# Highly overcomplete sparse coding

Bruno A. Olshausen

Redwood Center for Theoretical Neuroscience, Helen Wills Neuroscience Institute and School of Optometry, University of California, Berkeley

## ABSTRACT

This paper explores sparse coding of natural images in the highly overcomplete regime. We show that as the overcompleteness ratio approaches 10x, new types of dictionary elements emerge beyond the classical Gabor function shape obtained from complete or only modestly overcomplete sparse coding. These more diverse dictionaries allow images to be approximated with lower L1 norm (for a fixed SNR), and the coefficients exhibit steeper decay. We also evaluate the learned dictionaries in a denoising task, showing that higher degrees of overcompleteness yield modest gains in peformance.

**Keywords:** sparse coding, natural images, overcomplete dictionaries, denoising

## 1. INTRODUCTION

An important goal of image representation is to make explicit the structure contained in a scene. Sparse coding achieves this by adapting a dictionary to the statistics of the data so that any given image may be represented by a small number of active units out of a large population. When the number of elements in the representation exceeds the dimensionality of the data, the representation is said to be *overcomplete*. Overcomplete representations are desired for a number of reasons: they allow for "shiftability," such that translation or other transformations in the image result in smooth and easily predictable changes among the coefficients;[1] they provide robustness in situations where coding precision is limited;[2] and they can provide highly compact, sparse representations suitable for compression.[3,4] From the point of view of neuroscience, overcomplete codes are of interest because the neural representation layer 4 of V1 is highly overcomplete (more neurons than input fibers), by factor of at least 100.[5] Thus an overcomplete sparse coding model provides a possible hypothesis for cortical image representation.

One of the central questions that arises in the design of an overcomplete representation is the choice of dictionary. Ideally, one would like the elements of the dictionary to match the structures contained in images. For this reason Gabor or Gaussian atoms of various aspect ratios have been used to capture lines and edges in images.[3,4] However, for the diverse forms of structure that occur in natural images it is difficult to know *a priori* what class of functions is most appropriate. The optimal choice of dictionary ultimately depends upon image statistics as well as task demands. Here, we focus on the contribution from image statistics, and we explore what dictionaries emerge as the degree of overcompleteness increases from 1.25x to 10x.

Our first efforts on this problem were reported at the HVEI conference in 1996,[6] however at that time we were only able to explore representations up to 2x overcomplete due to the limitations of computational resources at the time. The intervening years have seen dramatic increases in computational speed and memory which now make it feasible to explore much higher degrees of overcompleteness. In more recent work[7] we showed that as either overcompleteness or sparsity is increased, the dictionaries that emerge exhibit greater diversity, resulting in ridge-like functions, circularly symmetric functions, and gratings. This paper builds on that work by focusing on how the learned dictionaries and the resulting sparsity change as a function of increasing overcompletness only, while keeping SNR fixed. We also preprocess the data at somewhat higher resolution allowing more detailed structures to emerge. Finally, we evaluate the dictionaries on a denoising task.

Author contact information: baolshausen@berkeley.edu, (510) 642-7250

Human Vision and Electronic Imaging XVIII, edited by Bernice E. Rogowitz, Thrasyvoulos N. Pappas, Huib de Ridder,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 8651, 86510S • © 2013 SPIE-IS&T
CCC code: 0277-786X/13/$18 • doi: 10.1117/12.2013504



Downloaded From: http://spiedigitallibrary.org/ on 04/20/2013 Terms of Use: http://spiedl.org/terms

