

Super-Resolving Noisy Images

Abhishek Singh¹

University of Illinois at Urbana-Champaign
asingh18@illinois.edu

Fatih Porikli¹

Australian National University / NICTA
fatih.porikli@anu.edu.au

Narendra Ahuja

University of Illinois at Urbana-Champaign
n-ahuja@illinois.edu

Abstract

Our goal is to obtain a noise-free, high resolution (HR) image, from an observed, noisy, low resolution (LR) image. The conventional approach of preprocessing the image with a denoising algorithm, followed by applying a super-resolution (SR) algorithm, has an important limitation: Along with noise, some high frequency content of the image (particularly textural detail) is invariably lost during the denoising step. This ‘denoising loss’ restricts the performance of the subsequent SR step, wherein the challenge is to synthesize such textural details. In this paper, we show that high frequency content in the noisy image (which is ordinarily removed by denoising algorithms) can be effectively used to obtain the missing textural details in the HR domain. To do so, we first obtain HR versions of both the noisy and the denoised images, using a patch-similarity based SR algorithm. We then show that by taking a convex combination of orientation and frequency selective bands of the noisy and the denoised HR images, we can obtain a desired HR image where (i) some of the textural signal lost in the denoising step is effectively recovered in the HR domain, and (ii) additional textures can be easily synthesized by appropriately constraining the parameters of the convex combination. We show that this part-recovery and part-synthesis of textures through our algorithm yields HR images that are visually more pleasing than those obtained using the conventional processing pipeline. Furthermore, our results show a consistent improvement in numerical metrics, further corroborating the ability of our algorithm to recover lost signal.

1. Introduction

Noise corruption is a ubiquitous phenomenon that affects many image processing tasks. Image denoising algorithms have evolved from local averaging based techniques to non-local, patch similarity driven state-of-the-art approaches [3, 1, 14]. In methods such as BM3D [3] and non-local means (NLM) [1], each noisy patch is denoised by seeking several similar patches within the noisy image and computing their mean, with the intention of averaging out the noise, while retaining the underlying image structure. Such approaches are justified by studies on statistics of natural images which suggest that patches in an image tend to recur within the

¹A part of the research was done while A. Singh interned with F. Porikli at Mitsubishi Electric Research Labs (MERL), Cambridge, MA, USA.



Figure 1. (a) Noisy low resolution image as input. (b) Result obtained using the conventional processing approach of denoising followed by super-resolving, using state-of-the-art methods [3, 7]. (c) Result obtained by super-resolving first, followed by denoising. (d) Our result.

same image [23].

Like denoising, the single image super-resolution (SR) problem is also commonly addressed using patch-similarity. Many current state-of-the-art SR algorithms are based on seeking high-resolution (HR) versions of each low-resolution (LR) image patch, using a training database of LR-HR pairs [22, 6]. In [7, 4, 20], refined versions of this approach are proposed wherein the LR-HR training database is created using scaled down version(s) of the given LR image itself. Such self-similarity based approaches are again driven by natural image statistics which suggest that patches recur in an image not just at one scale but at multiple scales [7, 23].

While both denoising and SR use patch-similarity based priors, they are used towards different objectives. The goal in denoising is to seek a large number of similar patches so as to average out noise. On the other hand, SR increases the level of similarity required, to seek more similar, usually fewer, patches, at different scales, so as to obtain the best estimate of the high frequency textural content for each patch. While denoising seeks similar patches among noise corrupted patches, SR assumes a noise free database. In a noisy image, the SR algorithm would tend to match even the noise part, and would thus ‘overfit’ while searching for similar patches, in an effort to preserve textural details. Due to these conflicting objectives, it is difficult to simultaneously perform denoising and SR of a noisy LR image under a uni-

fied patch-recurrence driven algorithm.

To super-resolve an image with considerable noise, the conventional approach is therefore to first preprocess with a denoising algorithm, followed by using an SR algorithm of choice². However, being an ill-posed problem, denoising is subject to inherent performance bounds [2, 11, 10]. Some components of the underlying signal are bound to be attenuated or lost by any denoising algorithm. In general, such losses are more severe in areas containing complex structures such as fine textures. This loss of textural detail is particularly detrimental if the subsequent operation to be performed is super-resolution, since the synthesis of such high frequency details is the challenge in SR algorithms.

Our Contributions. In this paper, we propose a framework for obtaining a clean, HR image from a noisy LR image. We attempt to overcome the signal loss caused by denoising as a preprocessing step, when the subsequent operation is super-resolution. Our algorithm begins by obtaining two HR images from the given noisy LR image. The first image is obtained by denoising the given LR image followed by super-resolving it (as is conventionally done). We call this the *denoised HR* image. The second image is obtained by directly super-resolving the noisy LR image. We call this the *noisy HR* image. While also containing noise, the noisy HR image contains some of the textural components which are not present in the denoised HR image due to the denoising loss. In order to obtain a noise-free image that also contains these textural details, we propose a linear framework that obtains the desired HR image as a convex combination of the denoised HR image and the noisy HR image. This linear combination is performed on orientation and frequency selective bands of the two images, such as those obtained using the steerable pyramid decomposition [17, 18]. As we show in Section 2, on doing so we can obtain a desired HR image where 1) a part of the denoising loss is *recovered* in the HR domain and 2) additional textures can be *synthesized* by appropriately constraining the parameters of the linear combination. These parameters are determined based on our experiments which reveal where (in spatial and oriented frequency domains) signal loss is most prevalent. We describe these constraints and procedures to obtain the parameters in Sections 3, 4, 5.

