Fast edge-directed single-image super-resolution

Mushfiqur Rouf[*,+] ***Dikpal Reddy***[+,o] ***Kari Pulli***[+,o] ***Rabab Ward***[*]
[*] ***University of British Columbia*** [+] ***NVidia Inc*** [o] ***light co.***

Abstract

We present a novel method for single-image super-resolution (SR). In natural images, spatial edges usually have smooth contours. From this observation, we derive a fast edge-preserving natural image prior using our proposed fast edge-directed interpolation (EDI) method, and combine this prior with the well-known sparse gradient prior into a maximum-a-posteriori (MAP) formulation of the SR problem. We develop an efficient primal-dual algorithm to solve the inverse problem. The application of our edge-preserving prior adds little computational overhead and the output produced by our method demonstrates that results are better than those of the state-of-the-art conventional methods.

Introduction

A single-image super-resolution (SR) method takes a low-resolution (LR) image as input and produces a high-resolution (HR) image. This is an underdetermined inverse problem because the input LR image does not contain the full HR image information. The missing information is crucial in making the HR image look sharp. In order to reconstruct the unknown HR image from the input LR image, SR methods must therefore fill in the missing information, using prior knowledge.

It is well-known that SR involves three tasks: upsampling, deconvolution, and denoising. The upsampling task uses the input LR image data to form the target HR image. However, the upsampled image appears blurry because the upsampling step does not account for the optical anti-aliasing employed by cameras. This anti-aliasing is implemented via an optical low-pass filter—an optical element such as a diffuser is introduced on the optical path. This filter slightly blurs the signal incident on the camera's image sensor and suppresses the spatial frequencies above the Nyquist limit. Another reason for blur is that the camera optics may not be perfectly focused at the imaged target. The deconvolution step accounts for this anti-aliasing blur when restoring the high spatial frequencies in the captured image. Since these high spatial frequencies have been suppressed during capture, the signal-to-noise ratio tends to be poor in the high spatial frequencies, and as a result the deconvolution step ends up enhancing the high-spatial-frequency noise. A denoising task can reduce this noise. While these three tasks can be applied successively to solve the SR problem, a better alternative is a global optimization approach that addresses all of these three aspects at the same time.

We argue that an edge-aware anisotropic filtering component is desirable for SR. Many existing SR techniques use an isotropic upsampling component, such as a bilinear or bicubic resampler. However, since isotropic upsampling is essentially a convolution with a sampling kernel, such techniques introduce additional blur to the upsampled image on top of the anti-aliasing blur already present. This additional blur makes it harder to deconvolve the image. On the contrary, an edge-aware anisotropic filter can reduce the blurring of strong edges in the image, which makes SR a better-posed problem.

The strength of our proposed method lies in its use of a prior that assumes that image edges are smooth along their contours. We call this prior a smooth contour prior. This prior is fast to compute. We employ this prior in conjunction with the widely-used sparse gradient prior. The sparse gradient prior is applicable to many inverse problems in imaging, including SR. The sparse gradient prior assumes that natural images are piecewise smooth and thus prefers sharp spatial edges. However, it does not explicitly model the edges to be smooth (not jaggy) along their contours. We find that our proposed smooth contour prior complements the sparse gradient prior and helps to reconstruct the unknown pixel values by interpolating along the contours of strong spatial edges. This results in a strong combined prior. As a result, edges reconstructed with our proposed method are both sharp and smooth along their contours.

The proposed smooth contour prior uses an edge-directed interpolation operator as the main building block. While any edge-aware anisotropic filter would do, we propose a new method which improves New Edge Directed Interpolation [16] in terms of speed and stability. This improvement directly translates into a reduction of the time complexity of our proposed smooth contour prior. Since edge-preserving methods estimate scene edge-directions explicitly or implicitly at every pixel, these methods use sliding windows, i.e, they process a window around every pixel. Larger windows provide more accurate estimates but generally require more processing. The implementation of our proposed fast edge-directed interpolation (EDI) method (and therefore the evaluation of our proposed smooth contour prior) has a time-complexity that is linear with the number of image pixels and is independent of window size.

We formulate the SR problem as a convex optimization problem, and for the solution we develop an efficient algorithm based on the primal-dual optimization framework of Chambolle and Pock [1]. Primal-dual optimization algorithms are not only efficient, but are also easy to implement and have good convergence properties. We demonstrate the performance of our algorithm on a number of images. To summarize, our contributions are:

- A smooth contour prior which enforces smoothness along the contours of the image edges and is complementary to the sparse gradient prior.
- A fast edge-directed interpolation operator which applies the smooth contour prior with a time complexity that is linear in number of image pixels and independent of the window size (Figure 1).
- A primal-dual optimization algorithm incorporating the smooth contour prior with results better than the state-of-the-art conventional methods and on par with the very recent methods based on convolutional neural networks.

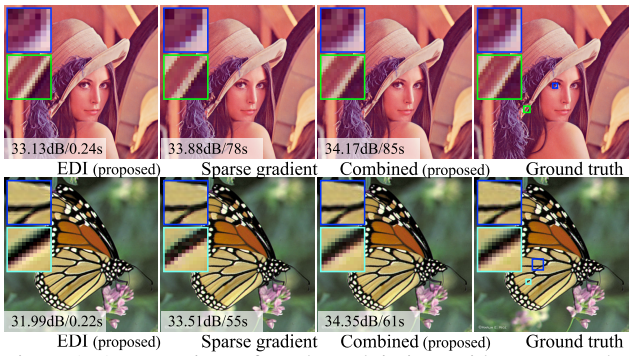

Figure 1: A comparison of results and timings with respect to the priors used in this paper. The SR reconstruction with the well-known sparse gradient prior looks jaggy and over-sharpened. Our fast edge-directed interpolation (EDI) algorithm can compute an HR image very fast (~ 0.2sec) but shows a few artifacts. Our proposed combination of the two produces results with the highest PSNR with a small time overhead added to the sparse gradient prior reconstruction. PSNR in dB and computation times in seconds are shown.

