# ON NONLOCAL IMAGE COMPLETION USING AN ENSEMBLE OF DICTIONARIES

*Ehsan Miandji* and *Jonas Unger*

Department of Science and Technology, Linköping University, Sweden

## ABSTRACT

In this paper we consider the problem of nonlocal image completion from random measurements and using an ensemble of dictionaries. Utilizing recent advances in the field of compressed sensing, we derive conditions under which one can uniquely recover an incomplete image with overwhelming probability. The theoretical results are complemented by numerical simulations using various ensembles of analytical and training-based dictionaries.

*Index Terms*— compressed sensing, image completion, nonlocal, inverse problems, uniqueness conditions

## 1. INTRODUCTION

Nonlocal image representations have attracted a lot of attention in the image processing community for solving inverse problems such as inpainting, denoising, and deblurring. The success of state-of-the-art methods such as BM3D [1], LSSC [2], PLE [3], SAIST [4], and their variants, can mainly be attributed to the utilization of such a model. Nonlocal image representations exploit the intrinsic self-similarity of natural images to cluster image patches into groups, each with an associated dictionary. In this manner, images can be represented by an ensemble of dictionaries and inverse problems are solved locally in the transformation domain. These models are closely related to structured sparsity in the Compressed Sensing (CS) literature [5, 6, 7]. In Figure 1 we compare the nonlocal image completion method presented in [8] with similar methods using *one* dictionary. These results, also supported by [1, 2, 3, 4], show the advantage of nonlocal approaches. However, it is not clear whether one can exactly recover an incomplete image using nonlocal models.

In this paper, we derive uniqueness conditions for the problem of nonlocal image completion. In other words, we answer the question that whether an incomplete image can be exactly recovered using a nonlocal sparse model. We assume that the incomplete input image is a random sub-sampling of a natural image. The presented analysis provides upper bounds for sparsity and probability of exact recovery by considering the structure of a dictionary ensemble and the number of measurements. We believe that the theoretical results of this paper provide insight into training or selecting suitable
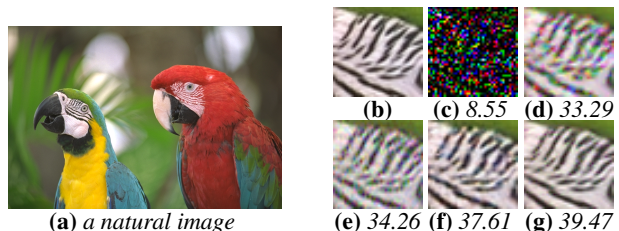
**(a)** *a natural image*  **(b)**  **(c)** *8.55*  **(d)** *33.29*  **(e)** *34.26*  **(f)** *37.61*  **(g)** *39.47*

**Fig. 1**: **(b)** reference, **(c)** the incomplete patch with $75\%$ of pixels removed randomly. **(d)-(g)** show the reconstruction of **(c)** obtained by: **(d)** linear interpolation, **(e)** a two-times overcomplete DCT dictionary, **(f)** a two-times overcomplete K-SVD dictionary [11], **(g)** an ensemble of dictionaries as presented in [8]. Numbers represent PSNR.

dictionaries, as well as required number of measurements for the problem of nonlocal image completion.

**Notations** - The $n$th column and row of a matrix, X, are denoted $X_{.,n}$ and $X_{n,.}$, respectively. Given an index set, $I$, the sub-matrix $X_{.,I}$ is formed from columns of X indexed by $I$.

## 2. PROBLEM DEFINITION

Compressed sensing (CS) [9] is a relatively new field in signal processing. In essence, CS states that if a signal is sufficiently sparse in a dictionary, it can be reconstructed exactly from only a few measurements, much less that what is required by the Nyquist criterion (see [10] for a tutorial on this topic). Consider a signal vector, $p \in \mathbb{R}^n$, and define a linear sampling operator, $\Phi \in \mathbb{R}^{m \times n}$, $m < n$, operating on p, i.e. $\hat{p} = \Phi p$. The matrix $\Phi$ is called a *measurement* or *sensing* matrix. We now define a sparse signal:

