Compressive Image Super-resolution

Pradeep Sen and Soheil Darabi
Advanced Graphics Lab, Department of Electrical and Computer Engineering
University of New Mexico, New Mexico, USA

***Abstract*— This paper proposes a new algorithm to generate a super-resolution image from a single, low-resolution input without the use of a training data set. We do this by exploiting the fact that the image is highly compressible in the wavelet domain and leverage recent results of compressed sensing (CS) theory to make an accurate estimate of the original high-resolution image. Unfortunately, traditional CS approaches do not allow direct use of a wavelet compression basis because of the coherency between the point-samples from the downsampling process and the wavelet basis. To overcome this problem, we incorporate the downsampling low-pass filter into our measurement matrix, which decreases coherency between the bases. To invert the downsampling process, we use the appropriate inverse filter and solve for the high-resolution image using a greedy, matching-pursuit algorithm. The result is a simple and efficient algorithm that can generate high quality, high-resolution images without the use of training data. We present results that show the improved performance of our method over existing super-resolution approaches.**

I. Introduction

Super-resolution (SR) algorithms attempt to generate a single high resolution (HR) image from one or more low-resolution (LR) images of the same scene. The main challenge is to recover the high-frequency information that was lost in the process of generating the low-resolution inputs. If the low-resolution images were directly captured by a camera, for example, this information was eliminated by the band-limitting filter of the photographic process due to imperfections in the optics and integration over the pixels of the sensor. If the low-resolution images were the result of software downsampling, this information was lost through the filtering process of anti-aliasing. The goal of the SR algorithms is to recover this missing information in a way that approximates the original high-resolution image as closely as possible. This is an important problem in several communities and has applications which include object recognition, video transmission, image compression, etc.

Traditionally, SR approaches have used a set of low-resolution images that were captured with sub-pixel accuracy to try to solve for the missing high-frequency information [1]. Recently, however, there has been growing interest in recovering this information from a single, low-resolution image, an area of research also known as *image hallucination* [2] or *image upsampling* [3]. The "single-image super-resolution" problem is particularly important because there are many applications in which only a single, low-resolution image is available and the upsampling must be applied as a post-process.

This work was funded by the NSF CAREER award #0845396 "A Framework for Sparse Signal Reconstruction for Computer Graphics."

In this paper, we focus on the problem of single-image super-resolution and present a novel algorithm for reconstructing the high-resolution image based on the recently-introduced field of *compressed sensing* (CS). The basic idea is that after reconstruction, the high-resolution result will be sparse in a transform domain (e.g., wavelet) and we can therefore use compressed sensing theory to directly solve for the sparse coefficients from the low-resolution input. By recovering an approximation to the wavelet transform of the high-resolution image, we can then compute the final result in the spatial domain.

This paper makes two specific contributions to the field of image super-resolution. First, we pose the super-resolution problem within the framework of compressed sensing, which allows us to apply the tools developed for CS (e.g., simple, greedy algorithms such as ROMP [4]) to solve for the high-resolution image. Second, we propose a novel way of using the wavelet basis (which works well for image compression) in our formulation by incorporating the blur filter from the downsampling process into our framework. This allows us to increase the incoherency between the sampling and compression basis and results in an algorithm that yields better results than current approaches. We begin by giving a brief overview of existing upsampling methods in the next section.

II. Previous Work

There has been significant research in super-resolution and upsampling algorithms in previous years and a complete description of all of the previous work is beyond the scope of a conference paper. Many of these algorithms utilize multiple low-resolution input images which are registered together and then reconstructed into a high-resolution image using constraints (e.g., Huber Markov Random Fields (MRF) [1], [5] and Bilateral Total Variation [6]), which is done through a regularization technique such as maximum a-posteriori (MAP). In this work, however, we are interested in the problem of single-image super-resolution. Traditionally, there have been two kinds of algorithms proposed for this problem:

1) Learning-based super-resolution algorithms which use a dictionary generated from an image database to invert the process of down-sampling in a reasonable manner [7], [8], [9], [10], [11].
2) Reconstruction-based super-resolution algorithms which do not use a training set but rather define constraints for

the target high-resolution image to improve the quality of the reconstruction [12], [3], [13].

Examples of the first kind of algorithm include Sun et al.'s work on image hallucination using priors [2], a Bayesian method which tries to preserve edge continuity in the HR image by relying on an image database. Another approach is Chang et al.'s neighbor embedding method [7], which extracts features from a database of small patches and finds correspondences between them and the LR image. These correspondences are used to generate the HR result from their high-resolution counterparts in the database. Finally, the recent work of Yang et al. [11] exploits the fact that the patches in the LR input can be sparse when represented in the basis of existing patches in the database. Specifically, they use an optimization algorithm to determine how to linearly combine the patches in the database to form the low-resolution image. They then combine their high-resolution counterparts with an added global reconstruction constraint to generate the final result.

Traditionally, algorithms of the second type try to reconstruct image details by interpolating the low-resolution input while making edges sharper. In addition to the standard bilinear and bicubic interpolation techniques, the most well-known algorithm in this category is the Back Projection (BP) method [13] which iterates on the image and gradually sharpens the edges in each iteration. Unfortunately this algorithm often has the unappealing artifact of ringing or jagginess. Another example is Dai et al.'s approach [12], which extracts the edges of the image in order to enforce their continuity and blends them with the interpolated result to yield the final image. More recent approaches include the work of Fattal [3], which uses edge statistics to reconstruct the missing high frequency information. We compare our approach to Fattal's in Sec. V.

