Class-Specific Image Deblurring

Saeed Anwar[1], Cong Phuoc Huynh[1,2], Fatih Porikli[1,2]
[1]The Australian National University* Canberra ACT 2601, Australia
[2]NICTA†, Locked Bag 8001, Canberra ACT 2601, Australia

Abstract

In image deblurring, a fundamental problem is that the blur kernel suppresses a number of spatial frequencies that are difficult to recover reliably. In this paper, we explore the potential of a class-specific image prior for recovering spatial frequencies attenuated by the blurring process. Specifically, we devise a prior based on the class-specific subspace of image intensity responses to band-pass filters. We learn that the aggregation of these subspaces across all frequency bands serves as a good class-specific prior for the restoration of frequencies that cannot be recovered with generic image priors. In an extensive validation, our method, equipped with the above prior, yields greater image quality than many state-of-the-art methods by up to $5\ dB$ in terms of image PSNR, across various image categories including portraits, cars, cats, pedestrians and household objects.

1. Introduction

Image deblurring is an important research topic in low-level vision, dating back to the middle of the previous century [29]. The amount of image data captured by hand-held devices, *e.g.* smart phones or tablet computers, has increased rapidly thanks to the recent advent of built-in high-resolution imaging sensors. In the last decade, there has been enormous research effort in image quality restoration from blur due to camera shake, camera motion and defocus [4, 23, 2, 31, 15, 18, 32, 21, 30].

By nature, the deblurring problem is ill-posed, as there exists an infinite number of pairs of latent image and kernel that result in the same blurred observation. To resolve this ambiguity, many previous works have imposed constraints on the sparsity of image gradients via the hyper-Laplacian

prior [14, 19], the L_2 prior [2], the L_1/L_2 prior [15], the Gaussian [16] or Gaussian mixture model [4]. As these constraints favour constant intensity regions, the resulting image only contains few spatial details in addition to strong edges. This is certainly not the case for many object categories with gradual changes in surface orientation such as faces, animals, cars *etc*.

Alternatively, image deblurring methods rely on the implicit or explicit extraction of edge information for kernel computation [26, 32]. The works in [2] and [31] are related to the restoration of strong edges for kernel estimation via edge-preserving image smoothing, shock filtering and gradient magnitude thresholding. Joshi *et al.* [11] predicted the underlying step edges for the estimation of spatially varying sub-pixel point-spread functions. A concern about these approaches is that wrong edges can be mistakenly selected based on only local information, due to the presence of multiple false edges induced by a large kernel width.

More recently, class-specific information has been employed up to some extent in image deblurring. Joshi *et al.* [10] proposed a method of image enhancement, including image deblurring, given a personal photo collection. HaCohen *et al.* [9] approached the personal photo deblurring problem with a strong requirement for the dense correspondence between the blurred image and a sharp reference image with similar contents. Pan *et al.* [21] introduced a face image deblurring method by selecting an exemplar with the closest structural similarity to the blurred image. The above methods are either cumbersome, involving manual image annotations for the training phase, or of limited applications due to strict requirements for reference or exemplar images. In another work, Sun *et al.* [27] investigated context-specific priors to transfer mid and high frequency details from example scenes for non-blind deconvolution.

In this paper, we address the problem of restoring image frequency components that have been attenuated during the convolution process using class-specific training examples. Our approach is inspired by the success of using image statistics for image categorisation [28, 7]. Image statistics traditionally constitute the power spectra of high-frequency image gradients. Here, we extend the notion of image statis-

tics to the distribution of band-pass filter responses across the frequency spectrum. This frequency statistics modelling approach is complementary to the representation of multiscale image subbands, *e.g.* multiple resolutions in the wavelet domain, as fields of Gaussian scale mixtures [20].

We propose a new image prior for blind deblurring based on the following conjectures. Firstly, for every frequency band, the distribution of band-pass filter responses is characteristic of the image class. Secondly, the band-pass responses of an image class span a low-dimensional linear subspace. As demonstrated later, this prior is proven to be effective in recovering frequencies attenuated by blur kernels.

Instead of imposing a general sparsity constraint, we focus on modeling a class-specific prior in each band of the frequency spectrum. Specifically, we represent each frequency band of an image as a sparse linear combination of the filter responses of training (sharp) images in the same class. Aggregating this representation over the entire frequency spectrum, we can capture the characteristics of a given image class. The spirit of this work is to learn a more detailed prior than those based on edges or high-frequency image gradients. As a natural choice of the frequency domain, we choose to analyse images in the Fourier domain due to the convenient transformation of the blur model between the spatial and the frequency domain. With these ingredients in hand, we can incorporate them into the deblurring process in much the same way as previous works.

In figure 1, we demonstrates our approach to the restoration of the spatial frequencies attenuated by a blurred kernel. In each row, we show images from the cat dataset [34] in the first column and the magnitudes of their Fourier components in several frequency bands in the subsequent columns. The frequency components in each band are obtained by a convolution of the input image with a Butterworth band-pass filter. Although most of the frequency components of the blurred image in the second row have been annihilated, the recovered components (shown in the last row) highly resemble those of the original using the characteristic Fourier magnitude spectrum specific to the training images (in the third row).

2. Problem Formulation

Our problem is stated as follows. Given a set of N sharp training images $\{\mathbf{z}_i | i = 1, \ldots, N\}$ and an arbitrary blurred image $\mathbf{y}$ that belong to the same class, we aim to recover the latent image $\mathbf{x}$ and the kernel $\mathbf{k}$.