Previous work

Classical SR approaches such as the bilinear and bicubic resampling methods reconstruct the unobserved pixels via isotropic interpolation, and as a result these methods produce blurry HR edges. More recent edge-aware anisotropic filtering approaches perform interpolation along spatial edges so that the strong edges do not appear blurry in the HR image. Wang and Ward [22] used an explicit per-pixel estimation of the angle of the local isophote or the equi-intensity contours. In contrast, other techniques such as the edge-directed interpolation techniques [16, 19] make statistical estimations of the dominant directions of spatial edges in the neighborhood. However, since these anisotropic interpolation methods do not account for the spatial band limit of the LR image signal, the resulting images appear slightly out of focus.

Edges provide strong visual cues, therefore it is important that SR methods restore the sharpness of edges. Most approaches focus on the edge structure "across" natural image edges (i.e., not "along" their natural contours), and as a result they perform the edge-aware reconstruction only indirectly. Farsiu et al. [6] combined the bilateral filter with the sparse gradient prior. Recently, Venkataraman et al. [21] used the Bilateral filter to regularize their multiview SR algorithm. Markov random fields (MRF) are also known to preserve edges in reconstruction [18, 23]. Dai et al. [4] used a soft edge prior for alpha-matte SR. Their method obtains soft edge reconstructions but gives up sharpness to obtain edge smoothness along the alpha matte cut-edges. In very specific cases, such as building facades and other man-made objects showing a repetitive structure, Fernandez-Granda and Candes [7] apply a global transform to axis-align all scene edges so that the sparse gradient prior can be applied without producing jaggies. This method is not applicable to natural scenes with scene edges in random configurations; however, it asserts that the detection of dominant edge directions can help the sparse gradient prior, which is otherwise oblivious of the edge directions. This further justifies our claim that our proposed smooth contour prior works in a complementary fashion to the sparse gradient prior.

Example-based, or more generally learning-based, methods build an implicit prior knowledge base from preprocessed training examples. These methods aim at learning image-patch-based SR rules. Early work such as that by Freeman et al. [9] used nearest neighbors search for looking up similar LR-HR examples for reconstruction. Later work aimed at leveraging various sources of sparsity in the data. Methods such as Yang et al. [25, 26] and He et al. [12] employ sparse coding and simultaneous LR-HR dictionary learning. More recent related work investigated improved nearest-neighbor strategies, and machine learning (ML) techniques in general [14, 15]. The performance of such ML-based techniques will always depend on the training on the previously seen or learned examples. Recently Zhu et al. [28] at least partly overcame this limitation by introducing a deformation model that allows patches to be deformed so that the learned dictionary can be more expressive. Since a map from HR patches to LR patches is many-to-one, a successful ML technique still requires local image priors to avoid high-frequency artifacts. SR methods using deep convolutional neural networks [5, 20] are in some cases able to produce results that are a little better than ours but require a huge amount of data for training as opposed to our method which requires no training and is therefore applicable to problems where training data is hard to obtain.

Recently, nonlocal self-similarity has been proven to be a powerful natural image prior for denoising [3] and similar image reconstruction problems [13]. A few recent work have leveraged local self-examples as the source of the training data for SR. Glasner et al. [10] utilized cross-scale self-similarities on an image pyramid. Zhang et al. [27] combined nonlocal means and steerable kernels in order to leverage both nonlocal self-similarities and edge preservation. He and Siu's method [11] can be loosely described as a non-local version of edge-directed interpolation followed by a deconvolution: they perform a Gaussian process regression to mine the structural similarities across 3×3 patches. However, for SR, local methods are expected to perform just as well as the more expensive nonlocal methods: Freedman and Fattal [8] and more recently Yang et al. [24] have shown that self-examples from the exact same location of an image patch but from a different scale of the same image can produce plausible SR reconstructions. An intuition behind this finding is that, while nonlocal approaches can potentially strengthen the denoising component of an SR solution, an edge-directed upsampling component can reduce noise at the source by avoiding additional blurring due to isotropic resampling. This has motivated us to develop the proposed smooth contour prior using the proposed fast edge-directed interpolation (EDI) algorithm.

Single-image super-resolution

We propose a new method for single-image super-resolution (SR). Our method takes a low-resolution (LR) image as input and produces a high-resolution (HR) image. In the following discussion, we first describe the *forward* model: how an LR image is formed from the latent HR image. We invert the forward model in order to find the unobserved HR image from the input LR image. This *inverse* problem, under the maximum-a-posteriori (MAP) sense, becomes a convex optimization problem. We solve this convex optimization problem with the primal-dual method [1]. This inverse problem is underdetermined, and we make it well-posed by using two complementary priors: the well-known sparse gradient prior, and a novel image prior that we propose, namely,

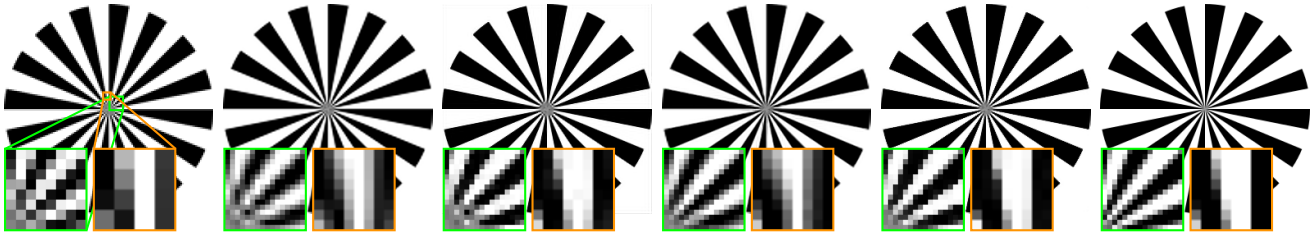

Figure 2: A demonstration of the combined effect of the smooth contour prior and the sparse gradient prior with a Siemens star chart. Enlarged parts of each image are shown are in the insets. Bicubic upsampled edges show jaggy artifact, that are missing in the EDI-upsampled image. As EDI only performs upsampling and not deconvolution, its resulting edges are blurry, as seen by comparing the **orange insets**. Reconstruction with the sparse gradient prior restores the sharp edges comparably to the ground truth. However, the bottom-left corner of the **green insets** demonstrate that the sparse gradient reconstruction fails to restore high spatial frequencies around the center of the star chart) and this is only as good as the bicubic upsampling result slightly sharpened. On the other hand, EDI can infer higher frequencies because of its edge-directed interpolation. It is evident from the insets that our method combines the strengths of the both approaches discussed above.

the smooth contour prior. We discuss the intuition behind this proposed smooth contour prior. At the core of this novel prior is our proposed fast edge-directed interpolation (EDI) algorithm we develop later in this section.