Our proposed algorithm is of the second type since we do not require a training data set. Rather, we enforce the constraint that the high-resolution image be sparse in the wavelet domain and use this property to solve for the desired image. To do this, we pose our single-image super-resolution problem in the framework of compressed sensing (Sec. III) and then use a greedy algorithm to select the sparse wavelet coefficients that represent our result.

Our work is similar in spirit to the methods that attempt to solve the ill-posed problem of SR through regularization methods such as TV [14], [15]. Indeed, enforcing the sparsity in the wavelet domain is a form of regularization that gives us a way to overcome the ill-posedness of the SR problem. The fundamental difference between our approach and these previous methods is that by posing it within the framework of compressed sensing, we have access to an ample set of new tools to tackle this problem. One of these is the greedy Regularized Orthogonal Matching Pursuit (ROMP) algorithm we use in this paper.

We note that the aforementioned work of Yang et al. [11] also uses sparsity to regularize the problem of super-resolution. However, the main difference between the two approaches is that we use a general wavelet basis to sparsify the image, not a dictionary of images. This means that our approach does not require an existing data set and is potentitally more widely applicable, in the same way wavelet-based image compression methods are more general than dictionary methods such as vector quantization. In the next section, we present the key theoretical results in compressed sensing that form the foundation of our algorithm.

n	size of high resolution signal
m	size of the low resolution signal
k	size of the support of the signal in the transform domain
$\mathbf{x}$	$n \times 1$ high-resolution image vector in spatial domain
$\tilde{\mathbf{x}}$	$m \times 1$ low-resolution image vector in spatial domain
$\hat{\mathbf{x}}$	$n \times 1$ hi-resolution image in transform domain
$\mathbf{S}$	$m \times n$ downsampling matrix, s.t. $\tilde{\mathbf{x}} = \mathbf{S}\mathbf{x}$
$\mathbf{\Psi}$	$n \times n$ compression basis, a.k.a. "synthesis" matrix
$\mathbf{\Psi}^T$	compression transform matrix ($\mathbf{\Psi}^T = \mathbf{\Psi}^{-1}$), so $\hat{\mathbf{x}} = \mathbf{\Psi}^T\mathbf{x}$
$\mathcal{F}$	$n \times n$ Fourier transform matrix
$\mathcal{F}^H$	$n \times n$ conjugate transpose of $\mathcal{F}$ (inverse Fourier transform)
$\mathbf{G}$	$n \times n$ diagonal matrix with Gaussian function in the diagonal
$\mathbf{A}$	$m \times n$ "measurement" matrix, $\mathbf{A} = \mathbf{S}\mathbf{\Psi}$

TABLE I
NOTATION USED IN THIS PAPER

III. Compressed Sensing Theory

The theory of compressed sensing (CS) demonstrates how a subsampled signal can be faithfully reconstructed through non-linear optimization techniques [16], [17]. Suppose that we represent our desired high-resolution image as an n-dimensional vector $\mathbf{x} \in \mathbb{R}^n$ where n is large. In theory, $\mathbf{x}$ can represent any 1-D signal, but for this discussion we assume it to be an n-pixel grayscale (scalar) image which has been converted into an $n \times 1$ vector (with trivial extension to vector-valued signals, e.g., RGB images). We want to estimate this high-resolution signal from the low-resolution input $\tilde{\mathbf{x}} \in \mathbb{R}^m$, where $m \ll n$. We assume that $\tilde{\mathbf{x}}$ has been acquired from the original through a linear downsampling measurement process, written as:

$$\tilde{\mathbf{x}} = \mathbf{S}\mathbf{x}, \tag{1}$$

where $\mathbf{S}$ is a sampling matrix that performs the linear measurements on $\mathbf{x}$. Our goal is to recover the high-resolution $\mathbf{x}$ using only $\tilde{\mathbf{x}}$ as input.

Initially, this seems like an impossible feat since the m samples of $\tilde{\mathbf{x}}$ yield a $(n - m)$-dimensional subspace of possible solutions for the original $\mathbf{x}$ that would match our given observations. How do we know which one of those possible solutions is our desired $\mathbf{x}$? This is where we apply a key assumption of compressed sensing: we assume that the transformed version of the signal, $\hat{\mathbf{x}}$, is k-sparse under some basis $\mathbf{\Psi}$, meaning that it has at most k non-zero coefficients in that basis (e.g., $\|\hat{\mathbf{x}}\|_0 \leq k$, where $\|\cdot\|_0$ denotes the ℓ_0 quasi-norm). This is not an unreasonable assumption, since we know that the high-resolution image will be a "real-world image" (as opposed to random white noise), and so it will be compressible in a transform domain, e.g., wavelet. We can now write our measurement process from Eq. 1 as:

$$\tilde{\mathbf{x}} = \mathbf{S}\mathbf{\Psi}\hat{\mathbf{x}} = \mathbf{A}\hat{\mathbf{x}}, \tag{2}$$

