

000
001
002054
055
056003 **External Patch Group Prior Guided Internal Prior Learning for Real Image
004 Denoising**057
058
059005
006
007060
061
062008
009
010063
064
065011
012
013066
067
068

014

069

Abstract

For image denoising problem, the external and internal priors are playing key roles in many different methods. External priors learn from external images to restore noisy images while internal ones exploit priors of given images for denoising. The external priors are more generative and efficient on recovering structures existing in most images while the internal priors are more adaptive on recovering details existed in given noisy images. In this paper, we propose to employ the external patch group prior of images to guide the clustering of internal patch groups, and develop an external dictionary guided internal orthogonal dictionary learning algorithm for real image denoising. The internal orthogonal dictionary learning process has closed-form solutions and hence very efficient for online denoising. The experiments on standard datasets demonstrate that, that the proposed method achieves better performance than other state-of-the-art methods on real image denoising.

1. Introduction

Most vision systems, such as medical imaging and surveillance, need accurate feature extraction from high-quality images. The camera sensors and outdoor low light conditions will unavoidably bring noise to the captured images. The impact is that the image details will be lost or hardly visible. As a result, image denoising is an essential procedure for the reliability of these vision systems. In the research area, image denoising is also an ideal platform for testing natural image models and provides high-quality images for other computer vision tasks such as image registration, segmentation, and pattern recognition, etc.

For several decades, there emerge numerous image denoising methods [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11], and all of them focus mainly on dealing with additive white Gaussian noise (AWGN). In real world, the cameras will undertake high ISO settings for high-speed shots on actions, long exposure for low light on night shots, etc. Under these

situations, the noise is generated in a complex form and also been changed during the in-camera imaging pipeline [12, 13]. Therefore, the noise in real images are much more complex than Gaussian [13, 14]. It depends on camera series, brands, as well as the settings (ISO, shutter speed, and aperture, etc). The models designed for AWGN would become much less effective on real noisy images.

In the last decade, the methods of [15, 16, 17, 18, 19, 20, 13] are developed to deal with real noisy images. Almost all these methods employ a two-stage framework: estimating the parameters of the assumed noise model (usually Gaussian) and performing denoising with the help of the noise modeling and estimation in the first stage. However, the Gaussian assumption is inflexible in describing the complex noise on real noisy images [17]. Although the mixture of Gaussians (MoG) model is possible to approximate any noise distribution [21], estimating its parameters is time consuming via nonparametric Bayesian techniques [20]. To evaluate the performance of these methods on dealing with complex real noise, we apply these methods, with corresponding default parameters, on a real noisy image provided in [13]. The testing image is captured by a Nikon D800 camera when ISO is 3200. The "ground truth" image is also provided with which we can calculate objective measurements such as PSNR and SSIM [22]. The denoised images are listed in Figure 1, from which we can see that these methods either remove the noise or oversmooth the complex details in real noisy image.

The above mentioned methods can be categorized into external methods which learn priors from external images to recover noisy images, and internal ones which exploit priors of given images for denoising. The external priors in natural images are free of the high correlation between noise and signals in real noisy images, while the internal prior is adaptive to the image and can recover better the latent clean image. Combining the priors of external clean images and adaptively of internal testing images can naturally improve the performance of denoising methods, especially on real noisy images. Based on these observations, in this paper, we propose to employ the external patch group prior [10]

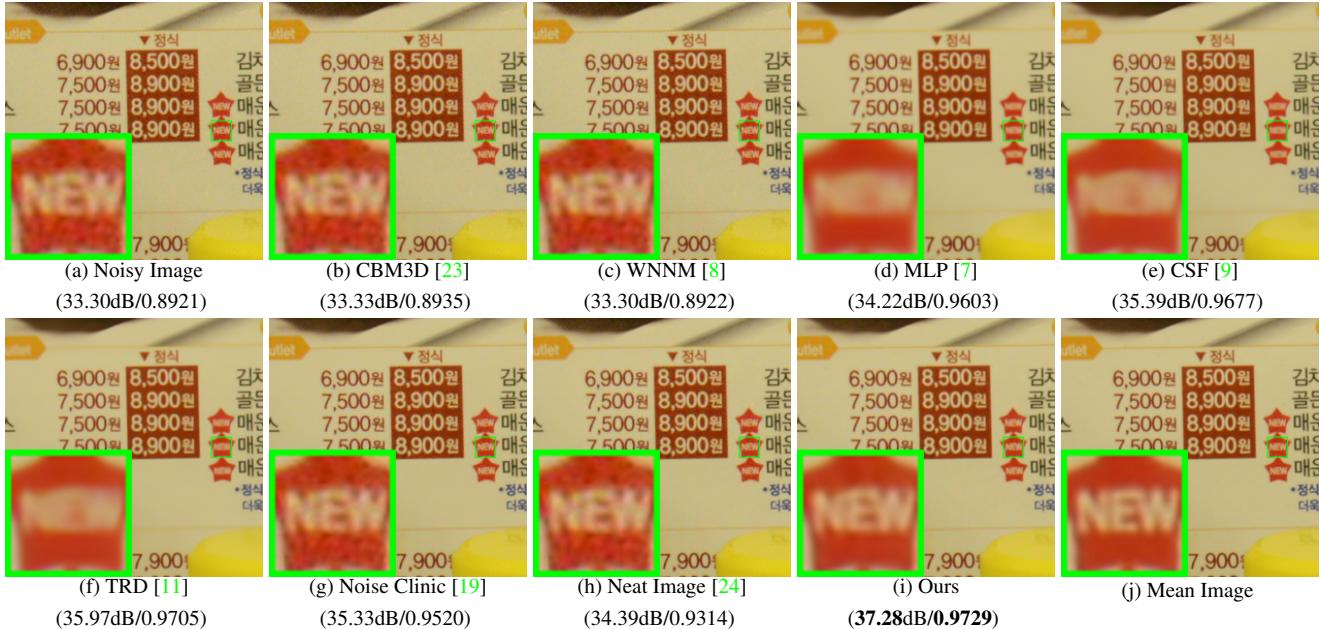


Figure 1. Denoised images of the real noisy image "Nikon D800 ISO 3200 A3" from [13] by different methods. The images are better viewed by zooming in on screen.

of natural clean images to guide the clustering of internal patch groups in given noisy image, and develop an external prior guided internal orthogonal dictionary learning (DL) algorithm for real image denoising. The internal orthogonal DL process includes two alternating stages: updating sparse coefficients and updating orthogonal dictionary. Both of the two stages have closed-form solutions. Hence, our internal DL process is very efficient for online internal denoising. Through comprehensive experiments on real noisy images captured by different cameras and settings, we demonstrate that the proposed method achieves better performance on real image denoising

1.1. Our Contributions

The contributions of this paper are summarized as follows:

- We propose a novel model to learn internal priors adaptive to given images. This model employs the external patch group (PG) prior learned from clean images to guide the internal PG prior learning of given images. The external prior benefits the internal learning on subspace selection and orthogonal dictionary learning.
- The proposed guided internal prior learning method is very efficient. The reason is that both the subspace selection and orthogonal dictionary learning have explicit solutions.
- For real image denoising problem, the proposed method achieves much better performance than other competing methods.

