Super-Resolution Using Sub-Band Self-Similarity

Abhishek Singh and Narendra Ahuja

University of Illinois at Urbana-Champaign

Abstract. A popular approach for single image super-resolution (SR) is to use scaled down versions of the given image to build an *internal* training dictionary of pairs of low resolution (LR) and high resolution (HR) image patches, which is then used to predict the HR image. This self-similarity approach has the advantage of not requiring a separate *external* training database. However, due to their limited size, internal dictionaries are often inadequate for finding good matches for patches containing complex structures such as textures. Furthermore, the quality of matches found are quite sensitive to factors like patch size (larger patches contain structures of greater complexity and may be difficult to match), and dimensions of the given image (smaller images yield smaller internal dictionaries). In this paper we propose a self-similarity based SR algorithm that addresses the abovementioned drawbacks. Instead of seeking similar patches directly in the image domain, we use the self-similarity principle independently on each of a set of different sub-band images, obtained using a bank of orientation selective band-pass filters. Therefore, we allow the different directional frequency components of a patch to find matches independently, which may be in different image locations. Essentially, we decompose local image structure into component patches defined by different sub-bands, with the following advantages: (1) The sub-band image patches are simpler and therefore easier to find matches, than for the more complex textural patches from the original image. (2) The size of the dictionary defined by patches from the sub-band images is exponential in the number of sub-bands used, thus increasing the effective size of the internal dictionary. (3) As a result, our algorithm exhibits a greater degree of invariance to parameters like patch size and the dimensions of the LR image. We demonstrate these advantages and show that our results are richer in textural content and appear more natural than several state-of-the-art methods.

1   Introduction

The single image super-resolution (SR) problem has received significant attention in recent years. Due to its ill-posed nature, this problem has fueled research in various statistical properties of natural images, which are used as priors for regularizing the SR problem. Learning based approaches, which attempt to predict high-resolution (HR) features corresponding to the low resolution (LR) features of the given image, have become the state-of-the-art in the field [1–5].
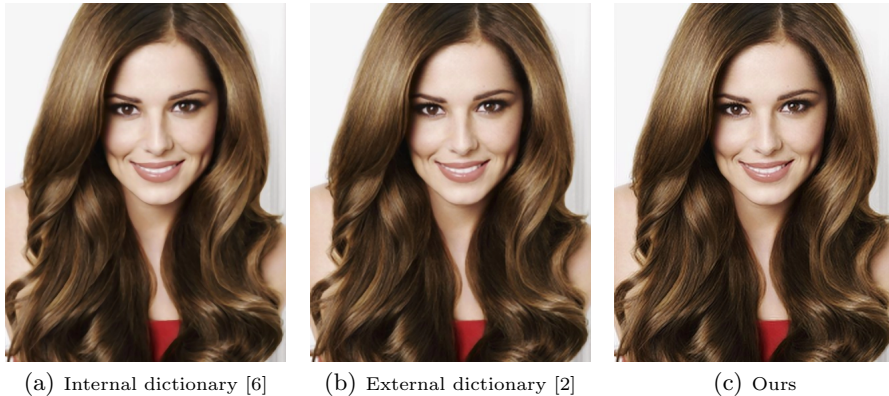
(a) Internal dictionary [6]   (b) External dictionary [2]   (c) Ours

Fig. 1. *Woman (2X):* Notice the hair. Methods based on internal dictionaries tend to smooth out fine details while preserving high contrast edges. External database driven methods can appear soft overall if the training patches are not relevant enough for super-resolving the given image. Our algorithm does not require external training images, and is designed to address the limitation of internal dictionaries. This helps enhance performance of internal dictionary based methods by better synthesizing complex image structures such as fine textures. Our results appear sharper, with more realistic details than those produced by the other schemes.

Many learning based methods first construct an *external* dictionary of LR-HR patch pairs using a database of several generic LR-HR image pairs. Given a test image to be super-resolved, this dictionary is used to predict the HR patch corresponding to each patch in the LR given image. [7, 5, 3, 2, 8, 9].

The quality of results of such methods, however, depends heavily on the construction of the external dictionary. The type and number of training images required for obtaining satisfactory SR quality is not clear. If a small database is used (for faster computations), then the results often have a strong dependence on the specific training images chosen. If a large training database is used (for better generalization), methods often have to rely on data reduction techniques such as sparsification [2] or clustering [9] of the training set for computational feasibility. This may cause a drop in performance due to loss of information in the compact representation. If a different scaling factor is desired, re-training is needed to learn new SR prediction functions for the new scaling factor.

To avoid external training databases and these problems associated with them, several methods have been proposed that exploit self-similarities within the *given* image. Similar patches are sought across scales of the given image to build an *internal* dictionary of LR-HR patch pairs. This dictionary is then used to predict the HR patch corresponding to each patch of the given image using nearest neighbor patch-matching, linear regression, etc. [10, 6, 11–14]. The general principle involved in such a self-similarity based SR algorithm is illustrated in Fig. 2(a). Self-similarity methods find their roots in fractal image coding from the 1990s [15, 16], and are based on the idea that natural image patches tend to recur within and across scales in the same image [6, 17]. It has been shown that internal dictionaries tend to contain more *relevant* training patches, and, in general, yield nearest-neighbor matches with lower error as compared to external dictionaries, with or without compact representation [17].