SR as an inverse problem

Let x denote the unknown latent HR image, defined over the set of pixels whose locations span $\Omega \equiv [1,h] \times [1,w]$. Let z denote the (partially observed) LR image, where the observed pixels are centered at image locations $\Gamma \subset \Omega$.

We follow the standard LR image formation model [17]. We assume that the observed image has been properly anti-aliased before it was captured, i.e., the spatial signal of the observed image has been band-limited via a *known* low-pass filter s so as to reduce aliasing in the captured data. We also assume that the observed image has been corrupted by independent and identically distributed (i.i.d.) additive white Gaussian noise (AWGN) η,

$$z = (s \otimes x)|_\Gamma + \eta, \tag{1}$$

where $\otimes$ denotes two-dimensional convolution, and $\cdot|_\Gamma$ denotes the restriction operator that selects the subset Γ.

The image formation model (1) above describes the relationship between the LR image z and the latent HR image x. In order to estimate the latent HR image, we solve (1) for x. Since the forward model constitutes convolution, downsampling, and corruption by noise, the inverse problem involves denoising, upsampling, and deconvolution.

The maximum-a-posteriori (MAP) estimate of x can be obtained via the minimization of an energy function composed of two parts: an ill-posed data-fitting term corresponding to the forward model, and prior terms for well-posedness. The data-fitting term follows directly from (1) while for prior terms we apply the sparse gradient prior and our proposed smooth contour prior,

$$\min_x \underbrace{\|z - (s \otimes x)|_\Gamma\|_2^2}_{\text{data-fitting}} + \underbrace{\lambda_{\text{TV}} \|\nabla x\|_{\text{TV}}}_{\text{sparse gradient}} + \underbrace{\frac{\lambda_{\mathscr{E}}}{2} \|x - \mathscr{E}(x|_\Gamma)\|_2^2}_{\text{smooth contour}}, \tag{2}$$

where λ_{TV} and $\lambda_{\mathscr{E}}$ are regularization weights, $\|\cdot\|_{\text{TV}}$ is total variation (TV), and $\mathscr{E} : \Gamma \to \Omega$ is an edge-aware anisotropic filter.

Global optimization for reconstruction

We use the primal-dual method [1] to solve the convex optimization problem (2). The primal-dual method is more efficient in optimizing for the various norms involved. This iterative method uses general projection operators, known as the proximal operators and denoted by "prox", to iteratively converge to the optimum solution. In order to derive the primal-dual algorithm, we first rewrite our problem in the primal-dual form. We then develop the corresponding algorithm from [1] and derive the proximal operators employed by this algorithm.

We first rewrite our convex optimization problem (2) as

$$\min_x \mathbf{G}(x) + \mathbf{F}_{\text{TV}}(K_{\text{TV}} x) + \mathbf{F}_{\mathscr{E}}(K_{\mathscr{E}} x), \tag{3}$$

where the data-fitting term is captured by $\mathbf{G}(\cdot)$ defined as

$$\mathbf{G}(x) \equiv \|z - (s \otimes x)|_\Gamma\|_2^2, \tag{4}$$

K_{TV} and $K_{\mathscr{E}}$ are linear transformations from the primal (image) domain to respective dual domains defined as

$$K_{\text{TV}} x \equiv \lambda_{\text{TV}} \nabla x \tag{5}$$

$$K_{\mathscr{E}} x \equiv x - \mathscr{E}(x|_\Gamma), \tag{6}$$

and $\mathbf{F}_{\text{TV}}$ and $\mathbf{F}_{\mathscr{E}}$ are functions defined as

$$\mathbf{F}_{\text{TV}} \equiv \|\cdot\|_{\text{TV}} \tag{7}$$

$$\mathbf{F}_{\mathscr{E}} \equiv \frac{\lambda_{\mathscr{E}}}{2} \|\cdot\|_2^2. \tag{8}$$

Although (3) is convex, solving it directly is not easy because of the mixed ℓ_1 and ℓ_2 norms. Instead, we solve the corresponding primal-dual saddle point problem which is an equivalent problem and has the same optimum solution,

$$\min_x \mathbf{G}(x) + \max_{y_{\text{TV}}} \langle K_{\text{TV}} x, y_{\text{TV}} \rangle - \mathbf{F}^*_{\text{TV}}(y_{\text{TV}}) + \max_{y_{\mathscr{E}}} \langle K_{\mathscr{E}} x, y_{\mathscr{E}} \rangle - \mathbf{F}^*_{\mathscr{E}}(y_{\mathscr{E}}), \tag{9}$$

where $\cdot^*$ denotes the convex conjugate of a function, and y_{TV} and $y_{\mathscr{E}}$ are slack variables defined over respective dual domains. In

order to solve (9), we have developed an algorithm, based on [1], which is presented below.

Initializing $\bar{x}^{(0)} = x^{(0)} = \mathscr{E}(z), y_{\text{TV}}^{(0)} = 0, y_{\mathscr{E}}^{(0)} = 0$, and,

repeating until convergence, we get,

$$y_{\text{TV}}^{(k+1)} = \text{prox}_{\sigma \mathbf{F}_{\text{TV}}^*}\left(y_{\text{TV}}^{(k)} + \sigma K_{\text{TV}} \bar{x}^{(k)}\right) \tag{10}$$

$$y_{\mathscr{E}}^{(k+1)} = \text{prox}_{\sigma \mathbf{F}_{\mathscr{E}}^*}\left(y_{\mathscr{E}}^{(k)} + \sigma K_{\mathscr{E}} \bar{x}^{(k)}\right) \tag{11}$$

$$x^{(k+1)} = \text{prox}_{\tau \mathbf{G}}\left(x^{(k)} - \tau\left(K_{\text{TV}}^T y_{\text{TV}}^{(k+1)} + K_{\mathscr{E}}^T y_{\mathscr{E}}^{(k+1)}\right)\right) \tag{12}$$

$$\bar{x}^{(k+1)} = x^{(k+1)} + \theta(x^{(k+1)} - x^{(k)}), \tag{13}$$

where $\sigma > 0$, $\tau > 0$ and θ are parameters of the algorithm that determines the step sizes for the iterates. Convergence is guaranteed [1] when: (i) $0 < \theta < 1$, and (ii) $\sigma\tau L^2 < 1$ where L is the operator norm of the combined primal-to-dual linear map, i.e., $L = \left\| \begin{bmatrix} K_{\text{TV}} \\ K_{\mathscr{E}} \end{bmatrix} \right\|_{op}$. The proximal operator (prox) represents a generalized projection [1] on to a feasible set. The three proximal operators used in our algorithm are discussed below:

(a) $\text{prox}_{\tau \mathbf{G}}$, the proximal operator of the data fitting term, follows directly from the definition of proximal operators,

$$\text{prox}_{\tau \mathbf{G}}(x_0) = \arg\min_x \frac{\|x - x_0\|_2^2}{2\tau} + \mathbf{G}(x) \tag{14}$$

$$= \arg\min_x \frac{\|x - x_0\|_2^2}{2\tau} + \|z - (s \otimes x)|_\Gamma\|_2^2. \tag{15}$$

This is a linear least-squares minimization problem, which we solve using the conjugate gradient method.

(b) $\text{prox}_{\sigma \mathbf{F}_{\text{TV}}^*}$, the proximal operator of the TV term, reduces to pointwise shrinkage [1],

$$\text{prox}_{\sigma \mathbf{F}_{\text{TV}}^*}(y_0) = \frac{y_0}{\max(1, |y_0|)}. \tag{16}$$

(c) $\text{prox}_{\mathbf{F}_{\mathscr{E}}^*}$, the proximal operator of the convex conjugate function $\mathbf{F}_{\mathscr{E}}^*$ can be derived using Moreau's Identity [1]. Moreau's Identity relates the proximal operator of a convex conjugate function (e.g. $\mathbf{F}_{\mathscr{E}}^*$) with the proximal operator of the original function (e.g. $\mathbf{F}_{\mathscr{E}}$), and we get,

$$\text{prox}_{\sigma \mathbf{F}_{\mathscr{E}}^*}(y_0) \equiv y_0 - \sigma\, \text{prox}_{\frac{1}{\sigma}\mathbf{F}_{\mathscr{E}}}\left(\frac{y_0}{\sigma}\right), \tag{17}$$

where the proximal operator of the original function $\text{prox}_{\mathbf{F}_{\mathscr{E}}}$ follows directly from the definition, and we get,

$$\text{prox}_{\sigma \mathbf{F}_{\mathscr{E}}^*}(y_0) \equiv y_0 - \sigma\left(\arg\min_y \frac{\sigma}{2}\left\|y - \frac{y_0}{\sigma}\right\|_2^2 + \frac{\lambda_{\mathscr{E}}}{2}\|y\|_2^2\right)$$

$$= \frac{\lambda_{\mathscr{E}}}{\sigma + \lambda_{\mathscr{E}}} y_0. \tag{18}$$

This completes our algorithm. Next, we discuss the details of the proposed smooth contour prior: the intuition behind it, and the fast algorithm that computes it.

The smooth contour prior

The proposed smooth contour prior term promotes the reconstructed image edges to be smooth along the natural contours of edges, and thus it complements the sparse gradient prior. The EDI operator $\mathscr{E}$ produces smooth contour, i.e., for some HR estimate x^* with smooth contour, $\mathscr{E}(x^*|_\Gamma)$ is expected to produce the same image x^*. In other words, the smooth contour prior term penalizes reconstructed edges that are not smooth *along* their contours. Since, in contrast, the sparse gradient prior improves sharpness *across* edges, these two priors complement each other. A simple test case in Figure 2 demonstrates that the reconstruction quality when both of the priors are applied is better than that when either of them is applied separately.

Our choice of the anisotropic interpolation operator is made based on the following criteria:

(a) Local calculations: Since non-local methods employ a search for (patch or other) similarities in the image, the runtime complexity tends to be much higher than that of local methods. Although non-local methods are better for denoising [2] and similar image reconstruction problems [13], local methods produce comparable results for SR [8]. We therefore chose a local method for speed without sacrificing the quality of reconstruction.

(b) Direct estimation of anisotropic interpolation coefficients: Methods that depend on explicit edge detection [22] are prone to discontinuity artifacts in case of edge mis-estimation. Our proposed method can directly estimate the anisotropic interpolation coefficients. In difficult cases (e.g., when our method is not confident about a strong local edge) our method gracefully falls back to bilinear interpolation and produces no noticeable discontinuity.

(c) Fast implementation: Anisotropic interpolation methods are generally sliding-window algorithms. This means that the accuracy of such methods (such as the bilateral filter methods [6]) depends on the amount of data available, i.e., the window size; but a larger window usually results in a higher run-time. Our proposed method has a run-time that is independent of the window size, which makes it much faster than any other local method.

Our proposed fast edge-directed interpolation technique, which we refer to as EDI, meets all of these criteria.

Fast edge-directed interpolation (EDI)

In this section, we present a fast edge-directed interpolation (EDI) operator, the building block of our proposed smooth contour prior. Our proposed method is based on [16], but is more stable and much faster. The improved stability is due to the regularized regression (25), and the speedup is from the proposed fast two-pass calculations using summed column tables (29).

EDI copies the observed pixel values and estimates the unobserved values from the known neighbors using anisotropic interpolation. EDI derives the anisotropic interpolation coefficients via sliding-window linear least-squares regressions.