## 2. SPARSE CODING MODEL

In a sparse coding model,[8,9] an image $I(\vec{x})$ is approximated in terms of a set of dictionary elements or basis functions $\phi_i(\vec{x})$ as follows:

$$I(\vec{x}) = \sum_{i=1}^{M} a_i \, \phi_i(\vec{x}) + \epsilon(\vec{x}) \tag{1}$$

where $\vec{x}$ indexes position within the two-dimensional image. The residual term $\epsilon(\vec{x})$ is included to capture structure not well described by the dictionary and is usually small compared to the first term. The image is thus represented in terms of the coefficient activations, $a_i$, which are encouraged to be sparse by imposing a cost function on their absolute value. The coefficients are computed by minimizing an energy function that includes both this cost function and the squared L2 norm of residual, $\epsilon$:

$$E = \frac{1}{2} \sum_{\vec{x}} \left[ I(\vec{x}) - \sum_{i=1}^{M} a_i \, \phi_i(\vec{x}) \right]^2 + \lambda \sum_{i=1}^{M} |a_i| \tag{2}$$

The parameter $\lambda$ controls the tradeoff between reconstruction error and sparsity.

The energy function may be minimized by a neural circuit consisting of leaky integrators and threshold elements connected in a resistive grid, similar to a Hopfield network:[10]

$$\tau \, \dot{u}_i + u_i = b_i + \sum_{j \neq i} G_{ij} \, a_i \tag{3}$$

$$a_i = g(u_i) \tag{4}$$

with $b_i = \sum_{\vec{x}} \phi_i(\vec{x}) \, I(\vec{x})$, $G_{ij} = \sum_{\vec{x}} \phi_i(\vec{x}) \phi_j(\vec{x})$, and

$$g(u) = \begin{cases} \mathrm{sgn}(u)[|u| - \lambda] & |u| > \lambda \\ 0 & \text{otherwise} \end{cases}$$

Learning of the dictionary is accomplished by stochastic gradient descent of $E$ with respect to the dictionary $\{\phi_i(\vec{x})\}$, using the coefficients computed from eqs. 3 and 4.

$$\Delta\phi_i(\vec{x}) = \eta \left[ I(\vec{x}) - \sum_{j=1}^{M} a_j \, \phi_j(\vec{x}) \right] a_i \tag{5}$$

The learning rate $\eta$ is chosen to be sufficiently small so that the dictionary adapts on a very slow timescale in response to the statistics of many images. After each update the dictionary is renormalized to enforce $\sum_{\vec{x}} \phi_i^2(\vec{x}) = 1 \;\; \forall i$.

## 3. RESULTS

The model was adapted to $16 \times 16$ pixel patches extracted from a set of images of natural scenes drawn from the van Hateren database[11,12] (see Appendix A for a listing of all the images used and details of preprocessing). Very similar results were obtained with David Field's images of the northwest.[13] Each image was transformed to log-intensity and then whitened and lowpass filtered to equalize variance at all spatial-frequencies and to remove energy from the corners of the two-dimensional frequency domain, following the same procedure described previously.[9] This yields approximately 2 million distinct $16 \times 16$ image patches differing by a translation of at least 2 pixels in any direction. Because of the lowpass filtering which cuts out the corners of the frequency-domain, there are only about 200 significant dimensions in these data (this is confirmed by noting the point at which the eigenvalues of the covariance matrix begin to drop off sharply). Thus, a dictionary with $M$=256 basis functions is actually about 1.25x overcomplete.



Downloaded From: http://spiedigitallibrary.org/ on 04/20/2013 Terms of Use: http://spiedl.org/terms

Dictionaries of size $M$=256, 512, 1024, and 2048 were learned from the data, yielding overcompletness ratios of 1.25, 2.5, 5, and 10, respectively. For each dictionary, lambda was adjusted during learning so as to maintain an SNR of 16 dB in the reconstruction error, where SNR is computed as

$$\text{SNR (dB)} = 10 \log_{10} \frac{\text{pixel variance}}{\text{mean squared error}} \tag{6}$$

Figure 1 shows a random sampling of 100 elements from each learned dictionary. As the degree of overcompleteness increases, one sees a greater degree of specialization emerge. At 1.25x overcomplete, the basis functions resemble oriented Gabor functions at different spatial scales, as has been reported previously.[8,9,12,14] At 2.5x and 5x overcomplete, the basis functions diverge in two different directions: 1) elongated functions, and 2) compact (non-elongated) functions resembling gratings. At 10x overcomplete, one sees at least four new classes of basis functions emerge: 1) straight contour or "ridgelet" functions[15] that extend across the entire image patch, 2) circular functions resembling difference-of-Gaussians, 3) curved functions, and 4) gratings at different spatial frequencies. Representative examples of these four types of functions are shown in Figure 2. The full 10x dictionary is shown in Figure 7.