Internal dictionaries, however, have limitations while super-resolving textural regions. Indeed, [17] shows that the likelihood of finding a good internal match for a patch decreases as the gradient content of the patch increases. This suggests that textural details like hair, animal fur etc, often find suboptimal matches, using a self-similarity approach, and are thus averaged or smoothed out in the final SR result. The reason behind such a limitation is that the internal dictionary obtained from the given image generally has fewer number of LR-HR patch pairs than external dictionaries, which can potentially be as large as desired. Due to the limited size of the internal dictionary, textural patches (which contain complex structures) fail to find suitable representations. The size of the self-learned dictionary furthermore depends on the dimensions of the given image; smaller images consist of fewer patches and thereby yield fewer LR-HR patch pairs. Additionally, the quality of matches depends on the patch size chosen. For e.g., the complexity of structures in the patches increases with increase in patch size, making it difficult to find accurate matches.

Our Contributions: In this paper, we propose an SR algorithm that alleviates the abovementioned problems of self-similarity based approaches, *without* resorting to any external training database. We propose a self-similarity driven algorithm wherein, instead of seeking self-similar patches directly in the image domain, we use self-similarity based SR independently on images corresponding to different sub-bands. These sub-bands are the responses of the image to a bank of spatially localized, orientation selective band-pass filters. Effectively, we unravel the complexity of the structure by representing it in terms of simpler components, which, being simpler, are easier to find matches for. Unlike in the case of patch matching in the image domain, we allow the different directional frequency components of the patch to independently find their best matches in different locations in the image. Therefore, we synthesize HR patches by combining different frequency components from the best matches found at different locations. Such a combinatorial expansion of the internal dictionary allows for finding better (lower error) patch matches for a test image produces a better quality HR image. Our SR results appear richer in texture and more natural than those produced by state-of-the-art methods. We also show that our algorithm leads to improvements in two other important aspects of the SR problem that have not received much attention in the past. We show that our approach has greater degree of invariance to the choice of patch size, which can be a sensitive parameter, particularly for self-similarity methods. We also show that due to the ability of our algorithm to generate richer internal dictionaries, we are able to super-resolve extremely small images much better, thereby achieving greater invariance to the size of the input image, as compared to the existing self-similarity approach.

2   Previous Work

Among methods that rely on external dictionaries, Freeman *et. al.* [18, 7] propose an 'example-based' SR algorithm, wherein a Markov random field (MRF) model

is used to learn the relationship between LR and HR patches. [5] uses manifold learning to learn this relationship, by assuming the manifold of HR patches to be locally linear. Yang *et. al.* [3] express image patches as sparse linear combinations of atoms from a fixed dictionary of image patches. [9] proposes to cluster the patches in the dictionary to facilitate fast nearest neighbor searches. In [2, 19, 8], compact dictionaries are learnt from the raw patches, in order to support a sparse representation of the patches to be super-resolved. [20] uses an external database to learn first order regression functions to map 'in-place' (extremely localized neighborhood) patches to high resolution. Instead of learning LR-HR transformations using patches, several methods learn how edges transform across resolutions [21–23]. Primal sketches (ridges, corners, etc.) are used as primitives for super-resolution in [4]. External dictionaries are used to learn transform domain coefficients in [24, 25]. Higher level features are used in [26, 27] for learning the LR-HR mapping using external training databases.

Among internal dictionary based methods, Ebrahimi and Vrscay [10] combine ideas from fractal coding [15, 16] and example-based algorithms (such as non-local means filtering [28]), to propose a self-similarity based SR algorithm. Glasner *et al.* [6] fuse together multiple matched patches from the internal dictionary of the image to generate HR patches, in a way similar to traditional multiframe SR. Freedman and Fattal [11] show that patches tend to recur across scales within *local* spatial neighborhoods, which they exploit for computational speed-up. In [13], transform domain matching criteria are used along with the traditional L_2 distance for patch-matching. A framework for super-resolving noisy images based on self-similarity principles in proposed in [14].

Our proposed algorithm retains the advantages of existing self-similarity based approaches, but also overcomes some of their limitations described in the previous section. In the next two sections, we describe the steps involved in our algorithm, which is conceptually quite straightforward and easy to implement. In Section 5, we discuss a number of important implications and corollaries resulting out of the proposed algorithm, and discuss the key advantages it brings over existing schemes. We demonstrate our performance vis-a-vis several other state-of-the-art methods, and corroborate our claims through a number of systematic experiments in Section 6.
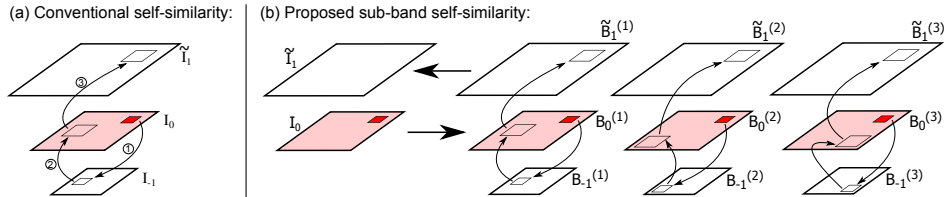

Fig. 2. (a) Conventional self-similarity based SR framework. Each patch of the given image I_0 is matched to a patch in I_{-1} in step 1. The corresponding patch (in the same location) in I_0 serves as the HR predictor (step 2). This patch is then pasted in the HR image $\tilde{I}_1$ (step 3). (b) The proposed sub-band self-similarity framework. Our method follows a series of similar steps as (a), but on each sub-band independently. Note that for super-resolving the patch shown in red, our algorithm allows for its various sub-bands to find matches in different spatial locations. See Sections 3, 4, 5 for details.