where $\mathbf{A} = \mathbf{S}\mathbf{\Psi}$ is a general $m \times n$ *measurement* matrix. If we can solve for $\hat{\mathbf{x}}$ given the measured $\tilde{\mathbf{x}}$, we could then apply the inverse transform $\mathbf{\Psi}\hat{\mathbf{x}}$ to get our desired high-resolution signal $\mathbf{x}$. Unfortunately, traditional techniques for solving for $\hat{\mathbf{x}}$ (e.g., inversion, least squares) do not work because Eq. 2 is severely under-determined (since $m \ll n$). However, recent breakthroughs in compressed sensing have shown that if $m \geq 2k$ and $\mathbf{A}$ meets certain properties (see Sec. III-A), then Eq. 2 can be solved uniquely for $\hat{\mathbf{x}}$ by looking for the sparsest $\hat{\mathbf{x}}$ that satisfies the equation (see complete proof in [16]). Therefore, we can find the desired $\hat{\mathbf{x}}$ by solving the following ℓ_0-optimization problem:

$$\min \|\hat{\mathbf{x}}\|_0 \text{ s.t. } \tilde{\mathbf{x}} = \mathbf{A}\hat{\mathbf{x}}, \tag{3}$$

The solution to this problem, however, involves a combinatorial algorithm in which every $\hat{\mathbf{x}}$ with $\|\hat{\mathbf{x}}\|_0 \leq k$ is checked to find the one that results in the measured samples $\tilde{\mathbf{x}}$. This problem is known to be *NP*-complete [18] and is intractable for any reasonably-sized signal. However, recent results [16] have spurred the growing excitement in the area of compressed sensing by showing that Eq. 2 can be solved by replacing the ℓ_0 with an ℓ_1-norm ($\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |\mathbf{x}_i|$):

$$\min \|\hat{\mathbf{x}}\|_1 \text{ s.t. } \tilde{\mathbf{x}} = \mathbf{A}\hat{\mathbf{x}}, \tag{4}$$

As long as the number of samples $m = O(k \log n)$ and the matrix $\mathbf{A}$ meets the RIC (described next) with parameters $(2k, \sqrt{2}-1)$, the ℓ_1 optimization of Eq. 4 will solve correctly for $\hat{\mathbf{x}}$ [19]. This can be done with methods such as linear programming [16] and basis pursuit [17].

A. Restricted Isometry Condition (RIC)

We cannot solve $\tilde{\mathbf{x}} = \mathbf{A}\hat{\mathbf{x}}$ for $\hat{\mathbf{x}}$ with any arbitrary $\mathbf{A}$ if $m \ll n$, even if $m \geq 2k$. However, we can apply the compressed sensing framework if matrix $\mathbf{A}$ meets the Restricted Isometry Condition (RIC) [4]:

$$(1-\epsilon)||\mathbf{v}||_2 \leq ||\mathbf{A}\mathbf{v}||_2 \leq (1+\epsilon)||\mathbf{v}||_2, \tag{5}$$

with parameters (z, ϵ), where $\epsilon \in (0, 1)$ for all z-sparse vectors $\mathbf{v}$. Essentially, the RIC states that a measurement matrix will be valid if every possible set of z columns of $\mathbf{A}$ forms an approximate orthogonal set. In effect, we want the sampling matrix $\mathbf{S}$ to be as incoherent to the compression basis $\mathbf{\Psi}$ as possible. Examples of matrices that have been proven to meet RIC include Gaussian matrices (where the entries are independently sampled from a normal distribution), Bernoulli matrices (binary matrices drawn from a Bernoulli distribution), and partial Fourier matrices (randomly selected Fourier basis functions) [20].

In this work, we would like to use wavelets as our compression basis $\mathbf{\Psi}$ because they are much better at sparsely representing images than non-localized bases such as Fourier. However, in super-resolution the downsampling matrix $\mathbf{S}$ involves point-sampled measurements, which could result in a measurement matrix $\mathbf{A}$ that does not meet the RIC because point-sampling measurements are not incoherent with the wavelet compression basis. Intuitively, we can see that the better a basis is at representing localized features (such as wavelet), the more coherent it will be to point sampling because it can represent small spatial features (e.g., point samples) with only a few coefficients, by definition. Therefore, in order to successfully apply a wavelet basis to our problem, we must find a way to increase the incoherence between the bases. We explain how to do this in Section IV.

B. Greedy Reconstruction algorithms

Although the ℓ_1 optimization is considerably more efficient than the ℓ_0, its running time can still be large because there is no known strongly polynomial-time algorithm for linear programming [4]. For this reason, the CS research community has started to investigate greedy algorithms to solve Eq. 3. Orthogonal Matching Pursuit (OMP) was one of the first such algorithms explored [21]. Given the measured vector $\tilde{\mathbf{x}}$ and the measurement matrix $\mathbf{A}$, we can find the coefficient of $\hat{\mathbf{x}}$ with the largest magnitude by projecting $\tilde{\mathbf{x}}$ onto each column of $\mathbf{A}$ and selecting the largest $|\langle \tilde{\mathbf{x}}, \mathbf{a}_j \rangle|$, where $\mathbf{a}_j$ is the j^{th} column of $\mathbf{A}$. Once we have identified the largest coefficient of $\hat{\mathbf{x}}$, we then solve a least-squares problem assuming it is the only non-zero coefficient. We can then use the new estimate for $\hat{\mathbf{x}}$ to compute the estimated signal $\mathbf{x}$ and subtract it from the original measurements. We then iterate the algorithm, using the residual to solve for the next largest coefficient of $\hat{\mathbf{x}}$ one at a time. By iterating k times, we find an k-sparse approximation of the transform domain vector.