**Definition 1.** *A signal,* $p \in \mathbb{R}^n$, *is called* $\tau$-sparse *under a dictionary* D *if* $p = Dx$ *and* $\|x\|_0 = \tau \ll n$. *The* support *of* x, *denoted* $\text{supp}(x)$, *is the index set of non-zero elements.*

Assuming sparsity of p, the measurement model using the above definition becomes $\hat{p} = \Phi D x$. Recovering the coefficients, x, involves solving the following problem:

$$\hat{x} = \min_x \|x\|_0 \quad \text{s.t.} \quad \hat{p} = \Phi D x. \tag{1}$$

According to (1), uniqueness of the solution implies that it is also the most sparse one, i.e. the global minimizer. Various conditions have been proposed for the sensing matrix and the dictionary to achieve uniqueness for (1). We will discuss these conditions in Section 3.

The problem of image completion requires a special sensing matrix, which we define here. Given an index set $\{1, \ldots, n\}$, associate with each index a vector $e_i \in \mathbb{R}^n$,

where $(e_i)_j = 0$, $\forall i \neq j$, and $(e_i)_j = 1$, if $i = j$ (i.e. a canonical basis vector). Let $\{k_1, \ldots, k_n\}$ be a uniform random permutation of the index set $\{1, \ldots, n\}$.

**Definition 2.** *The sensing matrix* $I_{\Lambda,.} \in \mathbb{R}^{m \times n}$, $\Lambda = \{k_1, \ldots, k_m\}$, *called a spike ensemble, is constructed by stacking* $\{e_i\}_{i=k_1}^{k_m}$ *as rows of a matrix.*

The permutation is called uniform since all $n!$ permutations happen with equal probability. The operator $I_{\Lambda,.}$ can be interpreted as a uniform sampling without replacement too. A clear advantage of the spike ensemble is simplicity of hardware implementation. For instance, a light field camera using a binary mask was proposed in [12].

Using the spike ensemble, the problem of nonlocal image completion can be formulated as:

$$\hat{x} = \min_x \|x\|_0 \quad \text{s.t.} \quad \hat{p} = I_{\Lambda,.} D_k x, \tag{2}$$

where $\Psi = \{D_k\}_{k=1}^K$ is an ensemble of dictionaries and the signal is sparse at least in one of the dictionaries. Since no information on the underlying signal is available, one has to solve (2) for all the dictionaries in the ensemble and choose the "best" result, e.g. the most sparse one. However, when the signal is equally sparse in two or more dictionaries, uniqueness of the solution cannot be guaranteed. Our analysis resolves this issue for an spike ensemble, see section 3.

## 3. UNIQUENESS

In this section, we first introduce the challenges associated with the uniqueness of (2) and then present our main result. As it can be seen in (2), we consider noiseless images and leave further analysis in presence of noise to future work.

The problem of uniqueness for an ensemble of dictionaries is indeed two-fold. First, equation (2) should have a unique solution for every dictionary in the ensemble, which leads to $K$ solutions. Second, additional conditions have to be met for selecting *one* solution among $K$ solutions. Uniqueness for one dictionary has been extensively studied. For instance, the solution of (2) is unique for a dictionary $D_k$ if [13]:

$$\|\hat{x}\|_0 = \tau < \frac{1}{2}\text{spark}(\Phi D_k), \tag{3}$$

where spark is defined as a smallest number of linearly dependent columns of a matrix. Although the obtained result is by definition sharp [13], computing spark is a combinatorial process and often intractable. A popular alternative is mutual coherence [14], defined for an arbitrary matrix $X \in \mathbb{R}^{a \times b}$ as:

$$\mu(X) = \max_{1 \leq i \neq j \leq b} \frac{|X_{.,i}^T X_{.,j}|}{\|X_{.,i}\|_2 \|X_{.,j}\|_2}. \tag{4}$$

Using mutual coherence, if

$$\|\hat{x}\|_0 = \tau < \frac{1}{2}\left(1 + \frac{1}{\mu(\Phi D_k)}\right), \tag{5}$$

then $\hat{x}$ is the unique solution of (2) for a fixed dictionary $D_k$, see [15]. Although mutual coherence does not provide sharp bounds, it is computationally efficient. Note that in practice, one solves the convex relaxation of (2) using the $\ell_1$ norm.