3 Overview of Proposed Method

Notation: We denote the given image to be super-resolved as I_0. By I_1 we denote the HR version of I_0, whose linear dimension, or scale, is larger by a factor of s. Similarly, we denote by I_{-1}, the smaller version of I_0, by the scaling factor of $1/s$. We denote the super-resolved image(s) obtained using our algorithm using a *hat* (ˆ) symbol. Therefore, our objective is to super-resolve I_0 to obtain an HR image $\hat{I}_1$, that best approximates the true HR image I_1. We use scripted letters to denote sets, we use lowercase boldface letters to denote image patches, and lowercase italicized letters to denote scalars and indices.

3.1 Algorithm Summary

To super-resolve the image I_0, our algorithm consists of the following steps, also summarized in Fig. 2(b):

1. We decompose the image I_0 into N sub-bands $\{B_0^{(j)}\}_{j=1}^N$, which are obtained as the responses of the image I_0 to a bank of spatially localized, orientation selective, bandpass filters. We use the steerable pyramid decomposition [29, 30] for our work, although other schemes such as contourlet transform [31] may also be used.
2. We then apply a self-similarity based SR algorithm to each of the sub-bands $\{B_0^{(j)}\}_{j=1}^N$, independently, to yield the set of HR sub-bands $\{\tilde{B}_1^{(j)}\}_{j=1}^N$. We describe this step in detail in Section 4 and discuss the key advantages it brings in Section 5.
3. We then recombine the HR sub-bands $\{\tilde{B}_1^{(j)}\}_{j=1}^N$, by inverting the sub-band decomposition, to yield an HR image $\tilde{I}_1$.
4. Finally, in order to ensure that the downsampled version of our estimated HR image is close to the given LR image, we enforce the backprojection constraint [32] by minimizing,
$$J(\hat{I}_1) = |(\hat{I}_1 * f_{psf}) \downarrow - I_0|_2^2 \tag{1}$$
Starting with $\tilde{I}_1$ as initialization, we minimize the above cost function using a few iterations of gradient decent, to yield our final HR image $\hat{I}_1$.

4 Sub-Band Self-Similarity

We independently super-resolve each sub-band $B_0^{(j)}$ of I_0, using a self-similarity approach adopted from previous work [6, 17], summarized below:

For the sub-band $B_0^{(j)}$, we first obtain its downsampled version,
$$B_{-1}^{(j)} = \left(B_0^{(j)} * f_{psf}\right) \downarrow \tag{2}$$
where f_{psf} is an assumed point spread function. We then create internal dictionaries $\mathcal{L}^{(j)}$ and $\mathcal{H}^{(j)}$ that contain patches from $B_{-1}^{(j)}$ and their corresponding (higher resolution) patches from $B_0^{(j)}$, respectively. The sets $\mathcal{L}^{(j)}$ and $\mathcal{H}^{(j)}$ serve as our internal training database of LR-HR training patches, for super-resolving

the sub-band $B_0^{(j)}$. To super-resolve $B_0^{(j)}$ to $\tilde{B}_1^{(j)}$, we do the following: For every patch $\mathbf{l}$ of $B_0^{(j)}$, we look for its $k = 5$ most similar patches $\{\mathbf{l}_i\}_{i=1}^k$ in the LR set $\mathcal{L}^{(j)}$, based on L_2 distances. Their corresponding HR patches $\{\mathbf{h}_i\}_{i=1}^k$ from the set $\mathcal{H}^{(j)}$ serve as individual predictors for the patch $\mathbf{l}$. We compute a weighted average of $\{\mathbf{h}_i\}_{i=1}^k$ to estimate the HR patch $\tilde{\mathbf{h}}$ of $\mathbf{l}$ as follows,

$$\tilde{\mathbf{h}} = \frac{\sum w_i \cdot \mathbf{h}_i}{\sum w_i}, \text{ where, } w_i = exp\left(\frac{-||\mathbf{l} - \mathbf{l}_i||_2^2}{2\sigma^2}\right). \tag{3}$$

Using a larger number of patch matches (k) tends to cause oversmoothing, whereas very small values such as $k = 1$ or 2 produces sharper images but with some artifacts. We repeat the above procedure for every patch $\mathbf{l}$ of $B_0^{(j)}$, to get the corresponding HR patches. These together constitute the super-resolved sub-band $\tilde{B}_1^{(j)}$.

5   Implications

Matching image patches based on intensity differences is often difficult if the patches contain complex structures such as textural detail [17]. Using sub-band decomposition, our algorithm essentially aims at decomposing complex textural structures into relatively simpler ones, that are easier to find matches for. For each image patch, our algorithm allows each of its sub-band components to find its optimal matches at *different* spatial locations in the image. This is illustrated in Fig. 2(b). The sub-bands of the red patch are allowed to find their optimal matches in different spatial locations in the LR sub-bands $B_{-1}^{(1)}, B_{-1}^{(2)}, B_{-1}^{(3)}$. This is in contrast to the conventional way of matching raw patches as shown in Fig. 2(a), where all frequency components of the matched patch are restricted to be from the same spatial location, since no sub-band decomposition is performed.

