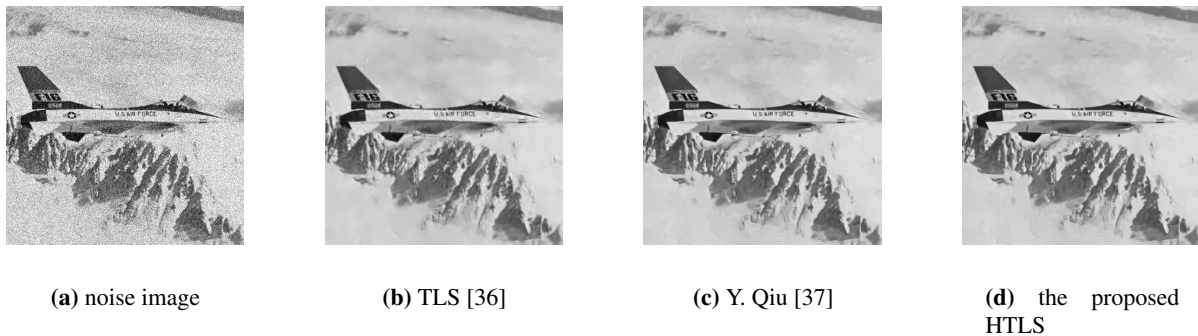
Figure 6. Denoising results.

We conducted a large number of experiments and the results are consistent with the above findings.

6. Conclusions

For the problem of noise estimation of CMOS images, a noise estimation model based on maximum likelihood is given in this paper. The error of the model is given using Fisher information and the results are formed into a theorem. We present two applications of the theorem: one is a guide to image representation. Images represented by standard pixels have higher accuracy for signal-dependent noise and higher error for additive noise. The pixel-normalised image, on the other hand, has the opposite accuracy of its parameters. Therefore, the normalised image is generally chosen for noise estimation when the additive noise is large, and the standard image when the multiplicative noise is dominant. The second application of the theorem is a guide to algorithm design. Based on the results of the theoretical analysis, we propose a divide-and-conquer hybrid total least square (HTLS) algorithm for CMOS image restoration. The method has two steps. In the first step, the estimated noise parameters and TLS are used to remove the signal-dependent noise of the image. In the second step, the additive noise parameter is updated by PCA, and the remaining noise is removed using BM3D. Theoretical analysis and experiments show that this hybrid algorithm can preserve the edge and other details of the image while removing the mixed noise. Compared with the state-of-the-art methods, the new

method shows superiority in subjective visual quality, objective image reconstruction metrics, and other aspects.

Not only traditional denoising algorithms but also most neural network-based denoising methods require that the noise parameters are known. The theory of noise estimation and accuracy analysis can be applied to AI-based image processing to increase the generalizability and robustness of the models.

In our future research, we will focus on the optimisation of the proposed blind denoising algorithm. In fact, the algorithm needs to traverse all the blocks to find similar blocks. Compared with TLS, HTLS adds a new step BM3D. All these operations increase the running time. To obtain better performance, BM3D can also be replaced by other methods such as WNNM [19], deep network-based denoising method [24] and so on. We will study these issues in the future.

Acknowledgments

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, *Sixth international conference on computer vision (IEEE Cat. No. 98CH36271)*, (1998), 839–846. http://dx.doi.org/10.1109/ICCV.1998.710815

2. S. Osher, M. Burger, D. Goldfarb, J. Xu, E. Yin, An iterative regularization method for total variation-based image restoration, *Multiscale Model. Sim.*, **4** (2005), 460–489. http://dx.doi.org/10.1137/040605412

3. Y. zhang, S. Li, B. Wu, S. Du, Image multiplicative denoising using adaptive Euler's elastica as the regularization, *J. Sci. Comput.*, **90** (2022), 69. http://dx.doi.org/10.1007/s10915-021-01721-7

4. L. Rudin, P. L. Lions, S. Osher, *Geometric level set methods in imaging, vision, and graphics*, New York, USA: Springer, 2003. http://dx.doi.org/10.1007/0-387-21810-6_6

5. J. Shi, S. Osher, A nonlinear inverse scale space method for a convex multiplicative noise model, *SIAM J. Imaging Sci.*, **1** (2008), 294–321. http://dx.doi.org/10.1137/070689954

6. K. Bredies, K. Kunisch, T. Pock, Total generalized variation, *SIAM J. Imaging Sci.*, **3** (2010), 492–526. http://dx.doi.org/10.1137/090769521