In this case, if (5) holds for the solution obtained from the convex relaxation, then it is the true solution of (2).

Let us assume that (3) or (5) is satisfied for all $D_k \in \Psi$. Solving (2) for all the dictionaries leads to coefficient vectors $\{\hat{x}_a\}_{a=1}^K$. However, the uniqueness cannot be guaranteed when an image patch is equally sparse in two or more dictionaries. To setup the notation, assume that the patch p can be equally sparse in two dictionaries, say $A \in \Psi$ and $B \in \Psi$; i.e. there exist two vectors $x_1$ and $x_2$ such that $p = Ax_1$ and $p = Bx_2$, where $\|x_1\|_0 = \|x_2\|_0 = \tau$.

Our analysis is based on the Blind Compressed Sensing (BCS) framework [16], which defines the conditions for obtaining a unique solution given a finite set of orthonormal dictionaries. Instead of analyzing the uniqueness of coefficients obtained from (2), BCS considers the uniqueness of the signal p. To elaborate on this, assume that $p_1 = Ax_1$ is a solution of (2), where $\|x_1\|_0 \leq \tau$ and $\hat{p} = \Phi p_1$. The BCS uniqueness implies that there is no $p_2 = Bx_2$, such that $\|x_2\|_0 \leq \tau$ and $\hat{p} = \Phi p_2$. To ensure this, the authors define a condition on the sensing matrix $\Phi$ for uniqueness. Formally,

**Definition 3.** *A sensing matrix* $\Phi$ *is* $\tau$-*rank preserving of the dictionary ensemble* $\Psi$ *if*

$$\text{rank}(\Phi[A_{.,I}, B_{.,J}]) = \text{rank}([A_{.,I}, B_{.,J}]), \tag{6}$$

*for any* $A \in \Psi$ *and* $B \in \Psi$, *and any index set I and J, defining the support, with cardinality* $\tau$.

The index sets $I$ and $J$ define the support of the signal in A and B, respectively; i.e. we have $I = \text{supp}(x_1)$ and $J = \text{supp}(x_2)$, where $p = Ax_1$ and $p = Bx_2$. The following theorem establishes the conditions for global uniqueness of (2) using Definition 3.

**Theorem 1** (Blind CS [16]). *Assume that* (3) *or* (5) *is satisfied for all dictionaries in* $\Psi$. *If* $\Phi$ *is* $\tau$-*rank preserving of any two dictionary in* $\Psi$, *then solving* (2) *for all the dictionaries in* $\Psi$ *and choosing the sparsest solution, denoted* $x_*$, *gives a unique signal* $p_* = D_* x_*$, *where* $D_*$ *is the dictionary corresponding to the* $x_*$.

The proof can be found in [16]. We observe that the key aspect of uniqueness is (6). Gleichman and Eldar [16] show that an i.i.d. Gaussian sensing matrix is $\tau$-rank preserving for any finite set of orthonormal dictionaries with probability 1 when $m \geq 2\tau$. The proof of this result follows trivially by observing that $\text{spark}(\Phi) = m + 1$, see [16]. Unfortunately, the spike ensemble $I_{\Lambda,.}$ does not satisfy this property. Although $I_{\Lambda,.}$ is full-rank, the spark is not maximal.

In this paper, we show that the spike ensemble is $\tau$-rank preserving of a dictionary ensemble with high probability if the signal is sufficiently sparse. Unlike [16], our analysis does not assume orthonormal dictionary ensembles; we also consider overcomplete dictionaries. Although overcomplete dictionary ensembles have not been utilized in practice, our numerical results show superiority of such ensembles.

