

Image Recovery in the Infrared Domain via Path-Augmented Compressive Sampling Matching Pursuit

Tegan H. Emerson

U.S. Naval Research Laboratory
Optical Sciences Division
4555 Overlook Ave. S.W.
Washington, DC 20375
tegan.emerson@nrl.navy.mil

Colin C. Olson

U.S. Naval Research Laboratory
Optical Sciences Division
4555 Overlook Ave. S.W.
Washington, DC 20375
colin.olson@nrl.navy.mil

Anthony Lutz

Sotera/KeyW Defense Solutions, Inc.
2121 Cooperative Way, Suite 400
Herndon, VA 20171
anthony.lutz.ctr@nrl.navy.mil

Abstract

We consider compressive sensing as a means of acquiring high-resolution images from low-cost, low-resolution sensors in the infrared domain. In particular, we reduce errors arising from basis mismatch between the observed image and the signal model by modifying a baseline matching pursuit recovery algorithm. Specifically, we introduce a modification to the analysis step which seeks to find more representative image atoms by searching over a 2-Wasserstein geodesic formed between the two most-correlated atoms at that step. We test our extension by quantifying recovery performance on an ensemble of representative infrared maritime scenes and find improvement over baseline when measured using PSNR, SSIM, and a metric that quantifies global edge recovery performance. We find that the most notable gains occur for very low sparsity levels which favors reduced computational load for the recovery.

1. Introduction

Advances in sensor technology have yielded higher image resolutions than ever before. The obvious advantages of higher resolution images include increased performance in problems such as automated detection, identification, and classification. However, there is an inevitable initial increase in cost associated with incorporating new, state-of-the-art components into deployed imaging systems. While the benefits may outweigh

the cost when limited numbers of imaging systems are needed, there are many settings in which networks of imaging systems are required and cost scales prohibitively. In addition, high-resolution (HR) sensors in wavebands outside the visible spectrum (e.g., infrared) can also be too expensive for reasonable application.

There has been a push to develop computational approaches that seek to improve the effective resolution of images acquired by low-resolution (LR) sensors. Of these approaches, compressive sensing (CS) [13, 19, 40] and super-resolution (SR) [33] have received considerable attention in recent years due to the availability of increased computational power. Although CS has shown promise in the lab, its adoption in deployed systems has been limited in some part due to the need to include a dynamic mask within the optical train. Traditional multi-frame SR [24] also requires a change in the acquisition model by requiring the accumulation of multiple frames from which a single higher resolution image may be constructed—a mandate which effectively trades time for the possibility of increased resolution. In both cases, requiring a modification to the acquisition model and trading time for resolution can limit adoption by user communities.

As a result, alternatives to multi-frame acquisition are of interest. One possible approach is single-frame SR [25, 1] where one means of improving resolution involves using image models or reconstruction processes that have been trained on images that are statistically similar to the images one expects to up-sample and

where recent advances in sparse coding models and convolutional neural networks (CNNs) have yielded significant performance improvements in applications where access to a set of high-resolution training images is feasible [46, 38, 17, 18]. There are, however, applications that preclude the acquisition of a sufficiently representative set of HR images. In particular, it can be dramatically more expensive to acquire HR infrared (IR) images (especially labeled images) that have similar resolution to images acquired in the visible band. Thus, a key challenge arising in the IR domain is limited access to HR training data which disqualifies many of the state-of-the-art approaches referenced above. Imaging applications subject to this constraint are the motivation behind the methodology proposed herein.

Here we investigate single-shot CS recovery as a means of acquiring HR images with lower-cost, LR infrared imagers. We consider an acquisition architecture introduced by Romberg [41] and assume access to a structured dictionary (e.g., discrete cosine transform (DCT)) where the resolution can be precisely controlled and use of HR images for training is not required. For comparison, we also test the method assuming access to learned, unstructured dictionaries [2, 15, 26] because they often yield superior image reconstructions (at the cost of requiring HR training images).

In our approach, we utilize a greedy matching pursuit (MP) algorithm [7, 32] to reconstruct local LR image patches from a LR dictionary produced by applying the CS acquisition model to the HR dictionary. Once a sparse code describing a LR patch has been found, we synthesize the HR patch using the HR atoms and the same set of sparse coefficients found in the LR synthesis. This approach of performing an analysis step in the compressed space followed by synthesis in the high-resolution space is known as CoSaMP [34]. Here, we extend the method by implementing a path-augmentation step [23, 21] which helps address the known issue of basis mismatch arising in sparse CS reconstructions [8, 14]. In addition, we implement the path-augmented CoSaMP algorithm on image patches with a global stitching of the recovered patches to produce the final image reconstruction.

A background on compressed sensing, the CoSaMP algorithm, and path-augmentation is provided in Section 2. In Section 3 we will present our novel path-augmented image reconstruction algorithm (PAIR). Section 4 describes an experimental setup for evaluating PAIR using paths between atoms and a description of the three metrics used to benchmark PAIR against the standard CoSaMP algorithm. Results of the experiment and a discussion of the findings are presented in Section

5 and Section 6, respectively.

