Super-Resolution via Recapture and Bayesian Effect Modeling

Neil Toronto, Bryan S. Morse, Kevin Seppi, Dan Ventura
Brigham Young University, Provo, Utah
{ntoronto,morse,kseppi,ventura}@cs.byu.edu

Abstract

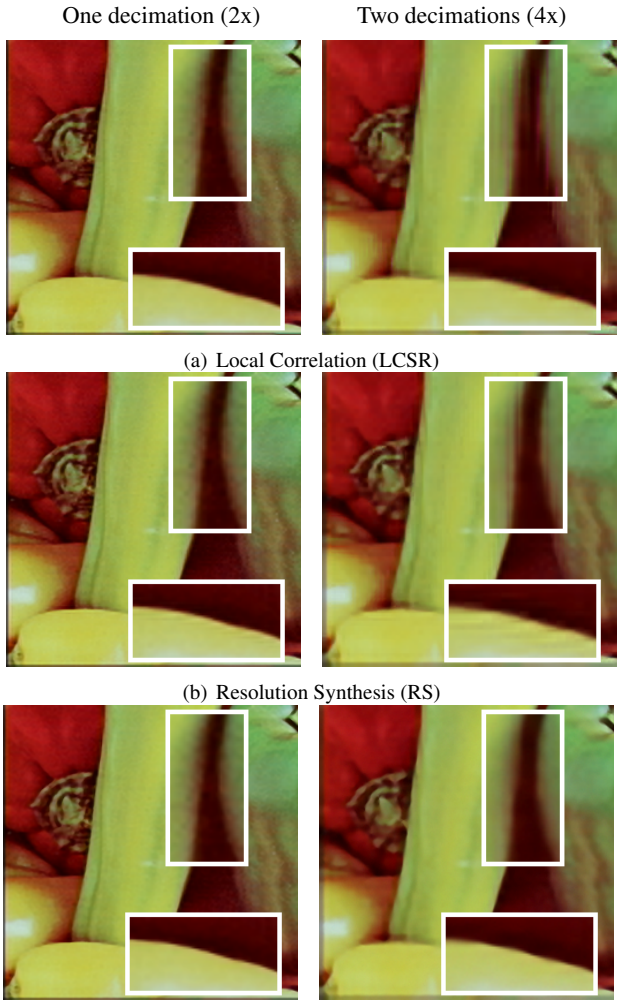
This paper presents Bayesian edge inference *(BEI), a single-frame super-resolution method explicitly grounded in Bayesian inference that addresses issues common to existing methods. Though the best give excellent results at modest magnification factors, they suffer from gradient stepping and boundary coherence problems by factors of 4x. Central to BEI is a causal framework that allows image capture and* recapture *to be modeled differently, a principled way of undoing downsampling blur, and a technique for incorporating Markov random field potentials arbitrarily into Bayesian networks. Besides addressing gradient and boundary issues, BEI is shown to be competitive with existing methods on published correctness measures. The model and framework are shown to generalize to other reconstruction tasks by demonstrating BEI's effectiveness at CCD demosaicing and inpainting with only trivial changes.*

1. Introduction

Many image processing tasks, such as scaling, rotating and warping, require good estimates of between-pixel values. Though this research may be applied to interpolation in any task, we restrict our attention to single-frame super-resolution, which we define as scaling a digital image to larger than its original size.

While recent methods give excellent results at moderate scaling factors [26], all show significant artifacts[1] by scaling factors of 4x (Figure 1). We contribute a new method explicitly grounded in Bayesian inference that preserves edges and gradients and is agnostic to scale. Central to this is an image reconstruction framework adapted from supervised machine learning. Certain aspects of BEI require modeling unknown causes with known effects, which we show can be incorporated easily into an otherwise causal model.

The simplest existing super-resolution methods are non-adaptive, which make the fewest assumptions and are easiest to implement. Function-fitting variants regard the image as samples from a continuous function and fit basis functions to approximate it [17]. Frequency-domain variants re-



(a) Local Correlation (LCSR)

(b) Resolution Synthesis (RS)

(c) Bayesian edge inference (BEI)

Figure 1. Comparison of three super-resolution methods on a region of "Peppers" at factors of 2x and 4x. Insets are subregions magnified an additional 2x using bilinear interpolation for display only. LCSR [7] (a) and RS [2] (b) are arguably the best published methods, as measured in [26]. Note that artifacts are almost entirely absent in 2x but show up clearly in 4x, namely steps in steep gradient areas (upper inset) and boundary incoherence (lower inset). BEI (c) does not exhibit these artifacts even at 4x.

1. Artifacts may be difficult to discern on a printed copy. A color PDF and supplementary material are available in the conference proceedings.

978-1-4244-3991-1/09/$25.00 ©2009 IEEE

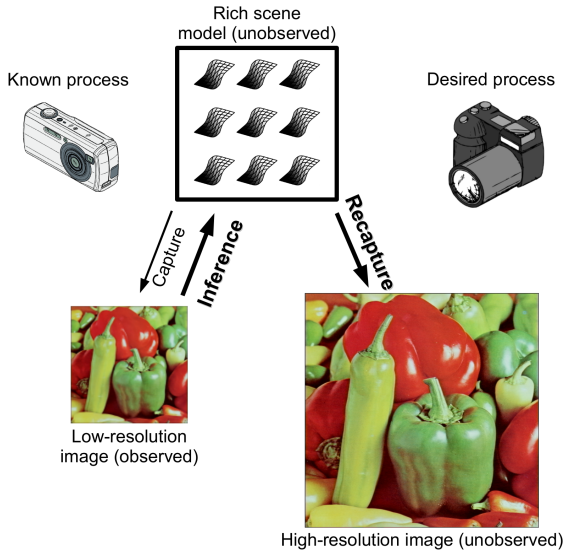


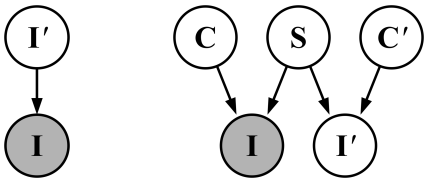
Figure 2. The recapture framework applied to super-resolution. The original low-resolution image is assumed to have been generated by a capture process operating on an unobserved scene. Inference recovers the scene, which is used to capture a new image.

gard the image as a sample of a bandlimited signal and target perfect reconstruction [19, 25]. All of these suffer from blockiness or blurring with moderate magnification. The reason is simple: the scene is effectively not bandlimited.