When dictionaries in $\Psi$ are orthonormal, we observe that even though $A_{.,I}$ and $B_{.,J}$ have orthonormal columns, $[A_{.,I}, B_{.,J}]$ is not necessarily orthonormal. Form the Gram

matrix of $[A_{.,I}, B_{.,J}]$:

$$W = [A_{.,I}, B_{.,J}]^T [A_{.,I}, B_{.,J}] \tag{7}$$

$$= \begin{pmatrix} A_{.,I}^T A_{.,I} = I & A_{.,I}^T B_{.,J} \\ (A_{.,I}^T B_{.,J})^T & B_{.,J}^T B_{.,J} = I \end{pmatrix}, \tag{8}$$

implying that $W = I$ only when $A_I^T B_J = 0$. Since (6) is a condition on rank, it is more convenient to transform $[A_{.,I}, B_{.,J}]$ into a matrix $G \in \mathbb{R}^{n \times r}$ with orthonormal columns, such that $\text{rank}([A_{.,I}, B_{.,J}]) = \text{rank}(G) = r$ and $G^T G = I$, where $I$ is the identity matrix. One can apply a Gram-Schmidt process on the columns of $[A_{.,I}, B_{.,J}]$ to get $G$. Therefore, without loss of generality, our goal is to verify:

$$\text{rank}(I_{\Lambda,.} G) = \text{rank}(G) = r, \tag{9}$$

instead of (6). Note that $r$ does not need to be known, but we include it to clarify the dimensionality of $G$. Moreover, when the dictionaries in $\Psi$ are overcomplete, (3) implies that every $\tau$ column in $D_k$ forms a linear subspace of dimensionality $\tau$ [17]. Therefore the same procedure can be used to form $G$, and we can verify (9) instead of (6). The following theorem establishes the conditions for a spike ensemble to be $\tau$-rank preserving.

**Theorem 2.** *Under the setting described above, if:*

$$m \geq \left\lceil -n\alpha \log(\frac{1}{r}) \right\rceil, \tag{10}$$

*where*

$$\alpha = \max_{j=1,\ldots,n} \|G_{j,.}\|_2^2, \tag{11}$$

*then with probability at least:*

$$1 - r\,(0.3679)^{\frac{m}{n\alpha}}, \tag{12}$$

*the spike ensemble* $I_{\Lambda,.}$ *is* $\tau$-*rank preserving of the dictionary ensemble* $\Psi$.

*Proof.* Since $G$ is full rank, it suffices to show that $I_{\Lambda,.} G$ is also full rank. The proof follows from the Matrix Chernoff inequality [18], therefore we verify the conditions of this theorem. Expand the matrix $I_{\Lambda,.}$ by placing zero vectors at rows $\{1,\ldots,n\} \setminus \Lambda$, and denote the resulting matrix $\hat{I}_{\Lambda,.}$. It can be seen that $I_{\Lambda,.} G = (\hat{I}_{\Lambda,.} G)_{\Lambda,.}$. Expanding the Gram matrix of $\hat{I}_{\Lambda,.} G$, denoted $Y$, we have:

$$Y = (\hat{I}_{\Lambda,.} G)^T (\hat{I}_{\Lambda,.} G) = \sum_{i=1}^{n} G^T \hat{I}_{i,.}^T \hat{I}_{i,.} G = \sum_{j \in \Lambda} \underbrace{G_{j,.}^T G_{j,.}}_{X_j}. \tag{13}$$

The matrices $X_j$, $j \in \Lambda$, are symmetric and positive semidefinite. Recall that $\Lambda$ is $m$ first indices obtained from a uniform random permutation of the index set $\{1,\ldots,n\}$. We can assume that the matrices $X_j$, $j \in \Lambda$, are randomly sampled, without replacement, from a finite set of positive semidefinite matrices:

$$\mathcal{X} = \{X_j = G_{j,.}^T G_{j,.} \mid j = 1,\ldots,n\}. \tag{14}$$

Therefore, $Y$ is a sum of random positive semidefinite matrices. We have that

$$\max_{j \in \Lambda} \lambda_{\max}(X_j) = \max_{j \in \Lambda} \|G_{j,.}\|_2^2 \leq \alpha. \tag{15}$$

In addition,

$$\mathbb{E}\{X_1\} = \frac{1}{n} \sum_{j=1}^{n} G_{j,.}^T G_{j,.} = \frac{1}{n} G^T G = \frac{1}{n} I, \tag{16}$$

and therefore

$$\lambda_{\max}(\mathbb{E}\{X_1\}) = \lambda_{\min}(\mathbb{E}\{X_1\}) = \frac{1}{n}. \tag{17}$$