The EDI operator $\mathscr{E} : \Gamma \to \Omega$ takes as input some partial observations z, and produces a 2×2 edge-aware anisotropic-upsampled image $\hat{x}$,

$$\hat{x} \equiv \mathscr{E}(z). \tag{19}$$

Since pixel values are already known over Γ, EDI obtains these values directly from the input (Figure 4(a)),

$$\hat{x}|_\Gamma = z. \tag{20}$$

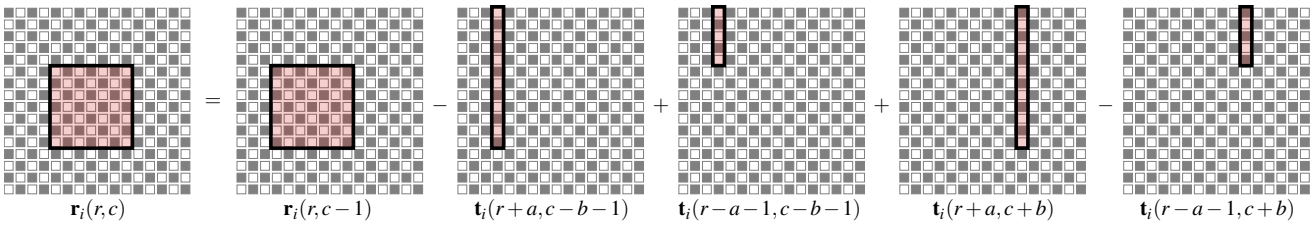

Figure 3: An illustration of the recurrence that lets us compute the area-sums incrementally and very fast. We calculate $\mathbf{r}_i(r,c)$ from $\mathbf{r}_i(r,c-1)$ by adding and subtracting precomputed partial column sums $\mathbf{t}_i$, marked above with the red rectangles.

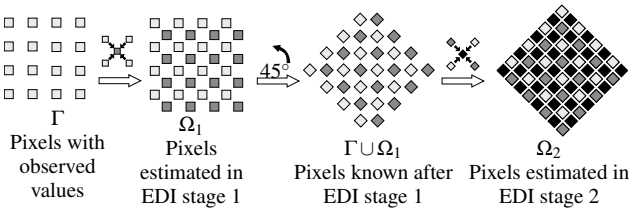

Figure 4: An illustration of the two stages of EDI. The input LR image pixels contribute to one-fourth of the target HR image pixel grid Γ. The rest of the pixels are calculated in two stages: First, the pixels in Ω_1 are estimated via an anisotropic interpolation of the diagonal neighbors. And second, half of the pixels are known $\Gamma \cup \Omega_1$, from which the remaining pixels Ω_2 are estimated. The second stage is algorithmically identical to the first rotated by 45°.

For the unknown pixels in $\Omega - \Gamma$, EDI runs the same anisotropic filtering algorithm twice (Figure 4). The first stage estimates pixels in Ω_1. In the second stage, the output of the first stage, i.e., values at pixels in $\Gamma \cup \Omega_1$ rotated by 45° are input back to the same algorithm which produces estimates of pixels in Ω_2. Therefore, without loss of generality, we limit our discussion to one stage.

EDI estimates each unobserved pixel separately. Let $p \equiv (r,c) \in \Omega_1$ denote the current pixel to estimate the value of. Let $\boxdot p$ denote the list of neighbors of p to interpolate from, where $|\boxdot| = 4$ and $\boxdot_i p \in \Gamma$, $1 \le i \le 4$. In particular, $\boxdot_1 p$ denotes the top-left neighbor (the closest known pixel to the top-left of p), $\boxdot_2 p$ the top-right neighbor, and so on. Then we obtain an estimate of the unobserved pixel $\hat{x}(p)$ via a weighted interpolation of its neighbors [16],

$$\hat{x}(p) = \textstyle\sum_{i=1}^{|\boxdot|} \alpha_i(p)\hat{x}(\boxdot_i(p)), \tag{21}$$

where $\alpha(p) \in \mathscr{R}^{|\boxdot|}$ are the edge-aware anisotropic interpolation coefficients. We determine $\alpha(p)$ by a linear least-squares regression over a window of size $(2a+1)\times(2b+1)$ centered around p denoted by $W(p)$,

$$\alpha(p) = \mathbf{R}(p)^{-1}\mathbf{r}(p), \tag{22}$$

where $\mathbf{R} \in \mathscr{R}^{|\boxdot|\times|\boxdot|}$ is a 4×4 matrix defined as

$$\mathbf{R}_{ij}(p) \equiv \textstyle\sum_{q\in W(p)} \hat{x}(\boxdot_i q)\hat{x}(\boxdot_j q) \tag{23}$$

and $\mathbf{r} \in \mathscr{R}^{|\boxdot|}$ is a 4-vector defined as

$$\mathbf{r}_i(p) \equiv \textstyle\sum_{q\in W(p)} \hat{x}(\boxdot_i q)\hat{x}(q). \tag{24}$$

In order to make the regression more stable, we propose to perform regularized regression instead of (22),

$$\hat{\alpha}(p) = (\mathbf{R}(p) + \mu\mathbf{I})^{-1}\left(\mathbf{r}(p) + \mu|\boxdot|^{-1}\right), \tag{25}$$

where $|\boxdot| = 4$, and μ is a regularization parameter.

Since (25) is a small 4×4 linear system, the overall complexity is $\mathbf{O}(N)$, where $N = hw$ is the number of pixels in the image. The main bottleneck is (23) and (24): when implemented in a straightforward manner, the complexity is $\mathbf{O}(N|W|)$. In contrast, our proposed algorithm has a time complexity of $\mathbf{O}(N)$, i.e., independent of the window size, which allows us to use a large window size for obtaining accurate estimates of the interpolation coefficients α.

We observe that (23) and (24) are sums over partially-overlapping (sliding) windows, and therefore partial sums can speed up the process. Without loss of generality, we show calculations for $\mathbf{r}_i(p)$, $1 \le i \le 4$. $\mathbf{R}_{ij}$ can be computed in a similar fashion. We use two $\mathbf{O}(N)$ passes: first we obtain an intermediate data structure, the summed column table $\mathbf{t}_i$, and then we calculate $\mathbf{r}_i$, for all pixels in Ω_1.