Adaptive methods make strong assumptions about the nature of scenes to obtain more plausible results. Parametric variants attempt to preserve strong edges by fitting edges [1, 14] or adapting basis functions [16, 18]. Nonparametric variants discover features that should be preserved using training images [2, 7] or use training images both as samples from the distribution of all images and as primitives for reconstruction [12, 24].

Ouwerkerk recently surveyed adaptive methods [26] and applied correctness measures to their outputs on test images at 2x and 4x magnification factors. Methods that give excellent results on 2x super-resolution tend to show artifacts at higher factors. For example, Figure 1 shows the results of applying the two methods found to be best to a region of "Peppers". Artifacts that are nearly absent at 2x become noticeable at 4x, namely steps in gradient areas and boundary incoherence. Because artifacts show up so well at those scales, our research focuses on factors of 4x or more.

Optimization methods [21, 23, 29] formulate desirable characteristics of images as penalties or priors and combine this with a reconstruction constraint to obtain an objective function. The reconstruction constraint ensures that the result, when downsampled, matches the input image. Though BEI is similar in many respects, it does not model downsampling of a high-resolution image, but models image capture of a detailed scene and *recapture* with a fictional, higher-resolution process (Figure 2). For this we adapt a Bayesian framework from supervised machine learning [8].



(a) Optimization framework (b) Recapture framework

Figure 3. Bayesian frameworks for image reconstruction. Shaded nodes are observed. Optimization methods (a) model an assumed high-resolution original $\mathbf{I}'$ generating the input image $\mathbf{I}$. This requires prior knowledge about $\mathbf{I}'$ and a degradation process $\mathbf{I}|\mathbf{I}'$. The proposed framework (b) models *capture* and *recapture* rather than (or including) degradation. This requires a rich scene model $\mathbf{S}$ and capture process models $\mathbf{I}|\mathbf{C}, \mathbf{S}$ and $\mathbf{I}'|\mathbf{C}', \mathbf{S}$.

For scale-invariance we model a projection of the scene as a piecewise continuous function, much like a facet model [11]. To address blur analytically, we construct it such that it approximates the continuous blurring of step edges with a spatially varying PSF. We address stepping in gradients by carefully modeling minimum blur.

Outside of modeling the scene hierarchically, some notion of compatibility among scene primitives [12, 24] is required to ensure that object boundaries are coherent. We show that Markov random field compatibility functions can be incorporated into Bayesian networks in a way that is direct, intuitive, and preserves independence relationships, and then incorporate compatibility into our model.

2. Reconstruction by recapture

Optimization methods apply to reconstruction tasks in general. These assume that an original image $\mathbf{I}'$ existed, which was degraded to produce $\mathbf{I}$. They are either motivated or formulated explicitly in Bayesian terms, in two parts. First is a prior on $\mathbf{I}'$, which encodes knowledge about images such as gradient profiles [23] or isophote curvature [21]. The second part is often called a *reconstruction constraint* [3], *back-projection* [13] or *sensor model* [5]: a conditional distribution $\mathbf{I}|\mathbf{I}'$ that favors instances of $\mathbf{I}'$ that, when degraded, match $\mathbf{I}$. The result is usually found by maximizing the joint probability $\mathrm{P}(\mathbf{I}|\mathbf{I}')\mathrm{P}(\mathbf{I}')$ to obtain the most probable $\mathbf{I}'|\mathbf{I}$. Figure 3(a) shows the framework as a simple Bayesian network.

Consider two tasks that have been addressed in the optimization framework. First, a super-resolution task: scaling up a full-resolution digital photo for printing. Second, CCD demosaicing: a camera has filtered light before detecting it with a single-chip sensor. Two-thirds of the color data is missing and must be inferred. Both violate assumptions made by optimization methods. There was no pristine original image that was degraded, and the only thing that can be reconstructed is the scene. In these cases and many others, *the true objective is to produce a novel image of the same scene as if it had been captured using a better process*.

Based on this objective, we propose the more general re-

capture framework shown in Figure 3(b). Here, a process (*e.g.* a camera) with parameters $\mathbf{C}$ is assumed to have captured the scene $\mathbf{S}$ as the original image $\mathbf{I}$. This process may include degradation. A fictional process (*e.g.* a better camera) with parameters $\mathbf{C}'$ *recaptures* the same scene as the result $\mathbf{I}'$. As in optimization methods, $\mathbf{I}$ is observed and inference recovers $\mathbf{I}'$, but through $\mathbf{S}$ rather than directly. This requires a scene model rich enough to reconstruct an image.

There is also a practical advantage to recapture. With the right scene model, if only recapture parameters are changed, $\mathbf{I}'$ can be recaptured at interactive speeds.

3. Effect modeling in Bayesian networks

Our super-resolution method models the scene using overlapping primitives, which must be kept locally coherent. This has been done using ad-hoc compatibility [12] and Markov random field (MRF) clique potentials [24]. However, converting recapture models to MRFs would hide independence relationships and the notion of causality—and image capture is obviously causal in nature. Graphical models rarely mix causal and noncausal dependence. Chain graphs do [6] but are less well-known and more complex than Bayesian networks and MRFs.

It is worth noting that MRFs are not used in reconstruction to model noncausal dependence for its own sake, but to *model unknown causes that have known effects* on an image or scene, which are usually symmetric. This is appropriate when inferring causes is cumbersome or intractable.