2. Background

2.1. Compressed Sensing

Compressive sensing (CS) is a methodology that enables higher-resolution digital sampling of natural phenomena via (1) the design of an acquisition process that satisfies the restricted isometry property (RIP) [13, 12, 10] and (2) the selection of good signal models to reconstruct higher-resolution signals that are undersampled according to classical Nyquist sampling theory [3, 13, 19, 40]. In this case, “good” models are those that can sparsely represent signals as linear combinations of relatively few atoms drawn from a dictionary. Signals that can be represented to within some acceptable error tolerance using at most k atoms are defined as k -sparse relative to that dictionary. Here, we will use CS and sparse representation/recovery interchangeably.

CS theory predicts, to a level of probabilistic certainty, successful reconstruction of an undersampled signal when the underlying true signal satisfies upper limits on sparsity relative to a given dictionary [4, 11, 12, 28]. For a fixed dictionary (Ψ) and a signal to reconstruct (\vec{x}) the k -sparse representation can be mathematically written as

$$\min_{\vec{s}} \|\vec{x} - \Psi \vec{s}\|_2^2 \quad \text{s.t.} \quad \|\vec{s}\|_0 = k \quad (1)$$

where $\|\cdot\|_2$ and $\|\cdot\|_0$ denote the vector 2-norm and 0-norm, respectively. This problem is non-convex and no closed form solution exists.

Within the CS framework the search for a k -sparse representation takes place in the downsampled, sensed space that results from the acquisition process. Finding solutions to Equation 1 quickly becomes intractable due to its non-convex, combinatorial nature. For this reason greedy algorithms have been developed to find good (even if non-optimal) reconstructions.

2.2. Matching Pursuit for Sparse Recovery

Matching Pursuit (MP) algorithms are a set of greedy algorithms that are used extensively for k -sparse recovery problems. MP algorithms are comprised of four steps which are repeated until a desired level of sparsity is obtained or a suitable stopping criterion has been met. These four steps are:

- Identify the dictionary element that is most correlated to the signal (first step only) or the current residual.
- Append the current most-correlated element to the support of the reconstruction.

- Determine the weight of each atom in the current support.
- Update the residual by subtracting the current estimate (the weighted sum of the current support) from the original observed signal.

Differences in MP variants arise from the number of correlated elements identified, number of atoms appended to the support, and how the weights are determined at each iteration [9, 37, 43, 45, 35, 20, 16, 31].

A variant of the MP algorithm called Compressive Sampling Matching Pursuit (CoSaMP) has been employed for a variety of CS tasks [34]. CoSaMP requires a dictionary from which signals can be reconstructed (referred to as the *measurement matrix*) and a known forward model for the data acquisition process (called the *sensing matrix/function*). Let $\Psi = \{\psi_1, \dots, \psi_D\}$ be a dictionary composed of high-resolution atoms. Given a sensing function, denoted F , we can generate a sensed dictionary $\hat{\Psi} = \{\hat{\psi}_1, \dots, \hat{\psi}_D\}$ where $\hat{\psi}_i = F(\psi_i)$.

CoSaMP uses the same general MP steps when searching for a good k -sparse representation of the observed/sensed signal with respect to the sensed dictionary. The full resolution signal is then approximated using the weights identified in the sensed space but using the corresponding atoms in the high resolution space. A visual summary of the CoSaMP algorithm is shown in Figure 1.

2.3. Addressing Basis Mismatch

Basis mismatch refers to the challenge encountered when discretization of the dictionary used for signal recovery introduces reconstruction errors. While significant effort has been spent on designing dictionaries which reduce modeling error (e.g. learned, overcomplete dictionaries constructed from ensembles of expected images), these prior efforts have not completely solved the problem because even a perfect signal model will still have basis mismatch error arising from the discretization. Although there is a reduction in basis mismatch error when using a learned overcomplete dictionary, this reduction due to overcompleteness comes at the expense of weaker guarantees on the uniqueness of the k -sparse representation. Overall, a number of authors have considered the basis mismatch issue [8, 14, 22, 39, 42, 36] but the problem remains open.

Here, we will utilize a method for constructing better patch exemplars from an underlying dictionary that relies on constructing a geodesic from the optimal transport map between said atoms (the 2-Wasserstein geodesic) [44]. This “path-based” augmentation can be applied to any reconstruction algorithm that relies on the selection and sorting of high-correlation atoms

during an analysis or identification phase [23] and has shown recent success in other applications of MP algorithms for sparse recovery and reconstruction [21]. Optimal transport has proven useful in other domains [29, 21, 27, 30, 5] but this work is the first time that path-based augmentation appears within the CoSaMP framework.

The difference between overcompleteness and path-augmentation can be conceptualized in terms of sampling an underlying model space. Overcomplete dictionaries more densely sample the space so as to increase the likelihood of sampling a point in the model space “close” to the observed point. With path-augmentation the sampling is done in two stages. The first stage is a coarse sampling of the model space that is used to identify a region of the model space “near” the observed point. Using the identified region of the model space a denser sampling of the local region is performed by generating samples along the geodesic between the two pin points in the model space.

3. Path-Augmented Image Recovery

We now present our proposed algorithm for image recovery based on the k -Sparse Compressive Sampling Matching Pursuit which we refer to as Path-Augmented Image Recovery (PAIR). For ease of notation a bold-faced letter will indicate the variable is a matrix and a vector marker over a lowercase letter will indicate the variable is a vector. Additionally, for dimensions we will use lower(upper)case to indicate a dimension pertaining to the low(high) resolution components of the model. We assume, without loss of generality, that both the low- and high-resolution images are square.

