# COMPRESSED SENSING USING A GAUSSIAN SCALE MIXTURES MODEL IN WAVELET DOMAIN

*Yookyung Kim[1], Mariappan S. Nadar[2], and Ali Bilgin[3,1]*

[1]Department of Electrical and Computer Engineering,[3]Biomedical Engineering
University of Arizona, Tucson, AZ, USA 85721
[2]Siemens Corporation, Corporate Research, Princeton, NJ, USA 08540

## ABSTRACT

Compressed Sensing (CS) theory has gained attention recently as an alternative to the current paradigm of sampling followed by compression. Early CS recovery techniques operated under the implicit assumption that the transform coefficients in the sparsity domain are independently distributed. Recent works, however, demonstrated that exploiting the statistical dependencies between transform coefficients can further improve the recovery performance of CS. In this paper, we propose the use of a Gaussian Scale Mixtures (GSM) model in CS. This model can efficiently exploit the statistical dependencies between wavelet coefficients during CS recovery. The proposed model is incorporated into several recent CS techniques including Reweighted $l_1$ minimization (RL1), Iteratively Reweighted Least Squares (IRLS), and Iterative Hard Thresholding (IHT). Experimental results show that the proposed method improves reconstruction quality for a given number of measurements or requires fewer measurements for a desired reconstruction quality.

***Index Terms***— Compressed sensing, Wavelets, Gaussian scale mixtures

## 1. INTRODUCTION

The recently introduced Compressed Sensing (CS) theory provides an attractive alternative to conventional Nyquist sampling [1, 2]. CS theory demonstrated that, under certain conditions, signals with sparse representations can be reconstructed from much fewer measurements than suggested by the Nyquist theory. This is very desirable in many applications where acquiring data at the desired Nyquist rate is either expensive or impossible due to physical/hardware constraints. Although early CS methods did not explicitly exploit the dependencies that often exist between coefficients in the sparsity domain, there has been recent interest in incorporating additional prior knowledge (statistical dependencies, structure, etc.) about the transform coefficients into the CS framework [3–8]. These recent results indicate that incorporating prior knowledge into the recovery process through the use of accurate models can further improve the performance of CS methods.

In this paper, we propose to incorporate a Gaussian Scale Mixtures (GSM) model into the wavelet-based CS framework. The GSM model has been previously shown to accurately describe the non-stationary behavior of the wavelet coefficients in images [9]. A significant advantage of our proposed approach is that it is flexible enough to be incorporated into several existing CS algorithms such as Reweighted $l_1$ minimization (RL1) [10], Iteratively Reweighted Least Squares (IRLS) [11], and Iterative Hard Thresholding (IHT) [12].

The remainder of this paper is organized as follows. In the next section, we review the statistical dependencies that exist among wavelet coefficients of natural images and discuss the use of the GSM to model wavelet coefficients. In Section 3, we describe the RL1, IRLS, and IHT algorithms and illustrate how the proposed model can be incorporated into each algorithm. Experimental results are provided in Section 4, where we compare the performances of the proposed model-based CS techniques to the original versions of the algorithms. In addition, we also provide comparisons to other state-of-the-art CS algorithms, including a recent Bayesian CS approach that also exploits dependencies between wavelet coefficients. Conclusions are provided in Section 5.

## 2. MODELING WAVELET COEFFICIENTS USING GAUSSIAN SCALE MIXTURES

The wavelet transform produces efficient representations of natural images by decomposing them into multi-scale subbands having different frequencies and orientations. The wavelet coefficients at different scales corresponding to a particular spatial location can be ordered in a hierarchical quadtree structure. In this hierarchical structure, a coefficient at the coarser scale (often referred to as *parent*) is correlated with the spatially co-located coefficients at the finer scale (referred to as *children*). In addition, wavelet coefficients are also statistically dependent on their spatially adjacent neighbors within the same subband. Over the past two decades, there has been significant research in modeling and exploiting such dependencies in different applications such as image denoising [9] and image compression [13].