Using (16) and (17), we define the following quantity used in the matrix Chernoff theorem:

$$\omega_{\min} \equiv \lambda_{\min}\left(\sum_{j \in \Lambda} \mathbb{E}\{X_j\}\right) \leq m\lambda_{\min}(\mathbb{E}\{X_1\}) = \frac{m}{n} \tag{18}$$

We can now invoke Corollary 5.2 in [18], which states that given (15) and (18), for a finite sequence of independent random positive semidefinite matrices $X_j \in \mathbb{R}^{r \times r}$, $j \in \Lambda$, and for any $0 \leq \delta < 1$ we have:

$$\mathbb{P}\left\{\lambda_{\min}\left(\sum_{j \in \Lambda} X_j\right) \leq (1-\delta)\omega_{\min}\right\} \leq r\left[\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right]^{\frac{\omega_{\min}}{\alpha}}, \tag{19}$$

Substituting (18) into (19) and inverting the probability event we get:

$$\mathbb{P}\left\{\lambda_{\min}\left(\sum_{j \in \Lambda} X_j\right) \geq \frac{m}{n}(1-\delta)\right\} \geq 1 - r\left[\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right]^{\frac{m}{n\alpha}}. \tag{20}$$

Since we only need to show that:

$$\lambda_{\min}\left(\sum_{j \in \Lambda} X_j\right) > 0, \tag{21}$$

we set $\delta$ such that $1 - \delta$ is close to zero. Therefore, the right-hand side of (20) becomes, approximately, $1 - r\,(0.3679)^{\frac{m}{n\alpha}}$. This completes the proof. ∎

A similar analysis was presented in [19], however, for bounding the condition number of a matrix with orthonormal columns. In contrast, our goal is to verify (9) for blind-CS and the uniqueness of (2). In addition, observe that Theorem 2 only states a necessary condition for exact recovery and one has to also verify (3) or (5), or a similar condition for uniqueness of (2) using one dictionary, e.g. RIP [20].
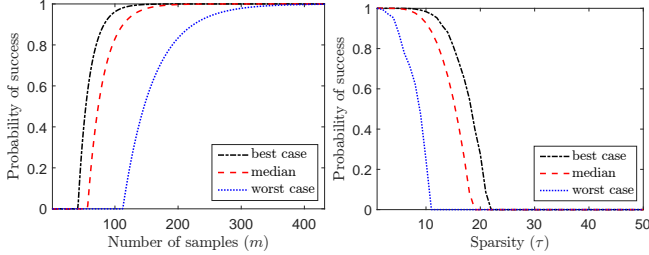
## 4. NUMERICAL RESULTS

To evaluate Theorem 2, we report numerical results for three case studies associated with different dictionary ensembles. We consider two signal lengths, $n = 12 \times 12 \times 3$ and $n = 9 \times 9 \times 5 \times 5 \times 3$, which are commonly used patch sizes for color images [1, 8, 21], and 4D lightfields [8, 22], respectively. The sparsity for image and lightfield patches are assumed to be $\tau = 5$ and $\tau = 10$, respectively; these values are typically used by dictionary training algorithms [11, 8, 22].

The probability of success for exact recovery, using Theorem 2, is shown in Figures 2-4. To compute this quantity, we perform $10^5$ trials, where in each trial a spike ensemble
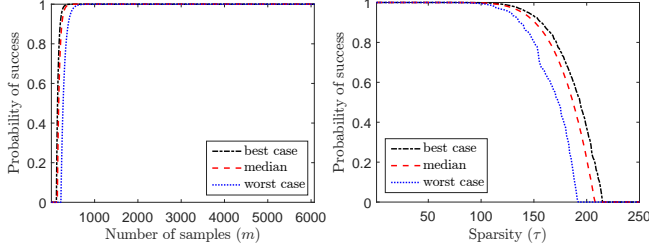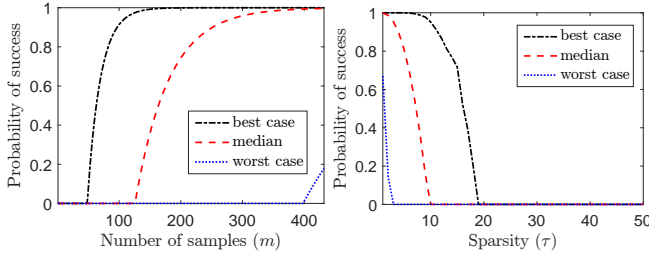
**(a)** $n = 25 \times 81 \times 3$, $\tau = 10$    **(b)** $n = 25 \times 81 \times 3$, $m = n/2$
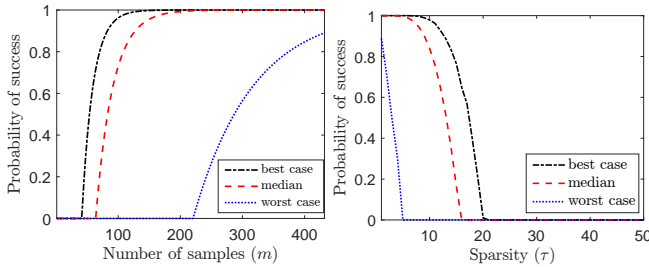
**Fig. 2**: Evaluation of Theorem 2 using an ensemble with a DCT dictionary and orthonormal dictionaries constructed from random data. **(a)**-**(b)** image completion and **(c)**-**(d)** 4D lightfield completion