Let \mathbf{Y} be the observed low-resolution image given by

$$\mathbf{Y} = \begin{bmatrix} y_{1,1} & \cdots & y_{1,l} \\ \vdots & \ddots & \vdots \\ y_{l,1} & \cdots & y_{l,l} \end{bmatrix} \in \mathbb{R}^{l \times l}. \quad (2)$$

The vectorized low-resolution, observed image is then defined as

$$\vec{y} = [y_{1,1} \cdots y_{1,l} \ y_{2,1} \cdots y_{2,l} \ y_{l,1} \cdots y_{l,l}]^\top \in \mathbb{R}^{l^2 \times 1}. \quad (3)$$

The unknown high-resolution image we want to recover, as well as its vectorized counter-part, are given by

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,L} \\ \vdots & \ddots & \vdots \\ x_{L,1} & \cdots & x_{L,L} \end{bmatrix} \in \mathbb{R}^{L \times L} \quad (4)$$

and

$$\vec{x} = [x_{1,1} \cdots x_{1,L} \ x_{2,1} \cdots x_{2,L} \ x_{L,1} \cdots x_{L,L}]^\top \in \mathbb{R}^{L^2 \times 1}, \quad (5)$$

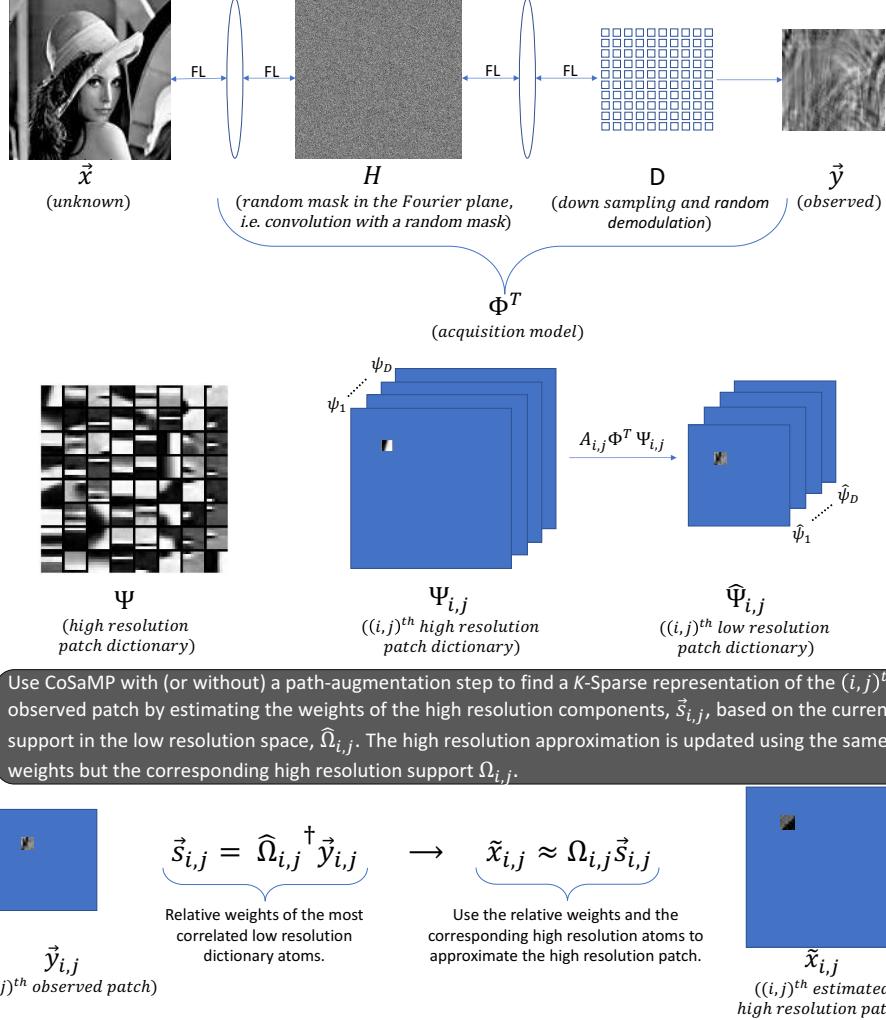


Figure 1. Graphical abstract for using CoSaMP and a known acquisition model to achieve high-resolution image recovery. The sensing model is assumed to be known and is applied to the atoms of a zero-padded, high-resolution, patch-based dictionary. Matching pursuit is applied in the sensed space where a set of dictionary elements and their corresponding weights are identified. PAIR is distinguished from CoSaMP by the addition of a path-augmentation step when constructing the k -sparse representation which can be seen in Figure 2. A high-resolution image patch is then constructed using the weights from the sensed space but using the corresponding high-resolution atoms corresponding to the sensed space support. These high-resolution patches are then stitched together to form a global estimate for the unknown image.

respectively.

We assume an acquisition model based on convolution with a random mask and a random demodulation. By placing the random mask in the Fourier plane, i.e. between two lenses of our imaging system, we are able to write our forward model as a linear operator. In order to leverage the theory of CS we assume that our forward model satisfies RIP, see [41] for verification of this assumption. Our model then becomes

$$\vec{y} = \Phi^\top \vec{x}, \quad \Phi^\top = \mathbf{D} \mathcal{F}^{-1} \mathbf{H} \mathcal{F} \in \mathbb{R}^{L^2 \times l^2} \quad (6)$$

where Φ is the forward model capturing convolution with the random mask, \mathbf{H} , in the Fourier plane (where

\mathcal{F} and \mathcal{F}^{-1} indicate the Fourier transform and inverse transforms, respectively) followed by a downsampling and random demodulation (indicated by \mathbf{D}). For more information regarding this formulation we refer the reader to [41].