We show that this part-recovery and part-synthesis of textures using our approach yields HR images that are visually more pleasing and richer in textural content than those obtained using the conventional processing pipeline. To corroborate our hypothesis that our algorithm does indeed recover lost textural components and not just hallucinates them, we also compute quantitative metrics (PSNR and SSIM [25]) over several test images and observe a consistent improvement in these metrics.

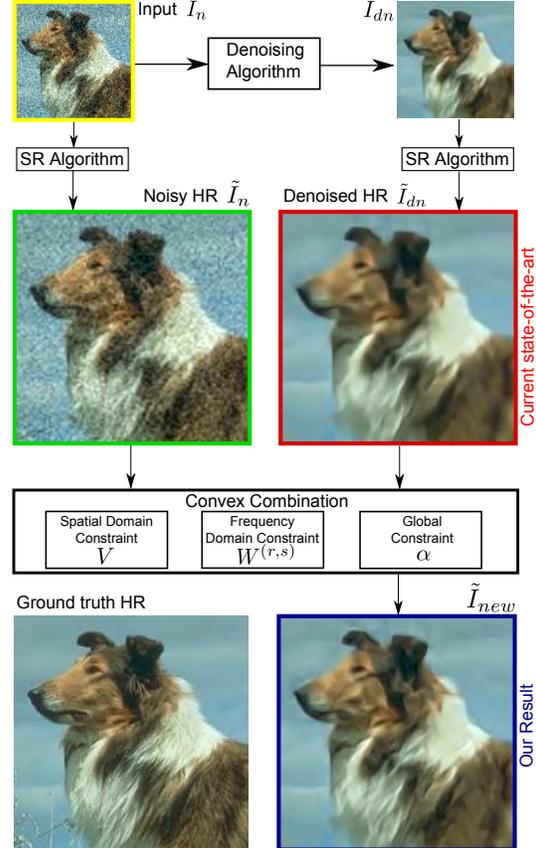


Figure 2. Summary of proposed approach for obtaining a noise-free HR image from a noisy LR image. Using a linear convex combination framework, our algorithm facilitates part-recovery and part-synthesis of lost textures. Our result (blue box) appears richer in texture as compared to the current state-of-the-art (red box).

2. Proposed Model

Notation. We use capital letters to denote images/matrices, as well as scalar constants, as appropriately defined. We use scripted letters ($\mathcal{S}, \mathcal{U}, \mathcal{B}$ etc.) to denote operators, and/or sets, as appropriate. We use the *tilde* symbol to denote HR versions of LR images. Therefore, if \mathcal{S} is a super-resolution operator, $\tilde{I} = \mathcal{S}(I)$. We denote indices using super-scripts.

Consider a noisy observation $I_n = I + N(\sigma)$ of an LR image I under additive white Gaussian noise $N(\sigma)$ of variance σ . Our goal is to obtain the best estimate of the HR version \tilde{I} of the noise free image I .

Let \mathcal{D} be a denoising operator such that $I_{dn} = \mathcal{D}(I_n)$. If $D(\sigma)$ is the signal loss caused by \mathcal{D} , we can write, $I_{dn} = I - D(\sigma)$. Now, on super-resolving this denoised image I_{dn} (and assuming the SR operation to be linear³), we get,

$$\tilde{I}_{dn} = \tilde{I} - \tilde{D}(\sigma) \quad (1)$$

Here \tilde{I}_{dn} denotes the denoised HR image. \tilde{I}_{dn} is the result obtained using the conventional approach of denoising as a preprocessing step before super-resolving. Such an approach results in loss of signal, given by $\tilde{D}(\sigma)$.

³Although most SR algorithms are not linear, we make this assumption to simplify our analysis and clarify the motivation of our algorithm.

²Note that the reverse approach of super-resolving first, followed by denoising, yields unacceptable results as shown in Fig. 1. This happens because SR introduces spatial correlation in the noise, and most denoising algorithms fail at removing correlated noise.

Can we obtain a better estimate for \tilde{I} than (1)? To answer this, let us now super-resolve the noisy LR image I_n ,

$$\tilde{I}_n = \tilde{I} + \tilde{N}(\sigma) \quad (2)$$

Now, consider a new estimate \tilde{I}_{new} of \tilde{I} that is obtained by taking a convex combination of \tilde{I}_{dn} and \tilde{I}_n ,

$$\tilde{I}_{new} = (1 - A) \cdot \tilde{I}_{dn} + A \cdot \tilde{I}_n \quad (3)$$

where ‘ \cdot ’ denotes Hadamard or entry-wise product, and the weighting matrix A contains values in $[0, 1]$. Substituting \tilde{I}_{dn} and \tilde{I}_n from (1) and (2),

$$\tilde{I}_{new} = (1 - A) \cdot [\tilde{I} - \tilde{D}(\sigma)] + A \cdot [\tilde{I} + \tilde{N}(\sigma)] \quad (4)$$

$$= \tilde{I} - (1 - A) \cdot \tilde{D}(\sigma) + A \cdot \tilde{N}(\sigma) \quad (5)$$

$$= \tilde{I}_{dn} + A \cdot \tilde{D}(\sigma) + A \cdot \tilde{N}(\sigma) \quad (6)$$