These properties of our algorithm have useful implications discussed below:

1) Lower matching error: We expect our approach to find nearest neighbor (NN) matches with lower error, as compared to the traditional image domain patch matching. To verify this, we compute the NN error map for the image shown in Fig. 3. The error map is the error produced by a given LR image I_0 while reconstructing itself using its internal dictionary $\mathcal{L}$. In the SR algorithm, the NN error map therefore denotes the "training error" (in pattern recognition parlance). For the conventional self-similarity based SR, we obtain the error map

Fig. 3. *Left:* Input image. *Center:* Image indicating the errors obtained using conventional nearest neighbor search for each image patch, in the internal LR dictionary $\mathcal{L}$. *Right:* Corresponding error map obtained using the proposed sub-band based patch matching approach. Our approach yields lower matching errors, particularly around textural regions such as the fur around the faces.

as follows: Given the image I_0, we first obtain its reconstructed version $\tilde{I}_0$, by replacing each patch of I_0 with its closest patch in the LR internal dictionary $\mathcal{L}$. We then compute the pixelwise difference between I_0 and $\tilde{I}_0$ to obtain the error map. To obtain the error map for our approach, instead of reconstructing $\tilde{I}_0$ directly, we reconstruct its sub-bands $\{\tilde{B}_0^{(j)}\}_{j=1}^N$ using nearest neighbor searches in the internal sub-band dictionaries $\{\mathcal{L}^{(j)}\}_{j=1}^N$. We combine the reconstructed sub-bands $\{\tilde{B}_0^{(j)}\}_{j=1}^N$ to obtain $\tilde{I}_0$ and compute its difference with the original image I_0 to obtain our error map.

Fig. 3, *Center* and *Right* show the error maps obtained by the conventional approach and by our approach, respectively. Clearly, the errors are much lower for our algorithm, particularly in textured regions such as the fur surrounding the faces. We show in our results in Section 6 that this lower NN error translates to better reconstruction of textural details.

2) Invariance to patch size: The choice of patch size has an important effect on the quality of the SR results, particularly for self-similarity based methods. Using larger patch sizes for conventional patch matching leads to greater difficulty in matching textural regions since the complexity of image structures is larger. On the other hand, using extremely small patch sizes is also not expected to improve results since very small patches may not contain enough structural information to learn their transformations across resolutions. For a given image, the optimal patch size to use is difficult to determine *a priori*. Using the proposed approach, complex patches are broken down into relatively simpler sub-bands. The simpler structure of the sub-bands decreases the variety of the sub-band patches and thus reduces the error of the best matching patch for a given dictionary size. Therefore, we expect our algorithm to suffer less if the patch size chosen is sub-optimal. Indeed, as compared to traditional self-similarity based SR, we find our results to be less sensitive to the choice of patch size. We show this in our experiments later in Section 6.

3) Exponentially larger internal dictionary: Allowing different sub-bands of the HR patch to come from different spatial locations of the LR image has an important corollary. Combining sub-bands from different locations effectively allows us to synthesize *new* patches, originally not present in the dictionary of raw image patches. This, in a sense, leads to a combinatorial expansion of the internal dictionaries $\mathcal{L}$ and $\mathcal{H}$, resulting in a dictionary whose size increases exponentially with the number of sub-bands. Further, this is achieved *without* the
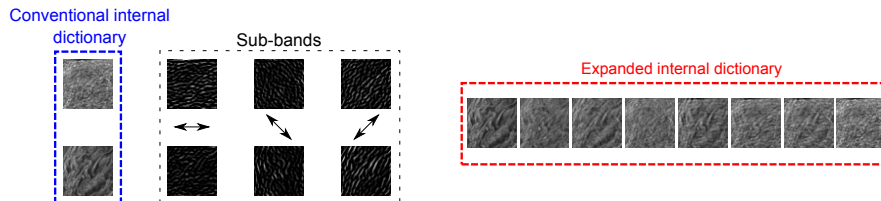

Fig. 4. An example showing a conventional self-similarity based training dictionary containing just two patches (blue box), along with their sub-band decompositions. Combining sub-bands of different patches effectively allows us to *synthesize* new patches, as shown in the expanded dictionary in the red box. Note that the size of the patches here is chosen to be quite large for illustration purposes.
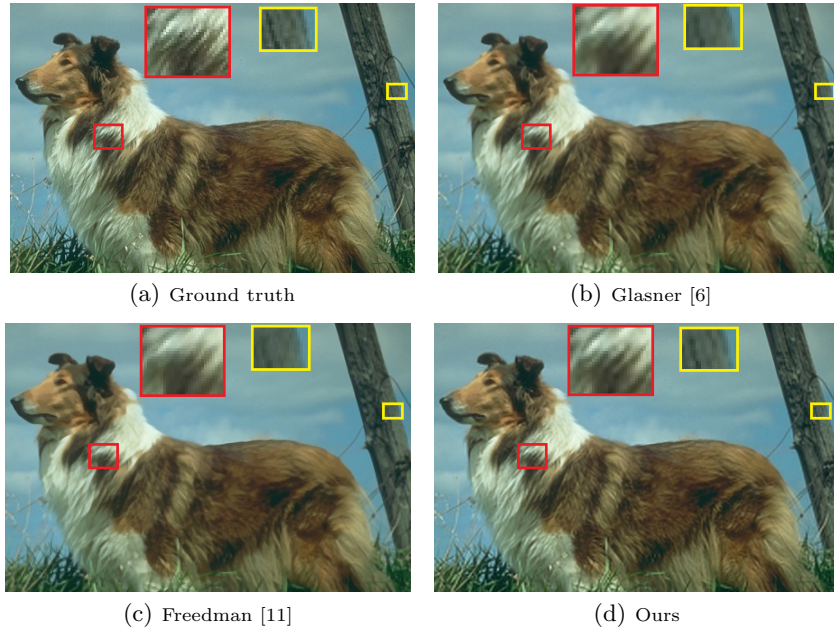
(a) Ground truth   (b) Glasner [6]