2.1. Image prior

Now we formulate the class-specific image prior, which states that, the frequency components in each band span a sparse linear subspace in the Fourier domain. To proceed, we use the notation $\mathcal{F}_{\mathbf{x}}(\omega)$ to denote the Fourier coefficient
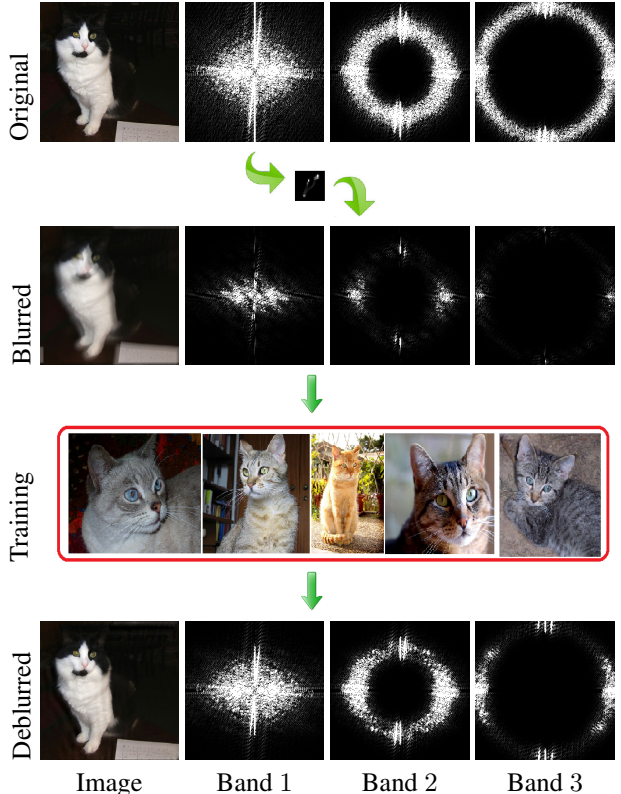

Figure 1: Recovering spatial frequencies that have been attenuated by a blur kernel using bandpass frequency components from the training data.

of the 2D function/image $\mathbf{x}$ at the spatial frequency ω. Dividing the frequency spectrum into a set of M frequency bands, we formulate the linear subspace constraint for band b_j as $\mathcal{F}_{\mathbf{x}}(\omega) = \sum_{i=1}^{N} w_{i,j} \mathcal{F}_{\mathbf{z}_i}(\omega), \forall \omega \in b_j, j = 1, \ldots, M$, where $w_{i,j}$ is a weight associated with the training image $\mathbf{z}_i$ and the band b_j in the representation of the latent image $\mathbf{x}$. This coefficient correlates to the similarity between the frequency components of the training and the latent image in the band b_j.

In addition, we enforce sparsity on the weight vector $\mathbf{w}_j \triangleq [w_{1,j}, \ldots, w_{N,j}]$ for each band b_j. The sparsity constraint emphasizes the major contributions from a few training images to the representation of the latent image $\mathbf{x}$ for each separate band. Here, we express this constraint as a minimisation of the L_1-norm $\|\mathbf{w}_j\|_1$ due to its well-known robustness. Combining the linear subspace constraint and the sparsity constraint on $\mathbf{w}_j$ over all the frequency bands, we define the prior function $P(\mathbf{x}, \mathbf{w})$ as

$$
\begin{aligned}
P(\mathbf{x}, \mathbf{w}) \triangleq & \gamma \sum_{j=1}^{M} \sum_{\omega \in b_j} |\mathcal{F}_{\mathbf{x}}(\omega) - \sum_{i=1}^{N} w_{i,j} \mathcal{F}_{\mathbf{z}_i}(\omega)|^2 \\
& + \tau \sum_{j=1}^{M} \|\mathbf{w}_j\|_1,
\end{aligned}
\tag{1}
$$

where γ and τ are the balance factors of the reconstruction error and the sparsity term, respectively, and $|\cdot|$ denotes the modulus of a complex number.

For each band b_j, we define a corresponding band-pass filter $\mathbf{f}_j$, such as a Butterworth filter [8], whose Fourier transform is a non-zero constant c within b_j and zero elsewhere. With this filter, let us consider the 2D function $\mathbf{g} = \mathbf{x} \otimes \mathbf{f}_j - \sum_{i=1}^{N} w_{i,j}(\mathbf{z}_i \otimes \mathbf{f}_j)$, where $\otimes$ denotes the convolution operator. The Fourier transform of this function is

$$\mathcal{F}_{\mathbf{g}}(\omega) = \begin{cases} c\left(\mathcal{F}_{\mathbf{x}}(\omega) - \sum_{i=1}^{N} w_{i,j}\mathcal{F}_{\mathbf{z}_i}(\omega)\right) & \forall \omega \in b_j, \\ 0 & \text{otherwise} \end{cases}$$ (2)

Applying the Parseval's theorem to the function $\mathbf{g}$, we have $\int \mathbf{g}(u)^2 du = \int |\mathcal{F}_{\mathbf{g}}(\omega)|^2 d\omega$. Noting that $\int |\mathcal{F}_{\mathbf{g}}(\omega)|^2 d\omega$ is a multiple of the reconstruction error in equation 1, we rewrite it as follows

$$P(\mathbf{x}, \mathbf{w}) = \beta \sum_{j=1}^{M} \|\mathbf{x} \otimes \mathbf{f}_j - \sum_{i=1}^{N} w_{i,j}(\mathbf{z}_i \otimes \mathbf{f}_j)\|_2^2 + \tau \sum_{j=1}^{M} \|\mathbf{w}_j\|_1,$$ (3)

where we use the variable substitution $\beta \triangleq \frac{\gamma}{c^2}$.

2.2. Objective function

In image deblurring, the aim is to minimise the data fidelity term associated with the blur model $\mathbf{y} = \mathbf{x} \otimes \mathbf{k} + \mathbf{n}$, where $\mathbf{n}$ is the image noise. In addition, several deblurring approaches have utilised image gradients to enforce an *a priori* image gradient distribution, *i.e.* natural image statistics [4] and to better capture the spatial randomness of noise [23]. Following these previous approaches, we also exploit the derivative form of the blur model and aim to minimise the error $\|\nabla_d \mathbf{x} \otimes \mathbf{k} - \nabla_d \mathbf{y}\|_2$, where ∇_d denotes the gradient operator in the direction $d \in \{x, y\}$.

In addition, we employ a regulariser on the blur kernel using the conventional L_2-norm $\|\mathbf{k}\|_2^2$ as in previous works [2, 33]. Combining all the above components, we arrive at a minimisation of the objective function

$$J(\mathbf{x}, \mathbf{w}, \mathbf{k}) = \|\mathbf{x} \otimes \mathbf{k} - \mathbf{y}\|_2^2 + P(\mathbf{x}, \mathbf{w})$$
$$+ \sum_{d \in \{x,y\}} \|\nabla_d \mathbf{x} \otimes \mathbf{k} - \nabla_d \mathbf{y}\|_2^2 + \alpha \|\mathbf{k}\|_2^2,$$ (4)

where α is the balancing factor for the kernel regulariser.