The emergence of new types of basis functions begs the question of whether each type forms a complete tiling within its own feature space. I do not have a rigorous answer to this question, but in previous work[7] we showed that the circular functions form a complete and uniform tiling of the image patch, and that the ridgelet functions span all orientations and positions. I have not performed such an analysis here, but for now I will simply note that the number of functions of each type is similar that previous study and so it seems plausible that they would form a complete tiling in this case as well.

The specializations that emerge at higher degrees of overcompleteness would seem to make it easier to encode these different forms of structure within an image. For example, an image patch containing a straight contour can be more compactly described in terms of one of the ridgelet functions. But if the dictionary consisted only of these functions then it would be difficult to approximate more complex shapes such as T-junctions, textures, and so forth. For this reason, the circular, curved and grating functions are likely needed. Thus, the Gabor-like functions found in the regime of low overcompleteness may be better understood as a "one size fits all" compromise that emerges to approximate the diverse forms of structure that occur in natural images. As the sparse coding model is given more degrees of freedom, the Gabor functions fade away and the representation is dominated by more specialized classes of functions.

To investigate whether the more overcomplete representations allow for a simpler or more compact encoding of images, we examine how the coefficient values of a given dictionary decay for each image, and we evaluate their average L1 norm. If the more specialized dictionary elements emerging at higher overcompleteness ratios are better matched to image structure, they should decay more rapidly and exhibit a lower average L1 norm as overcompleteness increases. Figure 3 confirms that this is the case.

Finally, we evaluated the performance of each dictionary on a denoising task. Again, the reasoning here is that a dictionary better matched to the structure of natural images makes a better model, and thus it should be more adept at removing artifacts such as noise. We test this by adding Gaussian i.i.d. noise of variance 1.0 (same as the image variance) to the images to give an SNR of 0 dB. We then infer the coefficients using eqs. 3 and 4, sweeping through a range of values for $\lambda$ in order to find the optimal value. The results are shown in Figure 4 and examples of noisy and denoised image patches are shown in Figure 5. As expected, the higher overcomplete dictionaries yield a measurable increase in performance, though the gain is modest and barely perceptually noticeable in the denoised images. Overall, the overcomplete dictionaries recover about 5-6 dB of signal from the noisy images.

## 4. DISCUSSION

That "Gabor functions emerge from sparse coding of natural images" is by now regarded as a well known fact. Here we show that this statement must be qualified with regard to the overcompleteness ratio of the dictionary. Only sparse representations that are complete or modestly overcomplete result in a Gabor-like solution. As the overcompleteness ratio increases to 10x, the solution differs dramatically from the classical Gabor solution,



Downloaded From: http://spiedigitallibrary.org/ on 04/20/2013 Terms of Use: http://spiedl.org/terms