(c) Freedman [11]   (d) Ours

Fig. 5. *Dog (2X):* The dog fur, and the details on the wooden pole are better reconstructed using our method, and bear closer resemblance to the ground truth.

use of external databases. We illustrate this with a simple example in Fig. 4. We assume here that our raw patch dictionary $\mathcal{L}$ consists of only two patches as shown in the blue box. In this example we decompose these patches into $N = 3$ sub-bands as depicted in the black dotted box. Now, if using traditional image domain patch-matching, one is restricted to choosing among only two possible candidate matches while searching for a nearest neighbor match. However, if patch-matching is done independently for each sub-band, the number of unique combinations possible is $2^N = 8$. In Fig. 4 on the right, we show the patches resulting from each of the unique sub-band combinations. Clearly, in addition to the original two patches, several more new textural patches have been synthesized in this expanded dictionary. Note that one never has to explicitly obtain such an expanded dictionary. Such an expansion is an implicit consequence of independently finding best matches for the different sub-band patches.

4) Invariance to image size: We have shown that super-resolving sub-bands independently has the overall effect of performing conventional patch-similarity based SR, but using a much larger internal dictionary, whose elements are generated by combining different sub-band patches from different locations in the scene. While the use of a larger dictionary is expected to be always beneficial in general, it becomes particularly useful in cases where the original internal dictionary is small, such as while super-resolving extremely small images. Indeed, as we show in Section 6, in such cases we observe a much greater improvement in our results over the conventional self-similarity approach. Our algorithm there-
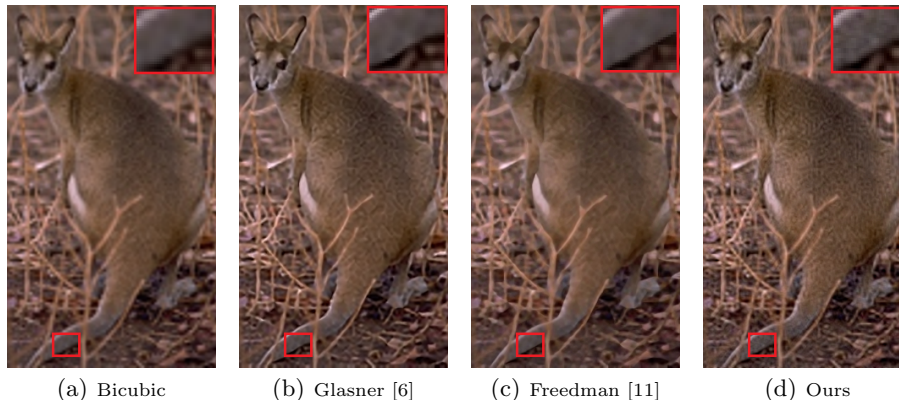
<center>(a) Bicubic   (b) Glasner [6]   (c) Freedman [11]   (d) Ours</center>

Fig. 6. *Kangaroo (3X):* Both Glasner [6] and Freedman [11] almost completely lose the textural details of the kangaroo's tail. Our algorithm is able to better synthesize this. Ground truth for this image is not available so absolute error cannot be obtained.

fore yields relatively more consistent levels of performance across different image sizes. We corroborate this claim in Section 6.

6   Experiments and Results

Implementation Details: For the steerable pyramid, we use eight different orientation bands, and a single scale decomposition. Using more orientation bands improved results in general, but the improvements became marginal beyond eight bands. We use only a single (highest) scale decomposition since the lower scale bands contain lower frequency information which does not pose much challenge for SR. We perform SR in two steps. Therefore, for 3X SR, we perform $\sqrt{3}$X SR twice. Our algorithm is used only on the luminance channel of color images. The chroma components are separately upscaled using bicubic interpolation and combined with our output to obtain the final color image.

We compare our results to eight popular single image SR methods[1] [6, 11, 2, 9, 33, 23, 34, 32], as described in the paragraphs below. Additional results are included as supplementary material.

Comparison with self-similarity methods: Our most important comparison is with other self-similarity methods. We compare our results to [6] and [11], which are two very popular self-similarity based SR methods in the literature. Fig. 5 shows results on the *Dog* image. Our result shows more detail and richer texture in the dog fur and the wooden pole. The self-similarity methods [6, 11] in general are quite good at preserving sharpness of high contrast edges, but tend to smooth out finer details. [11] tends to smooth details more than [6] since it performs only a very localized search for nearest neighbors, for computational reasons. Our result bears closer resemblance to the ground truth.

---

[1] The software for many of these methods were provided by the respective authors, while we implemented the others.