We now compare this new estimate \tilde{I}_{new} of (6), with the conventionally obtained image \tilde{I}_{dn} in (1). We observe that in addition to \tilde{I}_{dn} that is obtained by conventional processing, (6) contains two more terms: The first additive term, $A \cdot \tilde{D}(\sigma)$, recovers a fraction (A) of the underlying textural signal that is lost during the denoising step. The second term, $A \cdot \tilde{N}(\sigma)$, introduces high frequency (noisy) components into \tilde{I}_{new} . As we describe later, appropriately filtering the noisy components to align with underlying local image structure serves as a way to *synthesize* additional texture. In order to facilitate such texture synthesis, we reformulate the convex combination model of (3) in terms of orientation and frequency selective bands of the images [17]. Given an image I , let $\{B^{(r,s)}\}$, $r = 1, \dots, R$, $s = 1, \dots, S$ denote its responses to a filter bank consisting of S scales and R orientation bands per scale. We rewrite the model of (3) in terms of frequency bands as,

$$\tilde{B}_{new}^{(r,s)} = (1 - A^{(r,s)}) \cdot \tilde{B}_{dn}^{(r,s)} + A^{(r,s)} \cdot \tilde{B}_n^{(r,s)} \quad (7)$$

Note that we have now replaced the weighting matrix A , with a set of weighting matrices $A^{(r,s)}$, one for each band (r, s) . We propose a further re-parameterization of $A^{(r,s)}$ to the form,

$$A^{(r,s)} = \alpha V \cdot W^{(r,s)} \quad (8)$$

As we discuss below, such a re-parameterization allows for incorporation of several prior constraints, without which, determining the optimal coefficients for the convex linear combination of \tilde{I}_{dn} and \tilde{I}_n is difficult.

The matrix V with values in $[0, 1]$ is called the *variance map*, and for every pixel location in the scene, it measures the “textureness” of the local neighborhood. We explain our procedure for its estimation in detail in Section 3. The variance map allows us to perform the linear mixing of (7) in a spatially selective manner. In smooth, textureless regions, V favors greater influence from the denoised HR image, since there is little textural loss expected in such regions.

Our convex combination model presents a tradeoff: We see in (6) and (7) that choosing high values of the mixing weights would help recover more of the signal lost during denoising, but would also introduce more noise through

$\tilde{N}(\sigma)$ (present in \tilde{I}_n). We show through experiments in Section 4 that at any location in the image, denoising loss is prevalent only in the most dominant orientation bands. Therefore, instead of uniformly combining all orientation bands of \tilde{I}_n and \tilde{I}_{dn} , it would suffice to combine only those bands corresponding to the dominant local texture orientation. The advantage of doing so is that only a filtered version of the noisy components from \tilde{I}_n would be introduced in the resulting image \tilde{I}_{new} . Such orientation selective addition of noisy components in fact serves to perceptually enhance the local texture. Indeed, this has been the key idea behind several “texture-from-noise” synthesis algorithms in the literature [8, 15, 19]. The matrices $W^{(r,s)}$ allow us to perform the linear mixing in such a band selective manner. We elaborate more on this in Section 4.

The scalar parameter $\alpha \in [0, 1]$ globally controls the relative weights of the overly smooth \tilde{I}_{dn} and the noisy \tilde{I}_n in the resultant linear combination \tilde{I}_{new} . While V and $W^{(r,s)}$ determine *where* to blend and *which frequencies* to blend, the scalar parameter $\alpha \in [0, 1]$ determines *how much* to blend \tilde{I}_{dn} and \tilde{I}_n . We choose an optimal α such that the resultant image \tilde{I}_{new} obeys the *kurtosis invariance* properties of noise-free natural images [24]. We elaborate this procedure in Section 5.

Once we have determined the weights of the linear combination, we use (7) to combine the bands of \tilde{I}_{dn} and \tilde{I}_n . The resulting bands are used to invert the bandpass decomposition to obtain our final result. Fig. 2 summarizes our algorithm.

3. Spatial Constraint

In this section we discuss the estimation of the variance map V , which can be easily computed as a by-product of any patch-based SR algorithm, without significant overhead. We first briefly outline the SR algorithm that we use, and then explain how we obtain V from it.

Super-Resolution Algorithm. The SR algorithm we use follows the self-similarity principles described in [7, 4]. Given an LR image I , we first create a database of LR-HR image patches, from the image I as follows: We first create a smoothed version of the input image $I_L = \mathcal{U}(\mathcal{L}(I))$, by downsampling the input image I using the operator \mathcal{L} , and then upsampling using a simple (bicubic or bilinear) interpolator \mathcal{U} . We then create two sets of image patches \mathcal{P} and \mathcal{P}_L , that contain corresponding patches extracted from I and I_L respectively. The sets \mathcal{P}_L and \mathcal{P} serve as our database of LR-HR training patches.

To super-resolve the given image I to \tilde{I} , we first create a bicubic upsampled image $I_U = \mathcal{U}(I)$, which has missing high frequency components. Now, for every patch p_U in I_U , we find its most similar patch p'_U in the LR set \mathcal{P}_L . Let $p \in \mathcal{P}$ be HR patch corresponding to p'_U . We replace p_U by p in I_U , and repeat for all patches to obtain the HR image \tilde{I} .