**(c)** $n = 12 \times 12 \times 3$, $\tau = 5$    **(d)** $n = 12 \times 12 \times 3$, $m = n/2$



**(a)** $n = 12 \times 12 \times 3$, $\tau = 18$    **(b)** $n = 12 \times 12 \times 3$, $m = n/2$

**Fig. 3**: Evaluation of Theorem 2 using E2DOD.



**(a)** $n = 12 \times 12 \times 3$, $\tau = 18$    **(b)** $n = 12 \times 12 \times 3$, $m = n/2$

**Fig. 4**: Evaluation of Theorem 2 using a K-SVD ensemble.

is constructed as described in Definition 2. Moreover, each trial constructs a uniformly random support in each dictionary, denoted $I$ and $J$ for two dictionaries in Definition 3; i.e. we assume that the signal is equally sparse in each dictionary, however, with different support (a weaker assumption than group-sparsity [2]). Since Theorem 2 is only a necessary condition, we set the probability of success to zero for trials that do not satisfy the mutual coherence condition in (5).

For the first case study, we construct an ensemble with:

1. A DCT dictionary and 2. A basis for column space of randomly generated image patches stacked as columns of a matrix. Random image patches are constructed by sampling uniformly from the interval $[0, 255]$. As described in section 3, we first create two sub-dictionaries using the random support sets, $I$ and $J$. Then a QR decomposition is applied to the concatenated dictionaries, obtaining the orthonormal matrix G. Results are summarized in Figure 2, where we plot (12) against the number of measurements and sparsity for both signal lengths. Each plot reports the best, median, and worst values for the probability of success among all trials. It can be seen that a 10-sparse lightfield can be exactly recovered with as low as $5\%$ of measurements with probability at least 0.98. For a 5-sparse image, median of the results suggest that with $35\%$ of measurements one can achieve exact reconstruction with the same probability.

Gurumoorthy et al. [21] present a method for learning an ensemble of 2D orthonormal dictionaries (E2DOD) with applications in image compression. The ensemble, $\Psi = \{U_k, V_k\}_{k=1}^K$, is trained such that an image patch can be represented as $P_i = U_k S_{ik} V_k^T$, where $S_{ik}$ is a sparse coefficient matrix. A CS-based method was proposed in [8] for image completion using E2DOD. The authors form the ensemble $\Psi = \{V_k \otimes U_k\}_{k=1}^K$ to solve an $\ell_1$ recovery problem, where $\otimes$ denotes the Kronecker product. In this paper we analyze the exact recovery conditions for E2DOD, shown in Figure 3. We train an ensemble of 16 dictionaries with sparsity $\tau = 5$ on a set of natural images. In each trial, two dictionaries are randomly selected for evaluating Theorem 2. The same procedure is used, as described earlier, for calculating the probability of success.

The third ensemble consists of five K-SVD dictionaries trained on different sets of natural images. The sparsity was set to 5 and, similar to E2DOD, we randomly select two dictionaries in each trial. Results are summarized in Figure 4. Although E2DOD typically performs better than one K-SVD dictionary [8], our results show that a K-SVD ensemble can outperform E2DOD. This suggests new research directions for dictionary training as overcomplete dictionary ensembles have not been utilized in practice. The inferior performance of E2DOD can be associated to the coherence among dictionaries, which can be resolved by improving the clustering stage of the training.