Let p and P be the size of image patches from the low- and high-resolution images, respectively. We can associate the $(i, j)^{th}$ patches of the observed and unknown images, denoted $\vec{y}_{i,j}$ and $\vec{x}_{i,j}$ respectively, to their global image by

$$\vec{y}_{i,j} = \mathbf{A}_{i,j} \vec{y} \quad \text{and} \quad \vec{x}_{i,j} = \mathbf{B}_{i,j} \vec{x} \quad (7)$$

where $\mathbf{A}_{i,j}$ and $\mathbf{B}_{i,j}$ are block matrices defined as follows.



Figure 2. An example of the path-augmented procedure within CoSaMP. The two most-correlated atoms to the current residual in the sensed space are identified. The corresponding atoms in the high-dimensional space are identified. The 2-Wasserstein geodesic between the atoms is sampled. These samples are then passed through the forward model. Once in the sensed space the current residual is compared to each of path samples and the optimal t_* , $\hat{\psi}_*$, and ψ_* are identified. Finally $\hat{\psi}_*$, and ψ_* are augmented to the support in the corresponding space, $\hat{\Omega}_{i,j}$ and $\Omega_{i,j}$ respectively.

First we define

$$\mathbf{A}_{i,j} = \begin{bmatrix} \mathbf{a}_{1,1} & \cdots & \mathbf{a}_{1,l} \\ \vdots & \ddots & \vdots \\ \mathbf{a}_{l,1} & \cdots & \mathbf{a}_{l,l} \end{bmatrix} \in \mathbb{R}^{l^2 \times l^2} \quad (8)$$

such that

$$\mathbf{a}_{m,n} = \begin{bmatrix} \mathbf{I}_p & \mathbf{0}_{p,l-p} \\ \mathbf{0}_{l-p,p} & \mathbf{0}_{l-p,l-p} \end{bmatrix} \quad (9)$$

when $i, j \leq m, n \leq i+p-1, j+p-1$ and $\mathbf{b}_{m,n} = \mathbf{0}_{l,l}$ otherwise. Similarly, we write

$$\mathbf{B}_{i,j} = \begin{bmatrix} \mathbf{b}_{1,1} & \cdots & \mathbf{b}_{1,L} \\ \vdots & \ddots & \vdots \\ \mathbf{b}_{L,1} & \cdots & \mathbf{b}_{L,L} \end{bmatrix} \in \mathbb{R}^{L^2 \times L^2} \quad (10)$$

where

$$\mathbf{b}_{m,n} = \begin{bmatrix} \mathbf{I}_P & \mathbf{0}_{P,L-P} \\ \mathbf{0}_{L-P,P} & \mathbf{0}_{L-P,L-P} \end{bmatrix} \quad (11)$$

when $i, j \leq m, n \leq i+P-1, j+P-1$ and $\mathbf{b}_{m,n} = \mathbf{0}_{L,L}$ otherwise. We note that Φ , $\mathbf{A}_{i,j}$, and $\mathbf{B}_{i,j}$ are rank-deficient matrices.

Consider a patch-based signal model, Ψ , for high-resolution image patches. We can then build a signal model for the $(i, j)^{th}$ patch by appropriately zero padding the elements to spatially coincide with the $(i, j)^{th}$ patch. Thus, we define $\Psi_{i,j} = [\psi_1, \psi_2, \dots, \psi_D] \in \mathbb{R}^{L^2 \times D}$ to be signal model for $\vec{x}_{i,j}$. $\vec{x}_{i,j}$ can be represented sparsely with respect to this signal model, that is $\vec{x}_{i,j} = \Psi_{i,j} \vec{s}_{i,j}$ such that $\vec{s}_{i,j} \in \mathbb{R}^D$ is sparse.

We can first relate the information in a low-resolution patch to the full high-resolution image by

$$\vec{y}_{i,j} = \mathbf{A}_{i,j} \vec{y} = \mathbf{A}_{i,j} \Phi^T \vec{x}. \quad (12)$$

In this way we see that each low-resolution patch captures information regarding the full high-resolution image. Next, we identify the relationship between the low-

and high-resolution patches by

$$\vec{y}_{i,j} = \mathbf{A}_{i,j} \Phi^T \vec{x} = \mathbf{A}_{i,j} \Phi^T \mathbf{B}_{i,j}^\dagger \vec{x}_{i,j} \quad (13)$$

where \dagger denotes the pseudoinverse. Finally, using the assumption that each $\vec{x}_{i,j}$ can be written sparsely with respect to an appropriate signal model we arrive at

$$\vec{y}_{i,j} = \mathbf{A}_{i,j} \Phi^T \mathbf{B}_{i,j}^\dagger \Psi_{i,j} \vec{s}_{i,j}. \quad (14)$$

Using the singular-value decomposition construction for approximating $\mathbf{B}_{i,j}^\dagger$ it can be shown that

$$\mathbf{B}_{i,j}^\dagger \Psi_{i,j} = \Psi_{i,j}. \quad (15)$$

The model in the low resolution domain takes the form

$$\vec{y}_{i,j} = \hat{\Psi}_{i,j} \vec{s}_{i,j} \quad (16)$$

where $\hat{\Psi}_{i,j} = \mathbf{A}_{i,j} \Phi^T \Psi_{i,j}$ is rank-deficient and $\vec{s}_{i,j}$ is sparse. The columns of $\hat{\Psi}_{i,j} = [\hat{\psi}_1, \hat{\psi}_2, \dots, \hat{\psi}_D]$ are therefore a model for the low-resolution, observed patch.