The rest of this paper will be summarized as follows: in Section 2, we briefly introduce the related work; in Section 3, we develop the proposed external prior guided internal prior learning model and formulate the overall image denoising algorithm; in Section 3, we demonstrate extensive experiments on real image denoising probelm; in Section 4, we conclude our paper and give future work.

2. Related Work

2.1. Patch Group Prior of Natural Images

The Patch Group (PG) prior [10] is proposed to directly model the non-local self similar (NSS) property of natural images. The NSS property is commonly used in image restoration tasks [1, 4, 5, 8, 10]. The PG prior largely reduces the space of images to be modeled when compared to the patch prior [6]. In [10], only the PGs of clean natural images is utilized, while the PGs of noisy input images are ignored. In this paper, we make use of PGs both from external clean images and internal given real noisy image for better denoising performance.

2.2. Internal v.s. External Prior Learning

Learning priors to represent images has been successfully used in image modeling [3, 6, 10, 25, 26]. There are mainly two categories of prior learning methods: 1) External methods pre-learned priors (e.g., dictionaries) from a set of clean images, and the learned priors are used to recover the noisy images [6, 10]. 2) Internal methods directly learned priors from the given noisy image, and the image denoising is simultaneously done with the learning process



Figure 2. Denoised images of the image "Nikon D600 ISO 3200 C1" by different methods. The images are better to be zoomed in on screen.

[3, 25, 26]. Both the two categories of methods have limitations. The external methods is not adaptive to the noisy image, while the internal methods ignores the information hidden in clean images. In this paper, our goal is to employ the external prior to guide the internal prior learning.

2.3. Real Image Denoising

In the last decade, there are many methods [15, 16, 17, 18, 19, 20, 13] proposed for real image denoising problem. In the seminar work of BLS-GSM [27] for real image denoising, Portilla et al. proposed to use scale mixture of Gaussian in overcomplete oriented pyramids to estimate the latent clean images. In [15], Portilla proposed to use a correlated Gaussian model for noise estimation of each wavelet subband. The work of Rabie [16] modeled the noisy pixels as outliers which are removed via Lorentzian robust estimator [28]. Liu et al. [17] proposed to use 'noise level function' to estimate the noise and then use Gaussian conditional random field to obtain the latent clean image. Gong et al. [18] models the noise by mixed ℓ_1 and ℓ_2 norms and remove the noise by sparsity prior in the wavelet transform domain. Later, Lebrun el al. proposed a multiscale denoising algorithm called 'Noise Clinic' [19]. This method generalizes the NL-Bayes model [29] to deal with blind noise and achieves state-of-the-art performance. Recently, Zhu et al. proposed a Bayesian model [20] which approximates and removes the noise via Low-Rank Mixture of Gaussians.

3. External Patch Group Prior Guided Internal Prior Learning for Image Denoising

In this section, we formulate the framework of external patch group (PG) prior guided internal prior learning. We first introduce the external PG prior leaning on natural clean RGB images. Then we propose to employ the learned external PG prior to guide the internal prior (subspace selection and dictionary learning (DL)) learning of given degraded (such as noisy) images. Under the weighted sparse coding framework, the internal prior learning process has alternative closed-form solutions in term of updating sparse co-

efficients and orthogonal dictionary. Finally, we discuss in details how external prior learned from natural clean images guide the internal prior learning of given degraded (noisy) images.

3.1. External Patch Group Prior Learning

Natural images often demonstrate repetitive local patterns, this nonlocal self-similarity (NSS) property is a key successful factor for many image denoising methods [1, 4, 5, 26, 8, 10]. In this section, we formulate the Patch Group prior learned on natural color images. Similar to [10], the patch group (PG) is defined as a group of similar patches to the local patch. The patch group mean is distracted, and hence different groups patches can share similar PGs. In this way, the space natural image patches to be modeled is largely reduced.

In this work, each local patch extracted from RGB images is of size $p \times p \times 3$. Then we search the M most similar patches $\{\mathbf{x}_m\}_{m=1}^M$ around each local patch through Euclidean distance, in a local window of size $W \times W$. The $\mathbf{x}_m \in \mathbb{R}^{3p^2 \times 1}$ is a patch vector formed by combining the 3 patch vectors (of size $p^2 \times 1$) in R, G, B channels. The mean vector of this PG is $\boldsymbol{\mu} = \frac{1}{M} \sum_{m=1}^M \mathbf{x}_m$, and the group mean subtracted PG is defined as $\bar{\mathbf{X}} \triangleq \{\bar{\mathbf{x}}_m = \mathbf{x}_m - \boldsymbol{\mu}\}, m = 1, \dots, M$. Assume we have extracted N PGs from a set of external natural images, and the n -th PG is defined as $\bar{\mathbf{X}}_n \triangleq \{\bar{\mathbf{x}}_{n,m}\}_{m=1}^M, n = 1, \dots, N$. We employ the Gaussian Mixture Model (GMM) to learn the external patch group based NSS prior. In this model, the likelihood of the n -th PG $\{\bar{\mathbf{X}}_n\}$ can be calculated as

$$P(\bar{\mathbf{X}}_n) = \sum_{k=1}^K \pi_k \prod_{m=1}^M \mathcal{N}(\bar{\mathbf{x}}_{n,m} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad (1)$$

where K is the number of Gaussians and the parameters π_k , $\boldsymbol{\mu}_k$, $\boldsymbol{\Sigma}_k$ are mixture weight, mean vector, and covariance matrix of the k -th Gaussian, respectively. By assuming that all the PGs are independently sampled, the overall objective log-likelihood function is

$$\ln \mathcal{L} = \sum_{n=1}^N \ln \left(\sum_{k=1}^K \pi_k \prod_{m=1}^M \mathcal{N}(\bar{\mathbf{x}}_{n,m} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right). \quad (2)$$