(a) Bicubic   (b) Glasner [6]   (c) Freedman [11]   (d) Ours
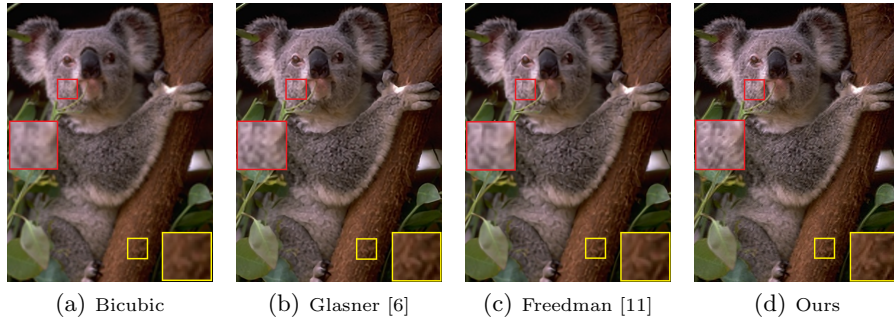
Fig. 7. *Koala (3X):* Our result shows richer texture in koala's fur and the tree trunk. Ground truth for this image is not available so absolute error cannot be obtained.

Fig. 6 shows results on the *Kangaroo* image. Notice here that both [6] and [11] almost completely lose the textural details of the tail. Our algorithm is able to better preserve this texture.

Fig. 7 shows results on the *Koala* image. Here as well, our algorithm is able to synthesize richer texture in the fur and the tree trunk, than both [6] and [11]. Note that the *Koala* and *Kangaroo* images do not have ground truth available.

Comparison with external dictionary based methods: We now compare our results with methods that use external dictionaries for SR. Specifically, we consider [2] which is a popular method based on dictionary learning and sparse representations, the method in [33] that uses ridge regression for predicting HR patches, and the more recent method [9], which is based on using simple regression functions on a pre-clustered training dictionary. We also compare to the classic iterative backprojection algorithm [32] for reference.
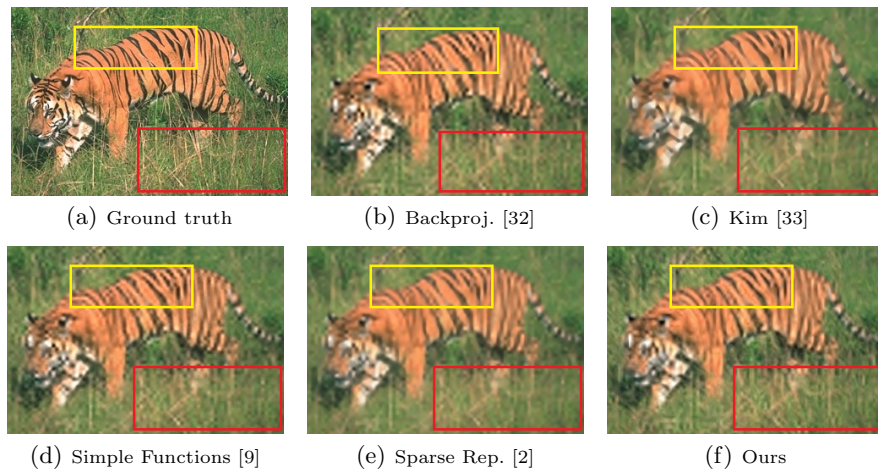

(a) Ground truth   (b) Backproj. [32]   (c) Kim [33]

(d) Simple Functions [9]   (e) Sparse Rep. [2]   (f) Ours

Fig. 8. *Tiger (4X):* Notice the grass above and below the tiger. Our result shows greater textural detail in the grass regions (red box), as compared to most methods. While [9] also seems to produce rich texture, it also produces ringing artifacts such as on the stripes of the tiger (yellow box).

(a) Ground truth   (b) Backproj. [32]   (c) Kim [33]

(d) Simple Fns. [9]   (e) Sparse Rep. [2]   (f) Ours

Fig. 9. *Sunlight (4X):* Notice the woman's hair. [33] and [19] do not produce sufficient detail in the hair, whereas [9] and [32] show excessive ringing artifacts. Our result appears more natural.

Fig. 8 shows the results on the *Tiger* image. While Kim [33] reconstructs high contrast edges almost as sharp as ours, textural details appear highly washed out. The result of [2] also appears a little soft, both along high contrast edges as well as in textural regions such as the grass (red box). [9] appears slightly more detailed than [2], but it shows excessive ringing artifacts such as along the stripes of the tiger (yellow box), much like the backprojection algorithm [32]. Overall, our result has richer textural details without excessive ringing artifacts.

Fig. 9 shows results on the *Sunlight* image. Notice that the woman's hair appears most natural in our result. [9] and [32] clearly show more ringing artifacts in the hair, whereas [2] and [33] are not able to reconstruct sufficient detail.

Comparison with other methods: We also compare our approach with two other methods popularly used in literature - the gradient profile prior (GPP) method [23], that is based on learning gradient profile transformations across resolutions, and the method in [34] that uses iterative feedback based upsampling, without any external databases. Fig. 10 shows our results on the *Red hair* image. The fine strands of hair in the blue box are clearly visible in our result, but is lost in the results of [23] and [34]. Our result appears almost indistinguishable from the ground truth in this example.

(a) Ground truth   (b) GPP [23]   (c) Shan [34]   (d) Ours

Fig. 10. *Red Hair (2X):* Notice the details of the hair as shown in the blue box. Fine strands of hair are discernible in our result, whereas they are smoothed out in the result of [23] and [34]. Our result seems almost indistinguishable from the ground truth in this example.