Variance Map Estimation. The above SR procedure replaces each patch of the LR image with its best match-

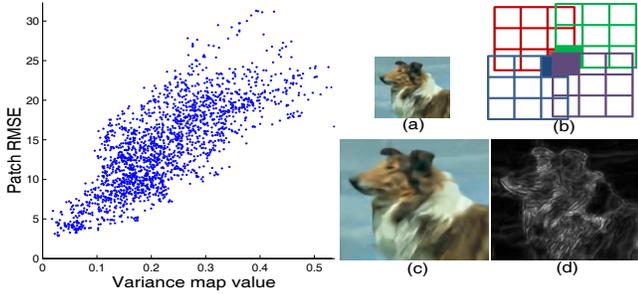


Figure 3. Given an LR image (a), we use a patch-similarity based SR algorithm to obtain the HR image (c). In the process, we obtain the variance map (d), by computing the variance across multiple predictions obtained through overlapping patches, for every HR pixel (shaded square in (b)).

ing HR patch in the database. To avoid blocking artifacts, overlap is allowed between the extracted patches. Therefore, if the patch size is 5-by-5, each pixel in the HR image would belong to 25 overlapping patches, and would receive 25 predictions during the SR process. In textural regions, these multiple explanations for the pixel are likely to be inconsistent since finding high quality patch-matches in textured regions is difficult [23]. Therefore, the variance of the multiple predictions of a pixel obtained during the SR procedure serves as a measure of the textural content of its local neighborhood. We compute this variance across all the pixels in the HR image and normalize the values to lie between 0 and 1 to thus obtain the *variance map* V . Fig. 3 illustrates this procedure with an example.

We now verify through an experiment that V does indeed indicate pixels where signal loss occurs in the denoised SR image. We obtain 50 images from the Berkeley segmentation database [13], downsample them by a factor of two, and add Gaussian noise. This creates set of noisy LR observations. We then denoise the images using the BM3D algorithm [3], and super-resolve the denoised images using the algorithm presented above, to yield the denoised HR images. In the process, we also obtain the variance maps for each image. We then extract around 1000 7-by-7 patches from all the variance maps. For each patch, we plot its average variance map value against its intensity domain RMSE value (difference between the denoised HR image and the ground truth image). In Fig. 3, we show the resulting scatter plot. Clearly, there is a strong correlation between the values in the variance map and the amount of signal lost in the denoised HR image. Regions with high values in the variance map lose more signal and are therefore expected to benefit more using our proposed convex combination model of (3), justifying our use of V in (8).

4. Frequency Domain Constraint

In this section we discuss the estimation of the parameters $W^{(r,s)}$ that facilitates frequency and orientation band selective blending.

We first examine the behavior of signal power in small, textured patches of \tilde{I}_{dn} and \tilde{I}_n , across oriented frequency bands. We again use 50 images from the Berkeley database

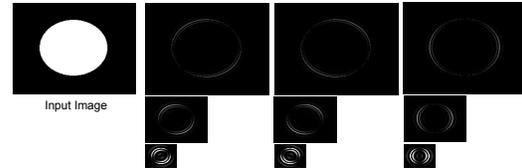


Figure 4. The steerable pyramid yields a jointly-localized (in space and frequency) invertible decomposition of an image into multi-orientation and multi-scale bands (3 orientations and 3 scales in this example).

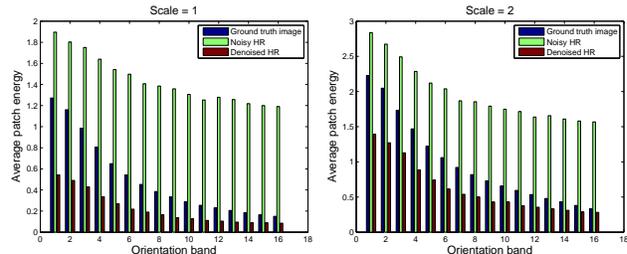


Figure 5. Average distribution of patch energy, across orientation and scale, for the denoised HR image (\tilde{I}_{dn}), noisy HR image (\tilde{I}_n) and the ground truth HR image. The signal loss in \tilde{I}_{dn} , as compared to the ground truth is primarily in the first few largest orientation bands.

and create sets of noisy HR and denoised HR images, along with their variance maps. We then compute a steerable pyramid decomposition for each image in the two sets. The steerable pyramid provides jointly-localized (space/frequency) representation of images using an invertible multi-scale, multi-orientation image decomposition [17, 18, 5], as shown in Fig. 4. We use $S = 4$ scales and $R = 16$ orientations for the decomposition. We then extract around a 1000 patches of size 7-by-7 across all bands from the 50 images, from areas containing significant textures ($V > 0.5$). We compute the average energies in these patches, in the different orientation and scale bands. To achieve rotation invariance, for each patch we sort the orientation bands in decreasing order of energy before averaging.

Fig. 5 shows the average distribution of energy across two scales and all orientation bands, for patches from the denoised HR images (red bars), the noisy HR images (green bars) and the corresponding ground truths (blue bars).

We make a simple yet important observation: Signal loss is most prevalent in the orientation bands with higher energies. In the high energy bands, we observe that the ground truth bands lie within the convex hulls of the corresponding denoised HR and noisy HR bands. This, in a way, further justifies our convex combination model of 3.