Although it is simple and fast, OMP has a major drawback because of its weaker guarantee of exact recovery than the ℓ_1 methods [4]. To overcome these limitations, a modification to OMP called Regularized Orthogonal Matching Pursuit (ROMP) was proposed which recovers multiple coefficients in each iteration, thereby accelerating the algorithm and making it more robust to meeting the RIC [4]. In this work, we use the ROMP algorithm for signal reconstruction and we describe it in more detail in Sec. IV-A.

C. Applications of Compressed Sensing

Since its inception just a few years ago, compressed sensing has been applied to problems in video processing [22], medical imaging [23], [24], bio-sensing [25], wireless channel mapping [26], and compressive imaging [27], [28], [29], [30], [31]. It is also starting to be used in the computer vision and graphics communities to solve problems in face recognition [32] and light transport acquisition [33], [34]. However, posing the problem of super-resolution within the CS framework by assuming sparsity in the wavelet domain like we are proposing is novel and has not been done before.

IV. Compressive Image Super-resolution

As discussed in Section III-A, although wavelet bases are very good at representing image data sparsely, we cannot use them directly in compressed sensing because they do not meet the RIC when combined with point-sampled measurements.

In order to fulfill this condition, we propose to modify Eq. 2 based on the observation that the high resolution image is filtered before downsampling, which happens both in the camera when the image is acquired or in software if the image is downsampled on a computer. In other words, we can write our desired high-resolution image as $\mathbf{x}_s$ (the sharp version), which is then filtered by matrix $\mathbf{\Phi}$ to result in a blurred, high-resolution version $\mathbf{x}_b = \mathbf{\Phi}\mathbf{x}_s$. This blurred version is then downsampled by Eq. 1:

$$\tilde{\mathbf{x}} = \mathbf{S}\mathbf{x}_b = \mathbf{S}\mathbf{\Phi}\mathbf{x}_s, \tag{6}$$

In this work, we choose a Gaussian filter as our filter $\mathbf{\Phi}$. Since we can think of this filter as a multiplication by a Gaussian in the frequency domain we can define $\mathbf{\Phi} = \mathcal{F}^H\mathbf{G}\mathcal{F}$, which makes Eq. 6:

$$\tilde{\mathbf{x}} = \mathbf{S}\mathcal{F}^H\mathbf{G}\mathcal{F}\mathbf{x}_s, \tag{7}$$

where $\mathcal{F}$ is the Fourier transform matrix and $\mathbf{G}$ is a Gaussian matrix with values of the Gaussian function along its diagonal and zeros elsewhere. With this formulation in hand, we can now solve for $\mathbf{x}_s$ by posing it as a compressed sensing problem by assuming that its transform $\hat{\mathbf{x}}_s$ is sparse in the wavelet domain:

$$\min \|\hat{\mathbf{x}}_s\|_0 \text{ s.t. } \tilde{\mathbf{x}} = \mathbf{S}\mathcal{F}^H\mathbf{G}\mathcal{F}\mathbf{\Psi}\hat{\mathbf{x}}_s, \tag{8}$$

As we discussed earlier, we can approximate a solution to this optimization problem using greedy methods. However, these greedy algorithms require having both the "forward" matrix $\mathbf{A}$ and "backward" matrix $\mathbf{A}^*$ in order to solve $\mathbf{y} = \mathbf{A}\mathbf{x}$, where typically $\mathbf{A}^* = \mathbf{A}^T$ with the assumption that $\|\mathbf{A}^T\mathbf{A}\mathbf{v}\| \approx \|\mathbf{v}\|$ [4]. In our formulation, the forward matrix $\mathbf{A} = \mathbf{S}\mathbf{\Phi}\mathbf{\Psi}$. Because the Gaussian matrix in $\mathbf{\Phi}$ cannot be inverted through transpose, i.e., $\mathbf{G}^{-1} \neq \mathbf{G}^T$, we use the backwards matrix of the form $\mathbf{A}^* = \mathbf{\Psi}^T\mathbf{\Phi}^{-1}\mathbf{S}^T = \mathbf{\Psi}^T\mathcal{F}^H\mathbf{G}^{-1}\mathcal{F}\mathbf{S}^T$. We note that $\mathbf{G}^{-1}$ is also a diagonal matrix that is supposed to have the inversion of the Gaussian function of $\mathbf{G}$ along its diagonal. However, we must be careful when inverting the Gaussian because of the well-known problem of noise amplification. To avoid this, we use a linear Weiner filter to invert the Gaussian function [35], which means that our inverse matrix $\mathbf{G}^{-1}$ has diagonal elements of the form $\mathbf{G}^{-1}_{i,i} = \mathbf{G}_{i,i}/(\mathbf{G}^2_{i,i} + \gamma)$.

To verify that the proposed formulation indeed increases the incoherence between the measurement and compression bases, we must compute the coherence with and without the blurring filter $\mathbf{\Phi}$. The coherence between two bases can be found by taking the maximum inner product between any two basis elements scaled by $\sqrt{n}$ [36]. Without the Gaussian filter, the coherence is 261.3 for $n = 512 \times 512$. On the other hand, when the Gaussian filter is introduced the coherence drops to 158.3. This reduction is enough to enable us to apply the CS framework to this problem and get the high-quality results shown in the paper. For comparison, we also show the results of leaving out the filter in Fig. 4. With this formulation in place, we are now ready to solve the compressed sensing problem of Eq. 8. We describe our reconstruction algorithm in the next section.