It can be observed empirically that the densities of wavelet coefficients of natural images are non-Gaussian with a sharper peak and heavier tails; Large magnitude coefficients are sparsely distributed within each wavelet subband and occur in clusters. The GSM model has been shown to account for the shape of the wavelet coefficient marginals as well as the correlation between magnitudes of neighboring coefficients [9]. Consider a neighborhood $\mathbf{v}$ of wavelet coefficients. This neighborhood of coefficients can be modeled as the product of a zero-mean Gaussian random variable $\mathbf{u}$ and a scalar variable $z$, i.e., $\mathbf{v} = \sqrt{z}\mathbf{u}$. This assumed GSM structure of wavelet coefficients enables accurate reproduction of the high kurtosis nature of their distribution. Now consider noisy observations of the coefficients in the neighborhood: $\mathbf{y} = \mathbf{v} + \mathbf{e}$, where $\mathbf{e}$ denotes the noise vector. If $\mathbf{e}$ is independent additive white Gaussian noise, the random variables $z$, $\mathbf{u}$ and $\mathbf{e}$ are independent. In [9], a Bayes least squares (BLS) estimate of the wavelet coefficient $v_c$ surrounded by the given neighborhood $\mathbf{y}$ was derived under these conditions as

$$\hat{v}_c = \mathrm{E}\{v_c|\mathbf{y}\} = \int_0^\infty p(z|\mathbf{y})\mathrm{E}\{v_c|\mathbf{y}, z\}dz \qquad (1)$$

This BLS-GSM estimate was then successfully employed for

wavelet-based image denoising. Our goal in this work is to use this model in CS.

## 3. BLS-GSM COMPRESSED SENSING

The emergence of the CS theory caused extensive research into developing signal recovery algorithms for solving the CS problem. Numerous algorithms were presented within the last few years [14]. Many of these CS algorithms are iterative in nature and explicitly or implicitly utilize weights during the iterations. In this section, we will review three such CS algorithms and discuss how they can be modified to incorporate the BLS-GSM model presented in the previous section. The reweighted $l_1$ minimization problem [10] is defined as

$$\mathbf{x}^k = \arg\min_{\mathbf{x}} \left\| \mathbf{W}^k \mathbf{x} \right\|_1 \text{ such that } \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 \leq \varepsilon \qquad (2)$$

where $\mathbf{x}$ and $\mathbf{b}$ denote the sparse coefficient vector and the measurement vector, respectively. $\mathbf{A}$ is the matrix that projects the sparse coefficients $\mathbf{x}$ into the measurement space, and $\mathbf{x}^k$ is the estimate of the sparse coefficients at iteration $k$. $\varepsilon$ is used to account for noise in the measurements. $\mathbf{W}^k$ is a diagonal weight matrix whose diagonal entries at iteration $k$ are defined as

$$w^k(i) = 1/|x^{k-1}(i) + \delta| \qquad (3)$$

where $\delta$ is a small value to avoid division by zero. In other words, the weights are initialized to ones and refined at each iteration as the inverse of the solution obtained at the previous iteration. We refer to this algorithm with the acronym RL1.

Daubechies *et al.* proposed the use of IRLS to solve the sparse recovery problem in CS [11]. Similar to the RL1, IRLS alternates between estimating the sparse vector $\mathbf{x}$ and redefining the weights. More specifically, estimation of the sparse vector at iteration $k$ is performed using

$$\mathbf{x}^k = \arg\min_{\mathbf{x}} \mathbf{W}^k \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 \qquad (4)$$

$\mathbf{W}^k$ is a diagonal weight matrix whose diagonal entries at iteration $k$ are defined as

$$w^k(i) = 1/\sqrt{|x^{k-1}(i)|^2 + \delta^2} \qquad (5)$$

The IHT algorithm, proposed by Blumensath and Davies [12], aims to solve the $S$-sparse problem

$$\arg\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 \text{ such that } \|\mathbf{x}\|_0 \leq S \qquad (6)$$

by using the following iterative method:

$$\mathbf{x}^k = H_S(\mathbf{x}^{k-1} + \mu \mathbf{A}^H(\mathbf{b} - \mathbf{A}\mathbf{x}^{k-1})) \qquad (7)$$

where $H_S$ is an operator retaining the $S$ largest elements and setting the remainder to zero, and $\mu$ denotes a step size. Unlike RL1 or IRLS, it is easy to see that IHT does not explicitly use weights. However, we can see that Eq. (7) can be rewritten as

$$\mathbf{x}^k = \mathbf{W}^k(\mathbf{x}^{k-1} + \mu \mathbf{A}^H(\mathbf{b} - \mathbf{A}\mathbf{x}^{k-1})) \qquad (8)$$

where the entries of the diagonal weight matrix are determined as

$$w^k(i) = \begin{cases} 1 & |g^k(i)| > 0 \\ 0 & \text{otherwise} \end{cases} \qquad (9)$$

where $\mathbf{g}^k = H_S(\mathbf{x}^{k-1} + \mu \mathbf{A}^H(\mathbf{b} - \mathbf{A}\mathbf{x}^{k-1}))$. In other words, the entries of the weight matrix corresponding to the $S$ largest elements are set to one and all other entries are set to zero.