Based on this, we propose the following technique for choosing the weight matrices $W^{(j)}$, given a noisy HR image \tilde{I}_n and the denoised HR image \tilde{I}_{dn} : For any spatial location \mathbf{x} , we first consider a patch centered at \mathbf{x} in the image \tilde{I}_{dn} . Let $\mathcal{B}_\lambda^{(s)}(\mathbf{x})$ be the set of the most dominant orientation bands in the scale s in this image patch, as shown in Fig. 6. This set is determined by a scalar parameter $\lambda \in (0, 1)$ that controls the fraction (in terms of energy) of the total number of orientation bands, that are present in the set $\mathcal{B}_\lambda^{(s)}(\mathbf{x})$. For the location \mathbf{x} , we assign $W^{(r,s)}(\mathbf{x})$ the following binary

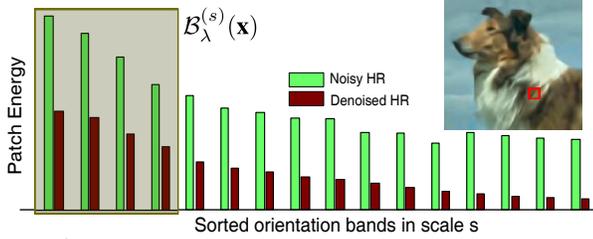


Figure 6. Given a patch (red box) at any location \mathbf{x} , and its oriented bandpass decomposition, the set $\mathcal{B}_\lambda^{(s)}(\mathbf{x})$ contains the most dominant orientation bands in the patch. The proposed convex combination is selectively done only on these bands. See text for details.

valued weight:

$$W^{(r,s)}(\mathbf{x}) = \begin{cases} 1 & \text{if } r \in \mathcal{B}_\lambda^{(s)}(\mathbf{x}) \\ 0 & \text{else} \end{cases} \quad (9)$$

The above weights effectively allow for blending the images \tilde{I}_{dn} and \tilde{I}_n only along the most dominant orientations of \tilde{I}_{dn} . As noted from the above experiment, these are the bands where maximum signal loss occurs. As far as the noise in \tilde{I}_n is concerned, it is also added into \tilde{I}_{new} , only along the direction of the underlying texture. Adding noise which is filtered along the texture orientation has the effect of perceptually enhancing the texture. We illustrate this in Fig. 7. In this simplified example, since the third orientation band has the highest energy among all bands in \tilde{I}_{dn} , it is combined with the corresponding band of \tilde{I}_n to obtain the band for \tilde{I}_{new} . The other bands of \tilde{I}_{new} are simply copied from \tilde{I}_{dn} . The resulting patch \tilde{I}_{new} appears richer in texture than \tilde{I}_{dn} .

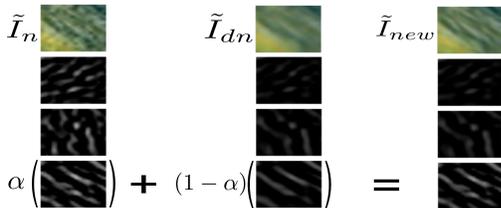


Figure 7. A simplified example showing how the orientation bands of \tilde{I}_{dn} and \tilde{I}_n are combined to get \tilde{I}_{new} . Since the third orientation band has the most energy in \tilde{I}_{dn} , the convex combination is performed on this band. Although, in this process, noise is also introduced from \tilde{I}_n , it is done so only along the texture orientation. This enhances the texture in \tilde{I}_{new} .

5. Global Constraint

We now discuss the estimation of the scalar parameter α of (8). A low value results in an overly smooth image (close to \tilde{I}_{dn}), whereas high values may result in excessive high frequency content.

To optimally choose α , we again resort to the statistical behavior of natural images across bandpass decompositions. It is well known that the marginal responses of natural images to bandpass filters is highly non-Gaussian [16, 21]. This deviation from Gaussian model can be measured by the kurtosis of the responses. In fact, studies have shown that the kurtosis of natural images remains constant across different frequency bands [9, 24].

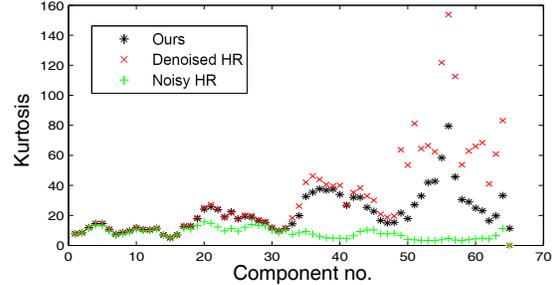


Figure 8. *Lena*. The plot shows the kurtosis values across bandpass decompositions, for the denoised HR image (\tilde{I}_{dn}), noisy HR image (\tilde{I}_n) and our result (\tilde{I}_{new}). Higher component numbers correspond to higher frequency bands. The images above the plot shows visual comparison of the results. Textures are better recovered in our image. The numbers below the images denote (PSNR in dB, SSIM).

Kurtosis of a distribution is defined as, $\kappa = \frac{\mu_4}{\sigma^4} - 3$, where μ_4 is the fourth moment about the mean, and σ is the standard deviation of the distribution. By this definition, the kurtosis of a Gaussian is zero. It has been shown in [24] that in noisy images, the kurtosis values in higher frequency bands are smaller than those in lower frequencies. This is indeed expected since noise (which predominantly affects higher frequency bands) has Gaussian statistics, and therefore has the overall effect of reducing kurtosis. We observe that on the other hand, excessive smoothing dramatically increases the kurtosis values of the high frequency bands.