7. Y. Lv, Total generalized variation denoising of speckled images using a primal-dual algorithm, *J. Appl. Math. Comput.*, **62** (2020), 489–509. http://dx.doi.org/10.1007/s12190-019-01293-8

8. A. Ben-Loghfyry, A. Hakim, A. Laghrib, A denoising model based on the fractional Beltrami regularization and its numerical solution, *J. Appl. Math. Comput.*, **69** (2023), 1431–1463. http://dx.doi.org/10.1007/s12190-022-01798-9

9. T. H. Ma, T. Z. Huang, X. L. Zhao, Spatially dependent regularization parameter selection for total generalized variation-based image denoising, *Comput. Appl. Math.*, **37** (2018), 277–296. http://dx.doi.org/10.1007/s40314-016-0342-8

10. H. Houichet, A. Theljani, M. Moakher, A nonlinear fourth-order PDE for image denoising in Sobolev spaces with variable exponents and its numerical algorithm, *Comput. Appl. Math.*, **40** (2021), 1–29. http://dx.doi.org/10.1007/s40314-021-01462-1

11. A. Hakim, A. Ben-Loghfyry, A total variable-order variation model for image denoising, *AIMS Math.*, **4** (2019), 1320–1335. http://dx.doi.org/10.3934/math.2019.5.1320

12. J. L. Starck, E. J. Candès, D. L. Donoho, The curvelet transform for image denoising, *IEEE T. Image Process.*, **11** (2002), 670–684. http://dx.doi.org/10.1109/TIP.2002.1014998

13. J. Yang, Y. Wang, W. Xu, Q. Dai, Image and video denoising using adaptive dual-tree discrete wavelet packets, *IEEE T. Circ. Syst. Vid.*, **19** (2009), 642–655. http://dx.doi.org/10.1109/TCSVT.2009.2017402

14. L. Fan, X. Li, H. Fan, Y. Feng, C. Zhang, Adaptive texture-preserving denoising method using gradient histogram and nonlocal self-similarity priors, *IEEE T. Circ. Syst. Vid.*, **29** (2019), 3222–3235. http://dx.doi.org/10.1109/TCSVT.2018.2878794

15. Z. Long, N. H. Younan, Denoising of images with multiplicative noise corruption, *2005 13th European Signal Processing Conference*, (2005), 1–4.

16. W. Dong, L. Zhang, G. Shi, X. Li, Nonlocally centralized sparse representation for image restoration, *IEEE T. Image Process.*, **22** (2012), 1620–1630. http://dx.doi.org/10.1109/TIP.2012.2235847

17. A. Buades, B. Coll, J. M. Morel, A non-local algorithm for image denoising, *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, (2005), 60–65. http://dx.doi.org/10.1109/CVPR.2005.38

18. K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, Image denoising by sparse 3-D transform-domain collaborative filtering, *IEEE T. Image Process.*, **16** (2007), 2080–2095. http://dx.doi.org/10.1109/TIP.2007.901238

19. S. Gu, L. Zhang, W. Zuo, X. Feng, Weighted nuclear norm minimization with application to image denoising, *2014 IEEE Conference on Computer Vision and Pattern Recognition*, (2014), 2862–2869. http://dx.doi.org/10.1109/CVPR.2014.366

20. J. Mairal, F. Bach, J. Ponce, G. Sapiro, A. Zisserman, Non-local sparse models for image restoration, *2009 IEEE 12th International Conference on Computer Vision*, (2009), 2272–2279. http://dx.doi.org/10.1109/ICCV.2009.5459452

21. W. Dong, L. Zhang, G. Shi, X. Li, Nonlocally centralized sparse representation for image restoration, *IEEE T. Image Process.*, **22** (2013), 1620–1630. http://dx.doi.org/10.1109/TIP.2012.2235847

22. Q. Guo, C. Zhang, Y. Zhang, H. Liu, An efficient SVD-based method for image denoising, *IEEE T. Circ. Syst. Vid.*, **26** (2016), 868–880. http://dx.doi.org/10.1109/TCSVT.2015.2416631

23. M. Yahia, T. Ali, M. M. Mortula, R. Abdelfattah, S. E. Mahdy, N. S. Arampola, Enhancement of SAR speckle denoising using the improved iterative filter, *IEEE Journal of*

Selected Topics in Applied Earth Observations and Remote Sensing, **13** (2020), 859–871. http://dx.doi.org/10.1109/JSTARS.2020.2973920

24. K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising, *IEEE T. Image Process.*, **26** (2017), 3142–3155. http://dx.doi.org/10.1109/TIP.2017.2662206