It can be observed that all of the conventional CS techniques discussed above utilize the magnitudes of the sparse domain coefficients in an independent manner while determining the weights to be used in the next iteration. In other words, the fact that there exist statistical dependencies between sparse domain coefficients is not exploited. The key idea in this paper is that exploitation of the statistical dependencies between sparse domain coefficients can be incorporated into these conventional CS techniques using a statistical model when determining the weights. In particular, we illustrate how the BLS-GSM model discussed in the previous section can be used to calculate weights that lead to significantly improved performance in all three CS algorithms.

| | | |
|---|---|---|
| Input: | Projection matrix $\mathbf{A}$, measurements $\mathbf{b}$, constant $r(r<1)$, CS problem $CS$, Eq. for weights $WT$, BLS-GSM model $M_{BLS-GSM}$ | |
| Output: | Estimate of sparse coeffs. $\hat{\mathbf{x}}$ | |
| $k = 0, \mathbf{x}^0 = \mathbf{0}, \mathbf{W}^1 = \mathbf{I}$ | | initialize |
| $\sigma = STD(\text{HH1})$ | | calculate $\sigma$ |
| `do` | | |
| 1. $k \leftarrow k+1$ | | increment iteration index |
| 2. $\mathbf{x}^k \leftarrow CS(\mathbf{A}, \mathbf{b}, \mathbf{x}^{k-1}, \mathbf{W}^k)$ | | estimate sparse coeffs. |
| 3. $\tilde{\mathbf{x}} \leftarrow M_{BLS-GSM}(\mathbf{x}^k, \sigma)$ | | impose BLS-GSM model |
| 4. $\mathbf{W}^k \leftarrow WT(\tilde{\mathbf{x}})$ | | update weights |
| 5. $\sigma \leftarrow r\sigma$ | | decrease $\sigma$ |
| **`while`** `(halting criterion false)` | | |
| `return` $\hat{\mathbf{x}} \leftarrow \mathbf{x}^k$ | | |

**Fig. 1**. BLS-GSM CS algorithm.

As discussed in the previous section, we model a neighborhood of the reconstructed wavelet coefficients $\mathbf{y}$ as the sum of the true neighborhood $\mathbf{v}$ and a random variable $\mathbf{e}$ accounting for the remaining artifacts (i.e. measurement noise and undersampling artifacts): $\mathbf{y} = \mathbf{v} + \mathbf{e}$. Under the assumption that $\mathbf{e}$ is a zero-mean Gaussian independent random variable with variance $\sigma^2$ , an estimate of the wavelet coefficient surrounded by the given neighborhood $\mathbf{y}$ can be obtained using Eq. (1)[1]. The parameter $\sigma$ is initialized using the standard deviation of the wavelet coefficients in the highest frequency diagonal orientation subband (HH1) prior to the start of the CS recovery. Since aliasing artifacts are reduced at each CS iteration, the value of $\sigma$ is adjusted to account for this reduction. Once estimates of the wavelet coefficients are determined using the BLS-GSM model, these estimates are used to determine the weights for the CS algorithms. Thus, instead of directly using the magnitudes of the wavelet coefficients obtained in the previous iteration, the estimates of the wavelet coefficients from the BLS-GSM model are used to determine the weights for the next iteration. The pseudo-code of the proposed algorithm is presented in Fig. 1.

## 4. EXPERIMENTS

Experiments were performed to compare the proposed method to existing CS recovery algorithms: The following CS algorithms were used in the comparison: Orthogonal Matching Pursuit (OMP) [15]; Stagewise Orthogonal Matching Pursuit (StOMP) [16]; IRLS, RL1, IHT, and the Tree-Structured Wavelet Compressive Sensing (TSW-CS) [5] (It is worth emphasizing that TSW-CS is a state-of-the-art model-based CS algorithm based on a hierarchical Bayesian model and, thus, is built on similar ideas as the proposed methods). Three of these techniques (IRLS, RL1, and IHT) were modified as described in the previous section to incorporate the BLS-GSM model. The modified algorithms are referred to as BLS-GSM IRLS, BLS-GSM RL1, and BLS-GSM IHT.