## 5. CONCLUSIONS

We presented the uniqueness conditions for recovering incomplete images using an ensemble of dictionaries. We believe that our theoretical results can be used for designing efficient dictionaries, as well as evaluating existing methods for image completion. We would like to extend our work by including the effect of noise. In the presence of noise, image patches are not exactly sparse, but rather compressible. In this case, exact recovery may not be feasible; however, one can derive an upper bound for the reconstruction error.

# 6. REFERENCES

[1] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *Image Processing, IEEE Transactions on*, vol. 16, no. 8, pp. 2080–2095, Aug 2007.

[2] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Computer Vision, 2009 IEEE 12th International Conference on*, Sept 2009, pp. 2272–2279.

[3] Guoshen Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity," *Image Processing, IEEE Transactions on*, vol. 21, no. 5, pp. 2481–2499, May 2012.

[4] Weisheng Dong, Guangming Shi, and Xin Li, "Nonlocal image restoration with bilateral variance estimation: A low-rank approach," *Image Processing, IEEE Transactions on*, vol. 22, no. 2, pp. 700–711, Feb 2013.

[5] R.G. Baraniuk, V. Cevher, M.F. Duarte, and C. Hegde, "Model-based compressive sensing," *Information Theory, IEEE Transactions on*, vol. 56, no. 4, pp. 1982–2001, April 2010.

[6] Y.C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *Information Theory, IEEE Transactions on*, vol. 55, no. 11, pp. 5302–5316, Nov 2009.

[7] Y.C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3042–3054, 2010.

[8] Ehsan Miandji, Joel Kronander, and Jonas Unger, "Compressive Image Reconstruction in Reduced Union of Subspaces," *Computer Graphics Forum*, vol. 34, no. 2, pp. 33–44, May 2015.

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

[10] E.J. Candès and M.B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[11] M. Aharon, M. Elad, and A. Bruckstein, "k -svd: An algorithm for designing overcomplete dictionaries for sparse representation," *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4311–4322, Nov 2006.

[12] Gordon Wetzstein, Ivo Ihrke, and Wolfgang Heidrich, "On plenoptic multiplexing and reconstruction," *International Journal of Computer Vision*, vol. 101, no. 2, pp. 384–400, 2012.

[13] David L. Donoho and Michael Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via 1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

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

[15] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *Information Theory, IEEE Transactions on*, vol. 49, no. 12, pp. 3320–3325, Dec 2003.

[16] S. Gleichman and Y.C. Eldar, "Blind compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6958–6975, 2011.

[17] Michal Aharon, Michael Elad, and Alfred M. Bruckstein, "On the uniqueness of overcomplete dictionaries, and a practical way to retrieve them," *Linear Algebra and its Applications*, vol. 416, no. 1, pp. 48–67, 2006.

[18] JoelA. Tropp, "User-friendly tail bounds for sums of random matrices," *Foundations of Computational Mathematics*, vol. 12, no. 4, pp. 389–434, 2012.

[19] Ilse C. F. Ipsen and Thomas Wentworth, "The effect of coherence on sampling from matrices with orthonormal columns, and preconditioned least squares problems," *SIAM Journal on Matrix Analysis and Applications*, vol. 35, no. 4, pp. 1490–1520, 2014.

[20] E.J. Candes and T. Tao, "Decoding by linear programming," *Information Theory, IEEE Transactions on*, vol. 51, no. 12, pp. 4203–4215, Dec 2005.

[21] K.S. Gurumoorthy, A. Rajwade, A. Banerjee, and A. Rangarajan, "A method for compact image representation using sparse matrix and tensor projections onto exemplar orthonormal bases," *IEEE Transactions on Image Processing*, vol. 19, no. 2, pp. 322–334, 2010.

[22] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar, "Compressive Light Field Photography using Overcomplete Dictionaries and Optimized Projections," *ACM Trans. Graph. (Proc. SIGGRAPH)*, vol. 32, no. 4, pp. 1–11, 2013.