We propose to choose the α that results in minimum variation of the kurtosis values across bands. Let $\kappa_{new}^{(r,s)}(\alpha)$ be the kurtosis value of the band $\tilde{B}_{new}^{(r,s)}$ of our image \tilde{I}_{new} . We obtain the optimum α as,

$$\alpha^* = \operatorname{argmin}_{0 \leq \alpha \leq 1} \sum_{r,s} \left[\kappa_{new}^{(r,s)}(\alpha) - \bar{\kappa}_{new}(\alpha) \right]^2. \quad (10)$$

$\bar{\kappa}_{new}(\alpha)$ is the mean kurtosis value across all bands. We numerically solve the above optimization problem. Alternatively, one may use MATLAB's *fminsearch* function.

6. Results

Implementation Details. We implement our method with both non-local means (NLM) [1] and BM3D [3] denoising algorithms. Both these algorithms require the noise variance as an input. Although most of our noisy images are simulated by adding noise of *known* variance, we use the algorithm of [12] to estimate noise variance from the noisy images. This is then fed to the denoising algorithms. In all our images, we found the estimated variance to be



Figure 9. *Lama*. Textures on the fur, and on rocks in the background are much better reconstructed in our result as compared to the conventional BM3D-SR. The numbers in bracket denote (PSNR in dB, SSIM).



Figure 10. *Baby*. The woolen cap in our result is significantly richer in texture as compared to the conventional NLM-SR approach. The numbers in bracket denote (PSNR in dB, SSIM).

within $\pm 5\%$ of the true variance. We run our algorithm with several different noise levels. We use $S = 4$ scales in the steerable pyramid decomposition. Scale levels of 5 or more required much larger images. For each scale, we computed decompositions along $R = 16$ orientation bands, which is the maximum allowable in the available implementation by Simoncelli. We set the band energy threshold parameter $\lambda = 0.6$ in most cases, but we also study the effects of changing it

Qualitative Results. Fig. 8 shows our result on the *Lenna* image. We plot the kurtosis values of our result across all frequency bands, and compare it to those of the denoised HR image (red markers) and the noisy HR image (green markers). Higher component numbers correspond to higher frequency bands. Due to noise, the kurtosis values in the higher frequencies of the noisy HR image are low, whereas they are very high for the denoised HR image. Subject to our constraints, our algorithm yields kurtosis values as shown by the black markers. Fig. 8 also shows our resulting image. Textural details are better preserved as compared to the denoised HR image, both visually and quantitatively. Noise variance was $\sigma = 20$ in this experiment.

Henceforth, we refer to the denoised HR images as BM3D-SR or NLM-SR, depending on the denoising algorithm used. This indicates the conventional processing approach of first denoising (BM3D or NLM), followed by super-resolution (SR).

Fig. 10 shows our results with the NLM denoising algorithm, on the *Baby* image. As compared to NLM-SR, we recover significantly more texture in the woolen cap. Fig. 9 shows our result while using the BM3D algorithm, on the *Lama* image. Again, textural details such as the fur, and the rocks behind are significantly well preserved as compared to BM3D-SR. Fig. 11 shows the results of both the NLM and BM3D based algorithms on the *Horse* image. Textures like the grass and the horse fur are visually and quantitatively better recovered by our approach, using either NLM or BM3D. Using BM3D gives slightly better results.

Fig. 12 shows results of using a different SR algorithm in our framework, on the *Dog* image. Our results improve over the corresponding baseline for both the self-similarity based SR (SsSR) algorithm as described in Section 3, as well as the sparse-coding based SR (ScSR) algorithm of [22]. Textures on the dog fur, grass, and the wooden pole are better in our results.

Quantitative Analysis. We use 50 natural images from the Berkeley segmentation database [13]. We run our algorithm(s) on these images and compute PSNR and SSIM [25]. We first analyze the quantitative performance of our algorithm(s) with different noise levels. Fig. 13 plots our results. We observe that our algorithms consistently improve over the conventional methods, across different noise levels. Fig. 15 shows the visual results of varying noise on a test image. All our images are visually better as well.

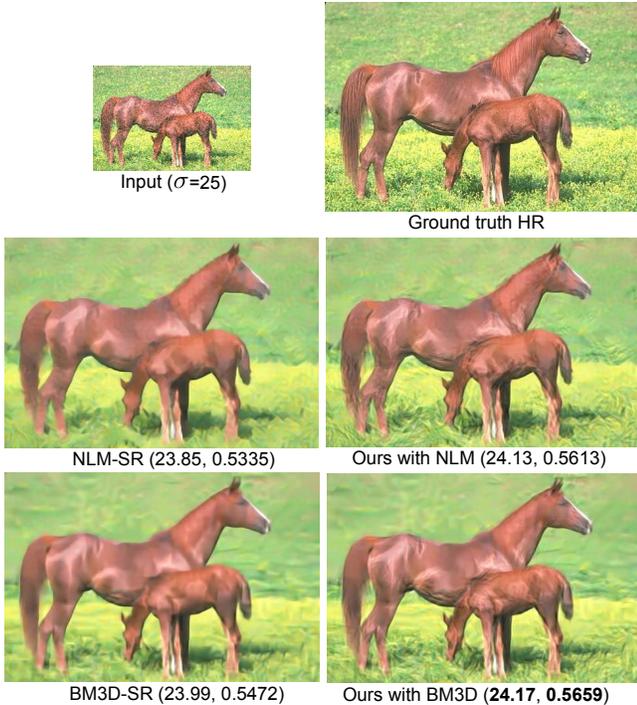


Figure 11. *Horse*. For both NLM and BM3D, our algorithms significantly improve over the respective baselines, both visually and numerically. The grass, flowers and horse fur show significant visual improvement. The numbers in brackets denote (PSNR in dB, SSIM).