Algorithm 1 Modified ROMP algorithm

Input: measured vector $\tilde{\mathbf{x}} \in \mathbb{R}^m$, target sparsity k, forward matrix $\mathbf{A}$, backward matrix $\mathbf{A}^*$, max num of coeffs added each iter q
Output: set of indices $I \subset \{1, ..., n\}$ of non-zero coeffs in $\hat{\mathbf{x}}$
Initialize: $I = \emptyset$ and $\mathbf{r} = \tilde{\mathbf{x}}$

while $\mathbf{r} \neq \mathbf{0}$ **do**
  /* *multiply residual by* $\mathbf{A}^*$ *to approx. larger coeffs of* $\hat{\mathbf{x}}$ */
  $\mathbf{u} \Leftarrow \mathbf{A}^*\mathbf{r}$, max_energy $\Leftarrow 0$
  $J \Leftarrow$ set of k largest magnitude coefficients of $\mathbf{u}$
  sort (J) in non-increasing order
  for $i = 0$ to k **do**
    $j \Leftarrow$ largest index where $J(j) \geq J(i)/2$ and $j - i < q$
    /* *compute energy in J from element i to j* */
    energy $\Leftarrow$ ComputeEnergy(J, i, j)
    /* *replace old set if new energy is greater* */
    if energy $>$ max_energy **then**
      max_energy $\Leftarrow$ energy
      $J_0 \Leftarrow \{i, ..., j\}$
    end if
  end for
  $I \Leftarrow I \cup J_0$ /* *add new indices to overall set* */
  /* *find vector of* I *coeffs that best matches measurement* */
  $\mathbf{y} \Leftarrow \underset{\mathbf{z}\,:\,\text{supp}\,(\mathbf{z})=I}{\text{argmin}} \|\tilde{\mathbf{x}} - \mathbf{A}\mathbf{z}\|_2$
  $\mathbf{r} \Leftarrow \tilde{\mathbf{x}} - \mathbf{A}\mathbf{y}$ /* *recompute residual* */
end while
return I

A. Reconstruction algorithm

Given an initial low resolution image $\tilde{\mathbf{x}}$, we would like to solve for the wavelet transform of the sharp, high-resolution image $\hat{\mathbf{x}}_s$ as in Eq. 8. The idea is that once we solve for $\hat{\mathbf{x}}_s$, we can take its inverse wavelet transform $\mathbf{\Psi}\hat{\mathbf{x}}_s$ to recover our desired high-resolution image $\mathbf{x}_s$. To do this, we use the Regularized Orthogonal Matching Pursuit (ROMP) greedy algorithm we mentioned earlier [4]. ROMP is preferable over non-linear methods like linear programming [16] or Basis-pursuit [17] for our experiments because it is faster and can handle large vectors and matrices, which is necessary when working with images because the size of the matrices involved are $n \times n$ where n is on the order of 512^2.

ROMP is similar to the OMP algorithm described in Section III-B, in that it approximates the transform coefficients on every iteration and then sorts them in non-increasing order. However, the main difference is that instead of only selecting the largest coefficient, ROMP selects the continuous sub-group of coefficients with the largest energy, with the restriction that the largest coefficient in the group cannot be more than twice as big as the smallest member. These coefficients are then added to a list of non-zero coefficients and a least-squares problem is then solved to find the best approximation for these non-zero coefficients. The approximation error is then computed based on the measured results and the algorithm iterates again. In this work, we make the slight modification in that we limit the number of coefficients added in each iteration. We found experimentally that setting this limit to $m/60$ yielded better results than the conventional ROMP algorithm. A summary of our modified ROMP algorithm is shown in the previous page.