3. Optimisation

Given $\mathbf{y}$, $\{\mathbf{f}_b | b = 1, \ldots, M\}$ and $\{\mathbf{z}_i | i = 1, \ldots, N\}$, we address the minimisation of the objective function in equation 4. Since a simultaneous minimisation with respect to all the variables $\mathbf{x}, \mathbf{w}$ and $\mathbf{k}$ is computationally expensive,

we adopt an alternating minimisation scheme. In each iteration, we alternatively solve subproblems with respect to each of the variables $\mathbf{x}$, $\mathbf{w}$ and $\mathbf{k}$, while fixing the other variables.

3.1. Estimating $\mathbf{w}$ given $\mathbf{x}$ and $\mathbf{k}$

Assuming $\mathbf{x}$, and $\mathbf{k}$ have been obtained in a certain iteration, we now find the weights $w_{i,j}$ that minimise $J(\mathbf{x}, \mathbf{w}, \mathbf{k})$. We note that the part of $J(\mathbf{x}, \mathbf{w}, \mathbf{k})$ dependent on $\mathbf{w}_j$, *i.e.* $P(\mathbf{x}, \mathbf{w})$, contains separable terms across the bands b_j's. Therefore, this optimisation step can be decomposed into independent sub-problems, each of which aims to minimise the following function with respect to $\mathbf{w}_j$

$$J_{\mathbf{w}_j} = \|\mathbf{x} \otimes \mathbf{f}_j - \sum_{i=1}^{N} w_{i,j}(\mathbf{z}_i \otimes \mathbf{f}_j)\|_2^2 + \frac{\tau}{\beta}\|\mathbf{w}_j\|_1. \quad (5)$$

The above cost function can be solved by a standard L_1-regularized least-squares solver such as the one in [12]. The above problem is usually well-formed when the length of $\tilde{\mathbf{x}}_j$ and $\tilde{\mathbf{z}}_{i,j}$ exceeds that of $\mathbf{w}_j$, *i.e.* the number of image pixels is more than the number of training images N.

3.2. Latent image estimation

With the current update of the contributions $\mathbf{w}_j$, $j = 1, \ldots, M$, from the training images to each band, and the kernel $\mathbf{k}$, we now estimate the latent image so as to minimise $J(\mathbf{x}, \mathbf{w}, \mathbf{k})$ in equation 4. As above, we only consider the sum of the terms dependent on $\mathbf{x}$

$$J_{\mathbf{x}} = \|\mathbf{x} \otimes \mathbf{k} - \mathbf{y}\|_2^2 + \sum_{d \in \{x,y\}} \|\nabla_d \mathbf{x} \otimes \mathbf{k} - \nabla_d \mathbf{y}\|_2^2$$
$$+ \beta \sum_{j=1}^{M} \|\mathbf{x} \otimes \mathbf{f}_j - \sum_{i=1}^{N} w_{i,j}(\mathbf{z}_i \otimes \mathbf{f}_j)\|_2^2.$$ (6)

Next, we apply the Parseval's theorem to the right-hand side of equation 6 and rewrite it in the Fourier domain as

$$J_{\mathbf{x}} = \int |\mathcal{F}_{\mathbf{x}}(\omega)\mathcal{F}_{\mathbf{k}}(\omega) - \mathcal{F}_{\mathbf{y}}(\omega)|^2 d\omega$$
$$+ \sum_{d \in \{x,y\}} \int |\mathcal{F}_{\nabla_d}(\omega)\mathcal{F}_{\mathbf{x}}(\omega)\mathcal{F}_{\mathbf{k}}(\omega) - \mathcal{F}_{\nabla_d}(\omega)\mathcal{F}_{\mathbf{y}}(\omega)|^2 d\omega$$
$$+ \beta \sum_{j=1}^{M} \int |\mathcal{F}_{\mathbf{x}}(\omega)\mathcal{F}_{\mathbf{f}_j}(\omega) - \sum_{i=1}^{N} w_{i,j}\mathcal{F}_{\mathbf{z}_i}(\omega)\mathcal{F}_{\mathbf{f}_j}(\omega)|^2 d\omega,$$ (7)

where ω represents a spatial frequency, $|\cdot|$ signifies the modulus of a complex number and ∇_d is the derivative filter in the direction $d \in \{x, y\}$.

Since the function in equation 7 is a convex function of $\mathcal{F}_{\mathbf{x}}(\omega)$, a local optimisation method could be applied to obtain its global minimum. We set the partial derivative $\frac{\partial J_{\mathbf{x}}}{\partial \mathcal{F}_{\mathbf{x}}}$

to zero to obtain the minimiser $\mathcal{F}_{\mathbf{x}}$ as follows

$$
\mathcal{F}_{\mathbf{x}} = (\overline{\mathcal{F}_{\mathbf{k}}}\mathcal{F}_{\mathbf{y}} + \sum_d |\mathcal{F}_{\nabla_d}|^2 \overline{\mathcal{F}_{\mathbf{k}}}\mathcal{F}_{\mathbf{y}} + \beta \sum_{j=1}^{M} |\mathcal{F}_{\mathbf{f}_j}|^2
$$
$$
\times \sum_{i=1}^{N} w_{i,j}\mathcal{F}_{\mathbf{z}_i})./(|\mathcal{F}_{\mathbf{k}}|^2 + \sum_d |\mathcal{F}_{\nabla_d}\mathcal{F}_{\mathbf{k}}|^2 + \beta \sum_{j=1}^{M} |\mathcal{F}_{\mathbf{f}_j}|^2),
$$
(8)

where the ./ notation stands for a frequency-wise division in the Fourier domain. The latent image can then be obtained by an inverse Fourier transform of $\mathcal{F}_{\mathbf{x}}$.