Fortunately, modeling unknown causes with known effects in a Bayesian network is simple (Figure 4). In the interest of saving space, we note without giving details that motivation for the following comes from the conversion of MRFs to factor graphs to Bayesian networks [28]. Let $\mathbf{X} = \{X_1, X_2, ..., X_n\}$ be the set of random variables in a Bayesian network, and $\mathbf{\Phi} = \{\Phi_1, \Phi_2, ..., \Phi_m\}$ be a set of functions that specify an effect (such as compatibility). Let $\mathbf{x}$ be instances of $\mathbf{X}$, with $\mathbf{x}_{\{i\}},\ i \in 1..m$ denoting an indexed subset. Each Φ_i is a mapping from $\mathbf{x}_{\{i\}}$ to $\mathbb{R}^+$. For each Φ_i, add to the network a new real-valued observed variable Z_i with density f_{Z_i} such that

$$f_{Z_i}(z_i = 0|\mathbf{x}_{\{i\}}) = \Phi_i(\mathbf{x}_{\{i\}}) \tag{1}$$

Because Z_i is real-valued, f_{Z_i} does not have to be normalized. Because it will remain observed, its density does not have to be specified except at 0. (There are uncountably infinite candidates for f_{Z_i}; we will assume one of them.) Adding this new observed variable cannot create cycles or introduce unwanted first-order dependence.

Inference may proceed on joint density $p'(\mathbf{x})$:

$$\begin{aligned} p'(\mathbf{x}) &\equiv p(\mathbf{x}|\mathbf{z} = \mathbf{0}) = f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Z}}(\mathbf{z} = \mathbf{0}|\mathbf{x}) / p(\mathbf{z} = \mathbf{0}) \\ &\propto f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Z}}(\mathbf{z} = \mathbf{0}|\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}) \prod_{i=1}^{m} \Phi_i(\mathbf{x}_{\{i\}}) \end{aligned} \tag{2}$$

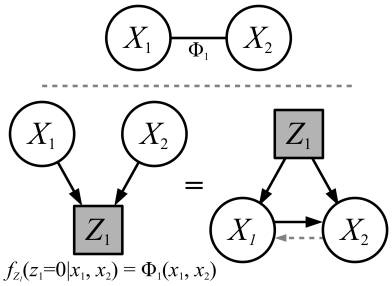


Figure 4. Bayesian effect modeling. X_1 and X_2 share an unknown cause with known effect Φ_1. This is modeled as an observed node in the network. An equivalent network exists in which X_1 and X_2 are directly dependent and the joint distribution X_1, X_2 is symmetric if Φ_1 is symmetric and $X_1 \sim X_2$.

For Gibbs sampling [10], Markov blanket conditionals are

$$\begin{aligned} p'(x_j|\mathbf{x}_{\{-j\}}) &\equiv p(x_j|\mathbf{x}_{\{-j\}}, \mathbf{z} = \mathbf{0}) \\ &\propto f_{X_j}(x_j|x_{\mathrm{par}(j)}) \prod_{k \in \mathrm{ch}(\mathbf{X},j)} f_{X_k}(x_k|x_{\mathrm{par}(k)}) \prod_{i \in \mathrm{ch}(\mathbf{Z},j)} \Phi_i(\mathbf{x}_{\{i\}}) \end{aligned} \tag{3}$$

where $\mathrm{par}(j)$ yields the indexes of the parents of X_j and $\mathrm{ch}(\mathbf{A}, j)$ yields the indexes of the children of X_j within $\mathbf{A}$.

4. Super-resolution model

The steps to using the recapture framework for reconstruction are: 1) define the scene model, expressing knowledge about the scene or scenes in general as priors; 2) define the capture and recapture processes; and 3) observe $\mathbf{I}$ and report $\mathbf{I}'$. Because the objective is to generate an image as if it had been captured by a fictional process, the proper report is a sample from (rather than say, a MAP estimate of) the *posterior predictive* distribution $\mathbf{I}'|\mathbf{I}$. This may be done by running a sampler such as Gibbs or MCMC on $\mathbf{S}|\mathbf{I}$, followed by sampling $\mathbf{I}'|\mathbf{S}$ once.

Definitions. An image $\mathbf{I}$, which is an $m \times n$ array of RGB triples normalized to $[0, 1]$, is observed. A real-valued scaling factor s is selected and an $\lfloor sm \rfloor \times \lfloor sn \rfloor$ image $\mathbf{I}'$ is reconstructed through an $m \times n$ scene model $\mathbf{S}$. Coordinates of triples, which are parameters of the capture process, are

$$\begin{aligned} \mathbf{C}^x_{i,j} &\equiv i + \tfrac{1}{2} \qquad i \in 0..m-1 \\ \mathbf{C}^y_{i,j} &\equiv j + \tfrac{1}{2} \qquad j \in 0..n-1 \end{aligned} \tag{4}$$

where i, j are image indexes. (In this paper, subscripts denote indexing an array-valued random variable and superscripts a named position in a tuple of random variables; *e.g.* $\mathbf{C}^x_{i,j}$ means "the i, j-th element of array x in tuple $\mathbf{C}$".)

The scene and capture models get the nine nearest neighbors of an integer- or real-valued coordinate with

$$\begin{aligned} \mathrm{N9}(x, y) \ \equiv &\{i \in \mathbb{Z} \mid -1 \le i - \lfloor x \rfloor \le 1\} \\ &\times \{j \in \mathbb{Z} \mid -1 \le j - \lfloor y \rfloor \le 1\} \end{aligned} \tag{5}$$

For clarity we omit here treatment of image borders.

The model uses an approximating quadratic B-spline kernel [9] to weight facet outputs, which we denote as w.