Performance vs. patch size: The chosen patch size can have a significant effect on the quality of the SR results particularly for internal dictionary based methods. We have shown earlier that using the proposed approach, complex patches are broken down into simpler sub-bands, that can find closer (lower error) matches. Therefore, our algorithm should suffer less if patch size is increased. To verify this empirically, we do the following: We super-resolve 100 natural images (with known ground truth) using our method and also using the conventional self-similarity method of [6], with several different patch sizes, ranging from 2×2 to 11×11. We then plot the average output image quality (in terms of PSNR and SSIM [35]) as a function of the patch size used. The plots in Fig. 11 show our results. As expected, the performance of our algorithm not only remains higher throughout the tested range, but the loss of PSNR and SSIM is also much slower than the conventional self-similarity approach.
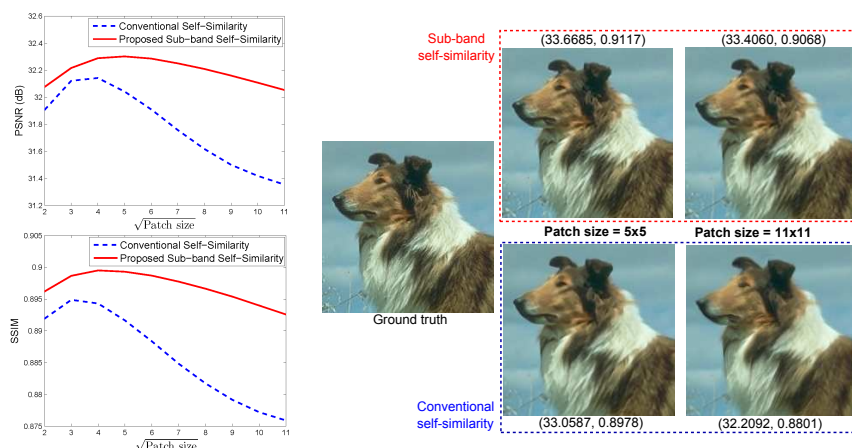

Fig. 11. *Left:* Plots of PSNR and SSIM as a function of the patch size used, for our algorithm as well as the conventional self-similarity method of [6]. *Right:* An example showing the effect of patch size on the results of both algorithms. Our result remains more consistent with patch size variation as compared to [6]. Numbers in brackets denote (PSNR in dB, SSIM [35]).

Performance vs. image size: Earlier, we showed that our algorithm has the effect of synthesizing a much larger internal dictionary, by combining sub-bands from different spatial locations in the image. We therefore expected our
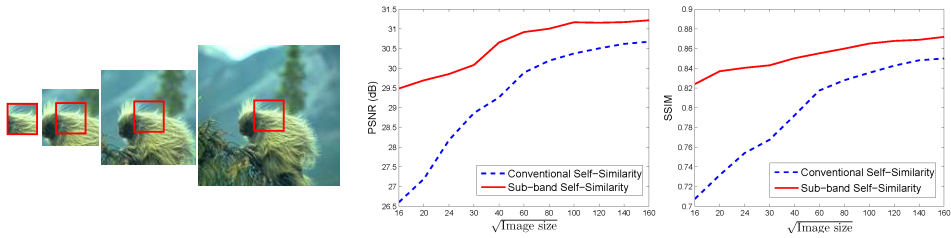
Fig. 12. *Left:* Data used for studying the performance of our algorithm as a function of the size of the input image. We use a series of cropped images as shown. We study how the common sub-image (red box) gets super-resolved in each of these images. *Right:* Plots showing the PSNR (in dB) and SSIM of the super-resolved sub-image as a function of the size of the image containing it. Our algorithm shows a much gradual decline in performance for smaller images, as compared to the conventional self-similarity method [6].

algorithm to perform significantly better than conventional self-similarity, if the input image size was very small. To verify this claim, we perform the following experiment: Consider super-resolving the set of images as shown in Fig. 12 on the left. Each image here is a cropped version of the image on its right. The leftmost (smallest) image, therefore, is a sub-image of all the other images, and appears in all of them, as marked by the red box. We now wish to see how well this sub-image gets super-resolved in each of these images. Clearly, in the rightmost (largest) image, the sub-image has access to all the patches from its surrounding regions as well, which should therefore result in better SR. We compute SR quality (in terms of PSNR and SSIM [35]) of this sub-image, as a function of the size of the image containing it, and plot the result in Fig. 12 on the right. As expected, the conventional self-similarity approach [6] shows a more drastic reduction in performance for smaller image sizes, as compared to our method.

To visualize this effect of image size in a more practical SR problem, we perform the following experiment: We consider super-resolving two input images, as shown in the black dotted box in Fig. 13. The first image shows a group photograph, whereas the second is a cropped version containing just one of the faces, measuring only 20×25 pixels. We super-resolve both these images using the method of [6] as well as our proposed algorithm and show the results in the blue and red dotted boxes respectively. We compare the quality of the super-resolved faces obtained using each method, in both the images. We make the following two observations: 1) In both images, the face is super-resolved better (visually) by our algorithm than the conventional internal dictionary based approach [6]. 2) There is a significant difference in the quality of the super-resolved faces from the bigger and the cropped images, using either method. Using our algorithm, however, this difference is smaller. Our algorithm is able to super-resolve the extremely small cropped image better than the conventional self-similarity approach.

In practice, small images are more commonly encountered as candidates for super-resolution than large ones. Our algorithm is therefore useful for practical applications like super-resolution of thumbnail images, detection/recognition of distant (small) faces in images/videos captured using surveillance cameras, etc. Like any self similarity based algorithm, our algorithm does not require manu-

Fig. 13. An example showing the performance of our algorithm for very small input images. We super-resolve the two images shown in the black dotted box, the right one being a one face sub-image cropped from the left image. Our algorithm is able to super-resolve this small face image much better than the conventional self-similarity approach of [6].

ally chosen training images, which makes it all the more attractive in terms of portability and ease of implementation.