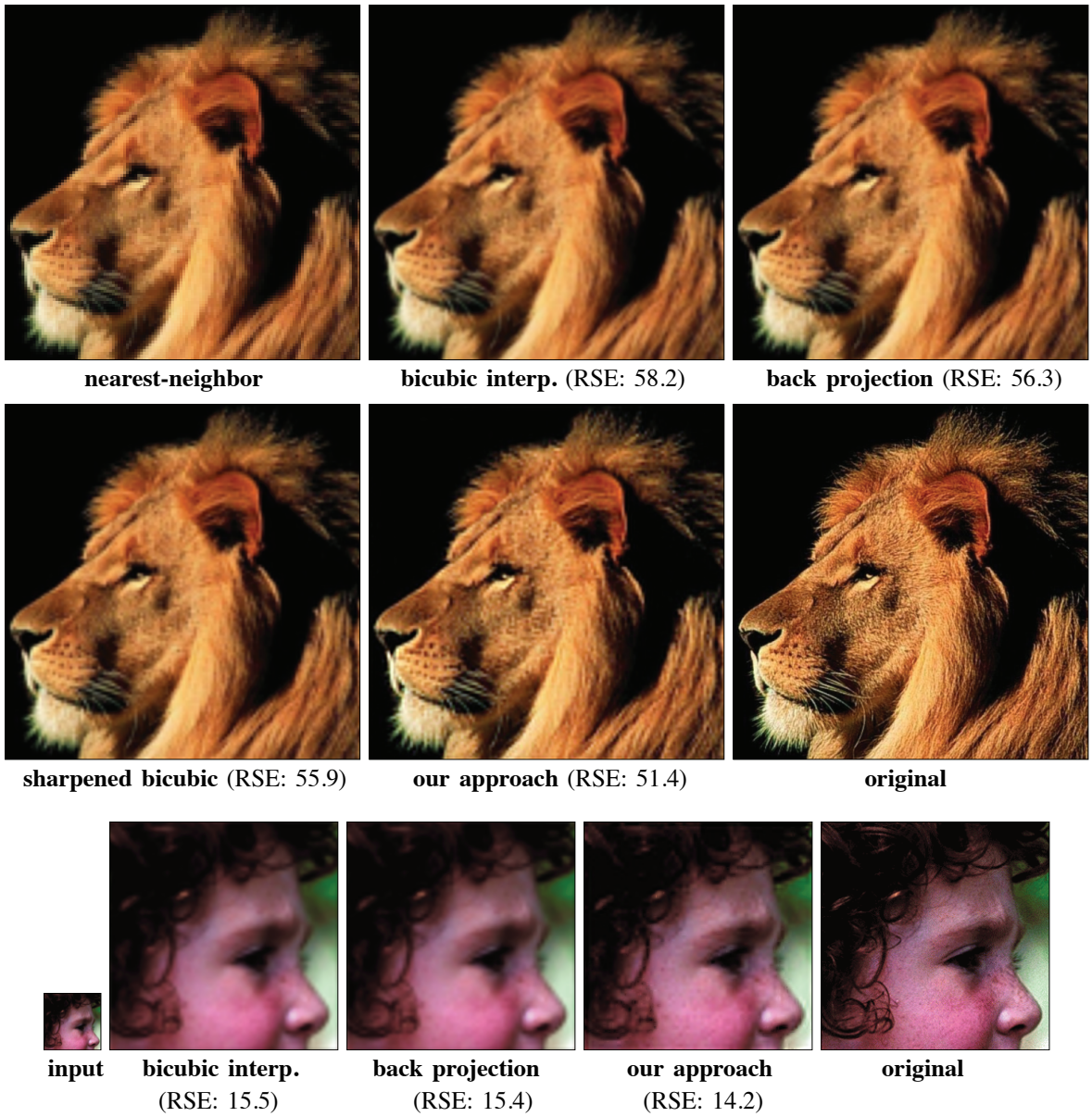

Fig. 1. These two images have been magnified by a factor of 4 and the results of various standard algorithms are compared. The proposed algorithm produces an upsampled image with more detail and lower RSE than the other approaches.

V. Implementation and Results

We implemented the proposed algorithm in C and tested it with the sample images shown in Figs. 1 – 5. To downsample the original images, we first blur them with a Gaussian filter $\mathbf{\Phi} = \mathcal{F}^H \mathbf{G} \mathcal{F}$ to bandlimit them for antialiasing. To avoid over-/under-blurring, the variance of the Gaussian along the diagonal of $\mathbf{G}$ is a function of the amount of downsampling. For example, if we downsample the original image by a factor of 4, we set the variance of $\mathbf{G}$ to $\sigma^2 = 1666.7$ (spatial-domain variance of 4).

Once we have computed the blurred, high-resolution image $\mathbf{x}_b = \mathbf{\Phi} \mathbf{x}_s$, we point-sample it to get low-resolution version $\tilde{\mathbf{x}} = \mathbf{S} \mathbf{x}_b$ which serves as the direct input to our algorithm without further transformation. Note that our algorithm only utilizes $\tilde{\mathbf{x}}$ to perform the reconstruction, which makes it simpler than other SR algorithms which use statistics of natural images (e.g., [3]) or libraries of patches (e.g., [11]) as additional input information. Given the input $\tilde{\mathbf{x}}$, we recover a high-resolution approximation by solving Eq. 8 using a Daubechies-8 wavelet for our compression basis $\mathbf{\Psi}$ and ROMP as the solver to approximate the solution in 30 iterations. Note that for color images, we solve Eq. 8 three times, once for each color channel. As mentioned in the last section, our solver uses a Weiner filter to invert the Gaussian filter, and in our experiments we set $\gamma = 1$. The resulting reconstruction algorithm is reasonably fast, and takes less than 40 seconds to upsample a 128×128 image to 512×512 on a 2.2 GHz laptop.

original | **bi-cubic interp.** (RSE: 28.34) | **example-based** (RSE: 33.26) | **edge stats** (RSE: 25.87) | **our approach** (RSE: 25.76)

Fig. 2. Comparison of our approach with recent work, specifically the example-based SR of Freeman et al. [8] and the edge statistics method of Fattal [3]. Image is being upsampled by $4\times$. Although Fattal's method is better than the other approaches, the result has a cartoon-like quality because of the emphasis on edges. Our result looks more natural and has a lower RSE. Data courtesy of Raanan Fattal.

To test our algorithm, we upsample images over a wide range of magnification factors using various algorithms. However, for clear visual comparison, the images in this paper are generated with 2, 3, and $4\times$ magnification. In Fig. 1 we magnify the two test images by a factor of 4 and compare the results of our technique against the standard approaches: bicubic interpolation, a "sharpened bicubic" filter computed in MATLAB by first performing a bicubic interpolation and then sharpening the result, and the back-projection method [13]. We quantify each algorithm's accuracy by computing the Root Square Error (RSE), a measure of the Euclidean distance, of their output to the original HR image. We note that for both images in Fig. 1, our algorithm produces results with sharper details and lower RSE. For example, one can see the freckles on the child's face in the image generated by our technique. Although we do not use an image database, the quality of our reconstructed image is visually comparable to the ones in [7] or in [8]. Additional results on more test images are shown in Fig. 5.