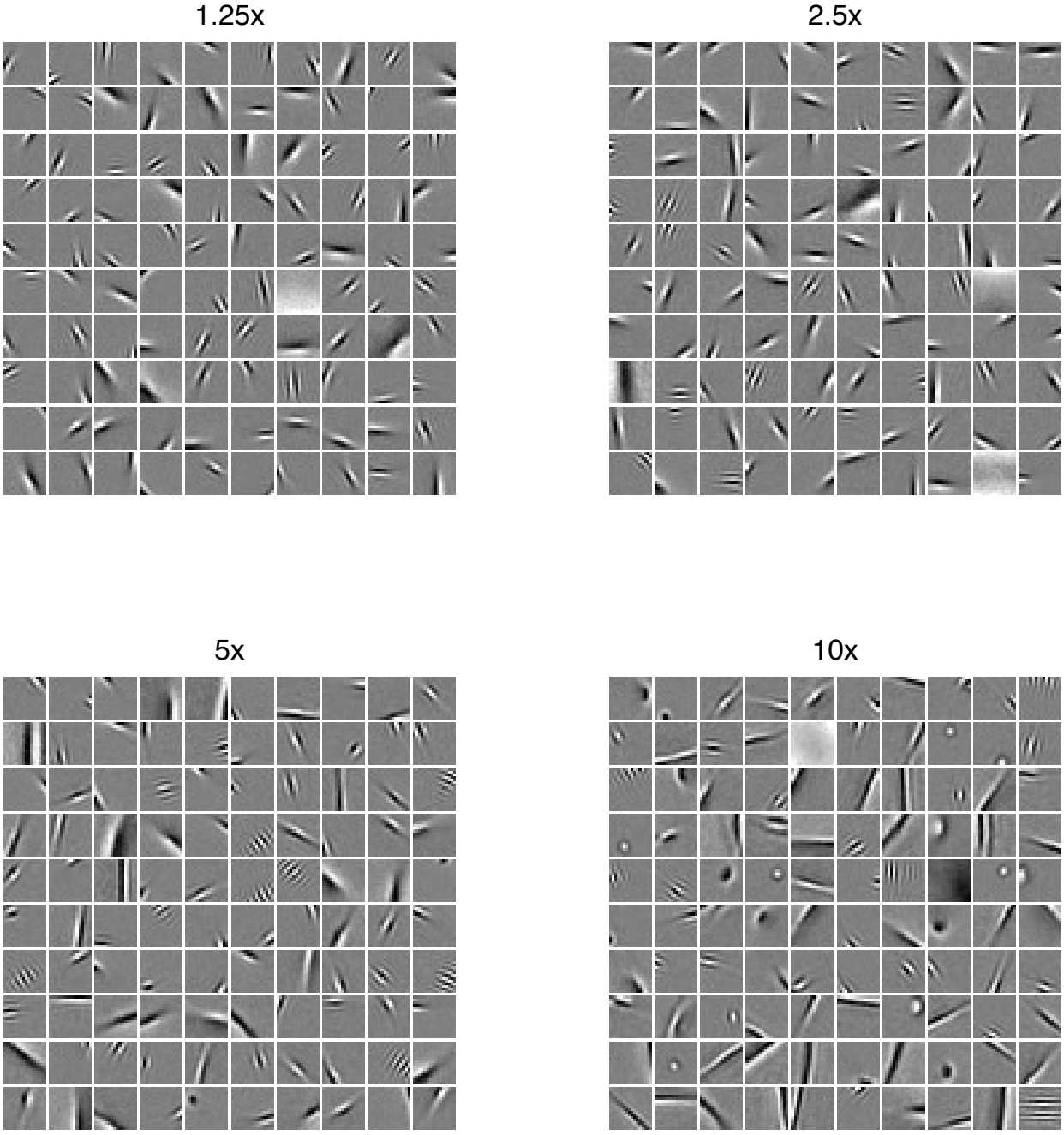



Figure 1. Learned dictionaries. Each panel shows 100 basis functions selected at random from the dictionary of a given overcompleteness ratio.

resulting in dictionaries containing more specialized elements such as straight contours, blobs, local curvature, and gratings. The specialized elements are better matched to the structures occurring natural images, as evidenced by the fact that they yield lower L1 norm representations, steeper coefficient decay, and better denoising. It seems plausible that they may also result in improved image compression though this remains to be seen.