For StOMP, the SparseLab toolbox [17] was used. For RL1 and

[1]While this simplified model may not accurately represent the reconstruction error, it performs very well in practice as illustrated in the next section. Refining this model through the development of a more accurate error model is the goal of our ongoing research.

BLS-GSM RL1, the NESTA toolbox [18] was used. For IRLS and BLS-GSM IRLS, the software provided by the authors of [11] on their web site was used. For OMP, IHT, and BLS-GSM IHT, the Sparsify toolbox [19] was used. The default settings were used for the parameters in all software. However, for those algorithms which required user specified values for certain parameters, we used the following: The false-alarm rate for StOMP with CFAR thresholding was set to 0.1. The values for $\mu$ and $\delta$ parameters used in NESTA were assigned to $10^{-5}$ and 0.1, respectively. The number of elements to keep, $S$, in IHT was set to one fifth of the number of measurements, and to one half of the number of measurements in BLS-GSM IHT. The reason for selecting different values of $S$ was because we observed that BLS-GSM IHT is capable of recovering more coefficients than IHT for a given set of measurements. Increasing $S$ beyond one fifth of the measurements resulted in performance degradation in IHT. In our model-based implementations, the value of $\sigma$ was initialized to the standard deviation of the (aliased) coefficients in the highest frequency diagonal subband (HH1). The value of $r$ which is used to decrease $\sigma$ at each iteration was set to 0.97. The neighborhood vector included the eight spatially adjacent neighbor coefficients as well as the parent coefficient.

We used the same experimental setup presented in the TSW-CS paper [5]: Ten images were selected randomly for each of the five image classes (flowers, cows, buildings, urban, and office) from the image database provided by Microsoft Research in Cambridge [20]. A random measurement matrix was formed by calculating an orthonormal basis from a random Gaussian matrix. All simulations are carried out using MATLAB on a PC with Intel Xeon 2.93 GHz CPU and 32GB RAM.

Table 1 summarizes the performances of different CS algorithms for 5000 measurements (which corresponds to an undersampling ratio of roughly 3.26 for the 128x128 images used in these experiments). Reconstruction accuracy was evaluated using the relative $l_2$ norm of reconstruction error defined as $\|\mathbf{x}-\hat{\mathbf{x}}\|_2/\|\mathbf{x}\|_2$, where $\mathbf{x}$ is the original image and $\hat{\mathbf{x}}$ is the reconstructed image. The means and standard deviations of the relative $l_2$ errors across ten images in each class are presented in Table 1. Images obtained using different algorithms are presented in Fig. 3. It can be observed that all three proposed algorithms (BLS-GSM IRLS, BLS-GSM RL1, and BLS-GSM IHT) had 24.9%, 13.6%, and 17.1% less average relative $l_2$ error than TSW-CS. Furthermore, the proposed model-based algorithms also achieved 42.4%, 29.9%, and 44.8% less average relative $l_2$ error compared to their original counterparts, IRLS, RL1 and IHT, respectively. In addition to these quantitative improvements, the proposed methods also provided qualitative improvements as illustrated in Fig. 3.

Figure 2 plots the relative $l_2$ norm of reconstruction errors and computation times for different number of measurements for the image shown in Fig. 3. The results in these plots were obtained by averaging the relative reconstruction errors obtained from five random Gaussian measurement matrices. As illustrated in the figure, the performances of the proposed algorithms are consistently superior to others for all numbers of measurements. In addition, BLS-GSM IHT was the fastest CS algorithm tested, averaging roughly 10 times faster than TSW-CS while providing better reconstruction results.

## 5. SUMMARY

We proposed a novel approach for exploiting the statistical dependencies between transform coefficients to improve the recovery performance of CS algorithms. The proposed approach was incorporated into three CS algorithms and improved the reconstruction performances of all algorithms significantly. Experimental results indicated that the proposed methods compared favorably to other CS algorithms, including a recent model-based algorithm that also exploits statistical dependencies among transform coefficients.

## 6. REFERENCES


[1] E. J. Candes *et al.*, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Image Process.*, vol. 52, no. 2, pp. 489–509, 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Image Process.*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] R. G. Baraniuk *et al.*, "Model-based compressive sensing," 2008, to appear in IEEE Trans. on Information Theory.

[4] M. N. Do and C. N. H. La, "Tree-based majorize-minimize algorithm for compressed sensing with sparse-tree prior," in *Computational Advances in Multi-Sensor Adaptive Processing*, 2007.