(a) A region of "Monarch," downsampled 2x and reconstructed. (b) The inferred PSF standard deviation $\mathbf{S}^{\sigma}$. Darker is narrower.

Figure 5. BEI's spatially varying PSF. It has correctly inferred a wider PSF for the flower petals, which are blurry due to shallow depth-of-field. These are still blurry in the final output.

4.1. Facets

$\mathbf{S}$ is similar to a facet model [11] in that it uses overlapping geometric primitives to represent a continuous function. It differs in these fundamental ways: 1) facets are blurred step edges, not polynomials; 2) it represents a scene rather than an image; 3) the *combined* output of the primitives is fit to the data through a capture model; and 4) facets are made compatible with neighbors where they overlap.

The assumption that the most important places to model well are object boundaries determines the shape of the facets. Each is based on an implicit line:

$$\mathrm{dist}(x, y, \theta, d) \equiv x \cos\theta + y \sin\theta - d \tag{6}$$

To approximate blurring with a spatially varying point-spread function (PSF) [15], we assign each facet a Gaussian PSF and convolve each analytically before combining outputs. For simplicity, PSFs are symmetric and only vary in standard deviation. The usefulness of modeling the PSF as spatially varying is demonstrated in Figure 5.

Convolving a discontinuity with a Gaussian kernel gives the profile of the step edge:

$$\begin{aligned}&\mathrm{prof}(d, \sigma, v^+, v^-)\\ &\quad\equiv v^+ \int_0^\infty G(d-t, \sigma)\, dt + v^- \int_{-\infty}^0 G(d-t, \sigma)\, dt\\ &\quad= \tfrac{v^+ - v^-}{2}\, \mathrm{erf}\left(\tfrac{d}{\sqrt{2}\sigma}\right) + \tfrac{v^+ + v^-}{2}\end{aligned} \tag{7}$$

where erf is the error function and v^+ and v^- are the values on the positive and negative sides. Because of the PSFs' radial symmetry, facets are defined in terms of profiles:

$$\begin{aligned}&\mathrm{edge}(x, y, \theta, d, v^+, v^-, \sigma)\\ &\qquad\equiv \mathrm{prof}(\mathrm{dist}(x, y, \theta, d), \sigma, v^+, v^-)\end{aligned} \tag{8}$$

An example step edge is shown in Figure 6.

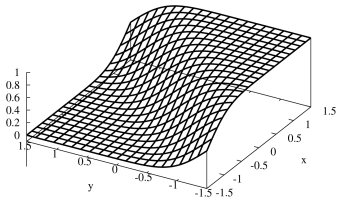


Figure 6. Scene facets are blurred step edges, or linear discontinuities convolved with blurring kernels. This has $\theta = -\pi/4, d = 0$ as the line parameters and a Gaussian kernel with $\sigma = 1/3$.

4.2. Scene model

The scene model random variables are a tuple of $m \times n$ arrays sufficient to parameterize an array of facets:

$$\mathbf{S} \equiv (\mathbf{S}^{\theta}, \mathbf{S}^{d}, \mathbf{S}^{v^+}, \mathbf{S}^{v^-}, \mathbf{S}^{\sigma}) \tag{9}$$

We regard the scene as an array of facet functions. Let

$$\begin{aligned}&\mathbf{S}^{\mathrm{edge}}_{i,j}(x, y) \equiv\\ &\quad\mathrm{edge}(x - \mathbf{C}^x_{i,j}, y - \mathbf{C}^y_{i,j}, \mathbf{S}^{\theta}_{i,j}, \mathbf{S}^{d}_{i,j}, \mathbf{S}^{v^+}_{i,j}, \mathbf{S}^{v^-}_{i,j}, \mathbf{S}^{\sigma}_{i,j})\end{aligned} \tag{10}$$

be an array of facet functions centered at $\mathbf{C}^x, \mathbf{C}^y$ and parameterized on the variables in $\mathbf{S}$.

A generalization of weighted facet output, *weighted expected scene value*, is also useful:

$$\mathrm{E}[h(\mathbf{S}_{x,y})] \equiv \sum_{k,l \in \mathrm{N9}(x,y)} w(x - \mathbf{C}^x_{k,l}, y - \mathbf{C}^y_{k,l})\; h(\mathbf{S}^{\mathrm{edge}}_{k,l}(x, y)) \tag{11}$$

When $h(x) = x$, this is simply weighted output. Weighted scene variance will be defined later using $h(x) = x^2$.

Priors. It seems reasonable to believe that, for each facet considered alone,

1. No geometry is more likely than any other.
2. No intensity is more likely than any other.
3. There are proportionally few strong edges [24].

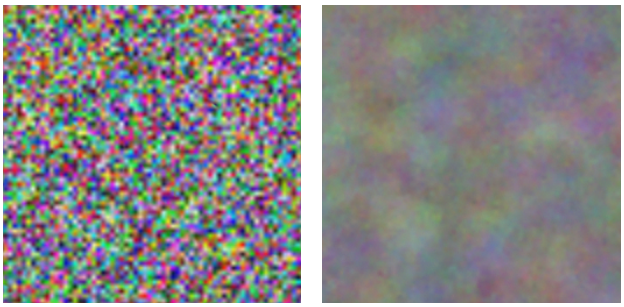
The priors are chosen to represent those beliefs:

$$\begin{array}{ll}\mathbf{S}^{\theta}_{i,j} \sim \mathrm{Uniform}(-\pi, \pi) & \mathbf{S}^{v^+}_{i,j} \sim \mathrm{Uniform}(0, 1)\\ \mathbf{S}^{d}_{i,j} \sim \mathrm{Uniform}(-3, 3) & \mathbf{S}^{v^-}_{i,j} \sim \mathrm{Uniform}(0, 1)\\ \mathbf{S}^{\sigma}_{i,j} \sim \mathrm{Beta}(1.6, 1) & \end{array} \tag{12}$$