25. Y. Meng, J. Zhang, A novel gray image denoising method using convolutional neural network, *IEEE Access*, **10** (2022), 49657–49676. http://dx.doi.org/10.1109/ACCESS.2022.3169131

26. G. Wang, Z. Pan, Z. Zhang, Deep CNN Denoiser prior for multiplicative noise removal, *Multimed. Tools Appl.*, **78** (2019), 29007–29019. http://dx.doi.org/10.1007/s11042-018-6294-9

27. H. Tian, B. Fowler, A. E. Gamal, Analysis of temporal noise in CMOS photodiode active pixel sensor, *IEEE J. Solid-St. Circ.*, **36** (2001), 92–101. http://dx.doi.org/10.1109/4.896233

28. J. Zhang, K. Hirakawa, Improved denoising via Poisson mixture modeling of image sensor noise, *IEEE T. Image Process.*, **26** (2017), 1565–1578. http://dx.doi.org/10.1109/TIP.2017.2651365

29. D. Chen, X. Teng, Novel variational approach for generalized signal dependent noise removal, *2018 11th International Symposium on Computational Intelligence and Design (ISCID)*, **2** (2018), 380–384. http://dx.doi.org/10.1109/ISCID.2018.10187

30. J. Zhang, Y. Duan, Y. Lu, M. K. Ng, H. Chang, Bilinear constraint based ADMM for mixed Poisson-Gaussian noise removal, *arXiv preprint arXiv:1910.08206*, 2019.

31. M. Ghulyani, M. Arigovindan, Fast roughness minimizing image restoration under mixed Poisson–Gaussian noise, *IEEE T. Image Process.*, **30** (2021), 134–149. http://dx.doi.org/10.1109/TIP.2020.3032036

32. S. Huang, T. Lu, Z. Lu, J. Rong, X. Zhao, J. Li, CMOS image sensor fixed pattern noise calibration scheme based on digital filtering method, *Microelectron. J.*, **124** (2022), 10543. http://dx.doi.org/10.1016/j.mejo.2022.105431

33. S. Lee, M. G. Kang, Poisson-Gaussian noise reduction for X-Ray images based on local linear minimum mean square error shrinkage in nonsubsampled contourlet transform domain, *IEEE Access*, **9** (2021), 100637–100651. http://dx.doi.org/10.1109/ACCESS.2021.3097078

34. J. Zhang, K. Hirakawa, Improved denoising via Poisson mixture modeling of image sensor noise, *IEEE T. Image Process.*, **26** (2017), 1565–1578. http://dx.doi.org/10.1109/TIP.2017.2651365

35. A. Repetti, E. Chouzenoux, J. Pesquet, A penalized weighted least squares approach for restoring data corrupted with signal-dependent noise, *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, (2012), 1553–1557.

36. K. Hirakawa, T. W. Parks, Image denoising using total least squares, *IEEE T. Image Process.*, **15** (2006), 2730–2742. http://dx.doi.org/10.1109/TIP.2006.877352

37. Y. Qiu, Z. Gan, Y. Fan, X. Zhu, An adaptive image denoising method for mixture Gaussian noise, *2011 International Conference on Wireless Communications and Signal Processing (WCSP)*, (2011), 1–5. http://dx.doi.org/10.1109/WCSP.2011.6096774

38. J. Byun, S. Cha, T. Moon, FBI-Denoiser: Fast blind image denoiser for Poisson-Gaussian noise, *IEEE Conference on Computer Vision and Pattern Recognition*, (2021), 5768–5777. http://dx.doi.org/10.1109/CVPR46437.2021.00571

39. D. L. Donoho, De-noising by soft-thresholding, *IEEE T. Inform. Theory*, **41** (1995), 613–627. https://doi.org/10.1109/18.382009

40. J. Immerkær, Fast noise variance estimation, *Comput. Vis. Image Und.*, **64** (1996), 300–302. http://dx.doi.org/10.1006/cviu.1996.0060

41. D. Zoran, Y. Weiss, Scale invariance and noise in natural images, *2009 IEEE 12th International Conference on Computer Vision*, (2009), 2209–2216. https://doi.org/10.1109/ICCV.2009.5459476

42. S. Zhu, Z. Yu, Self-guided filter for image denoising, *IET Image Prosess*, **14** (2020), 2561–2566. http://dx.doi.org/10.1049/iet-ipr.2019.1471

43. L. Lu, W. Jin, X. Wang, Non-local means image denoising with a soft threshold, *IEEE Signal Proc. Lett.*, **22** (2015), 833–837. http://dx.doi.org/10.1109/LSP.2014.2371332