[5] L. He and L. Carin, "Exploiting structure in wavelet-based bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 57, pp. 3488–3497, 2009.

[6] S. Mun and J. E. Fowler, "Block compressed sensing of images using directional transforms," in *IEEE Int. Conf. Image Process.*, 2009, pp. 3021–3024.

[7] X. Wu *et al.*, "Model-guided adaptive recovery of compressive sensing," in *Proc. Data Compression Conf.*, 2009, pp. 123–132.

[8] E. Vera *et al.*, "Bayesian compressive sensing of wavelet coefficients using multiscale laplacian priors," in *IEEE Workshop on Statistical Signal Processing*, 2009.

[9] J. Portilla *et al.*, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, 2003.

[10] E. J. Candes *et al.*, "Enhancing sparsity by reweighted $l_1$ minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, 2008.

[11] I. Daubechies *et al.*, "Iteratively re-weighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics*, vol. 63, pp. 1–38, 2010.

[12] T. Blumensath and M. Davies, "Iterative thresholding for sparse approximations," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 629–654, 2008.

[13] D. S. Taubman and M. W. Marcellin, *JPEG2000 : Image compression fundamentals, standards, and practice*, Kluwer Academic Publishers, Boston, 2002.

[14] Compressive Sensing Resources available at http://dsp.rice.edu/cs.

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

[16] D. L. Donoho *et al.*, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," *Stanford Statistics Technical Report 2006-2*, 2006.

[17] SparseLab toolbox available at http://sparselab.stanford.edu/.

[18] NESTA toolbox available at http://acm.caltech.edu/~nesta/.

[19] Sparsify toolbox available at http://www.personal.soton.ac.uk/tb1m08/.

[20] Microsoft Research image database available at http://research.microsoft.com/en-us/projects/.