Compatibility. It seems reasonable to believe that scenes are comprised mostly of regions of similar color, and that neighboring edges tend to line up. We claim that both can be represented by giving high probability to low variance in facet output. (Figure 7 demonstrates that this is the case.) Recalling that $\mathrm{E}[\mathbf{S}^2_{i,j}] - \mathrm{E}[\mathbf{S}_{i,j}]^2 = \mathrm{Var}[\mathbf{S}_{i,j}]$, define

$$\Phi_{i,j}(\mathbf{S}_{\mathrm{N9}(i,j)}) \equiv \exp\left(-\frac{\mathrm{Var}[\mathbf{S}_{i,j}]}{2\gamma^2}\right) \tag{13}$$

as the compatibility of the neighborhood centered at i, j, where γ is a standard-deviation-like parameter that controls the relative strength of compatibility. At values near ω (defined in the capture model as standard deviation of



(a) Two samples from the prior predictive $\mathbf{I}'$ (*i.e.* no data).

(b) Two samples from the posterior predictive $\mathbf{I}'|\mathbf{I}$ (*i.e.* with data).

Figure 7. The utility of compatibility. The right images include compatibility. The samples without data (a) show that it biases the prior toward contiguous regions. The samples with data (b) show that it makes coherent boundaries more probable.

the assumed white noise), compatibility tends to favor very smooth boundaries at the expense of detail. We use $\gamma = 3\,\omega = 0.015$, which is relatively weak.

In image processing, compatibility is usually defined in terms of pairwise potentials. We found it more difficult to control its strength relative to the capture model that way, and more difficult to reason about weighting. Weighting seems important, as it gives consistently better results than not weighting. This may be because weighted compatibility has a measure of freedom from the pixel grid.

4.3. Capture and recapture

The capture and recapture processes assume uniform white noise approximated by a narrow Normal distribution centered at the weighted output value:

$$\mathbf{I}_{i,j}|\mathbf{S}_{\mathrm{N9}(i,j)} \sim \mathrm{Normal}(\mathrm{E}[\mathbf{S}_{i,j}], \omega) \tag{14}$$

where i, j are real-valued coordinates and ω is the standard deviation of the assumed white noise. We use $\omega = 0.005$.

Recapture differs from capture in treatment of $\mathbf{S}^\sigma$ (defined in the following section) and in using a bilinear kernel to combine facet outputs. The bilinear kernel gives better results, possibly because it makes up for blur inadvertently introduced by the quadratic kernel w.

4.4. Minimum blur

To make the recaptured image look sharp, we assume the original capture process had a minimum PSF width C^σ and give the recapture process a narrower minimum PSF width $C^{\sigma\prime}$. Because variance sums over convolution of blurring kernels, these are accounted for in the capture model by adding variances. That is, rather than computing $\mathbf{S}^{\mathrm{edge}}$ using $\mathbf{S}^\sigma$, $\mathbf{I}|\mathbf{S}$ uses

$$\mathbf{S}^{\sigma\,*}_{k,l} \equiv \sqrt{(\mathbf{S}^\sigma_{k,l})^2 + (C^\sigma)^2} \tag{15}$$

The value of C^σ depends on the actual capture process. For recapture, we have found that $C^{\sigma\prime} \equiv C^\sigma/s$ tends to give plausible results. (Recall that s is the scaling factor.)

4.5. Decimation blur

In [26], as is commonly done, images were downsampled by decimation: convolving with a 2×2 uniform kernel followed by nearest-neighbor sampling. When decimation blur has taken place, regardless of how many times, the capture process can model it as a constant minimum PSF.

With image coordinates relative to $\mathbf{I}$, decimation is approximable as application of ever-shrinking uniform kernels. The last kernel was one unit wide, the kernel previous to that was a half unit wide, and so on. Let u_w be a uniform kernel with width w. Assuming no upper bound on the number of decimations, the upper bound on variance is

$$\begin{aligned} u &\equiv u_1 * u_{\frac{1}{2}} * u_{\frac{1}{4}} * u_{\frac{1}{8}} * u_{\frac{1}{16}} * \cdots \\ \mathrm{Var}[u] &= \mathrm{Var}[u_1] + \mathrm{Var}[u_{\frac{1}{2}}] + \mathrm{Var}[u_{\frac{1}{4}}] + \cdots \\ &= \sum_{n=0}^{\infty} \tfrac{(1/2)^{2n}}{12} = \tfrac{1}{12}\left(\tfrac{4}{3}\right) = \tfrac{1}{9} \end{aligned} \tag{16}$$

The series converges so quickly that $C^\sigma = \frac{1}{3}$ is a good estimate for any number of decimations.

4.6. Inference

The Markov blanket for $\mathbf{S}_{i,j}$ includes its nine children in $\mathbf{I}$ and $\mathbf{\Phi}$, and their nine parents each in $\mathbf{S}$.

Only the time to convergence seems to be affected by choice of initial values. We use the following:

$$\begin{aligned} \mathbf{S}^\theta &= \tan^{-1}((\nabla\mathbf{I})_y/(\nabla\mathbf{I})_x) & \mathbf{S}^d &= 0 \\ \mathbf{S}^{v^+} &= \mathbf{S}^{v^-} = \mathbf{I} & \mathbf{S}^\sigma &= \tfrac{1}{2} \end{aligned} \tag{17}$$

In this model, posterior density in $\mathbf{S}$ is so concentrated near the modes that samples of $\mathbf{S}$ after convergence are virtually indistinguishable. Therefore we find a MAP estimate of $\mathbf{S}|\mathbf{I}$ and sample $\mathbf{I}'|\mathbf{S}$ to approximate sampling $\mathbf{I}'|\mathbf{I}$.

Gibbs with stochastic relaxation [10] finds a MAP estimate quickly, but a deterministic variant of it is faster. It proceeds as Gibbs sampling, except that for each random variable X, it evaluates the Markov blanket conditional at x, $x + \sigma_X$ and $x - \sigma_X$, and keeps the argmax value.