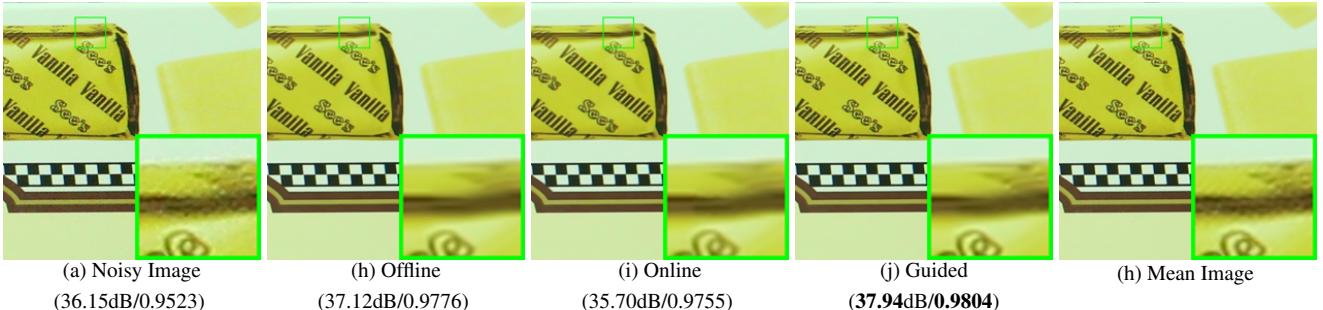


Figure 3. Denoised images of the image "Canon EOS 5D Mark3 ISO 3200 C1" by different methods. The images are better to be zoomed in on screen.

We maximize the above objective function via EM algorithm [30] and finally obtain the GMM model with learned parameters. Similar to [10], the mean vector of each cluster is natural zeros, i.e., $\mu_k = \mathbf{0}$.

Now, we have clustered the PGs extracted from external clean images into K Gaussians or subspaces. To better characterize each subspace, we perform singular value decomposition (SVD) on the covariance matrix:

$$\Sigma_k = \mathbf{U}_k \mathbf{S}_k \mathbf{U}_k^\top. \quad (3)$$

The singular vector matrices $\{\mathbf{U}_k\}_{k=1}^K$ are employed as the external orthogonal dictionary to guide the internal dictionary learning. The singular values in the diagonal of \mathbf{S}_k reflect the significance of the singular vectors in \mathbf{U}_k and utilized as prior weights for weighted sparse coding which will be discussed in next section.

3.2. External Prior Guided Internal Prior Learning

After the external patch group (PG) prior is learned, we can employ it to guide the internal PG prior learning for the given testing (real noisy) image. The guidance mainly comes from two aspects. One aspect is that the external prior can guide the internal noisy PGs to be assigned to most suitable Gaussians or subspaces. And for each subspace, the other aspect is to guide the orthogonal dictionary learning of internal noisy PGs.

3.2.1 Guided Internal Subspace Selection

Given a real noisy image, assume we can totally extract N local patches from it. Similar to the external prior learning stage, for the n -th local patch ($n = 1, \dots, N$), we extract its similar patches around it to form a noisy PG denoted by $\mathbf{Y}_n = \{\mathbf{y}_{n,1}, \dots, \mathbf{y}_{n,M}\}$. Then the group mean of \mathbf{Y}_n , denoted by $\boldsymbol{\mu}_n$, is calculated and subtracted from each patch, leading to the mean subtracted PG $\bar{\mathbf{Y}}_n$. Each mean subtracted PG is defined as $\bar{\mathbf{Y}}_n \triangleq \{\bar{\mathbf{y}}_{n,m}\}_{m=1}^M$ where $\bar{\mathbf{y}}_{n,m} = \mathbf{y}_{n,m} - \boldsymbol{\mu}_n$. For adaptivity, we project the PG $\bar{\mathbf{Y}}_n$ into its most suitable Gaussian component (subspace) of the GMM learned on external PGs. The subspace most suitable

for $\bar{\mathbf{Y}}_n$ is selected by firstly calculating the posterior probability of " $\bar{\mathbf{Y}}_n$ belonging to the k th Gaussian component":

$$P(k|\bar{\mathbf{Y}}_n) = \frac{\prod_{m=1}^M \mathcal{N}(\bar{\mathbf{y}}_{n,m}|\mathbf{0}, \Sigma_k)}{\sum_{l=1}^K \prod_{m=1}^M \mathcal{N}(\bar{\mathbf{y}}_{n,m}|\mathbf{0}, \Sigma_l)}, \quad (4)$$

and then choosing the component with the maximum A-posteriori (MAP) probability $\ln P(k|\bar{\mathbf{Y}}_n)$.

3.2.2 Guided Internal Orthogonal Dictionary Learning

Assume we have assigned all internal noisy PGs $\{\bar{\mathbf{Y}}_n\}_{n=1}^N$ to their corresponding most suitable Gaussians or subspaces in $\{\mathcal{N}(\mathbf{0}, \Sigma_k)\}_{k=1}^K$. Assume $\{\bar{\mathbf{Y}}_{k_n}\}_{n=1}^{N_k}$ are the noisy PGs assigned to the k -th subspace, such that $\sum_{k=1}^K N_k = N$ and $\bar{\mathbf{Y}}_{k_n} = [\bar{\mathbf{y}}_{k_n,1}, \dots, \bar{\mathbf{y}}_{k_n,M}]$. We consider to utilize the external orthogonal dictionary \mathbf{U}_k to guide the learning of an orthogonal dictionary $\mathbf{D}_k := [\mathbf{D}_{k,e} \ \mathbf{D}_{k,i}] \in \mathbb{R}^{3p^2 \times 3p^2}$. This dictionary has two parts: the external part $\mathbf{D}_{k,e} = \mathbf{U}_k(:, 1 : 3p^2 - r) \in \mathbb{R}^{3p^2 \times (3p^2 - r)}$ is directly obtained from the external dictionary in Equ. (3), and the internal part $\mathbf{D}_{k,i}$ is consisted of dictionary atoms adaptively learned from the internal noisy PGs $\mathbf{Y}_k = \{\bar{\mathbf{Y}}_{k_n}\}_{n=1}^{N_k}$. For notation simplicity, we ignore the subspace index k and denote as noisy PGs $\mathbf{Y} \triangleq \{\bar{\mathbf{Y}}_n\}_{n=1}^N = [\bar{\mathbf{y}}_{1,1}, \dots, \bar{\mathbf{y}}_{1,M}, \dots, \bar{\mathbf{y}}_{N,1}, \dots, \bar{\mathbf{y}}_{N,M}]$. The learning and denoising are simultaneously performed under the weighted sparse coding framework proposed as follows:

$$\begin{aligned} & \min_{\mathbf{D}_i, \{\boldsymbol{\alpha}_{n,m}\}} \sum_{n=1}^N \sum_{m=1}^M (\|\bar{\mathbf{y}}_{n,m} - \mathbf{D}\boldsymbol{\alpha}_{n,m}\|_2^2 + \sum_{j=1}^{3p^2} \lambda_j |\boldsymbol{\alpha}_{n,m,j}|) \\ & \text{s.t. } \mathbf{D} = [\mathbf{D}_e \ \mathbf{D}_i], \ \mathbf{D}_i^\top \mathbf{D}_i = \mathbf{I}_r, \ \mathbf{D}_e^\top \mathbf{D}_i = \mathbf{0}, \end{aligned} \quad (5)$$

where $\boldsymbol{\alpha}_{n,m}$ is the sparse coefficients vector of the m -th patch $\bar{\mathbf{y}}_{n,m}$ of the n -th PG $\bar{\mathbf{Y}}_n$ and $\boldsymbol{\alpha}_{n,m,j}$ is the j -th element of $\boldsymbol{\alpha}_{n,m}$. λ_j is the j -th regularization parameter defined as

$$\lambda_j = \lambda / (\sqrt{\mathbf{S}_j} + \varepsilon). \quad (6)$$

We employ square roots of the singular values in \mathbf{S} (please refer Equ. (3)) as external prior weights and add a small positive number ε to avoid zero denominator. Noted that

432 $\mathbf{D}_e = \emptyset$ if $r = 3p^2$ and $\mathbf{D}_e = \mathbf{U}_k$ if $r = 0$. The dictionary
 433 $\mathbf{D} = [\mathbf{D}_e \mathbf{D}_i]$ is orthogonal by checking that:
 434

$$435 \quad \mathbf{D}^\top \mathbf{D} = \begin{bmatrix} \mathbf{D}_e^\top \\ \mathbf{D}_i^\top \end{bmatrix} [\mathbf{D}_e \mathbf{D}_i] = \begin{bmatrix} \mathbf{D}_e^\top \mathbf{D}_e & \mathbf{D}_e^\top \mathbf{D}_i \\ \mathbf{D}_i^\top \mathbf{D}_e & \mathbf{D}_i^\top \mathbf{D}_i \end{bmatrix} = \mathbf{I} \quad (7)$$

437 Similar to K-SVD [3], we employ an alternating iterative
 438 framework to solve the optimization problem (5). Specifically,
 439 we initialize the orthogonal dictionary as $\mathbf{D}_{(0)} = \mathbf{U}_k$
 440 and for $t = 0, 1, \dots, T - 1$, alternatively do:

441 **Updating Sparse Coefficients:** given the orthogonal dic-
 442 tioanry $\mathbf{D}_{(t)}$, we update the sparse coefficients via solving

$$444 \quad \alpha_{n,m}^{(t)} := \arg \min_{\alpha_{n,m}} \|\bar{\mathbf{y}}_{n,m} - \mathbf{D}_{(t)} \alpha_{n,m}\|_2^2 + \sum_{j=1}^{3p^2} \lambda_j |\alpha_{n,m,j}| \quad (8)$$

$$445 \quad \text{s.t. } \mathbf{D}_{(t)} = [\mathbf{D}_e \mathbf{D}_i], \mathbf{D}_i^\top \mathbf{D}_i = \mathbf{I}_r, \mathbf{D}_e^\top \mathbf{D}_i = \mathbf{0},$$

446 Since dictionary $\mathbf{D}_{(t)} = [\mathbf{D}_e \mathbf{D}_i^{(t)}]$ is orthogonal, the prob-
 447 lems (8) has a closed-form solution [10]

$$448 \quad \alpha_{n,m}^{(t)} = \text{sgn}(\mathbf{D}_{(t)}^\top \bar{\mathbf{y}}_{n,m}) \odot \max(|\mathbf{D}_{(t)}^\top \bar{\mathbf{y}}_{n,m}| - \Lambda, 0), \quad (9)$$

449 where $\Lambda = [\lambda_1, \lambda_2, \dots, \lambda_{3p^2}]$ is the vector of regularization
 450 parameter and $\text{sgn}(\bullet)$ is the sign function, \odot means
 451 element-wise multiplication.

452 **Updating Orthogonal Dictionary:** given the sparse coeffi-
 453 cients $\mathbf{A}^{(t)} = [\alpha_{1,1}^{(t)}, \dots, \alpha_{1,M}^{(t)}, \dots, \alpha_{N,1}^{(t)}, \dots, \alpha_{N,M}^{(t)}]$, we up-
 454 date the internal orthogonal dictionary via solving

$$455 \quad \mathbf{D}_i^{(t+1)} := \arg \min_{\mathbf{D}_i} \sum_{n=1}^N (\|\bar{\mathbf{y}}_{n,m} - \mathbf{D} \alpha_{n,m}^{(t)}\|_2^2) \quad (10)$$

$$456 \quad = \arg \min_{\mathbf{D}_i} \|\mathbf{Y} - \mathbf{D} \mathbf{A}^{(t)}\|_F^2$$

$$457 \quad \text{s.t. } \mathbf{D} = [\mathbf{D}_e \mathbf{D}_i], \mathbf{D}_i^\top \mathbf{D}_i = \mathbf{I}_r, \mathbf{D}_e^\top \mathbf{D}_i = \mathbf{0},$$