Given this model in the low-resolution space we are now in a prime setting to utilize the CoSaMP framework which was conceptually described in Section 2.1. Here the dictionary is comprised of the vectorized, zero-padded, high-resolution dictionary patches, i.e. $\Psi_{i,j}$ in the above formulation. The sensed dictionary is $F(\Psi_{i,j}) = \hat{\Psi}_{i,j}$. The corresponding high-resolution patch will be sparsely represented with respect to the corresponding high-resolution dictionary atoms. Pseudocode for PAIR is given in Algorithm 1 and is described below.

A hat over a variable indicates that it is a variable in the low-dimension sensed space. The algorithm begins by initializing the residual to be represented in the sensed space, the estimate for the high-dimensional patch, and the iteration counter in Lines 3-5. In Lines 7-10 we identify the two sensed dictionary atoms that are most-correlated with the current residual as well as their

indices. The sign of the inner products between the most correlated atoms are then matched per the rationale presented in [21] in lines 11-15. The corresponding high-resolution dictionary atoms are identified in Lines 17-18. We next form the path between the two high resolution atoms (Line 19) and then pass it through the forward model (Line 20). An example of the path-augmentation phase can be seen in Figure 2. Next, the residual is compared to the sensed path atoms and we identify the most-correlated atom along the path, the weight of that atom, as well as the parameter value corresponding to it in Lines 21-24. Both the high-resolution patch and the low-resolution residual are updated using the identified weights and corresponding path sample in Lines 25-26. The PAIR algorithm concludes by stitching together the individual patches to form a global estimate (not shown in pseudocode).

4. Experiment

To evaluate the performance of PAIR we test the algorithm on a total of thirty images taken from an IR image database of ships. Examples from this data set are provided in Figure 3. Reconstructions are performed using both a Discrete Cosine Transform (DCT) dictionary and a k SVD learned dictionary trained on the high resolution IR images. We compare PAIR against CoSaMP (recall, no path-augmentation step) over several sparsity levels. Our high-resolution image patches, and high-resolution dictionary elements, are 16×16 . Patches in the observed image space are 8×8 and selected with a stride of one. The image patches are stitched together via averaging across overlap. In future work we will explore the effect of varying the patch sizes and the patch stitching algorithm.

We use three performance metrics to evaluate the algorithm. The first metric is a standard pixel-based performance metric measuring the Peak-Signal-to-Noise-Ratio (PSNR). Second, we use a local patch-based image quality metric called Structural SIMilarity (SSIM). Both PSNR and SSIM are well known and widely used. Finally, we propose to measure the global quality of the reconstruction by measuring the agreement of edge detection between the high-resolution image and the recovered images. We refer to this as the Percent Edge Agreement (PEA) metric. We compute the PEA by running an edge detection algorithm, the MATLAB implementation of Canny edge detection for the results shown, on the true high-resolution image and the reconstructed image resulting in two binary edge maps. The percentage of pixels that agree between the two edge maps is then computed and reported.

The path-augmentation step of PAIR is performed by discretely sampling along the 2-Wasserstein geodesic

Algorithm 1: Path-Augmented Image Recovery

Data: $\vec{y}_{i,j}$, the $(i,j)^{th}$ vectorized observed, low-resolution patch, $\Psi_{i,j}$ the high-resolution, zero-padded, spatially-coinciding dictionary, $\hat{\Psi}_{i,j}$ the modeled low-resolution dictionary, Φ the forward model, and K the number of iterations/sparsity level.