To compare against more recent research work in super-resolution and image upsampling, we test our technique against the example-based super-resolution of Freeman et al. [8] and the edge statistics method of Fattal [3] in Fig. 2. Fattal's method was presented at SIGGRAPH 2007 and is known to give very good results as compared to other leading algorithms and software packages. However, it produces results with a cartoon-like quality because it over-emphasizes the edges in the image. Our approach on the other hand, produces high quality results (lower RSE than the other approaches) and more natural images. We should also note that in this example the input image was downsampled by a bilinear filter, not by a Gaussian filter. Therefore, it also shows that our approach is robust even when the downsampling filter does not match the one used in Eq. 8.

In Fig. 3 we plot how the RSE error varies with magnification level for various approaches. We can see that our technique is better than both back-projection [13] and the sharpened bicubic at every level of magnification. Finally, it is interesting to consider what happens if we do not add the blur filter Φ we propose in our algorithm. As we said in Sec. IV, there would not be enough incoherence between the measurement and compression basis, but it is unclear how that would affect the quality of the reconstruction. As we can see in Fig. 4, the quality of the reconstruction is significantly reduced when the filter is removed, yielding results much worse than simple bicubic interpolation. This validates our proposed approach of incorporating the blur downsampling filter into the CS formulation.

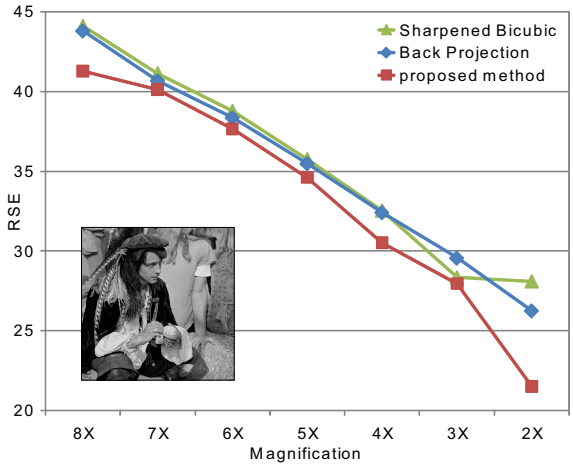

Fig. 3. Reconstruction error comparison as a function of magnification level for the given test image. As the magnification is decreased, more information is available for the algorithms so the error is reduced. The proposed algorithm works better than sharpened bicubic and back-projection for all magnification factors.

VI. Discussion

In earlier work, Donoho introduced the concept of image restoration with the wavelet transform under the name of Wavelet-Vaguelette Deconvolution (WVD) [37], in which an

Fig. 4. The problem of not using the Gaussian filter formulation. On the left, we show the result of simply point-sampling the initial image to downsample by a factor of 2 and then reconstructing by solving the CS problem using wavelet compression but without the blurring filter $\mathbf{\Phi}$ (RSE: 85.5). The result is unusable because the coherency between the point-samples and the wavelet basis breaks down the CS approach. On the right, we use our proposed formulation with a blur filter and the image quality is significantly improved (RSE: 40.6). For comparison, the RSE for bicubic interpolation is 51.9.

image that has been degraded by a filter is restored. This work was followed by several papers which tried to restore signals using the same idea. Since one of the weaknesses of the wavelet transform is its inability to efficiently represent different degrading convolution filters (which can be efficiently done in the Fourier domain), Neelamani et al. [38] proposed to combine these two transforms together to reconstruct the signal, a technique popularly known as ForWaRD, or Fourier-Wavelet Regularized Deconvolution.

Our proposed algorithm has a similar flavor to ForWaRD in that we are also trying to reconstruct the image in the wavelet domain while at the same time deconvolving the signal in the Fourier domain. The difference is that we use compressed sensing to solve the inverse equation for the missing signal values. Therefore, our algorithm is more general than Neelamani et al.'s because it can also handle the uncertainty due to down-sampling by assuming the sparsity of the signal in the wavelet domain.

VII. Future Work and Conclusion

We believe that there might be better-suited wavelet bases for the proposed application. As shown in the results, the DB-8 wavelets we used for compression worked well but there were sensitive to changes in the blurring filter $\mathbf{\Phi}$. For example, directed wavelet transforms such as complex wavelets might improve the effectiveness of our algorithm and make it more robust. In addition, the combination of our method with training-based techniques such as that of Yang et al. [11] would be interesting to explore.

In conclusion, we present a novel approach for single-image super-resolution by leveraging ideas of compressed sensing. In particular, we demonstrate how to pose the SR problem within the CS framework and use a greedy matching pursuit algorithm to solve for the high resolution image. The advantage of our technique is that it is very simple (it does not use a data dictionary for learning, for example), yet it can still produce results better than existing methods commonly used for this application. We hope that this approach encourages others to examine applications of compressed sensing in computer vision which would greatly benefit our community.

VIII. Acknowledgment

The authors would like to acknowledge Dr. Yasamin Mostofi for helpful discussions during the development of this work.

References

[1] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint map registration and high-resolution image estimation using a sequence of undersampled images," *Image Processing, IEEE Transactions on*, vol. 6, no. 12, pp. 1621–1633, Dec 1997.