44. D. -G. Kim, Y. Ali, M. A. Farooq, A. Mushtaq, M. A. A. Rehman, Z. H. Shamsi, Hybrid Deep Learning Framework for Reduction of Mixed Noise via Low Rank Noise Estimation, *IEEE Access*, **10** (2022), 46738–46752. http://dx.doi.org/10.1109/ACCESS.2022.3170490

45. D. H. Shin, R. H. Park, S. Yang, J. H. Jung, Block-based noise estimation using adaptive Gaussian filtering, *IEEE T. Consum. Electr.*, **51** (2005), 218–226. http://dx.doi.org/10.1109/TCE.2005.1405723

46. A. Danielyan, A. Foi, Noise variance estimation in nonlocal transform domain, *2009 International Workshop on Local and Non-Local Approximation in Image Processing*, (2009), 41–45. https://doi.org/10.1109/LNLA.2009.5278404

47. X. Liu, M. Tanaka, M. Okutomi, Noise level estimation using weak textured patches of a single noisy image, *2012 19th IEEE International Conference on Image Processing*, (2012), 665–668. http://dx.doi.org/10.1109/ICIP.2012.6466947

48. X. Liu, M. Tanaka, M. Okutomi, Estimation of signal dependent noise parameters from a single image, *2013 IEEE International Conference on Image Processing*, (2013), 79–82. http://dx.doi.org/10.1109/ICIP.2013.6738017

49. C. Sutour, A.-C. Deledalle, F.-J. Aujol, Estimation of the noise level function based on a non-parametric detection of homogeneous image regions, *SIAM J. Imaging Sci.*, **8** (2015), 1–31. http://dx.doi.org/10.1137/15M1012682

50. Z. Wang, Z. Huang, Y. Xu, Y. Zhang, X. Li, X. Li, et al., Image Noise Level Estimation by Employing Chi-Square Distribution, *2021 IEEE 21st International Conference on Communication Technology (ICCT)*, (2021), 1158–1161. http://dx.doi.org/10.1109/ICCT52962.2021.9657946

51. V. A. Pimpalkhute, R. Page, A. Kothari, K. M. Bhurchandi, V. M. Kamble, Digital image noise estimation using DWT coefficients, *IEEE T. Image Process.*, **30** (2021), 1962–1972. http://dx.doi.org/10.1109/TIP.2021.3049961

52. J. Sijbers, A. den Dekker, Maximum likelihood estimation of signal amplitude and noise variance from MR data, *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, **51** (2004), 586–594. http://dx.doi.org/10.1002/mrm.10728

53. M. W. Wu, Y. Jin, Y. Li, T. Song, P. Y. Kam, Maximum-likelihood, magnitude-based, amplitude and noise variance estimation, *IEEE Signal Proc. Lett.*, **28** (2021), 414–418. http://dx.doi.org/10.1109/LSP.2021.3055464

54. A. G. Vostretsov, S. G. Filatova, The Estimation of Parameters of Pulse Signals Having an Unknown Form That Are Observed against the Background of the Additive Mixture of the White Gaussian Noise and a Linear Component with Unknown Parameters, *Journal of Communications Technology and Electronics*, **66** (2021), 938–947. http://dx.doi.org/10.1134/S106422692108009X

55. R. A. Fisher, On the mathematical foundations of theoretical statistics, *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, **222** (1922), 309–368. http://dx.doi.org/10.1098/rsta.1922.0009

56. Y. Ouyang, S. Wang, L. Zhang, Quantum optical interferometry via the photon-added two-mode squeezed vacuum states, *J. Opt. Soc. Am. B*, **33** (2016), 1373–1381. https://doi.org/10.1364/JOSAB.33.001373

57. G. C. Knee, W. J. Munro, Fisher information versus signal-to-noise ratio for a split detector, *PHYS. REV. A*, **92** (2015), 012130. http://dx.doi.org/10.1103/PhysRevA.92.012130

58. J. Chao, E. S. Ward, R. J. Ober, Fisher information theory for parameter estimation in single molecule microscopy: tutorial, *J. Opt. Soc. Am. A*, **33** (2016), B36–B57. https://doi.org/10.1364/JOSAA.33.000B36

59. J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, et al., Overcoming catastrophic forgetting in neural networks, *Proceedings of the national academy of sciences*, **114** (2017), 3521–3526. https://doi.org/10.1073/pnas.1611835114

60. J. Martens, New insights and perspectives on the natural gradient method, *The Journal of Machine Learning Research*, **21** (2020), 5776–5851.