In our algorithm, we have introduced a parameter $\lambda \in (0, 1)$ that controls the fraction (in terms of energy) of the orientation bands that are involved in the blending procedure. A very low value would combine only the first few (largest) bands, resulting in improvement in texture only along these specific orientations. A higher value would combine more number of bands, resulting in better recovery of texture. However, an excessively high value of λ (e.g. close to 1), would tend to copy the noisy HR image ‘as is’, and may introduce noisy components in the resulting image. Indeed, this is what we observe quantitatively as well, as shown in Fig. 14; both PSNR and SSIM first increase with increasing λ , and then drop slightly at around $\lambda = 0.8$. Nevertheless, throughout the range, our performance remains significantly higher than the conventional results, as can be seen from the plots. Fig. 16 shows the result of varying λ on a test image.

Real-world Image. We use our algorithm to enlarge a part of an image captured with camera on a high ISO setting, resulting significant sensor noise in the image. Fig. 17 shows the result of our algorithm. While in smooth regions our result looks similar to the conventionally obtained result, in textured areas our image appears better.

7. Conclusion

We have presented an algorithm for denoising and super-resolving noise corrupted images. Contrary to the conven-

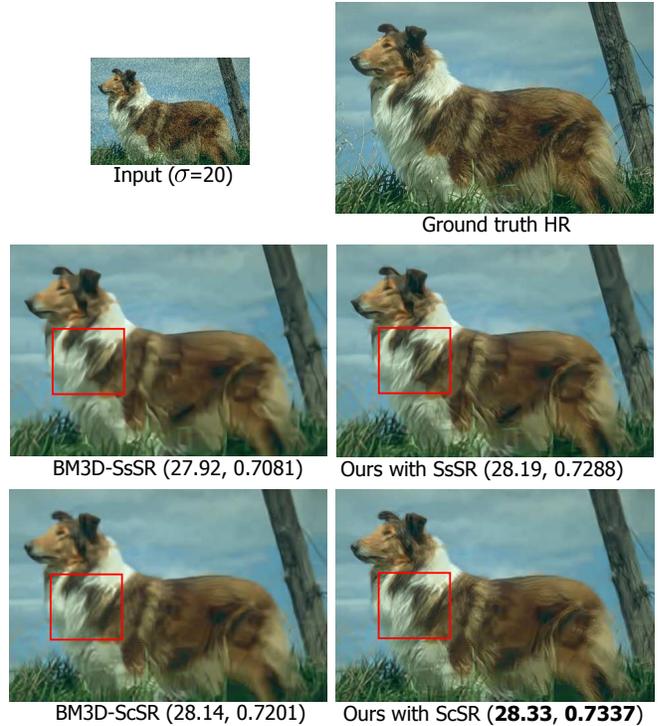


Figure 12. *Dog*. Using either self-similarity based SR (SsSR, described in Section 3), or sparse-coding based SR (ScSR [22]), our algorithm significantly improves over the respective baselines, both visually and numerically. The fur, grass and tree trunk show the most improvement visually. The numbers in brackets denote (PSNR in dB, SSIM).

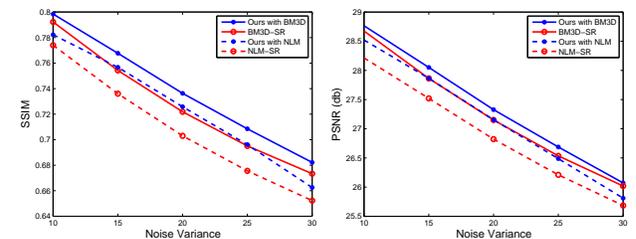


Figure 13. The plots show average SSIM (left) and PSNR (right) as functions of noise variance. Our algorithm(s) consistently improve over their corresponding baselines for all noise levels tested.

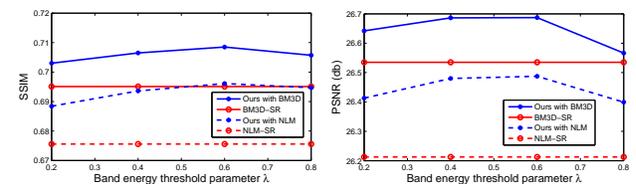


Figure 14. The figures plot average SSIM (left) and PSNR (right) as a function of the band energy threshold parameter λ . For a wide range of λ values, our performance remains significantly higher than the corresponding baselines.

tional strategy of denoising first and subsequently super-resolving the denoised image only, our algorithm, in addition, super-resolves the noisy image as well in order to extract some of the information missing in the denoised image. We have proposed an algorithm that partly restores missing texture, and partly synthesizes it from the

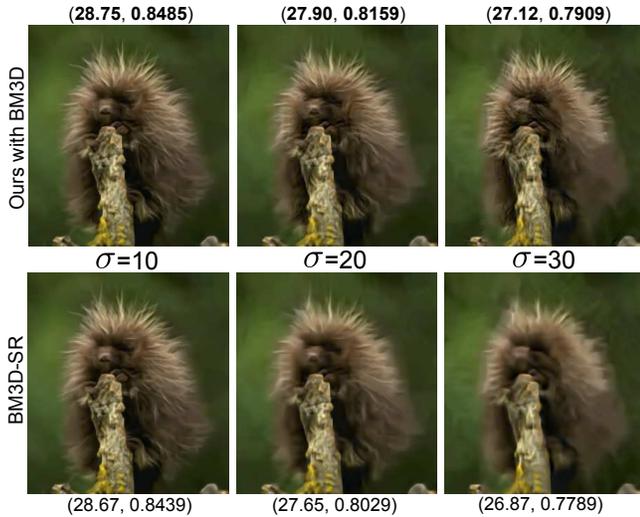


Figure 15. *Procupine*. Our algorithm can be seen to be both visually and quantitatively better than the conventional approach for a range of noise levels.