3.3. Blur kernel estimation

Once the latent image $\mathbf{x}$ is computed, the next step is to estimate the blur kernel $\mathbf{k}$. Based on equation 4, this optimisation step involves the following terms

$$
J_{\mathbf{k}} = \|\mathbf{x}\otimes\mathbf{k}-\mathbf{y}\|_2^2 + \sum_d \|\nabla_d\mathbf{x}\otimes\mathbf{k}-\nabla_d\mathbf{y}\|_2^2 + \alpha\|\mathbf{k}\|_2^2. \quad (9)
$$

Again, we leverage the Parseval's theorem and express the above function in the Fourier domain as

$$
J_{\mathbf{k}} = \int |\mathcal{F}_{\mathbf{x}}(\omega)\mathcal{F}_{\mathbf{k}}(\omega) - \mathcal{F}_{\mathbf{y}}(\omega)|^2 d\omega + \alpha \int |\mathcal{F}_{\mathbf{k}}(\omega)|^2 d\omega
$$
$$
+ \sum_{d\in\{x,y\}} \int |\mathcal{F}_{\nabla_d}(\omega)\mathcal{F}_{\mathbf{x}}(\omega)\mathcal{F}_{\mathbf{k}}(\omega) - \mathcal{F}_{\nabla_d}(\omega)\mathcal{F}_{\mathbf{y}}(\omega)|^2 d\omega.
$$
(10)

Since $J_{\mathbf{k}}$ is a quadratic function of $\mathcal{F}_{\mathbf{k}}(\omega)$, we can obtain the minimiser by setting $\frac{\partial J_{\mathbf{k}}}{\partial \mathcal{F}_{\mathbf{k}}}$ to zero. As a result, we arrive at the following closed-form for $\mathcal{F}_{\mathbf{k}}$

$$
\mathcal{F}_{\mathbf{k}} = (\overline{\mathcal{F}_{\mathbf{x}}}\mathcal{F}_{\mathbf{y}} + \sum_d |\mathcal{F}_{\nabla_d}|^2 \overline{\mathcal{F}_{\mathbf{x}}}\mathcal{F}_{\mathbf{y}})./
$$
$$
(|\mathcal{F}_{\mathbf{x}}|^2 + \sum_d |\mathcal{F}_{\nabla_d}\mathcal{F}_{\mathbf{x}}|^2 + \alpha).
$$
(11)

3.4. Implementation

Algorithm 1 summarises our optimisation approach. The algorithm initialises the latent image to the blurred one and the kernel to the Dirac delta function. Subsequently, it proceeds in an iterative manner. The update steps for $\mathbf{x}$ and $\mathbf{k}$ are undertaken by fast forward and inverse Fourier transforms according to equations 8 and 11. In practice, for stability reasons, the kernel estimation only involves the gradient terms in the numerator and denominator of equation 11. The algorithm terminates when it reaches a preset number of iterations or the changes in $\mathbf{x}$ and $\mathbf{k}$ over two successive iterations fall below preset tolerance thresholds.

To further enhance the estimation accuracy, we progressively increase the kernel size in a coarse-to-fine scheme.

Algorithm 1 Deblurring with the class-specific prior.

Require:
  $\mathbf{y}$: the given blurred image.
  $\mathbf{z}_i$, $i = 1, \ldots, N$: the class-specific training images.
  $\mathbf{f}_j$, $j = 1, \ldots, M$: a set of band-pass filters covering the frequency spectrum.
  α, β: the weights of the terms in Equation 4.
  ρ: the attenuation factor of the class-specific prior.

1: $\mathcal{F}_{\mathbf{x}} \leftarrow \mathcal{F}_{\mathbf{y}}$.
2: $\mathbf{k} \leftarrow \delta$ (Dirac delta kernel).
3: **while** $size(\mathbf{k}) \leq max\_size$ **do**
4:    $\beta \leftarrow \beta_0$.
5:    **repeat**
6:       Minimise $J_{\mathbf{w}_j}$ in 5 w.r.t. $\mathbf{w}_j$, $\forall j$, with solver in [12].
7:       Update $\mathbf{x}$ according to equation 8.
8:       $\beta \leftarrow \rho\beta$.
9:       Update $\mathbf{k}$ according to equation 11.
10:    **until** the maximum number of iterations is reached or $\mathbf{x}$ and $\mathbf{k}$ change by an amount below a relative tolerance threshold.
11:    $\mathbf{k} \leftarrow upsample(\mathbf{k})$ (Initialisation of kernel for the following scale) .
12: **end while**
13: **return** Latent image $\mathbf{x}$ and blur kernel $\mathbf{k}$.

Within a fixed kernel scale, we iterate between the estimation steps with respect to $\mathbf{w}$, $\mathbf{x}$ and $\mathbf{k}$ until convergence, before expanding the kernel size to the next scale. Starting at the coarsest size of 3×3, the kernel is expanded by a factor of $\sqrt{1.6}$ between two successive scales. The initialisation of the kernel in the following scale is obtained through a bicubic interpolation of the final kernel estimated in the previous scale (as in line 11).

In addition, we adjust the weight β of the class-specific prior incrementally over iterations (in line 8). The reason for this adjustment is that we initially prefer to obtain as much class information as needed to constrain the space of the latent image. As the iterations proceed, the resulting latent image and kernel will increasingly gather instance-specific details from the given blurred image, rather than the class prior. In our experiments, we initialise β to $\beta_0 = 50$, and attenuate it by a factor of $\rho = 1.3$ per iteration until it is below a threshold of 0.01. In addition, we preset $\alpha = 10$ and $\tau = 0.01$.

4. Results and Discussion

In this section, we provide a detailed performance comparison between our method and a number of state-of-the-art alternatives. Firstly, we demonstrate the influence of the class-specific prior on the overall performance. Next,
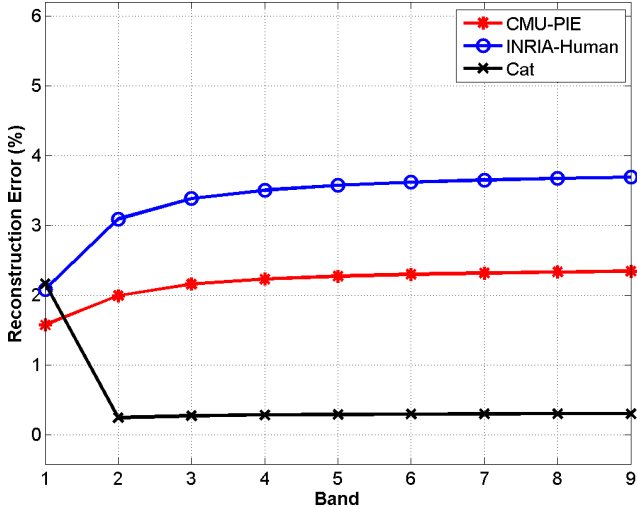
Figure 2: The relative representation errors (averaged over the test images) for different datasets.