458 The sparse coefficient matrix $\mathbf{A}^{(t)} = [(\mathbf{A}_e^{(t)})^\top (\mathbf{A}_i^{(t)})^\top]^\top$
 459 also has two parts: the external part $\mathbf{A}_e^{(t)}$ and the internal
 460 part $\mathbf{A}_i^{(t)}$ denote the coefficients over external dictionary
 461 \mathbf{D}_e and internal dictionary $\mathbf{D}_i^{(t)}$, respectively. According to
 462 the Proposition 2.2 in [33], the problem (10) has a closed-
 463 form solution $\mathbf{D}_i^{(t+1)} = \mathbf{U}_i \mathbf{V}_i^\top$, where \mathbf{U}_i and \mathbf{V}_i are the
 464 orthogonal matrices obtained by the following SVD
 465

$$466 \quad (\mathbf{I} - \mathbf{D}_e \mathbf{D}_e^\top) \mathbf{Y} (\mathbf{A}_i^{(t)})^\top = \mathbf{U}_i \mathbf{S}_i \mathbf{V}_i^\top \quad (11)$$

3.3. Discussions

467 Here we take a detailed analysis on the guiance of the
 468 external patch group (PG) prior for the internal noisy PGs
 469 of given real noisy images. The guidance comes from at
 470 least three aspects: 1) the external prior guides the internal
 471 PGs to be clustered into suitable subspaces through MAP in
 472 Equ. (5). The guided subspace clustering is more efficient
 473 than directly clustering the internal noisy PGs via k-means
 474 or Gaussian Mixture Model (GMM). The reason is, by guid-
 475 ance we only need calculate the probabilities via Equ. (5)

Alg. 1: External PG Prior Guided Internal PG Prior Learning for Real Image Denoising

486 **Input:** Noisy image \mathbf{y} , external prior GMM model
 487 **Output:** The denoised image $\hat{\mathbf{x}}$.
 488 **Initialization:** $\hat{\mathbf{x}}^{(0)} = \mathbf{y}$;
 489 **for** $t = 1 : IteNum$ **do**
 490 1. Extracting internal patch groups (PGs) from $\hat{\mathbf{x}}^{(0)}$;
 491 2. for each PG \mathbf{Y}_n **do**
 492 3. Calculate group mean μ_n and form PG $\bar{\mathbf{Y}}_n$;
 493 4. Gaussian component selection via Equ. (5);
 494 5. Internal Subspace Learning by (4);
 495 6. Recover each patch in all PGs via $\hat{\mathbf{x}}_m = \mathbf{D} \hat{\alpha} + \mu_y$;
 496 7. Aggregate the recovered PGs of all subspaces to form
 497 the recovered image $\hat{\mathbf{x}}^{(t)}$;
 498 **end for**

499 **for** each Internal Subspace **do**
 500 1. Internal Subspace Learning by (4);
 501 2. Recover each patch in all PGs via $\hat{\mathbf{x}}_m = \mathbf{D} \hat{\alpha} + \mu_y$;
 502 3. Aggregate the recovered PGs of all subspaces to form
 503 the recovered image $\hat{\mathbf{x}}^{(t)}$;
 504 **end for**

505 for all noisy PGs while by internal clustering via GMM we
 506 need perform EM algorithm [30]. 2) the external dictionary
 507 guides the internal dictionary learning, and the obtained dic-
 508 tionary consisted of the two parts is orthogonal and more
 509 adaptive to the given noisy image. The learning process has
 510 closed-form solutions and hence is very efficient. Besides,
 511 the learned orthogonal dictionary also makes the denoising
 512 process very efficient under sparse coding framework. 3)
 513 the singular values obtained by SVD in Equ. (3) reflect the
 514 prior weights of the atoms in learned dictionary and can be
 515 used as adaptive parameters for real image denoising.

3.4. The Denoising Algorithm

516 As discussed in previous section, the proposed frame-
 517 work employs the external prior learned from natural clean
 518 images to guide the internal prior learning for any given
 519 noisy image. The image denoising is simultaneously done
 520 with the learning process. In the denoising procedure, for
 521 each subspace (we ignore the indexes $k = 1, \dots, K$ of
 522 subspaces for notation simplicity), the group mean vectors
 523 $\{\mu_n\}_{n=1}^N$ of the internal noisy PGs $\{\mathbf{y}_n\}_{n=1}^N$ are saved for
 524 reconstruction. Until now, we obtain the solutions of sparse
 525 coefficients $\{\hat{\alpha}_n^{T-1}\}$ in (9) and the orthogonal dictionary
 526 $\mathbf{D}_{(T)} = [\mathbf{D}_e \mathbf{D}_i^{(T)}]$ in Equ. (10). Then the latent clean PGs
 527 $\{\hat{\mathbf{y}}_n\}_{n=1}^N$ are recovered by $\hat{\mathbf{y}}_n = \mathbf{D} \hat{\alpha}$. We add back its PG
 528 mean denoted as μ_y and reconstruct the clean image $\hat{\mathbf{x}}$ by
 529 aggregating all the estimated PGs. Similar to [10], we per-
 530 form the above denoising procedures for several iterations
 531 for better denoising outputs. But different from [10], we do
 532 not employ the iterative regularization strategy [35] since
 533 it is not effective in our method. The proposed denoising
 534 algorithm is summarized in Alg. 1.



Figure 4. Some cropped images of the dataset [13].

540
541
542
543
544
545
546
547

4. Experiments

In this section, we perform real image denoising experiments on three standard datasets. The first dataset is real noisy images with mean images as ground truths provided by [13], some samples are shown in Figure 5. The second dataset is provided by the website of Noise Clinic [19]. The third dataset is provided by the Commercial software Neat Image [24]. The second and third dataset do not have ground truth images.

558
559

4.1. Implementation Details

Our proposed method contains two stages, the external prior guided internal subspace learning stage and the adaptive denoising stage. In the learning stage, there are 4 parameters: the patch size p , the number of patches in a PG M , the window size W for PG searching and the number of clusters K . We set $p = 6$ (hence the patch size is $6 \times 6 \times 3$), $M = 10$, $W = 31$, $K = 32$. We extracted about 3.6 million PGs from the Kodak PhotoCD Dataset, which includes 24 high quality color images, to train the external prior via PG-GMM. In the denoising stage, the parameter $\lambda = 0.002$ is used to regularize the sparse term. The δ in iterative regularization is set as $\delta = 0.09$. Given a RGB image of size $256 \times 256 \times 3$, we can extract over 60 thousands local patches of size $6 \times 6 \times 3$. It is time-consuming to search patch groups (PG) for local patches one by one. To speed up the searching process, we employ the technique of 'Summed Area Table' [34] for efficient PG searching. The SAT permits to evaluate the sum of pixel values in rectangular regions of the image with four operations, regardless of the region size. Hence, we do not need to do distance measure for each local patch.