These results are of relevance to neuroscience because the input layer of V1 is thought to be at least 100x



Downloaded From: http://spiedigitallibrary.org/ on 04/20/2013 Terms of Use: http://spiedl.org/terms

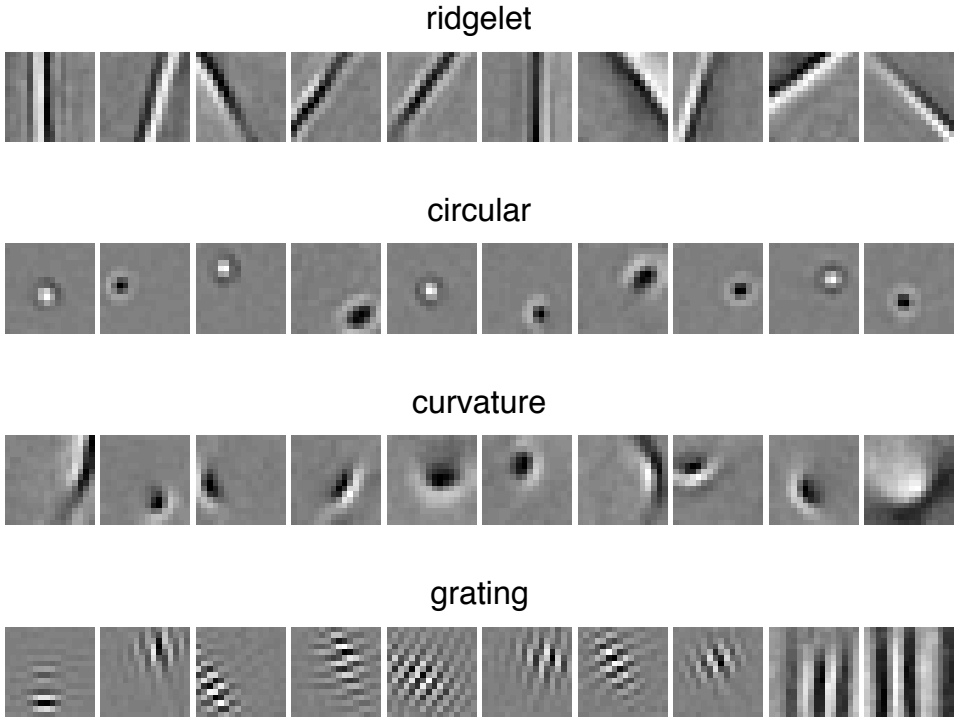



Figure 2. Representative examples of the four types of basis functions that emerge in the 10x overcomplete dictionary.

overcomplete[5] and firing rates are low as compared to their inputs from the LGN, suggesting that this system may also form a highly overcomplete, sparse representation of images. The results obtained here with highly overcomplete dictionaries thus suggest that these neurons may be representing more diverse types of features than previously thought. Indeed, as Rehn & Sommer have shown, the diversity seen in overcomplete sparse representations better matches the actual diversity seen in V1 neurons when their shapes are more accurately characterized.[7] However the Rehn & Sommer model was only about 3x overcomplete, and the feature diversity seen was rather modest as compared to what is seen here with 10x overcompleteness, which is still far short of the V1 regime. This begs the question then of what the extra degrees of freedom in cortex are being used for. Other dimensions such as time (motion), color, and disparity also need to be represented, which are not considered here. In any case, it may be worth re-examining the receptive field properties of these neurons, especially in response to natural scenes, in light of these findings.

## APPENDIX A. IMAGES USED IN TRAINING

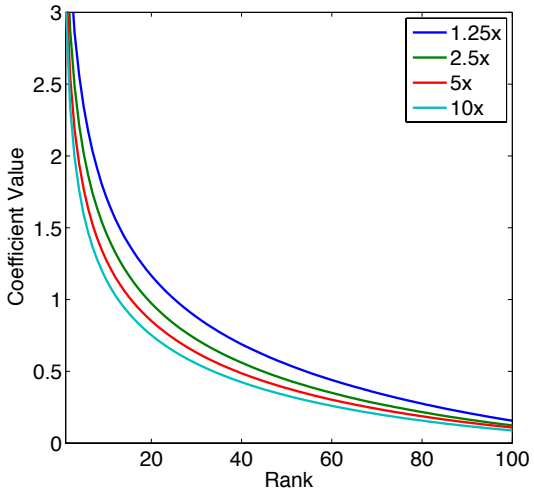
The training set consisted of 35 images selected from the van Hateren natural scenes database.[11,12] These were selected from an original set of 50 images that were randomly drawn from the database, from which we discarded images containing artifacts such as blur due to camera shake or excessive man made artifacts. The resulting set of 35 images are shown in Figure reffig:training-images. The filenames of the selected images are as follows:

| | | | | | | |
|---|---|---|---|---|---|---|
| imk00264.iml | imk00315.iml | imk00665.iml | imk00695.iml | imk00735.iml | imk00765.iml | imk00777.iml |
| imk00944.iml | imk00968.iml | imk01026.iml | imk01042.iml | imk01098.iml | imk01251.iml | imk01306.iml |
| imk01342.iml | imk01726.iml | imk01781.iml | imk02226.iml | imk02260.iml | imk02262.iml | imk02982.iml |
| imk02996.iml | imk03332.iml | imk03362.iml | imk03401.iml | imk03451.iml | imk03590.iml | imk03686.iml |
| imk03751.iml | imk03836.iml | imk03848.iml | imk04099.iml | imk04103.iml | imk04172.iml | imk04207.iml |

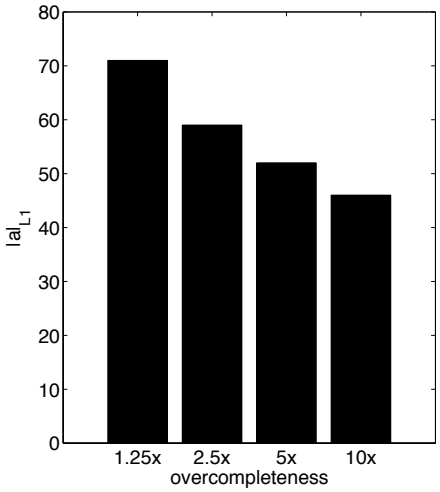
The original images are 1024 rows x 1536 columns, with pixel values linearly proportional to intensity. Following the same logic as Ruderman,[16] we transform each image to log-intensity. The central 1024x1024



Downloaded From: http://spiedigitallibrary.org/ on 04/20/2013 Terms of Use: http://spiedl.org/terms




a)




b)

Figure 3. a. Average coefficient decay. For each image patch, and for each dictionary, the coefficients are rank ordered by amplitude. The average decay as a function of rank for a random subsample of 10000 image patches is shown for each dictionary. b. Average L1 norm of the coefficients for each dictionary, using the same subsample of 10000 image patches.

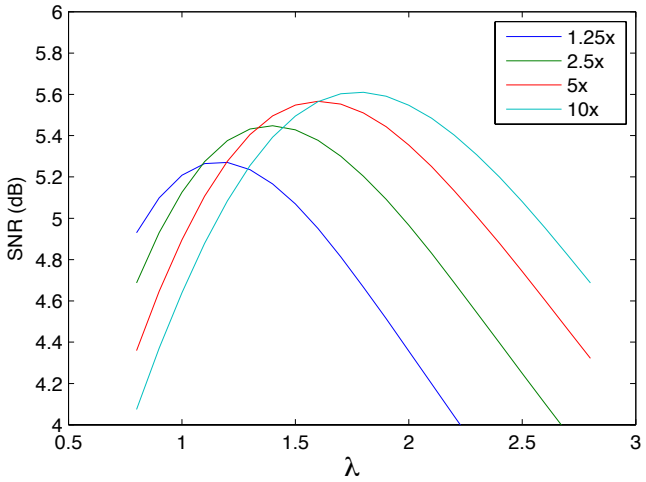



Figure 4. Denoising performance measured in terms of SNR as a function of *lambda* for each dictionary. The more overcomplete dictionaries yield better performance at higher values of $\lambda$.