Tuning σ_X online results in fast convergence. Every iteration, it is set to an *exponential moving standard deviation* of the values seen so far. This is computed by tracking an

exponential moving mean and moving squared mean separately and using $\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$. Let

$$\begin{array}{ll} \sigma^2_{X_i} = v_{X_i} - m^2_{X_i} & \\ m_{X_0} = x_0 & m_{X_i} = \alpha\, m_{X_{i-1}} + (1-\alpha)\, x_i \\ v_{X_0} = x_0^2 + \sigma^2_{X_0} & v_{X_i} = \alpha\, v_{X_{i-1}} + (1-\alpha)\, x_i^2 \end{array} \quad (18)$$

where σ_{X_0} is the initial standard deviation. The value of α denotes how much weight is given to previous values. Using $\sigma_{X_0} = 0.05$ for all X and $\alpha = 0.5$, we found acceptable convergence within 100 iterations on all test images.

5. Results

Ouwerkerk [26] chose three measures that tend to indicate subjective success better than mean squared error, and gave results for nine single-frame super-resolution methods on seven test images chosen for representative diversity. The original images were decimated once or twice, reconstructed using each method, and compared. Therefore we set minimum blur $C^\sigma = \frac{1}{3}$ as derived in Section 4.5.

Figure 8 shows that BEI keeps boundaries coherent even in difficult neighborhoods because of compatibility. Note the boundaries of the narrow black veins, which are especially easy to get wrong. Figure 9 is a comparison of some methods with BEI on a region of "Lena," which shows that BEI preserves gradients and sharpens edges. Note gradients on the nose and on the shadow on the forehead, and the crisp boundaries of the shoulder and brim of the hat.

Table 1 gives measures for BEI in 4x super-resolution along with linear interpolation for a baseline and the top two, resolution synthesis (RS) [2] and local correlation (LCSR) [7], for comparison.

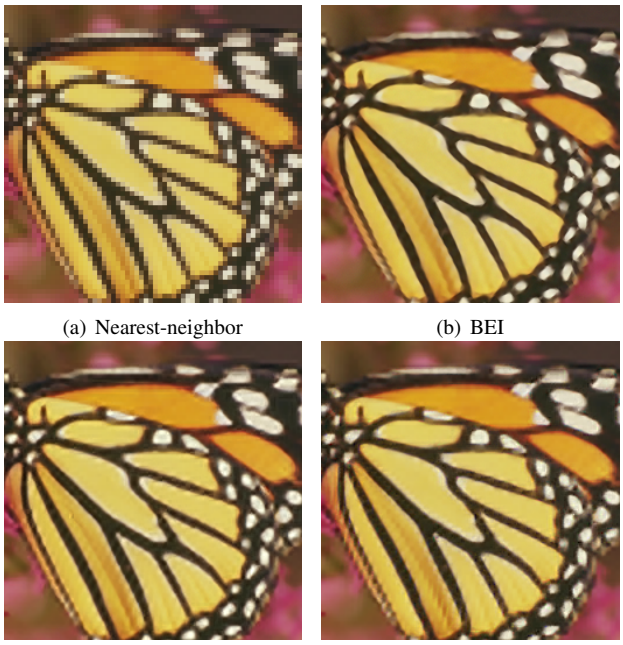
Unfortunately, a bug in computing ESMSE was not caught before publication of [26], making this measure suspect [27]. Also, it is questionable whether it measures *edge* stability, as the edge detector used falsely reports smooth, contiguous regions as edges. Therefore, Table 1 includes a corrected ESMSE measure using the same edge detector with its minimum threshold raised from 10% to 20%.

We give numeric results for noiseless recapture because the correctness measures are somewhat sensitive to noise. But in practice, a little noise usually increases plausibility.

6. Other applications

One advantage to Bayesian inference is that missing data is easy to deal with: simply do not include it.

In CCD demosaicing tasks [24], a Bayer filter, which is a checkerboard-like pattern of red, green, and blue, is assumed overlayed on the capture device's CCD array [4]. It could be said that the filtered two-thirds is missing data. We implemented this easily in BEI by not computing densities at missing values. The result of simulating a Bayer filter is shown in Figure 10. We also found it helpful to change the prior on $\mathbf{S}^\sigma$ to $\mathrm{Uniform}(0, 1)$ and set minimum blur to zero.



(a) Nearest-neighbor (b) BEI

(c) RS (d) LCSR

Figure 8. A difficult region of "Monarch". Most 3×3 neighborhoods within the black veins include part of the boundary on each side. While RS and LCSR have done well at avoiding artifacts here (much better than the others compared in [26]), BEI eliminates them almost entirely because of compatibility.

Inpainting can also be regarded as a missing data problem. By not computing densities in defaced regions, BEI returned the image shown in Figure 11. Again we flattened the prior on $\mathbf{S}^\sigma$ and set minimum blur to zero. We also set the initial values to the rather blurry output of a simple diffusion-based inpainting algorithm [22], which tends to speed convergence without changing the result.