581
582

4.2. Comparison on External and Internal methods

In this subsection, we compared the proposed external prior guided internal subspace learning model on real image denoising. The three methods are evaluated on the dataset provided in [13]. We calculate the PSNR, SSIM [22] and visual quality of these three methods. We also compare the speed. The PSNR and SSIM results on 60 cropped images from [13] are listed in Table 1. The images are cropped into size of 500×500 for better illustration. We also compare the three methods on visual quality in Figure 2. Compare the denoised images listed in Figure 2 and Figure 2, we can see that the Offline method is better at edges, smooth

Table 1. Average PSNR(dB)/SSIM results of external, internal, and guided methods on 60 cropped real noisy images in [13].

	Noisy	Offline	Online	Guided
PSNR	34.51	38.19	38.07	38.55
SSIM	0.8718	0.9663	0.9625	0.9675

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

regions while the Online method is good at complex textures. The reason is two folds. Firstly, the Offline method is learned on clean images and hence is better at representing edges, structures, and smooth area. The online method is influenced by the noise and hence some noise cannot be removed. Secondly, the Online method is better at recovering complex area since they could learn adaptive dictionaries for the specific area. The Offline method cannot recover the complex area since they did not learn the similar structures from the external natural clean images.

4.3. Comparison With other Competing Methods

We compare with previous state-of-the-art Gaussian noise removal methods such as BM3D [4], WNNM [8], MLP [7], CSF [9], and the recently proposed TRD [11]. We also compare with three competing real image denoising methods such as Noise Clinic, Neat Image, and the CCNoise method proposed recently. The commercial software Neat Image [24] first estimates the parameters of noise via a large flat area and then filters the noise accordingly. All these methods need noise estimation which is very hard to perform if there is no uniform regions available in the testing image. The NeatImage will fail to perform automatical parameters settings if there is no uniform regions.¹

We the competing denoising methods from various research directions on two datasets. Both the two datasets comes from the [13]. The first dataset contains 17 images of size over 7000×5000 . Since this dataset contains repetitive contents across different images, we crop 60 small images of size 500×500 from these 17 images in [13]. The PSNR and SSIM resluts are listed in Table 3. The number in red color and blue color means the best and second best results, respectively. From the Table 3, we can see that the external based method can already surpass largely the previous denoising methods. The improvement on PSNR over the second best method, i.e., TRD, is 0.44dB. The

4.4. Discussion on Parameter λ

The proposed method only has a key parameter, namely the regularization paramters λ . To demonstrate that the proposed method is robust to the variance of λ , we vary the parameter λ across a wide range and obtain the PSNR and

¹To compare with CCNoise, we first transform the denoised images into double format.



Figure 5. Denoised images of the image "Canon 5D Mark 3 ISO 3200 1" by different methods. The images are better to be zoomed in on screen.

Table 2. Average PSNR(dB) results of different methods on 60 cropped real noisy images captured in [13].

	Noisy	CBM3D	WNNM	MLP	CSF	TRD	NI	NC	Guided	Guided2
PSNR	34.51	34.58	34.52	36.19	37.40	37.75	36.53	37.57	38.72	38.90
SSIM	0.8718	0.8748	0.8743	0.9470	0.9598	0.9617	0.9241	0.9514	0.9694	0.9702

SSIM results as a function of the parameter λ . The results is shown in Figure 6, from which we can see that the proposed method can achieve a PSNR (SSIM) over 38.5dB (0.9660) when λ varies from 0.0015 to 0.0025. This shows that the proposed method is indeed robust to the chosen of the paramter λ .

5. Conclusion and Future Work

In the future, we will evaluate the proposed method on other computer vision tasks such as single image super-resolution, photo-sketch synthesis, and cross-domain image recognition. Our proposed method can be improved if we use better training images, fine tune the parameters via cross-validation. We believe that our framework can be useful not just for real image denoising, but for image super-resolution, image cross-style synthesis, and recognition tasks. This will be our line of future work.

References

- [1] A. Buades, B. Coll, and J. M. Morel. A non-local algorithm for image denoising. *CVPR*, pages 60–65, 2005. 1, 2, 3
- [2] S. Roth and M. J. Black. Fields of Experts. *International Journal of Computer Vision*, 82(2):205–229, 2009. 1

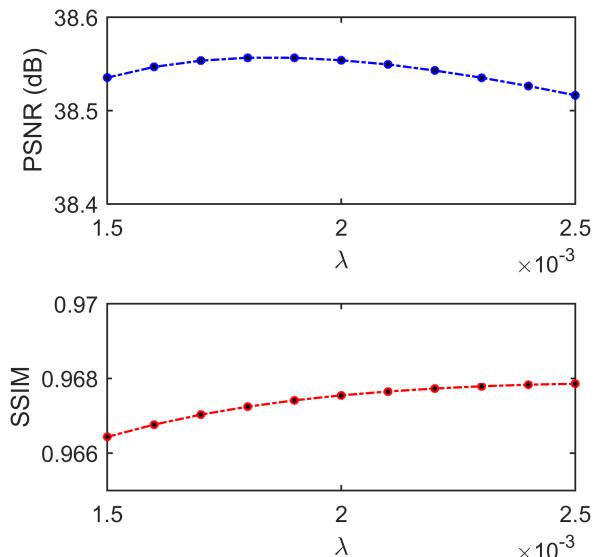


Figure 6. The PSNR/SSIM results as a function of the parameter λ .

756 Table 3. Average PSNR(dB) results of different methods on 15 cropped real noisy images used in [13]. 810
757 811
758 812
759 813
760 814
761 815
762 816
763 817
764 818
765 819
766 820
767 821
768 822
769 823
770 824
771 825
772 826
773 827
774 828
775 829
776 830
777 831
778 832
779 833
780 834
781 835
782 836
783 837
784 838
785 839
786 840
787 841
788 842
789 843
790 844
791 845
792 846
793 847
794 848
795 849
796 850
797 851
798 852
799 853
800 854
801 855
802 856
803 857
804 858
805 859
806 860
807 861
808 862
809 863