region is extracted and the mean is subtracted to yield a pixel distribution that is roughly symmetric around zero. This image is then whitened and lowpass filtered in the frequency domain by multiplying with the following filter:

$$W(\vec{f}) = |\vec{f}|\, e^{-\left(\frac{|\vec{f}|}{f_0}\right)^4}$$

where $\vec{f}$ denotes two-dimensional spatial-frequency. The exponent of 4 is chosen in the lowpass filter so as to provide a sharper cutoff. The cutoff frequency $f_0$ is set to 200 cycles/image and the central 512x512 region of the frequency domain is extracted and inverse Fourier transformed to yield a 512x512 image that is down sampled by a factor of two from the original. The set of images obtained this way is then multiplied by single scale factor so that the variance of the entire ensemble is 1.0.



Downloaded From: http://spiedigitallibrary.org/ on 04/20/2013 Terms of Use: http://spiedl.org/terms

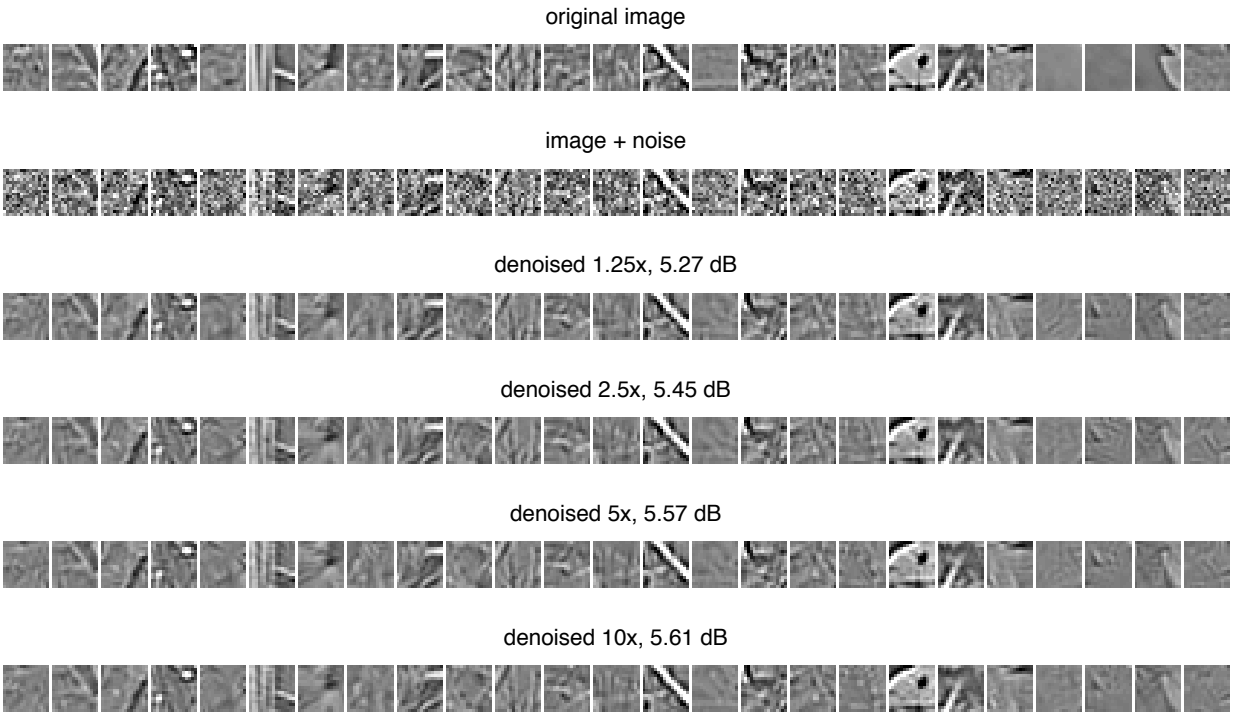



Figure 5. Denoising examples. Shown are 25 image patches in their original and noisy states, and their reconstructions for the optimal value of $\lambda$ for each dictionary.