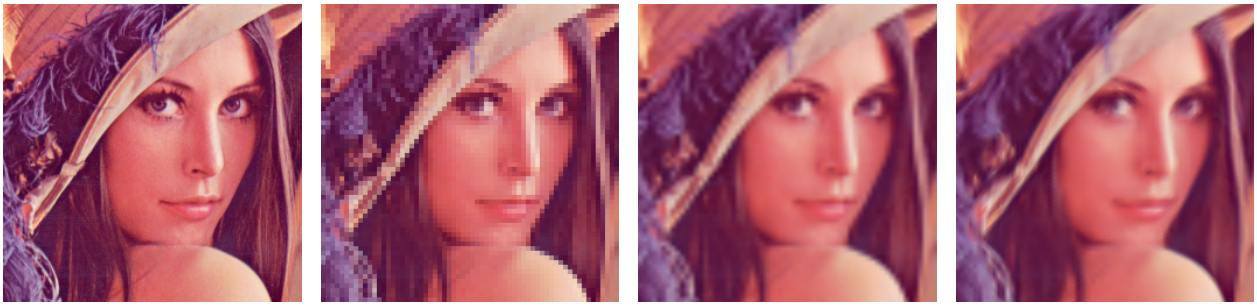
Super-resolution can be regarded as a missing data problem where the missing data is off the pixel grid. In fact, there is nothing specific to super-resolution in BEI's scene or capture model at all. Bayesian inference recovers the most probable scene given the scene model, capture process model, and whatever data is available. In this regard, super-resolution, CCD demosaicing, and inpainting are not just related, but are nearly identical.

7. Limitations and future work

BEI is computationally inefficient. Though inference is linear in image size, computing Markov blanket log densities for $9mn$ random variables is time-consuming. Our highly vectorized Python + NumPy implementation takes about 5 minutes on a 2.4GHz Intel CPU for 128×128 images. However, there is interpreter overhead, vectorization means BEI scales well in parallel, and nearly quadratic speedup could be gained by taking NEDI's hybrid approach [18], which restricts inference to detected edges. Also, while inference is inefficient, we believe recapture

	PSNR, higher is better				MSSIM, higher is better				ESMSE, lower is better				ESMSE fixed, 20% thresh.			
Image	Bilinear	RS	LCSR	BEI	Bilinear	RS	LCSR	BEI	Bilinear	RS	LCSR	BEI	Bilinear	RS	LCSR	BEI
Graphic	17.94	20.19	19.55	**20.87**	0.775	0.864	0.854	**0.898**	3.309	2.998	3.098	**2.151**	5.871	3.760	4.027	**3.571**
Lena	27.86	29.57	29.08	**29.60**	0.778	**0.821**	0.810	0.820	5.480	4.718	**4.706**	4.786	5.212	**4.472**	4.547	4.556
Mandrill	20.40	**20.71**	20.63	20.67	0.459	**0.536**	0.522	0.519	6.609	6.301	**6.278**	6.393	6.333	6.097	**6.075**	6.213
Monarch	23.91	26.41	25.90	**26.65**	0.848	0.896	0.889	**0.902**	5.448	**4.518**	4.606	4.547	5.260	**4.177**	4.445	4.214
Peppers	25.31	26.26	25.66	**26.27**	0.838	0.873	0.864	**0.876**	5.531	4.905	**4.864**	4.889	5.448	5.061	5.061	**5.043**
Sail	23.54	**24.63**	24.31	24.55	0.586	**0.679**	0.657	0.663	6.211	**5.776**	5.808	5.893	6.025	**5.305**	5.418	5.447
Tulips	25.43	28.19	27.56	**28.44**	0.779	0.843	0.831	**0.847**	5.994	5.198	5.286	**5.161**	5.679	4.569	4.769	**4.549**

Table 1. Comparison of bilinear, BEI, and the top two methods from [26], using correctness measures from the same, on 4x magnification. PSNR $= 10\log_{10}(s^2/\text{MSE})$, where s is the maximum image value and MSE is the mean squared error. MSSIM is the mean of a measure of local neighborhoods that includes mean, variance, and correlation statistics. ESMSE is the average squared difference in maximum number of sequential edges as found by a Canny edge detector with increasing blur. See the text for an explanation of "ESMSE fixed".



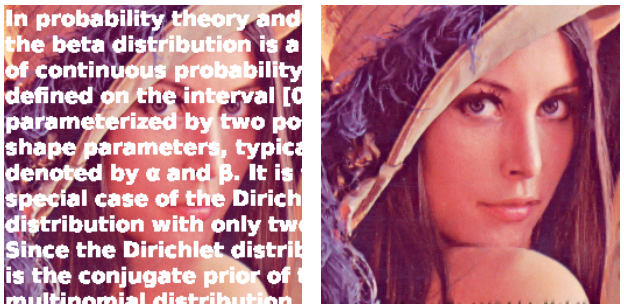
(a) Original (b) Nearest-neighbor (c) Bilinear (d) NEDI [18]

(e) LCSR (f) RS (g) BEI (h) BEI 8x

Figure 9. A 256×256 region of "Lena" (a) decimated twice and magnified 4x (b – g). Note the gradient steps in (e) and (f), especially in steep gradients such as on the nose and in the shadow on the forehead. Because BEI can model decimation blur explicitly in the capture and recapture processes, it preserves these gradients at 4x (g) and 8x (h) while keeping boundaries sharp.

(a) Cubic interpolation (b) BEI, flat $\mathbf{S}^\sigma$ priors, $C^\sigma = 0$

Figure 10. CCD demosaicing with a simulated Bayer filter. BEI, which was changed only trivially for this, treats it naturally as a missing data problem. Note the sharp edges and lack of ghosting.



(a) Defaced 33% (b) BEI, flat $\mathbf{S}^\sigma$ priors, $C^\sigma = 0$

Figure 11. Inpainting with BEI. As with CCD demosaicing, this requires only trivial changes to the model. Bayesian inference has recovered the most probable scene given the available data.

could be done at interactive speeds.

Almost all single-frame super-resolution methods tend to yield overly smooth results. Sharp edges and relative lack of detail combine to create an effect like the uncanny valley [20]. BEI, which does well on object boundaries and gradients, could be combined with methods that invent details like Tappen and Freeman's MRFs [24]. Good correctness measures for such methods could be difficult to find.

Many are confused by small amounts of noise, and would likely be even more confused by false but plausible details.

Ouwerkerk observed [26] that most methods could benefit from a line model, and we have observed that T-junctions are another good candidate.

Related to CCD demosaicing is undoing *bad* CCD demosaicing in images captured by devices that do not allow access to raw data. This may be as simple as modeling the naïve demosaicing algorithm in the capture process.