Camera Settings	Noisy	CBM3D	WNNM	MLP	CSF	TRD	NI	NC	CC	Guided2
Canon 5D Mark III ISO = 3200	37.00	37.08	37.09	33.92	35.68	36.20	37.68	38.76	38.37	40.50
	33.88	33.94	33.93	33.24	34.03	34.35	34.87	35.69	35.37	37.22
	33.83	33.88	33.90	32.37	32.63	33.10	34.77	35.54	34.91	37.13
Nikon D600 ISO = 3200	33.28	33.33	33.34	31.93	31.78	32.28	34.12	35.57	34.98	35.34
	33.77	33.85	33.79	34.15	35.16	35.34	35.36	36.70	35.95	36.69
	34.93	35.02	34.95	37.89	39.98	40.51	38.68	39.28	41.15	39.17
Nikon D800 ISO = 1600	35.47	35.54	35.57	33.77	34.84	35.09	37.34	38.01	37.99	38.82
	35.71	35.79	35.77	35.89	38.42	38.65	38.57	39.05	40.36	40.98
	34.81	34.92	34.95	34.25	35.79	35.85	37.87	38.20	38.30	38.90
Nikon D800 ISO = 3200	33.26	33.34	33.31	37.42	38.36	38.56	36.95	38.07	39.01	38.69
	32.89	32.95	32.96	34.88	35.53	35.76	35.09	35.72	36.75	36.82
	32.91	32.98	32.96	38.54	40.05	40.59	36.91	36.76	39.06	38.80
Nikon D800 ISO = 6400	29.63	29.66	29.71	33.59	34.08	34.25	31.28	33.49	34.61	33.31
	29.97	30.01	29.98	31.55	32.13	32.38	31.38	32.79	33.21	33.18
	29.87	29.90	29.95	31.42	31.52	31.76	31.40	32.86	33.22	33.35
Average PSNR	33.41	33.48	33.48	34.32	35.33	35.65	35.49	36.43	36.88	37.26
Average SSIM	0.8483	0.8511	0.8512	0.9113	0.9250	0.9280	0.9126	0.9364	0.9481	0.9505

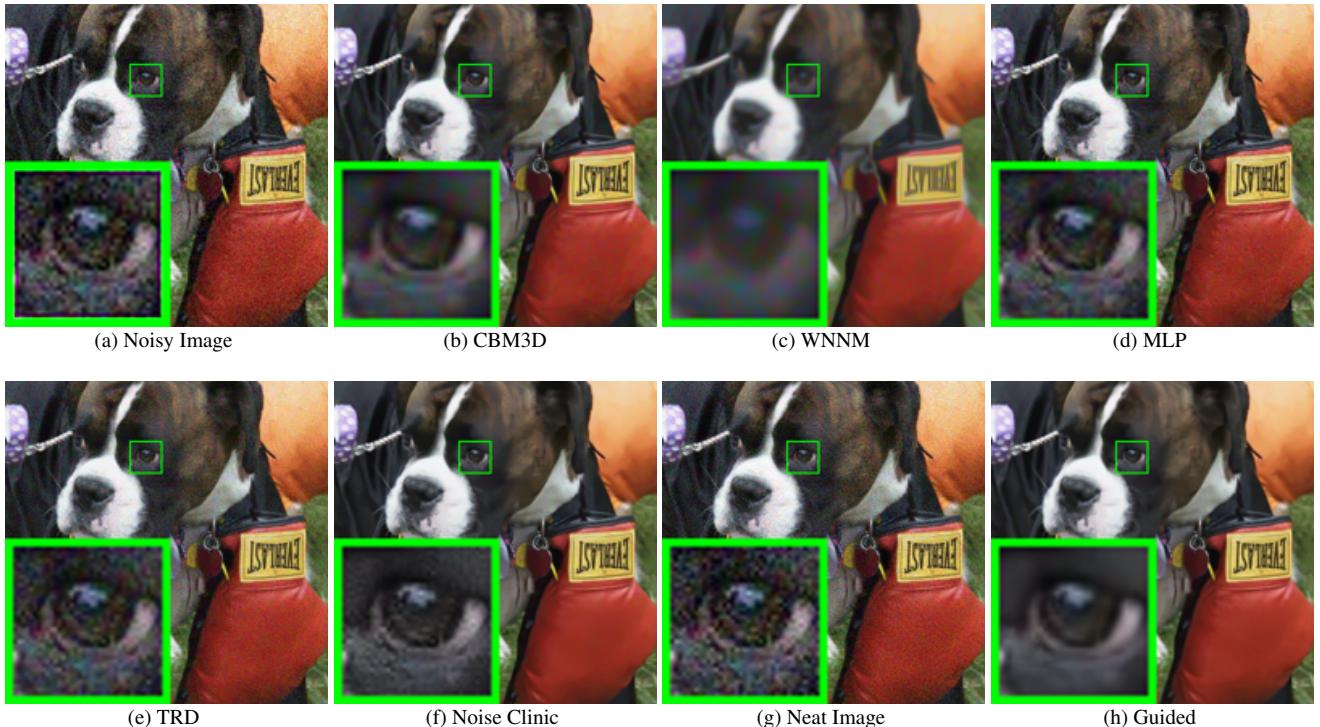


Figure 7. Denoised images of the image "5dmak3iso32003" by different methods. The images are better to be zoomed in on screen.

[3] 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. 1, 2, 3, 5

[4] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-D transform-domain collabora-

tive filtering. *Image Processing, IEEE Transactions on*, 16(8):2080–2095, 2007. 1, 2, 3, 6

[5] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. *ICCV*, pages 2272–2279, 2009. 1, 2, 3