Result: $\tilde{x}_{i,j}$, the estimate of the $(i,j)^{th}$ vectorized, high-resolution patch

```

1 begin
2    $\vec{r}_1 \leftarrow \vec{y}_{i,j};$ 
3    $\tilde{x}_{i,j} \leftarrow \text{zeros}(1, L^2);$ 
4    $k \leftarrow 1;$ 
5    $\hat{\Omega}_{i,j} = [];$ 
6    $\Omega_{i,j} = [];$ 
7   while  $k \leq K$  do
8      $\hat{\psi}_1 = \arg \max_{\hat{\psi} \in \hat{\Psi}_{i,j}} |\langle \hat{\psi}, \vec{r}_k \rangle|;$ 
9      $d^1 = \text{index}(\hat{\psi}^1);$ 
10     $\hat{\psi}_2 = \arg \max_{\hat{\psi} \in \hat{\Psi}_{i,j} \setminus \hat{\psi}^1} |\langle \hat{\psi}, \vec{r}_k \rangle|;$ 
11     $d^2 = \text{index}(\hat{\psi}^2);$ 
12    if  $\text{sign}(\langle \hat{\psi}_1, \vec{r}_k \rangle) \neq \text{sign}(\langle \hat{\psi}_2, \vec{r}_k \rangle)$  then
13       $\hat{\psi}_2 = -\hat{\psi}_2$ 
14    else
15       $\hat{\psi}_2 = \hat{\psi}_2$ 
16    end
17    ;
18     $\psi^1 = \psi_{d_1} \in \Psi_{i,j};$ 
19     $\psi^2 = \psi_{d_2} \in \Psi_{i,j};$ 
20     $\mathbf{P}(t) = \text{path}(\psi^1, \psi^2, t);$ 
21     $t \in [0, 1]$ 
22     $\hat{\mathbf{P}}(t) = \mathbf{A}_{i,j} \Phi^T \mathbf{P}(t);$ 
23     $t_* = \arg \max_{t \in [0, 1]} \langle \hat{\mathbf{P}}(t), \vec{r}_k \rangle;$ 
24     $\hat{\psi}_*^k = \hat{\mathbf{P}}(t_*);$ 
25     $\psi_*^k = \mathbf{P}(t_*);$ 
26     $\hat{\Omega}_{i,j} \leftarrow \hat{\Omega}_{i,j} \cup \hat{\psi}_*^k;$ 
27     $\Omega_{i,j} \leftarrow \Omega_{i,j} \cup \psi_*^k;$ 
28     $\vec{s}_{i,j} \leftarrow \hat{\Omega}_{i,j}^\dagger \vec{y}_{i,j};$ 
29     $\tilde{x}_{i,j} \leftarrow \Omega_{i,j} \vec{s}_{i,j};$ 
30     $\vec{r}_{k+1} \leftarrow \vec{r}_k - \mathbf{A}_{i,j} \Phi^\top \tilde{x}_{i,j};$ 
31     $k \leftarrow k + 1;$ 
32  end
33 end

```

between the two most-representative atoms at each iteration. Algorithms for approximating the 2-Wasserstein geodesic between image patches are based on solving or

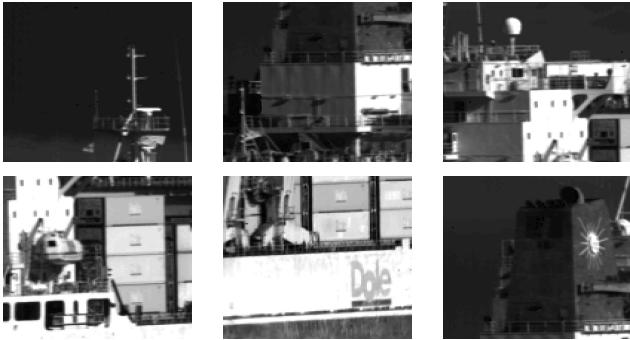


Figure 3. Examples of high-resolution IR images of boats we seek to recover from their observed, compressed counterparts (not shown). These images are included in the testing set for the results provided.

approximating a solution to the optimal transport path between two images. The results herein were generated using the Douglas-Rachford approach for solving the optimal transport problem [6] and geodesic samples selected uniformly with respect the parameterization of the path.

5. Results

The proposed PAIR algorithm is able to outperform standard CoSaMP with respect to the three performance metrics (PSNR, SSIM, and PEA) described in Section 4. Performance as a function of the sparsity of each patch’s reconstruction are shown in Figure 4. For PSNR and SSIM we report the average difference in Figure 4 and the average overall performance values in Table 1. Results shown are averaged over thirty images and the error bar correspond to one standard deviation of the performance. We report performance for both the DCT and k SVD dictionaries.

The PAIR approach achieves consistently higher values over all sparsity levels and across both dictionary types when compared using PSNR. Specifically, we see maximum performance gains of $\approx 0.5dB$ and $\approx 1.1dB$ for PAIR over CoSaMP using learned and DCT dictionaries, respectively. Interestingly the performance gain steadily grows (though not rapidly after the first few atoms) as more atoms are included in the reconstructions for the learned dictionary while the maximal performance difference occurs when fewer atoms are included and then decreases as more atoms are added despite still achieving higher overall PSNRs (reasons for which are discussed below).

The local region performance metric, SSIM, still shows improvement when using PAIR over CoSaMP. As was the case for PSNR, the SSIM difference between PAIR and CoSaMP grows and is maintained as more