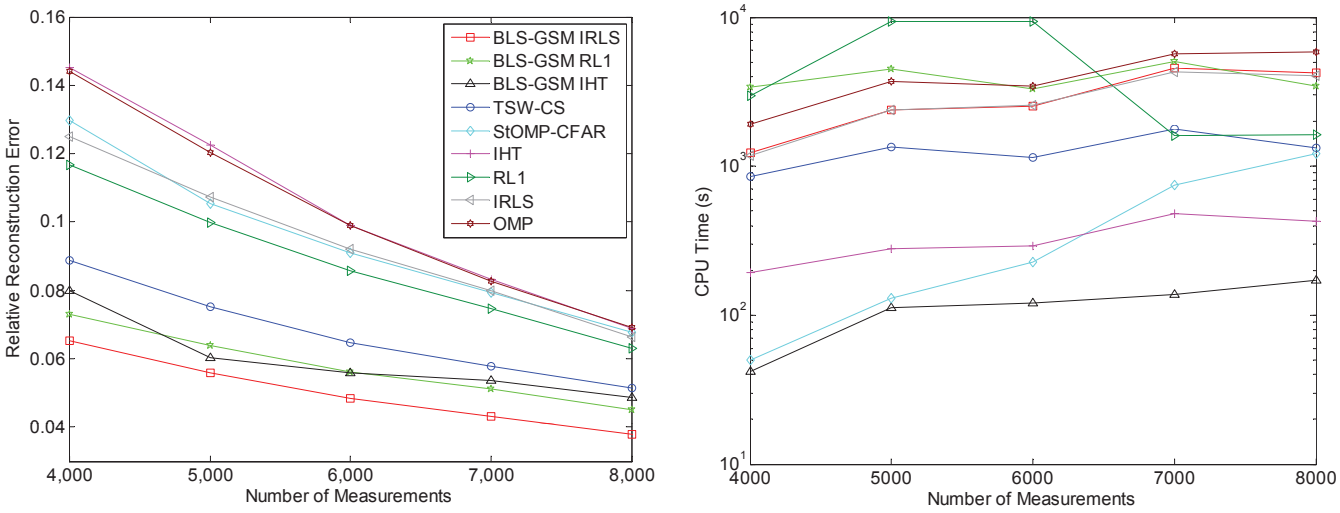


**Fig. 2**. Reconstruction accuracy and computation time for different number of measurements.

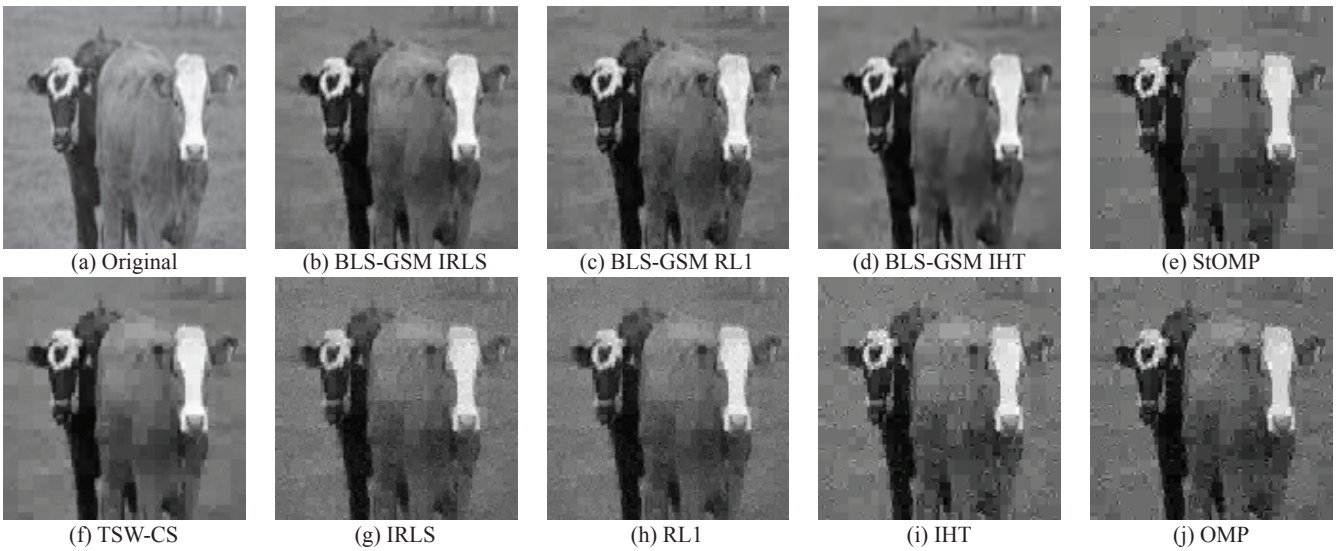


**Fig. 3**. Reconstructed images using different CS algorithms.

| Class | | BLS-GSM IRLS | BLS-GSM RL1 | BLS-GSM IHT | TSW-CS | StOMP | IRLS | RL1 | IHT | OMP |
|---|---|---|---|---|---|---|---|---|---|---|
| Flowers | MEAN | **0.1086** | 0.1254 | 0.1143 | 0.1396 | 0.1884 | 0.1795 | 0.1717 | 0.2087 | 0.2184 |
| | STD | 0.0627 | 0.0734 | **0.0625** | 0.0669 | 0.0756 | 0.0636 | 0.0660 | 0.0829 | 0.0977 |
| Cows | MEAN | **0.0653** | 0.0763 | 0.700 | 0.0853 | 0.1145 | 0.1126 | 0.1063 | 0.1294 | 0.1309 |
| | STD | **0.0202** | 0.0248 | 0.0211 | 0.0258 | 0.0334 | 0.0317 | 0.0295 | 0.0373 | 0.0383 |
| Buildings | MEAN | **0.0778** | 0.0901 | 0.0873 | 0.1054 | 0.1365 | 0.1296 | 0.1239 | 0.1524 | 0.1582 |
| | STD | **0.0158** | 0.0189 | **0.0158** | 0.0180 | 0.0202 | 0.0167 | 0.0172 | 0.0230 | 0.0263 |
| Urban | MEAN | **0.0756** | 0.0872 | 0.0874 | 0.1024 | 0.1372 | 0.1290 | 0.1230 | 0.1509 | 0.1569 |
| | STD | **0.0084** | 0.0100 | 0.0101 | 0.0134 | 0.0175 | 0.0162 | 0.0158 | 0.0183 | 0.0212 |
| Office | MEAN | **0.0669** | 0.0747 | 0.0751 | 0.0911 | 0.1330 | 0.1310 | 0.1223 | 0.1452 | 0.1470 |
| | STD | 0.0175 | 0.0217 | **0.0170** | 0.0266 | 0.0340 | 0.0292 | 0.0284 | 0.0343 | 0.0390 |

**Table 1**. Relative reconstruction errors for natural images using different CS algorithms.