[6] D. Zoran and Y. Weiss. From learning models of natural

- 864 image patches to whole image restoration. *ICCV*, pages 479–
865 486, 2011. 1, 2
- 866 [7] Harold C Burger, Christian J Schuler, and Stefan Harmeling.
867 Image denoising: Can plain neural networks compete with
868 bm3d? *Computer Vision and Pattern Recognition (CVPR),*
869 2012 IEEE Conference on, pages 2392–2399, 2012. 1, 2, 6
- 870 [8] S. Gu, L. Zhang, W. Zuo, and X. Feng. Weighted nu-
871 clear norm minimization with application to image denoising.
872 *CVPR*, pages 2862–2869, 2014. 1, 2, 3, 6
- 873 [9] U. Schmidt and S. Roth. Shrinkage fields for effective im-
874 age restoration. *Computer Vision and Pattern Recognition*
875 (*CVPR*), 2014 IEEE Conference on, pages 2774–2781, June
876 2014. 1, 2, 6
- 877 [10] J. Xu, L. Zhang, W. Zuo, D. Zhang, and X. Feng. Patch
878 group based nonlocal self-similarity prior learning for image
879 denoising. *2015 IEEE International Conference on Com-
880 puter Vision (ICCV)*, pages 244–252, 2015. 1, 2, 3, 4, 5
- 881 [11] Yunjin Chen, Wei Yu, and Thomas Pock. On learning
882 optimized reaction diffusion processes for effective image
883 restoration. *Proceedings of the IEEE Conference on Com-
884 puter Vision and Pattern Recognition*, pages 5261–5269,
885 2015. 1, 2, 6
- 886 [12] S. J. Kim, H. T. Lin, Z. Lu, S. Ssstrunk, S. Lin, and M. S.
887 Brown. A new in-camera imaging model for color computer
888 vision and its application. *IEEE Transactions on Pattern
889 Analysis and Machine Intelligence*, 34(12):2289–2302, Dec
890 2012. 1
- 891 [13] Seonghyeon Nam, Youngbae Hwang, Yasuyuki Matsushita,
892 and Seon Joo Kim. A holistic approach to cross-channel im-
893 age noise modeling and its application to image denoising.
894 *Proc. Computer Vision and Pattern Recognition (CVPR)*,
895 pages 1683–1691, 2016. 1, 2, 3, 6, 7, 8
- 896 [14] Glenn E Healey and Raghava Kondepudy. Radiometric ccd
897 camera calibration and noise estimation. *IEEE Transactions
898 on Pattern Analysis and Machine Intelligence*, 16(3):267–
899 276, 1994. 1
- 900 [15] J. Portilla. Full blind denoising through noise covariance
901 estimation using gaussian scale mixtures in the wavelet do-
902 main. *Image Processing, 2004. ICIP '04. 2004 Interna-
903 tional Conference on*, 2:1217–1220, 2004. 1, 3
- 904 [16] Tamer Rabie. Robust estimation approach for blind denois-
905 ing. *Image Processing, IEEE Transactions on*, 14(11):1755–
906 1765, 2005. 1, 3
- 907 [17] C. Liu, R. Szeliski, S. Bing Kang, C. L. Zitnick, and W. T.
908 Freeman. Automatic estimation and removal of noise from
909 a single image. *IEEE Transactions on Pattern Analysis and
910 Machine Intelligence*, 30(2):299–314, 2008. 1, 3
- 911 [18] Zheng Gong, Zuowei Shen, and Kim-Chuan Toh. Image
912 restoration with mixed or unknown noises. *Multiscale Mod-
913 eling & Simulation*, 12(2):458–487, 2014. 1, 3
- 914 [19] M. Lebrun, M. Colom, and J.-M. Morel. Multiscale image
915 blind denoising. *Image Processing, IEEE Transactions on*,
916 24(10):3149–3161, 2015. 1, 2, 3, 6
- 917 [20] Fengyuan Zhu, Guangyong Chen, and Pheng-Ann Heng.
918 From noise modeling to blind image denoising. *The IEEE
919 Conference on Computer Vision and Pattern Recognition
920 (CVPR)*, June 2016. 1, 3
- 921 [21] C. M. Bishop. *Pattern recognition and machine learning*.
922 New York: Springer, 2006. 1
- 923 [22] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli.
924 Image quality assessment: from error visibility to struc-
925 tural similarity. *IEEE Transactions on Image Processing*,
926 13(4):600–612, 2004. 1, 6
- 927 [23] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Color
928 image denoising via sparse 3d collaborative filtering with
929 grouping constraint in luminance-chrominance space. *IEEE
930 International Conference on Image Processing*, 1, 2007. 2
- 931 [24] Neatlab ABSoft. Neat image. <https://neatvideo.com/home>. 2, 6
- 932 [25] G. Yu, G. Sapiro, and S. Mallat. Solving inverse problems
933 with piecewise linear estimators: From Gaussian mixture
934 models to structured sparsity. *IEEE Transactions on Image
935 Processing*, 21(5):2481–2499, 2012. 2, 3
- 936 [26] W. Dong, L. Zhang, G. Shi, and X. Li. Nonlocally central-
937 ized sparse representation for image restoration. *Image Pro-
938 cessing, IEEE Transactions on*, 22(4):1620–1630, 2013. 2,
939 3
- 940 [27] J. Portilla, V. Strela, M.J. Wainwright, and E.P. Simoncelli.
941 Image denoising using scale mixtures of Gaussians in the
942 wavelet domain. *Image Processing, IEEE Transactions on*,
943 12(11):1338–1351, 2003. 3
- 944 [28] Peter J Huber. *Robust statistics*. Springer, 2011. 3
- 945 [29] M. Lebrun, A. Buades, and J. M. Morel. A nonlocal bayesian
946 image denoising algorithm. *SIAM Journal on Imaging Sci-
947 ences*, 6(3):1665–1688, 2013. 3
- 948 [30] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum
949 likelihood from incomplete data via the EM algorithm. *Journal
950 of the Royal Statistical Society. Series B (methodologi-
951 cal)*, pages 1–38, 1977. 4, 5
- 952 [31] David L Donoho and Michael Elad. Optimally sparse repre-
953 sentation in general (nonorthogonal) dictionaries via 1 mini-
954 mization. *Proceedings of the National Academy of Sciences*,
955 100(5):2197–2202, 2003.
- 956 [32] D. L. Donoho. De-noising by soft-thresholding. *IEEE Trans.
957 Inf. Theor.*, 41(3):613–627, 1995.
- 958 [33] Chenglong Bao, Jian-Feng Cai, and Hui Ji. Fast sparsity-
959 based orthogonal dictionary learning for image restoration.
960 *Proceedings of the IEEE International Conference on Com-
961 puter Vision*, pages 3384–3391, 2013. 5
- 962 [34] Franklin C. Crow. Summed-area tables for texture mapping.
963 *SIGGRAPH Comput. Graph.*, 18(3):207–212, January 1984.
964 6
- 965 [35] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An it-
966 erative regularization method for total variation-based image
967 restoration. *Multiscale Modeling & Simulation*, 4(2):460–
968 489, 2005. 5
- 969 [918] 919 [920] 921 [922] 923 [924] 925 [926] 927 [928] 929 [930] 931 [932] 933 [934] 935 [936] 937 [938] 939 [940] 941 [942] 943 [944] 945 [946] 947 [948] 949 [950] 951 [952] 953 [954] 955 [956] 957 [958] 959 [960] 961 [962] 963 [964] 965 [966] 967 [968] 969 [970] 971