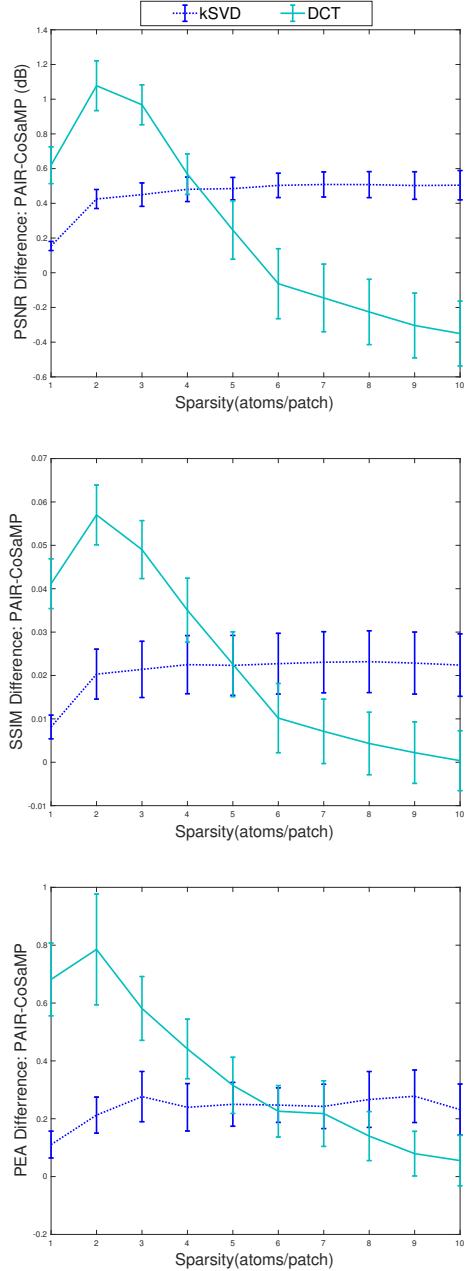


Figure 4. Average change in PSNR (top), SSIM (middle), and PEA performance (bottom) between PAIR and CoSaMP over the 30 IR boat images as a function of the number of atoms used for reconstruction of each patch in the image. Average base values for each metric are shown in Table 1.

atoms are added for the k SVD dictionary whereas diminishing returns are seen for the DCT dictionary. Our global performance metric appears to more strongly differentiate the overall quality of reconstruction obtained using a k SVD dictionary over the DCT dictionary.

kSVD/DCT	CoSaMP			PAIR		
	Atoms	PSNR	SSIM	PEA	PSNR	SSIM
1	29.38/25.24	0.75/0.61	96.38/93.57	29.53 /25.86	0.76 /0.65	96.49 /94.25
2	29.57/26.08	0.76/0.65	96.43/94.17	29.99 /27.15	0.78 /0.71	96.65 /94.96
3	29.69/26.74	0.76/0.68	96.51/94.65	30.14 /27.71	0.78 /0.73	96.70 /95.23
4	29.78/27.40	0.77/0.71	96.53/94.94	30.26 /27.97	0.79 /0.74	96.75 /95.38
5	29.87/27.94	0.77/0.73	96.53/95.15	30.36 /28.19	0.80 /0.75	96.78 /95.45
6	29.96/28.43	0.77/0.75	96.59/95.29	30.46 /28.37	0.80 /0.76	96.78 /95.52
7	30.03/28.69	0.78/0.76	96.60/95.35	30.54 /28.54	0.80 /0.77	96.83 /95.57
8	30.11/28.87	0.78/0.77	96.60/95.48	30.62 /29.65	0.80 /0.77	96.86 /95.62
9	30.20/29.05	0.78/0.77	96.63/95.61	30.70 /29.75	0.81 /0.78	96.91 /95.68
10	30.29/29.18	0.79/0.78	96.70/95.67	30.80 /29.83	0.81 /0.78	96.93 /95.73

Table 1. Average performances for PAIR and CoSaMP using either the DCT or k SVD dictionary at varying sparsity levels.

6. Discussion

Our computational experiments indicate that augmenting the CoSaMP algorithm with a search along the 2-Wasserstein geodesic at each analysis iteration always improves performance at when low numbers of atoms are used regardless of which dictionary is used for the reconstruction. As more atoms are added to the reconstruction there are diminishing returns from the augmentation, especially for the DCT dictionary. We hypothesize that the reason DCT improvement peaks and then diminishes as shown in Figure 4 is because the residual is characterized by information with higher spatial frequencies at higher iterations and that the 2-Wasserstein geodesic ceases to capture meaningful relationships between the two most-correlated atoms at those iterations.

There is, however, clear improvement for the DCT dictionary when k is small—a useful result in the case of limited HR training images. The fact that the largest relative improvements occur when very few atoms have been included in the reconstruction reduces the computational burden of repeated MP iterations. Not surprisingly, the results show that the learned dictionary is generally a better signal model for the tested IR scenes when compared to the DCT dictionary, however, we don’t expect to have a learned dictionary given our operational assumption of limited training data.

A number of avenues exist for improving our reported results. In particular, we use a naïve stitching algorithm that averages each reconstructed patch and blurs the reconstruction. More sophisticated algorithms that constrain the stitching according to image quality metrics such as edge preservation may yield improvements over our results. In addition, we have not considered the effects of patch size or stride (degree of overlap)—two variables that could strongly affect the reconstruction and overall quality of the stitching algorithm. Patch size also affects the estimated geodesic and should be

considered more closely in future work. Finally, additional performance improvements might arise from optimizing the mask employed in the CS architecture.

References

- [1] E. Agustsson and R. Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *IEEE Conference on Computer Vision and Pattern Recognition*, July 2017.
- [2] M. Aharon, M. Elad, and A. Bruckstein. rmk -SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. on Signal Processing*, 54(11):4311–4322, 2006.
- [3] R. Baraniuk. Compressive sensing. *IEEE Signal Processing Magazine*, 24(4):118–121, 2007.
- [4] R. G. Baraniuk, M. Davenport, R. A. DeVore, and M. B. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approx.*, 28(3):253–263, 2008.
- [5] S. Basu, S. Kolouri, and G. K. Rohde. Detecting and visualizing cell phenotype differences from microscopy images using transport-based morphometry. *Proc. of the National Academy of Sciences*, 111(9):3448–3453, 2014.
- [6] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.
- [7] F. Bergeaud and S. Mallat. Matching pursuit of images. In *Int. Conference on Image Processing*, volume 1, pages 53–56. IEEE, 1995.
- [8] P. Boufounos, V. Cevher, A. C. Gilbert, Y. Li, and M. J. Strauss. What’s the frequency, Kenneth?: Sub-linear Fourier sampling off the grid. *Lecture Notes in Computer Science*, 7408:61–72, 2012.
- [9] T. T. Cai and L. Wang. Orthogonal matching pursuit for sparse signal recovery with noise. *IEEE Trans. on Information theory*, 57(7):4680–4688, 2011.
- [10] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *C. R. Acad. Sci. Paris, Ser. I* 346, pages 589–592, 2008.