Ground truth   Blurred image   Without prior   With prior

Figure 3: Latent images and kernels recovered by our method with and without the proposed prior.

Prior	SSIM		PSNR (dB)	
	Without	With	Without	With
Image	0.3648	**0.7537**	16.8651	**25.7811**
Kernel	0.7065	**0.8552**	39.4240	**42.6563**

Table 1: A comparison of the accuracy achieved by our deblurring framework with and without the proposed prior.

we compare our method to a number of well-known deblurring algorithms with and without class-specific information/exemplars.

4.1. Datasets and experimental setup

We performed the experimental validation on six datasets including the CMU PIE face dataset [25], the car dataset [13], the cat dataset [34], the ETHZ dataset of shape classes [5], the cropped Yale face database B [6] and the INRIA person dataset [3]. For each dataset, we randomly selected half of the images as training data and between 10 to 15 sharp images from the remaining half as ground-truth test images for deblurring. To generate blurred images, we employed the eight complex blur kernels that emulate camera shakes in [17]. We use the method of [16] to recover the final image in the final non-blind deconvolution step. We have implemented our algorithm in MATLAB on an Intel Core^{TM} i7 machine with 16 GB of memory. Our unoptimised implementation takes approximately 33 seconds for 50 iterations to deblur an image of 240×320 pixels which has been blurred by a kernel with a size of 21×21 pixels. To model the class-specific image prior $P(\mathbf{x}, \mathbf{w})$, we employ 90 Butterworth bandpass filters occupying with nine different orders per band. The filters together span the full frequency spectrum, with an overlap between successive bands.

4.2. Analysis of the proposed framework

To validate the prior, we report the relative errors of representing test images using training images, which is evaluated for every band in the frequency domain. The plot in figure 2 shows the average relative representation error for ten test images across nine different bands, with respect to

100 training images per dataset. Overall, the error is reasonably low (below 4% across all the datasets and bands). This supports the claim that, with a sufficient number of training images, the bandpass responses of test images can be represented as a linear combination of those of the training images with a reasonably low error.

Now we demonstrate the effectiveness of the proposed class-specific prior within our deblurring framework. In Figure 3, we compare the visual quality of the recovered latent image and the kernel obtained with (last column) and without (third column) the prior. As observed, the image recovered with the prior does not contain visible artefacts, whereas the one obtained without the prior shows severe ringing and multiple false edges. Under close inspection, we observed a noisy kernel similar to the delta kernel with the prior excluded, as opposed to an accurate kernel estimate with the prior included. This suggests that the method may have not converged without the proposed prior.

Further, we have compared the accuracy of the recovered images and blur kernels with and without using the class-specific prior. In Table 1, we report the accuracy measured by the structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR in dB) averaged across the above datasets. These results demonstrate the accuracy of the recovered images and kernels improve significantly (by several orders of magnitude) with the proposed prior. This is consistent with the visual observations in Figure 3, suggesting that the proposed prior plays an important role in correctly guiding the estimates to the ground truth.

4.3. Comparisons with generic priors

In this section, we evaluate the performance of our method compared to several state-of-the-art deblurring methods that use non-class-specific priors, on the datasets mentioned earlier. The methods considered for this com-

Methods	SSIM						PSNR (dB)					
	Car [13]	Shape [5]	Cat [34]	CMU [25]	Person [3]	YaleB [6]	Car [13]	Shape [5]	Cat [34]	CMU [25]	Person [3]	YaleB [6]
Fergus [4]	0.411	0.415	0.598	0.559	0.207	0.535	16.99	16.69	19.88	18.26	14.81	19.56
Shan [23]	0.632	0.624	0.742	0.775	0.407	0.773	21.56	21.61	25.20	25.59	17.78	26.42
Cho [2]	0.559	0.595	0.627	0.699	0.293	0.678	19.99	20.47	22.54	24.38	15.05	22.99
Xu [31]	0.631	0.638	0.704	0.739	_	0.681	20.93	21.25	22.73	23.30	_	23.30
Krishnan [15]	0.544	0.544	0.668	0.693	0.296	0.755	19.75	19.73	22.79	23.54	15.41	24.09
Levin [18]	0.500	0.567	0.699	0.758	0.332	0.673	18.09	19.24	23.12	24.31	16.77	25.22
Cai [1]	0.298	0.358	0.292	0.178	_	0.205	13.89	14.86	14.63	11.72	_	12.37
Zhong [35]	0.485	0.520	0.643	0.641	_	0.655	17.23	18.00	20.73	20.93	_	22.16
Sun [26]	0.481	0.669	0.724	0.744	_	0.680	19.06	22.50	23.93	24.78	_	23.74
Ours	**0.765**	**0.715**	**0.864**	**0.881**	**0.509**	**0.788**	**24.51**	**23.43**	**30.10**	**30.75**	**18.56**	**27.35**

Table 2: Accuracy of deblurred images, measured by SSIM and PSNR. Best results are in bold.

Methods	Fergus [4]	Shan [23]	Cho [2]	Xu [31]	Krishnan [15]	Levin [18]	Cai [1]	Zhong [35]	Sun [26]	Our Method
PSNR	38.59	41.00	41.15	41.33	39.58	39.48	38.86	40.01	40.99	**42.92**

Table 3: Average kernel estimation accuracy, measured by PSNR, across the datasets under study. Best results are in bold.

No. images	50	125	250	500	1000	2000
Image	25.87	26.97	27.92	28.58	30.42	30.75
Kernel	41.15	42.11	42.80	42.99	44.01	44.13

Table 4: Deblurring performance (in PSNR) over different numbers of training images in the CMU PIE dataset.

parison include Fergus *et al*. [4], Shan *et al*. [23], Cho and Lee [2], Xu and Jia [31], Krishnan *et al*. [15], Levin *et al*. [18], Cai *et al*. [1], Zhong *et al*. [35] and Sun *et al*. [26].

In Table 2, we present the average image SSIM and PSNR measures across the above datasets. Among the considered methods, ours is the best performer for each dataset, in terms of both metrics. In fact, our method generally leads the second-best performer by more than 15% in terms of SSIM values, while our PSNR measure is several orders of magnitude higher. Similarly, Table 3 shows the average PSNR (in dB) of the kernel estimates for the same experiment. Again, our method outperforms all the competitors and leads the second best by several orders of magnitude. Since our method captures class information in every frequency band of the latent image, it is capable of estimating all the frequency components of the kernel with the latent image in hand. Unlike previous methods, this capability allows our method to cope with a wide range of kernels, irrespective of their sparsity property.