In the first pass, we precompute an intermediate summed column table $\mathbf{t}_i$ of size $h\times w$ to hold partial columnwise sums where every element is a sum of all the (known) products corresponding to that location and above it, and we get,

$$\mathbf{t}_i(r,c) = \sum_{1\le j\le r, (j,c)\in\Gamma} \hat{x}(\boxdot_i(j,c))\hat{x}(j,c). \tag{26}$$

We obtain the $\mathbf{O}(N)$ complexity by computing each element of $\mathbf{t}$ incrementally as,

$$\mathbf{t}_i(r,c) = \begin{cases} 0, & \text{if } r \le 0 \text{ or } c \le 0 \\ \mathbf{t}_i(r-1,c), & \text{if } (r,c) \notin \Gamma \\ \mathbf{t}_i(r-1,c) + \hat{x}(\boxdot_i(r,c))\hat{x}(r,c), & \text{otherwise.} \end{cases} \tag{27}$$

In the second pass, we process pixels in a row-major order to obtain $\mathbf{r}_i$. To achieve the $\mathbf{O}(N)$ overall complexity, we reorganize the terms in (24) to obtain each element $\mathbf{r}_i(r,c)$ with a constant number of operations: from the previously computed element $\mathbf{r}_i(r,c-1)$ and conjunctions and disjunctions of partial column sums (Figure 3) as

$$\mathbf{r}_i(r,c) \equiv \sum_{j=r-a}^{r+a}\sum_{k=c-b}^{c+b} z(\boxdot_i(j,k))z(j,k) \tag{28}$$

$$\begin{aligned} =&\mathbf{r}_i(r,c-1) \\ &-\mathbf{t}_i(r+a,c-b-1) \\ &+\mathbf{t}_i(r-a-1,c-b-1) \\ &+\mathbf{t}_i(r+a,c+b) \\ &-\mathbf{t}_i(r-a-1,c+b). \end{aligned} \tag{29}$$

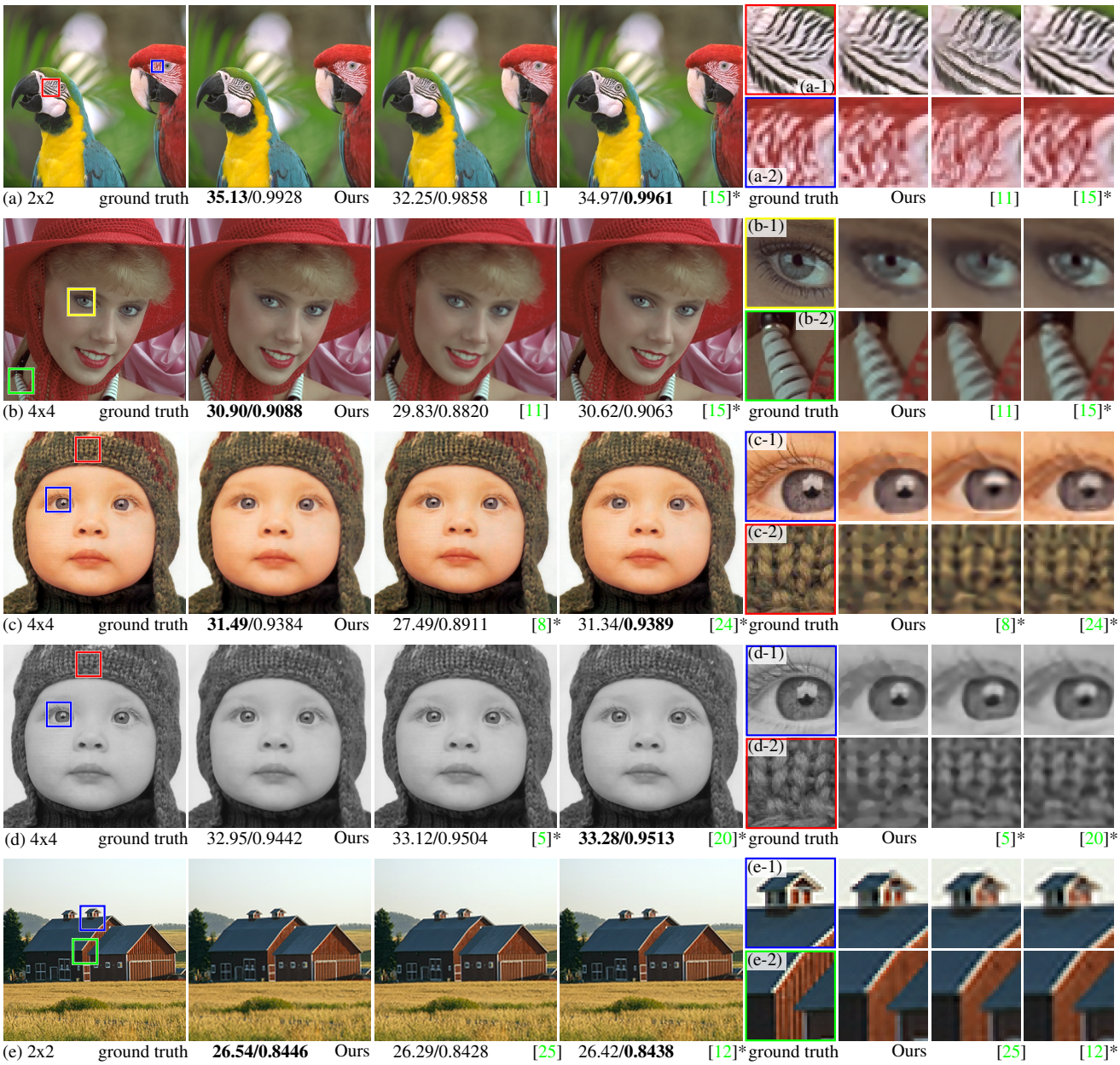

Figure 5: A few dyadic SR experiments with our algorithm. Each row represents one data set. On each set, the **first** image on the left is the ground truth (GT), the **second** image is our SR result, the **third** and the **fourth** images are results from baseline methods as cited below the image. The SR factor is listed under the ground truth image. PSNR and SSIM are listed under each image, the best numbers are shown in bold. Images courtesy of respective sources marked with a '*'. **Insets** on the right are presented in the same order as the full-size images. Insets are enlarged by a factor of 4 with point sampling.

Results

Implementation details. We have used 13×13 windows for estimating per-pixel EDI weights. This EDI method performs 2×2 upsampling. For other SR factors we re-apply our method: For 4×4 SR, we apply our algorithm twice. For 3×3 SR, we compute the 4×4 super-resolved image and bicubic-downsample it by a factor of $3/4$. The default parameter values in our implementation are: $\mu = 0.001$, $\sigma = 0.6$, $\theta = 0.9$ and $\lambda_{\mathrm{TV}} = \lambda_{\mathscr{E}} = 0.0025$.