[2] J. Sun, N.-N. Zheng, H. Tao, and H.-Y. Shum, "Image hallucination with primal sketch priors," in *IEEE CVPR*, vol. 2, June 2003.

[3] R. Fattal, "Image upsampling via imposed edge statistics," *ACM Trans. Graph.*, vol. 26, no. 3, p. 95, 2007.

[4] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," 2007, preprint.

[5] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *Image Processing, IEEE Transactions on*, vol. 5, no. 6, pp. 996–1011, Jun 1996.

[6] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *Image Processing, IEEE Transactions on*, vol. 13, no. 10, pp. 1327–1344, Oct. 2004.

[7] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," *IEEE CVPR*, vol. 1, pp. 275–282, 2004.

[8] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE CG & A*, vol. 22, no. 2, pp. 56–65, 2002.

[9] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low-level vision," *Int. J. Comput. Vision*, vol. 40, no. 1, pp. 25–47, October 2000. [Online]. Available: http://dx.doi.org/10.1023/A:1026501619075

[10] Q. Wang, X. Tang, and H.-Y. Shum, "Patch based blind image super resolution," *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 1, pp. 709–716, Oct. 2005.

[11] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," *IEEE CVPR*, pp. 1–8, June 2008.

[12] S. Dai, M. Han, W. Xu, Y. Wu, and Y. Gong, "Soft edge smoothness prior for alpha channel super resolution," *CVPR*, pp. 1–8, June 2007.

[13] M. Irani and S. Peleg, "Motion analysis for image enhancement: Resolution, occlusion, and transparency," *J. of Visual Communication and Image Representation*, vol. 4, no. 4, pp. 324–335, Dec 1993.

[14] H. Aly and E. Dubois, "Image up-sampling using total-variation regularization with a new observation model," *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1647–1659, Oct 2005.

[15] S. Osher, A. Solé, and L. Vese, "Image decomposition and restoration using total variation minimization and the H^{-1} norm," *SIAM J. of Multiscale Modeling and Simulation*, vol. 1, no. 3, pp. 349–370, 2003.

[16] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Info. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[17] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[18] E. J. Candès, M. Rudelson, T. Tao, and R. Vershynin, "Error correction via linear programming." in *IEEE Symposium on Foundations of Computer Science (FOCS)*, 2005, pp. 295–308.

[19] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus de l'Academie des Sciences, Paris, Serie I*, vol. 346, pp. 589–592, 2008.

[20] E. J. Candès and T. Tao, "Near optimal signal recovery from random projections: universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[21] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[22] R. Marcia and R. Willett, "Compressive coded aperture video reconstruction," in *European Sig. Proc. Conf.*, Aug. 2008.

[23] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

RSE: 50.0 vs. 60.0 (bicubic) RSE: 28.7 vs. 32.4 (bicubic) RSE: 30.9 vs. 34.3 (bicubic) RSE: 57.0 vs. 66.8 (bicubic)

Fig. 5. Sample results for more test images, magnified by a factor of 3. Top row shows the input image up-sampled with nearest-neighbor method, the bottom row is the output of our algorithm. We quote the RSE for our method compared to that of bicubic interpolation.

[24] J. Trzasko, A. Manduca, and E. Borisch, "Highly undersampled magnetic resonance image reconstruction via homotopic ell-0-minimization," *submitted*, 2008.

[25] M. A. Sheikh, O. Milenkovic, and R. G. Baraniuk, "Designing compressive sensing DNA microarrays," *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMPSAP 2007)*, pp. 141–144, Dec. 2007.

[26] Y. Mostofi and P. Sen, "Compressed signal strength mapping," *Milcom*, Nov 2008.

[27] K. Egiazarian, A. Foi, and V. Katkovnik, "Compressed sensing image reconstruction via recursive spatially adaptive filtering," in *IEEE Conference on Image Processing (ICIP)*, Sep. 2007.

[28] L. Gan, "Block compressed sensing of natural images," in *Conference on Digital Signal Processing (DSP)*, Jul. 2007.

[29] R. F. Marcia and R. M. Willett, "Compressive coded aperture superresolution image reconstruction," *IEEE ICASSP*, Apr. 2008.

[30] D. Takhar, J. Laska, M. Wakin, M. Duarte, D. Baron, S. Sarvotham, K. Kelly, and R. Baraniuk, "A new compressive imaging camera architecture using optical-domain compression," in *Proc. of Computational Imaging IV at SPIE Electronic Imaging*. SPIE, Jan. 2006.

[31] P. Sen and S. Darabi, "A novel framework for imaging using compressed sensing," *IEEE Conference on Image Processing (ICIP)*, Nov. 2009.

[32] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008.

[33] J. Gu, S. Nayar, E. Grinspun, P. Belhumeur, and R. Ramamoorthi, "Compressive structured light for recovering inhomogeneous participating media," in *ECCV*, Oct 2008.

[34] P. Sen and S. Darabi, "Compressive Dual Photography," *Computer Graphics Forum*, vol. 28, no. 2, pp. 609 – 618, 2009.

[35] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Addison-Wesley Longman Publishing Co., Inc., 2001.

[36] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. on Info. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov 2001.

[37] D. L. Donoho, "Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition," *Appl. Comput. Harm. Anal*, vol. 2, pp. 101–126, 1995.

[38] R. Neelamani, H. Choi, and R. Baraniuk, "Forward: Fourier-wavelet regularized deconvolution for ill-conditioned systems," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 418–433, Feb. 2004.