Next, we examine the variation of deblurring performance with respect to the number of training images. Table 4 shows that the image and kernel estimation accuracy for the CMU PIE dataset improves steadily with an increase in the number of training images. Even with only 50 train-

ing images, our method is able to produce an average image accuracy of 25.87 dB, outperforming all the others (whose results are in Table 2).

Next, we examine the visual quality constructed by our method as compared to the others by examples. In Figure 4, we show the deblurring results for a cat image from the dataset in [34]. Our recovered image, as shown in Figure 4n, exhibits remarkable sharpness and fine textures in places such as the cat's eyes, neck, mouth and whiskers. In contrast, these details are not perceivable in the results produced by the other methods. Further, a magnified view of the results in Figures 4e to 4m shows that the methods that rely on edges, *e.g*. [15], and patches with high contrast, *e.g*. [26], have failed to yield an accurate estimation of the kernel, because the blurred input image lacks texture and strong edges. Our method even outperforms [21], which uses class exemplars with an additional input mask around the cat's face, the mouth and eyes.

Subsequently, we take an example from the CMU PIE dataset [25]. As shown in Figure 5b, the blurred image is quite challenging due to the complex trajectory of the blur kernel. As a result, the kernels yielded by the algorithms in [4, 23, 15, 18, 32, 21] are less sparse than the ground truth one. Our deblurred image shows gradual changes in the intensity across the face, as opposed to the flatness on the other deblurred images. This result suggests that our algorithm recovers a wider range of spatial frequencies than the other methods.

Further, we present deblurring results on a challenging image from the INRIA person dataset [3]. In Figure 6, the input image has a low resolution of 64×80 and incurs se-
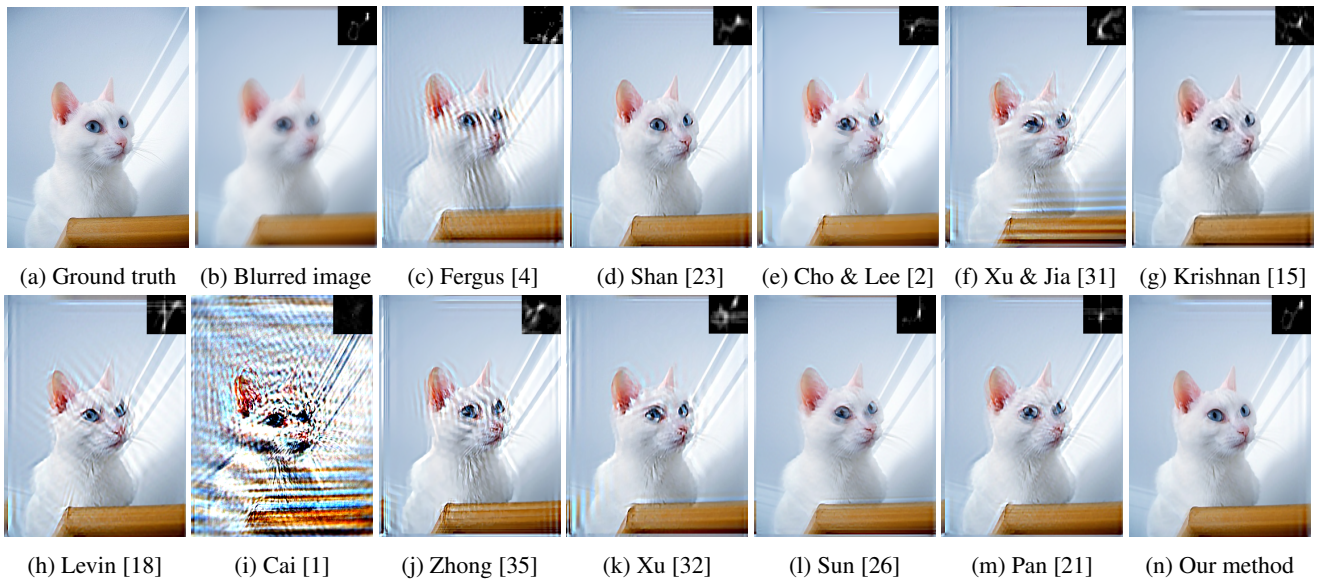
(a) Ground truth   (b) Blurred image   (c) Fergus [4]   (d) Shan [23]   (e) Cho & Lee [2]   (f) Xu & Jia [31]   (g) Krishnan [15]

(h) Levin [18]   (i) Cai [1]   (j) Zhong [35]   (k) Xu [32]   (l) Sun [26]   (m) Pan [21]   (n) Our method

Figure 4: Comparison of several deblurring methods on an image from the cat dataset[34].

(a) Ground truth   (b) Blurred image   (c) Fergus [4]   (d) Shan [23]   (e) Cho & Lee [2]   (f) Xu & Jia [31]   (g) Krishnan [15]