Computational Cost: Our algorithm applies a self-similarity SR algorithm (such as [6]) on R different sub-bands. A naive implementation would be R times slower than the corresponding self-similarity SR algorithm. But since each sub-band is super-resolved independently, they can be easily parallelized. Using such a parallelization, our algorithm is just around 1.5 times slower than the baseline self-similarity SR algorithm of [6].

7   Conclusion

While external dictionary based methods can produce good results in general, they are hindered by the problems associated with the choice and construction of the external training database. Internal dictionary based methods provide an attractive way to circumvent these issues, but also sacrifice some ability to reconstruct textural details well, particularly while super-resolving small sized images and/or when the optimal patch size not used. In this paper we have proposed a self-similarity based algorithm that overcomes these limitations. Our algorithm produces better SR results that remain fairly consistent across several scenarios commonly encountered in practice.

References

1. Freeman, W., Pasztor, E.: Learning low-level vision. In: ICCV. (1999)
2. Yang, J., Wright, J., Huang, T., Ma, Y.: Image super-resolution via sparse representation. IEEE Trans. Image Proc. (2010)
3. Yang, J., Wright, J., Huang, T., Ma, Y.: Image super-resolution as sparse representation of raw image patches. In: CVPR. (2008)
4. Sun, J., Zheng, N.N., Tao, H., Shum, H.Y.: Image hallucination with primal sketch priors. In: CVPR. (2003)
5. Yeung, D.Y., Yeung, D.Y., Xiong, Y.: Super-resolution through neighbor embedding. In: CVPR. (2004)
6. Glasner, D., Bagon, S., Irani, M.: Super-resolution from a single image. In: ICCV. (2009)
7. Freeman, W.T., Pasztor, E.C.: Learning low-level vision. IJCV (2000)
8. Wang, S., Zhang, D., Liang, Y., Pan, Q.: Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In: CVPR. (2012)
9. Yang, C.Y., Yang, M.H.: Fast direct super-resolution by simple functions. In: ICCV. (2013)
10. Ebrahimi, M., Vrscay, E.: Solving the inverse problem of image zooming using "self-examples". In: ICIAR. (2007)
11. Freedman, G., Fattal, R.: Image and video upscaling from local self-examples. ACM Trans. Graph. (2010)
12. Yang, C.Y., Huang, J.B., Yang, M.H.: Exploiting self-similarities for single frame super-resolution. In: ACCV. (2010)
13. Singh, A., Ahuja, N.: Sub-band energy constraints for self-similarity based super-resolution. In: ICPR. (2014)
14. Singh, A., Porikli, F., Ahuja, N.: Super-resolving noisy images. In: CVPR. (2014)
15. Barnsley, M.: Fractals Everywhere. Academic Press Professional, Inc. (1988)
16. Polidori, E., luc Dugelay, J.: Zooming using iterated function systems (1995)
17. Zontak, M., Irani, M.: Internal statistics of a single natural image. In: CVPR. (2011)
18. Freeman, W., Jones, T., Pasztor, E.: Example-based super-resolution. Computer Graphics and Applications, IEEE (2002)
19. Yang, J., Wang, Z., Lin, Z., Cohen, S., Huang, T.: Coupled dictionary training for image super-resolution. IEEE Trans. Image Proc. (2012)
20. Yang, J., Lin, Z., Cohen, S.: Fast image super-resolution based on in-place example regression. In: CVPR. (2013)
21. Fattal, R.: Image upsampling via imposed edge statistics. ACM Trans. Graph. (2007)
22. Sun, J., Sun, J., Xu, Z., Shum, H.Y.: Gradient profile prior and its applications in image super-resolution and enhancement. IEEE Trans. Image Proc. (2011)
23. Sun, J., Sun, J., Xu, Z., Shum, H.Y.: Image super-resolution using gradient profile prior. In: CVPR. (2008)
24. Jiji, C.V., Chaudhuri, S.: Single-frame image super-resolution through contourlet learning. EURASIP J. Appl. Signal Process. (2006)
25. Gajjar, P., Joshi, M.: New learning based super-resolution: Use of dwt and igmrf prior. IEEE Trans. Image Proc. (2010)
26. Sun, J., Zhu, J., Tappen, M.: Context-constrained hallucination for image super-resolution. In: CVPR. (2010)

27. HaCohen, Y., Fattal, R., Lischinski, D.: Image upsampling via texture hallucination. In: ICCP. (2010)
28. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: CVPR. (2005)
29. Simoncelli, E., Freeman, W.: The steerable pyramid: a flexible architecture for multi-scale derivative computation. In: ICIP. (1995)
30. Simoncelli, E., Freeman, W., Adelson, E., Heeger, D.: Shiftable multiscale transforms. IEEE Trans. Info. Theory (1992)
31. Do, M., Vetterli, M.: The contourlet transform: an efficient directional multiresolution image representation. IEEE Trans. Image Proc. (2005)
32. Irani, M., Peleg, S.: Improving resolution by image registration. CVGIP (1991)
33. Kim, K., Kwon, Y.: Single-image super-resolution using sparse regression and natural image prior. IEEE TPAMI (2010)
34. Shan, Q., Li, Z., Jia, J., Tang, C.K.: Fast image/video upsampling. ACM Trans. Graphics (2008)
35. Z.Wang, Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Proc. (2004)