Because Bayesian models are composable, any sufficiently rich causal model can model the scene. If not rich enough, it can be used as a scene prior. For example, parameterized shape functions that return oriented discontinuities or functions from region classifications to expected gradients and values can directly condition priors on $\mathbf{S}$. Even compatibility functions can be conditioned on these.

8. Conclusion

We have presented Bayesian edge inference (BEI), a single-frame super-resolution method that successfully addresses two problems common to the best existing methods, which are steps in steep gradients and boundary incoherence. It is based on a recapture framework: a general Bayesian reconstruction framework with a rich scene model and explicit capture and recapture processes. This explicitness has allowed us to correctly model downsampling blur, which sharpens edges and preserves steep gradients in the upscaled result. The rich scene model requires some notion of compatibility or noncausal dependence among scene primitives, and we have shown how to incorporate such into any Bayesian model. We have demonstrated that compatibility, as implemented in BEI, tends to keep object boundaries coherent in the upscaled result. These subjective assessments translate to good performance on correctness measures, competitive with the best existing methods and surpassing them in many cases.

We have also shown that BEI has good subjective performance on other missing data problems besides super-resolution, namely CCD demosaicing and inpainting. This suggests that in this framework, these three problems are nearly identical. It also suggests that BEI and the recapture framework should generalize to other reconstruction tasks.

Acknowledgments

We gratefully acknowledge Jos van Ouwerkerk for his time, test images, and code, and Mike Gashler for the exponential moving variance calculation.

References

[1] J. Allebach and P. W. Wong. Edge-directed interpolation. In *Proc. ICIP*, volume 3, pages 707–710, 1996.

[2] C. Atkins, C. Bouman, and J. Allebach. Optimal image scaling using pixel classification. In *Proc. ICIP*, pages 864–867, 2001.

[3] S. Baker and T. Kanade. Limits on super-resolution and how to break them. *IEEE Trans. PAMI*, 24(9):1167–1183, 2002.

[4] B. E. Bayer. Color imaging array. US Patent 3,971,065, 1976.

[5] T. E. Boult and G. Wolberg. Local image reconstruction and subpixel restoration algorithms. *CVGIP: Graphical Models and Image Proc.*, 55(1):63–77, 1993.

[6] W. L. Buntine. Chain graphs for learning. In *Uncertainty in Artificial Intelligence*, pages 46–54, 1995.

[7] F. M. Candocia and J. C. Principe. Super-resolution of images based on local correlations. *IEEE Trans. Neural Networks*, 10:372–380, 1999.

[8] J. L. Carroll and K. D. Seppi. No-Free-Lunch and Bayesian optimality. In *IEEE Int. Joint Conf. Neural Networks Workshop on Meta-Learning*, 2007.

[9] N. A. Dodgson. Quadratic interpolation for image resampling. *IEEE Trans. Image Proc.*, (9):1322–1326, 1997.

[10] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. PAMI*, pages 721–741, 1984.

[11] R. M. Haralick. Digital step edges from zero crossings of second directional derivatives. *IEEE Trans. PAMI*, 6, 1984.

[12] A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin. Image analogies. *ACM SIGGRAPH*, pages 327–340, 2001.

[13] M. Irani and S. Peleg. Improving resolution by image registration. *CVGIP: Graphical Models and Image Proc.*, 53(3):231–239, 1991.

[14] K. Jensen and D. Anastassiou. Spatial resolution enhancement of images using nonlinear interpolation. In *Proc. ICASSP*, volume 4, pages 2045–2048, 1990.

[15] N. Joshi, R. Szeliski, and D. J. Kriegman. PSF estimation using sharp edge prediction. In *Proc. CVPR*, 2008.

[16] S. Lee and J. Paik. Image interpolation using adaptive fast B-spline filtering. In *Proc. ICASSP*, volume 5, pages 177–180, 1993.

[17] T. M. Lehmann. Survey: Interpolation methods in medical image processing. *IEEE Trans. Med. Imaging*, 18(11):1049–1075, 1999.

[18] X. Li and M. T. Orchard. New edge-directed interpolation. *IEEE Trans. Image Proc.*, 10:1521–1527, 2001.

[19] E. H. W. Meijering, W. J. Niessen, and M. A. Viergever. Quantitative evaluation of convolution-based methods for medical image interpolation. *Med. Image Analysis*, 5:111–126, 2001.

[20] M. Mori. The uncanny valley. *Energy*, 7(4):33–35, 1970.

[21] B. S. Morse and D. Schwartzwald. Image magnification using level-set reconstruction. In *Proc. CVPR*, volume 1, pages 333–340, 2001.

[22] M. M. Oliveira, B. Bowen, R. McKenna, and Y. S. Chang. Fast digital image inpainting. In *Proc. VIIP*, pages 261–266, 2001.

[23] J. Sun, Z. Xu, and H. Shum. Image super-resolution using gradient profile prior. In *Proc. CVPR*, 2008.

[24] M. Tappen, B. Russell, and W. Freeman. Exploiting the sparse derivative prior for super-resolution and image demosaicing. In *IEEE Workshop on Stat. and Comp. Theories of Vision*, 2003.

[25] T. Theussl, H. Hauser, and E. Gröller. Mastering windows: Improving reconstruction. In *Proc. VolVis*, pages 101–108, 2000.

[26] J. D. van Ouwerkerk. Image super-resolution survey. *Image and Vision Computing*, 24(10):1039–1052, 2006.

[27] J. D. van Ouwerkerk. Personal correspondence, Nov. 2008.

[28] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Understanding belief propagation and its generalizations. Tech. report, MERL, 2002.

[29] X. Yu, B. S. Morse, and T. W. Sederberg. Image reconstruction using data-dependent triangulation. *IEEE Comp. Graph. and App.*, 21(3):62–68, 2001.