We have implemented EDI as a Matlab executable (MEX-file). It takes about 0.24 seconds to compute a 512×512 upsampled image from an input 256×256 image. This time complexity scales linearly with number of pixels as expected. We have implemented the primal-dual SR algorithm in Matlab. In most cases, for a 512×512 HR target image, our Matlab implementation takes less than one minute to reach within 0.2dB of the final solution, and takes less than three minutes to converge to the final solution. The most expensive operation is the computation of $\mathrm{prox}_{\tau \mathbf{G}}$, the proximal operator of the data-fitting term. The experiments were run on an Intel Core i7 1.9GHz machine.

Experiments and analysis. In order to validate the performance

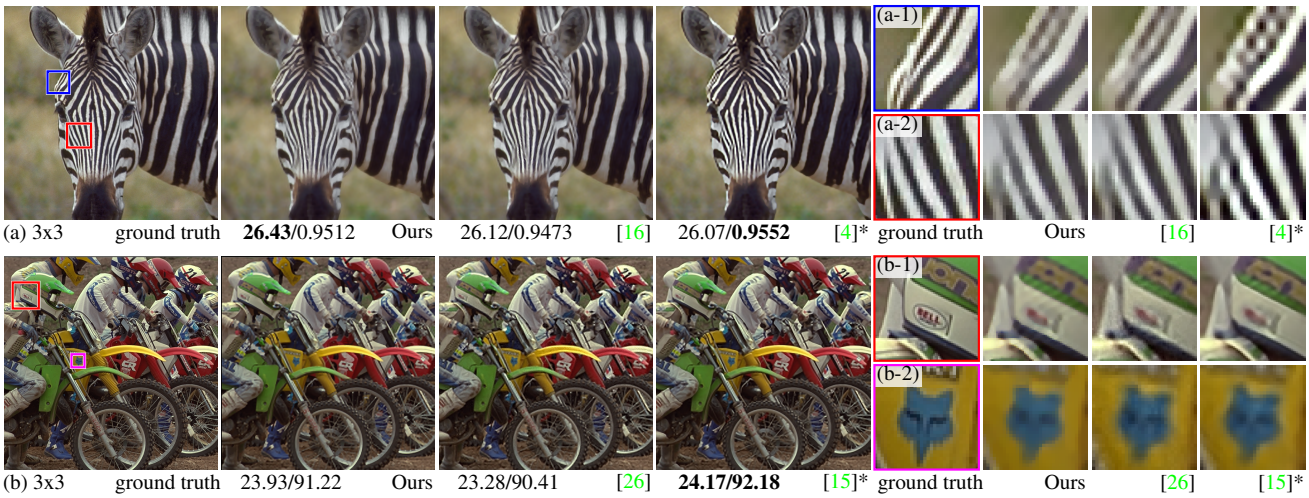

Figure 6: A few nondyadic SR experiments with our algorithm. Nondyadic SR ratios are not natively supported by our method but can be implemented by downsampling from SR with the smallest larger dyadic ratio. This comes with the potential cost of a small performance loss, which is why in (b) above our method falls a little behind the baseline methods despite producing plausible results. (The organization of the images and insets is similar to Figure 5).

of our SR method, we have run our algorithm on a number of test images and compared our results with several state-of-the-art methods. Figure 5 shows a few results of dyadic SR, with two baseline results per test case. Figure 6 shows a few results on nondyadic SR. The supplemental material may be consulted for full resolution images and additional results and comparisons.

Our method takes a reconstruction approach to SR, and therefore it does not produce high-frequency artifacts as produced by nearest-neighbor search-based methods such as [11] shown in Figure 5(a-1).

Methods based on convolutional neural networks such as [5, 20] produce slightly better results, although our result can reproduce details better in some cases as shown in 5(d-1).

Since our method is EDI-based, jaggies are easy to avoid, particularly when a dominant edge is present. This is evident from Figure 6(a). SoftCuts [4] has produced noticeable jaggies in 6(a-1) and 6(a-2) whereas both NEDI [16] and our method produced straight edge contours.

When compared to methods based on self-examples, in Figure 5(c), we see that all of [24], [8], and our method have been able to restore the strong edges, e.g., the outline of the face. The strength of the nonlocal methods is evident from the repetitive texture areas, such as the section of the woolen hat shown in inset (c-2). However, the dependence on self-similarity is a weakness as well; [8] has turned the eyelashes into sharp features on the eyelids in inset (c-1), whereas our method produces a plausible reconstruction free of such sharp artifacts. Our method also reconstructed the details of the eye and in the pupil better in (c-1).

Our method produces results that are slightly better than the state-of-the-art method due to Kwon et al. [15]. We present three comparisons with [15]; two dyadic cases in Figure 5, test cases (a) and (b), and a nondyadic case in Figure 6(b). For test case 5(a) and 5(b), our result has better PSNR and/or SSIM, although it is hard to visually identify much difference from the results presented in [15]. In test case 5(e), our method also outperforms SR methods that are based on joint LR-HR dictionary learning [12, 26]. Since many HR image patches can explain a LR image patch, strong local image priors such as our proposed combination of priors is needed for SR reconstruction.

Limitations. In case of more than one locally dominant edge or no dominant edge at all, e.g. a fine texture such as foliage or fur, our algorithm might not accurately estimate the anisotropic interpolation coefficients for upsampling as shown in Figure 5(c-2). The smooth contour prior falls back to bilinear interpolation to avoid discontinuity artfacts, and our method essentially becomes a sparse gradient-based reconstruction method in these image areas.

An SR factor of larger than 2×2 is not natively supported by our method since EDI is designed for upsampling by 2×2. Other dyadic factors are also possible via successive application of our method, but nondyadic factors such as 3×3 involve a bicubic downsampling step in which our result may lose sharpness. Our 3×3 reconstruction in Figure 6(b) suffers a small 0.24dB dip in PSNR, since we obtain the final result by downsampling our dyadic intermediate 4×4 reconstruction.