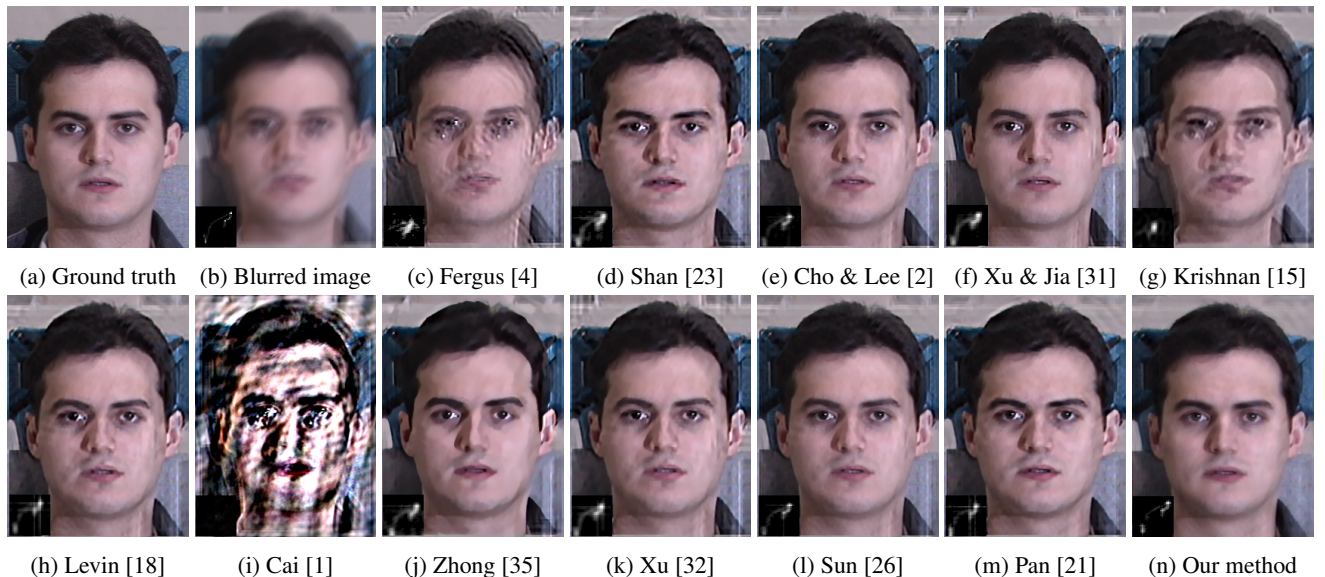(h) Levin [18]   (i) Cai [1]   (j) Zhong [35]   (k) Xu [32]   (l) Sun [26]   (m) Pan [21]   (n) Our method

Figure 5: Comparison of several deblurring methods on a face image selected from the CMU PIE dataset[25].

Original   Blurred   Fergus [4]   Shan [23]   Cho [2]   Xu & Jia [31]   Krishnan[15]   Levin [18]   Zhong [35]   Ours

Figure 6: Comparison of several deblurring methods on an image selected from the INRIA person dataset[3].

vere blurs due to the large kernel size relatively to the image size. Since all the edges are greatly distorted, methods using sparsity and gradient priors are not effective on this image, producing unrecognisable objects in the other deblurred images. However, our method successfully recovers several objects with great resemblance to the ground truth, such as
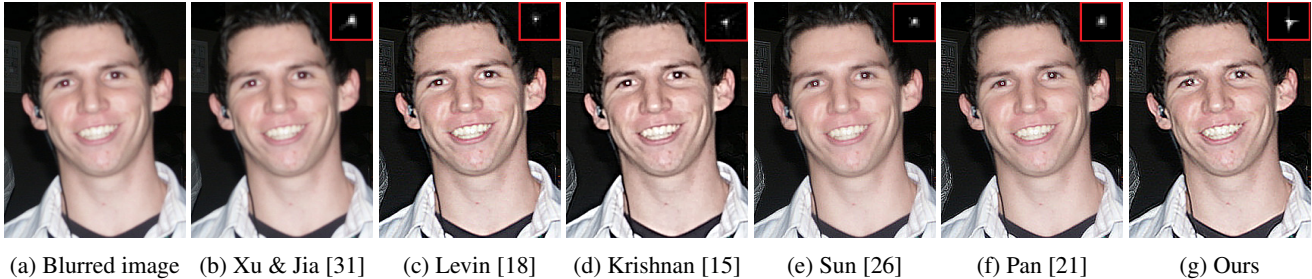
| (a) Blurred image | (b) Xu & Jia [31] | (c) Levin [18] | (d) Krishnan [15] | (e) Sun [26] | (f) Pan [21] | (g) Ours |

Figure 7: Deblurring a real-world blurred image from the dataset in [24].

(a) Original (b) Blurred image (c) Reference (d) HaCohen PSNR: 26.23

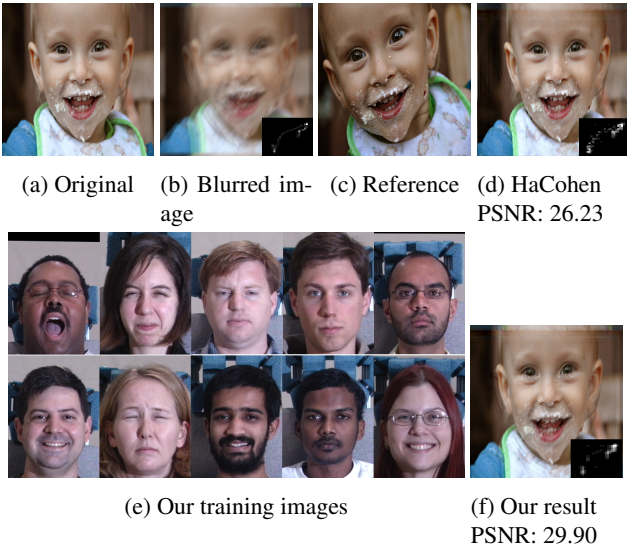(e) Our training images (f) Our result PSNR: 29.90

Figure 8: Comparison between our method and HaCohen *et al.*'s [9] on a blurred image taken from their paper.

the pedestrians and the bus in the background.

4.4. Comparison with exemplar-based methods

In Figure 8, we illustrate the advantage of our method over HaCohen *et al.* [9] on an example taking directly from their paper. Hacohen *et al.*'s [9] method requires a dense correspondence to be established between the input blurred image (Figure 8b) and a reference image (Figure 8c) of the same content and structure. Meanwhile, our method accepts flexible training images including a wide range of individuals and expressions. Although our deblurred image (in Figure 8f) appears to be visually similar to Hacohen *et al.*'s [9] (in Figure 8d), our PSNR measure is actually more than **3 dB** higher than the latter.

As another example, we compare our algorithm to a recent deblurring method by Pan *et al.* [22] on an image from the ETHZ shape classes dataset [5]. In Figure 9a, the original image consists of a large text printed on a cup in the foreground and a cluttered background. Our deblurred image in Figure 9d is nearly **4 dB** higher in PSNR than Pan *et al.*'s result in Figure 9c. Under magnification, our method

(a) Original (b) Blurred (c) Pan [22] PSNR:17.85 (d) Our results PSNR: 21.68

Figure 9: Comparison with Pan *et al.* [22] on an image containing foreground text and a complex background.

successfully restores the legibility of the background text within the marked red box, while Pan *et al.* [22] fails to achieve this.