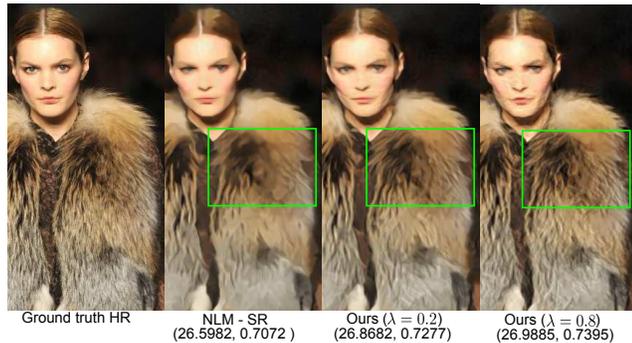


Figure 16. *Fur*. A small value of λ results in relatively smaller (but still noticeable) improvement in results. A very high value recovers more texture but may also yield a relatively more ‘noisier’ image.



Figure 17. *Real world example*. Although the estimated noise variance (5.3) in this real world image is quite lower than in any of our simulations, our result still shows perceivable improvement in visual quality as compared to the BM3D-SR baseline.

noise components present in the noisy image. We have demonstrated that our algorithm considerably outperforms the conventional processing methodology in regions containing stochastic textures. A possible limitation of our algorithm is in handling more regular (non-stochastic) textures (such as in Fig. 18). A possible solution would be



Figure 18. *Failure case*. While our algorithm yields better results in regions of irregular/stochastic texture (green box), our approach does not do as well in regions containing regular textures (red box), where our result appears slightly more noisy than the baseline.

to use more elaborate texture segmentation algorithms and use this additional information in our constraints. In this paper we have restricted ourselves to testing the basic idea wherein the weighting parameters can be easily estimated.

References

- [1] A. Buades, B. Coll, and J. M. Morel. A non-local algorithm for image denoising. In *CVPR*, 2005. 1, 5
- [2] P. Chatterjee and P. Milanfar. Is denoising dead? *IEEE Trans. Image Proc.*, 2010. 2
- [3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Trans. Image Proc.*, 2007. 1, 4, 5
- [4] G. Freedman and R. Fattal. Image and video upscaling from local self-examples. *ACM Trans. Graph.*, 2010. 1, 3
- [5] W. Freeman and E. Adelson. The design and use of steerable filters. *IEEE TPAMI*, 1991. 4
- [6] W. T. Freeman and E. C. Pasztor. Learning low-level vision. *IJCV*, 2000. 1
- [7] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In *ICCV*, 2009. 1, 3
- [8] D. J. Heeger and J. R. Bergen. Pyramid-based texture analysis/synthesis. In *SIGGRAPH*, 1995. 3
- [9] E. Lam and J. Goodman. A mathematical analysis of the dct coefficient distributions for images. *IEEE Trans. Image Proc.*, 2000. 5
- [10] A. Levin and B. Nadler. Natural image denoising: Optimality and inherent bounds. In *CVPR*, 2011. 2
- [11] A. Levin, B. Nadler, F. Durand, and W. T. Freeman. Patch complexity, finite pixel correlations and optimal denoising. In *ECCV*, 2012. 2
- [12] X. Liu, M. Tanaka, and M. Okutomi. Noise level estimation using weak textured patches of a single noisy image. In *ICIP*, 2012. 5
- [13] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, 2001. 4, 6
- [14] I. Mosseri, M. Zontak, and M. Irani. Combining the power of internal and external denoising. In *ICCP*, 2013. 1
- [15] J. Portilla and E. P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *IJCV*, 2000. 3
- [16] D. L. Ruderman. Origins of scaling in natural images. *Vision Research*, 37:3385–3398, 1997. 5
- [17] E. Simoncelli and W. Freeman. The steerable pyramid: a flexible architecture for multi-scale derivative computation. In *ICIP*, 1995. 2, 3, 4
- [18] E. Simoncelli, W. Freeman, E. Adelson, and D. Heeger. Shiftable multiscale transforms. *IEEE Trans. Info. Theory*, 1992. 2, 4
- [19] E. Simoncelli and J. Portilla. Texture characterization via joint statistics of wavelet coefficient magnitudes. In *ICIP*, 1998. 3
- [20] A. Singh and N. Ahuja. Sub-band energy constraints for self-similarity based super-resolution. In *ICPR*, 2014. 1
- [21] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S. c. Zhu. On advances in statistical modeling of natural images. *Journal of Mathematical Imaging and Vision*, 2003. 5
- [22] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Trans. Image Proc.*, 2010. 1, 6, 7
- [23] M. Zontak and M. Irani. Internal statistics of a single natural image. In *CVPR*, 2011. 1, 4
- [24] D. Zoran and Y. Weiss. Scale invariance and noise in natural images. In *ICCV*, 2009. 3, 5
- [25] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Proc.*, 2004. 2, 6