Conclusion

In this paper we have presented a new method for single-image super-resolution (SR). We observe that spatial edges usually are not only sharp across, but also smooth along the contours. We have proposed a novel smooth contour prior based on our observation. We have combined the proposed smooth contour prior with the sparse gradient prior; this combination models two complementary aspects of natural images and produces better HR reconstructions. We have also proposed a fast edge-directed interpolation algorithm that makes a fast application of the proposed smooth contour prior possible. Finally, we have also developed an efficient primal-dual SR algorithm that implements these priors. Our proposed SR method produces better results than the state-of-the-art conventional methods, and most of the methods that employ machine learning techniques and train on large datasets. We believe that this new image prior will be applicable to other similar image reconstruction problems.

References

[1] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Math. Imaging and Vision*, 40(1):120–145, 2011. 1, 2, 3, 4

[2] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising with block-matching and 3d filtering. In *Electronic Imaging*, pages 606414–606414. Int. Soc. for Optics and Photonics, 2006. 4

[3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process.*, 16(8):2080–2095, 2007. 2

[4] S. Dai, M. Han, W. Xu, Y. Wu, Y. Gong, and A. K. Katsaggelos. Softcuts: a soft edge smoothness prior for color image super-resolution. *IEEE Trans. Image Process.*, 18(5):969–981, 2009. 2, 7

[5] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In *Computer Vision–ECCV 2014*, pages 184–199. Springer, 2014. 2, 6, 7

[6] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super resolution. *IEEE Trans. Image Process.*, 13(10):1327–1344, 2004. 2, 4

[7] C. Fernandez-Granda and E. J. Candes. Super-resolution via transform-invariant group-sparse regularization. In *Proc. IEEE Int. Conf. on Comput. Vision*, pages 3336–3343. IEEE, 2013. 2

[8] G. Freedman and R. Fattal. Image and video upscaling from local self-examples. *ACM Trans. Graph.*, 28(3):1–10, 2010. 2, 4, 6, 7

[9] W. T. Freeman, T. R. Jones, and E. C. Pasztor. Example-based super-resolution. *IEEE Comput. Graph. Appl.*, 22(2):56–65, 2002. 2

[10] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In *Proc. IEEE Int. Conf. on Comput. Vision*, pages 349–356. IEEE, 2009. 2

[11] H. He and W.-C. Siu. Single image super-resolution using gaussian process regression. In *Proc. IEEE Conf. on Comput. Vision and Pattern Recognition*, pages 449–456. IEEE, 2011. 2, 6, 7

[12] L. He, H. Qi, and R. Zaretzki. Beta process joint dictionary learning for coupled feature spaces with application to single image super-resolution. In *Proc. IEEE Conf. on Comput. Vision and Pattern Recognition*, pages 345–352. IEEE, 2013. 2, 6, 7

[13] F. Heide, M. Steinberger, Y.-T. Tsai, M. Rouf, D. Pajak, D. Reddy, O. Gallo, J. Liu, W. Heidrich, K. Egiazarian, J. Kautz, and K. Pulli. FlexISP: A flexible camera image processing framework. *ACM Trans. on Graph.*, 33(6):231, 2014. 2, 4

[14] K. I. Kim and Y. Kwon. Example-based learning for single-image super-resolution. In *Pattern Recognition*, pages 456–465. Springer, 2008. 2

[15] Y. Kwon, K. I. Kim, J. Tompkin, J. H. Kim, and C. Theobalt. Efficient learning of image super-resolution and compression artifact removal with semi-local gaussian processes. *IEEE Trans. Pattern Anal. Mach. Intell.*, Dec. 2014. 2, 6, 7

[16] X. Li and M. T. Orchard. New edge-directed interpolation. *IEEE Trans. Image Process.*, 10(10):1521–1527, 2001. 1, 2, 4, 5, 7

[17] M. K. Ng and N. K. Bose. Mathematical analysis of super-resolution methodology. *IEEE Signal Process. Mag.*, 20(3):62–74, 2003. 3

[18] R. Pan and S. J. Reeves. Efficient huber-markov edge-preserving image restoration. *IEEE Trans. Image Process.*, 15(12):3728–3735, 2006. 2

[19] W.-S. Tam, C.-W. Kok, and W.-C. Siu. Modified edge-directed interpolation for images. *Electronic imaging*, 19(1):013011–013011, 2010. 2

[20] R. Timofte, V. De Smet, and L. Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Computer Vision–ACCV 2014*, pages 111–126. Springer, 2014. 2, 6, 7

[21] K. Venkataraman, D. Lelescu, J. Duparré, A. McMahon, G. Molina, P. Chatterjee, R. Mullis, and S. Nayar. Picam: an ultra-thin high performance monolithic camera array. *ACM Trans. on Graph.*, 32(6):166, 2013. 2

[22] Q. Wang and R. K. Ward. A new orientation-adaptive interpolation method. *IEEE Trans. Image Process.*, 16(4):889–900, 2007. 2, 4

[23] W. Wu, Z. Liu, X. He, and W. Gueaieb. Single-image super-resolution based on markov random field and contourlet transform. *Electronic Imaging*, 20(2):023005–023005, 2011. 2

[24] J. Yang, Z. Lin, and S. Cohen. Fast image super-resolution based on in-place example regression. In *Proc. IEEE Conf. in Comput. Vision and Pattern Recognition*, pages 1059–1066. IEEE, 2013. 2, 6, 7

[25] J. Yang, Z. Wang, Z. Lin, X. Shu, and T. Huang. Bilevel sparse coding for coupled feature spaces. In *Proc. IEEE Conf. on Comput. Vision and Pattern Recognition*, pages 2360–2367. IEEE, 2012. 2, 6

[26] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *Image Processing, IEEE Transactions on*, 19(11):2861–2873, 2010. 2, 7

[27] K. Zhang, X. Gao, D. Tao, and X. Li. Single image super-resolution with non-local means and steering kernel regression. *IEEE Trans. Image Process.*, 21(11):4544–4556, 2012. 2

[28] Y. Zhu, Y. Zhang, and A. L. Yuille. Single image super-resolution using deformable patches. In *Proc. IEEE Conf. on Comput. Vision and Pattern Recognition*, 2014. 2