As the last example, Figure 7 demonstrates a qualitative comparison with methods using generic priors and exemplars, for a real-world blurred image from the dataset in [24]. We fixed the size of the unknown kernel to 19×19 pixels. Our proposed method recovers all of the fine facial and hair textures and sharp highlights in the eyes while the methods in [26, 31] produce smoother images. Overall, our method is robust and our results are competitive, if not better, than the state-of-the-art methods.

5. Conclusion

We have introduced a novel class-specific image prior to further improve the performance of image deblurring. The key aim of this prior is to capture the Fourier magnitude spectrum of an image class across all frequency bands. With this aim, we formulate the prior via the subspace of band-pass filtered images, obtained through a Butterworth filter bank. Representing images in these subspaces, we have arrived at an effective deblurring framework for the reconstruction of spatial frequencies attenuated by the blurring process. In an extensive validation, we have verified the importance of the novel prior to our deblurring framework. Furthermore, our method has been demonstrated to significantly outperform state-of-the-art deblurring methods, both numerically and visually.

References

[1] J.-F. Cai, H. Ji, C. Liu, and Z. Shen. Framelet-based blind motion deblurring from a single image. *Image Processing, IEEE Transactions on*, 21(2):562–572, Feb 2012.

[2] S. Cho and S. Lee. Fast motion deblurring. In *ACM SIG-GRAPH Asia*, pages 145:1–145:8, New York, NY, USA, 2009. ACM.

[3] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, volume 1, pages 886–893, June 2005.

[4] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman. Removing camera shake from a single photograph. volume 25, pages 787–794. ACM, 2006.

[5] V. Ferrari, F. Jurie, and C. Schmid. From images to shape models for object detection. *IJCV*, 87:284–303, 2010.

[6] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *TPAMI*, 23:643–660, 2001.

[7] J.-M. Geusebroek and A. W. M. Smeulders. A six-stimulus theory for stochastic texture. *IJCV*, 62(1-2):7–16, 2005.

[8] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison-Wesley Longman Publishing Co., Inc., 2nd edition, 1992.

[9] Y. Hacohen, E. Shechtman, and D. Lischinski. Deblurring by example using dense correspondence. In *ICCV*, pages 2384–2391, Dec 2013.

[10] N. Joshi, W. Matusik, E. H. Adelson, and D. J. Kriegman. Personal photo enhancement using example images. *ACM Trans. Graph*, 29(2):12:1–12:15, 2010.

[11] N. Joshi, R. Szeliski, and D. Kriegman. PSF estimation using sharp edge prediction. *CVPR*, pages 1–8, 2008.

[12] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. An Interior-Point Method for Large-Scale L1-Regularized Least Squares. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):606–617, 2007.

[13] J. Krause, M. Stark, J. Deng, and L. Fei-Fei. 3D Object Representations for Fine-Grained Categorization. In *ICCVW*, pages 554–561, 2013.

[14] D. Krishnan and R. Fergus. Fast image deconvolution using hyper-laplacian priors. In *NIPS*, pages 1033–1041. Curran Associates, Inc., 2009.

[15] D. Krishnan, T. Tay, and R. Fergus. Blind deconvolution using a normalized sparsity measure. In *CVPR*, pages 233–240, June 2011.

[16] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. *ACM Trans. Graph.*, 26, 2007.

[17] A. Levin, Y. Weiss, F. Durand, and W. Freeman. Understanding and evaluating blind deconvolution algorithms. In *CVPR*, pages 1964–1971, June 2009.

[18] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman. Efficient marginal likelihood optimization in blind deconvolution. In *CVPR*, pages 2657–2664, 2011.

[19] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman. Understanding blind deconvolution algorithms. *TPAMI*, 33:2354–2367, Dec. 2011.

[20] S. Lyu and E. P. Simoncelli. Modeling Multiscale Subbands of Photographic Images with Fields of Gaussian Scale Mixtures. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(4):693–706, 2009.

[21] J. Pan, Z. Hu, Z. Su, and M. Yang. Deblurring face images with exemplars. In *ECCV*, pages 47–62, 2014.

[22] J. Pan, Z. Hu, Z. Su, and M. Yang. Deblurring text images via l0-regularized intensity and gradient prior. In *CVPR*, pages 2901–2908, 2014.

[23] Q. Shan, J. Jia, and A. Agarwala. High-quality motion deblurring from a single image. *ACM Transactions on Graphics (TOG)*, 27(3):73:1–73:10, 2008.

[24] J. Shi, L. Xu, and J. Jia. Discriminative blur detection features. In *Computer Vision and Pattern Recognition (CVPR)*, pages 2965–2972, 2014.

[25] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression (PIE) database. *Automatic Face and Gesture Recognition*, pages 46–51, 2002.

[26] L. Sun, S. Cho, J. Wang, and J. Hays. Edge-based blur kernel estimation using patch priors. In *ICCP*, pages 1–8, 2013.

[27] L. Sun, S. Cho, J. Wang, and J. Hays. Good Image Priors for Non-blind Deconvolution - Generic vs. Specific. In *European Conference on Computer Vision (ECCV), Part IV*, pages 231–246, 2014.

[28] A. Torralba and A. Oliva. Statistics of natural image categories. *Network: computation in neural systems*, 14(3):391–412, 2003.

[29] T. Trott. The effect of motion of resolution. *Photogrammetric Engineering*, 26:819–827, 1960.

[30] O. Whyte, J. Sivic, and A. Zisserman. Deblurring shaken and partially saturated images. *IJCV*, 110(2):185–201, 2014.

[31] L. Xu and J. Jia. Two-phase kernel estimation for robust motion deblurring. In *ECCV*, pages 157–170. Springer-Verlag, 2010.

[32] L. Xu, S. Zheng, and J. Jia. Unnatural l_0 sparse representation for natural image deblurring. In *CVPR*, pages 1107–1114, June 2013.

[33] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum. Image deblurring with blurred/noisy image pairs. In *ACM Transactions on Graphics (TOG)*, volume 26. ACM, 2007.

[34] W. Zhang, J. Sun, and X. Tang. Cat head detection-how to effectively exploit shape and texture features. In *ECCV*, pages 802–816. Springer, 2008.

[35] L. Zhong, S. Cho, D. Metaxas, S. Paris, and J. Wang. Handling noise in single image deblurring using directional filters. In *CVPR*, pages 612–619, 2013.