[11] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. on Information Theory*, 52(2):489–509, 2006.

[12] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Applied Mathematics*, 59(8):1207–1223, 2006.

[13] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. on Information Theory*, 51(12):4203–4215, 2005.

[14] Y. Chi, L. Scharf, A. Pezeshki, and A. R. Calderbank. Sensitivity to basis mismatch in compressed sensing. *IEEE Trans. on Signal Processing*, 59(5):2182–2195, 2011.

[15] R. R. Coifman and M. V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Trans. on Information Theory*, 38(2):713–718, 1992.

[16] W. Dai and O. Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Trans. on Information Theory*, 55(5):2230–2249, 2009.

[17] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In *Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision ECCV 2014. Lecture Notes in Computer Science, vol 8692*. Springer, Cham, 2014.

[18] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 38(2):295–307, 2016.

[19] D. L. Donoho. Compressed sensing. *IEEE Trans. on Information Theory*, 52(4):1289–1306, 2006.

[20] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck. Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit. *IEEE Trans. on Information Theory*, 58(2):1094–1121, 2012.

[21] T. Doster, T. Emerson, and C. Olson. Path orthogonal matching pursuit for sparse reconstruction and denoising of SWIR maritime imagery. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1161–1168. IEEE, 2018.

[22] C. Ekanadham, D. Tranchina, and E. P. Simoncelli. Recovery of sparse translation-invariant signals with continuous basis pursuit. *IEEE Trans. on Signal Processing*, 59(10):4735–4744, 2011.

[23] T. H. Emerson, T. Doster, and C. Olson. Path orthogonal matching pursuit for k-sparse image reconstruction. In *2018 26th European Signal Processing Conference*, Rome, Italy, September 2018.

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

[25] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In *IEEE Int. Conference on Computer Vision*, pages 349–356, Kyoto, 2009.

[26] R. Gribonval and M. Nielsen. Sparse representations in unions of bases. *IEEE Trans. on Information Theory*, 49(12):3320–3325, 2003.

[27] S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent. Optimal mass transport for registration and warping. *Int. Journal of Computer Vision*, 60(3):225–240, 2004.

[28] J. Haupt and R. Nowak. Signal reconstruction from noisy random projections. *IEEE Trans. on Information Theory*, 52(9):4036–4048, 2006.

[29] S. Kolouri, S. Park, M. Thorpe, D. Slepčev, and G. K. Rohde. Transport-based analysis, modeling, and learning from signal and data distributions. *preprint*, arXiv:1609.04767, 2017.

[30] S. Kolouri and G. K. Rohde. Transport-based single frame super resolution of very low resolution face images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4876–4884, 2015.

[31] S. Kwon, J. Wang, and B. Shim. Multipath matching pursuit. *IEEE Trans. on Information Theory*, 60(5):2986–3001, 2014.

[32] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. on Signal Processing*, 41(12):3397–3415, 1993.

[33] K. Nasrollahi and T. B. Moeslund. Super-resolution: a comprehensive survey. *Machine Vision and Applications*, 25:1423–1468, 2014.

[34] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, 2009.

[35] D. Needell and R. Vershynin. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. *Foundations of Computational Mathematics*, 9(3):317–334, 2009.

[36] J. M. Nichols, A. K. Oh, and R. M. Willett. Reducing basis mismatch in harmonic signal recovery via alternating convex search. *IEEE Signal Processing Letters*, 21(8):1007–1011, 2014.

[37] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Proc. of the 27th Asilomar Conference on Signals, Systems and Computers*, pages 40–44. IEEE, 1993.

[38] T. Peleg and M. Elad. A statistical prediction model based on sparse representations for single image super-resolution. *IEEE Trans. on Image Processing*, 23(6):2569–2582, 2014.

[39] N. Rao, P. Shah, S. Wright, and R. Nowak. A greedy forward-backward algorithm for atomic norm constrained minimization. In *Int. Conference on Acoustics, Speech and Signal Processing*, pages 5885–5889. IEEE, 2013.

[40] J. Romberg. Imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):14–20, 2008.

[41] J. Romberg. Compressive sensing by random convolution. *SIAM Journal of Imaging Science*, 2(4):1098–1128, 2009.

- [42] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht. Compressive sensing off the grid. In *50th Annual Allerton Conference on Communication, Control, and Computing*, pages 778–785. IEEE, 2012.
- [43] J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. on Information Theory*, 53(12):4655–4666, 2007.
- [44] C. Villani. *Optimal Transport: Old and New*, volume 338. Springer Science & Business Media, 2008.
- [45] J. Wang, S. Kwon, and B. Shim. Generalized orthogonal matching pursuit. *IEEE Trans. on Signal Processing*, 60(12):6202–6216, 2012.
- [46] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Trans. on Image Processing*, 19(11):2861–2873, 2010.