## APPENDIX B. FULL 10X DICTIONARY

See Figure 7.

## ACKNOWLEDGMENTS

I thank Charlies Cadieu for discussions and sharing with me the result of his 100x overcomplete representation, which motivated this study. Mayur Mudigonda provided assistance with computer simulations. This work builds on work started by David Warland,[7] which initially pointed me in this interesting direction. Work supported by grants from the National Geospatial Intellgence Agency (HM1582-08-1-0007), the National Science Foundation (IIS-1111765) and the Canadian Institute for Advanced Research.

## REFERENCES

[1] Simoncelli, E., Freeman, W., Adelson, E., and Heeger, D., "Shiftable multiscale transforms," *IEEE Transactions on Information Theory* **38**(2), 587–607 (1992).

[2] Doi, E., Balcan, D. C., and Lewicki, M. S., "Robust coding over noisy overcomplete channels," *IEEE Transactions on Image Processing* **16**, 442–52 (Feb 2007).

[3] Schmid-Saugeon, P. and Zakhor, A., "Dictionary design for matching pursuit and appliation to motion compensated video coding," *IEEE Transactions on Circuits and Systems for Video Technology* (6), 880 – 886 (2004).

[4] Rahmoune, A., Vandergheynst, P., and Frossard, P., "Flexible motion-adaptive video coding with redundant expansions," *IEEE Transactions on Circuits and Systems for Video Technology* **16**(2), 178–190 (2006).

[5] Barlow, H. B., "The ferrier lecture, 1980. critical limiting factors in the design of the eye and visual cortex.," *Proceedings of the Royal Society of London Series B* **212**, 1–34 (1981).



Downloaded From: http://spiedigitallibrary.org/ on 04/20/2013 Terms of Use: http://spiedl.org/terms

Figure 6. 35 images from the van Hateren database that were used in training


[6] Olshausen, B. A. and Field, D. J., "Learning efficient linear codes for natural images: the roles of sparseness, overcompleteness, and statistical independence," *Proc. SPIE 2657, Human Vision and Electronic Imaging, 132 (April 22, 1996)* (1996).

[7] Olshausen, B. A., Cadieu, C. F., and Warland, D. K., "Learning real and complex overcomplete representations from the statistics of natural images," *Wavelets XIII, Proc. of SPIE Vol. 7446* (2009).

[8] Olshausen, B. A. and Field, D. J., "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature* **381**(381), 607–609 (1996).

[9] Olshausen, B. A. and Field, D. J., "Sparse coding with an overcomplete basis set: a strategy employed by V1?," *Vision Research* **37**(23), 3311–25 (1997).

[10] Rozell, C. J., Johnson, D. H., Baraniuk, R. G., and Olshausen, B. A., "Sparse coding via thresholding and local competition in neural circuits," *Neural Computation* **20**, 2526–2563 (2008).

[11] van Hateren, H., "Natural image dataset," *http://bethgelab.org/datasets/vanhateren/* .

[12] van Hateren, J. H. and van der Schaaf, A., "Independent component filters of natural images compared with simple cells in primary visual cortex," *Proceedings: Biological Sciences* **265**, 359–366 (Mar 1998).

[13] Olshausen, B. A., "Sparsenet software and images (images.mat)," *http://redwood.berkeley.edu/bruno/sparsenet* .

[14] Bell, A. J. and Sejnowski, T. J., "The independent components of natural scenes are edge filters," *Vision Res.* **37**, 3327–3338 (1997).

[15] Candes, E. J. and Donoho, D. L., "Ridgelets: a key to higher-dimensional intermittency?," *Phil. Trans. R. Soc. Lond. A 1* **357**, 2495–2509 (1999).

[16] Ruderman, D. L., "The statistics of natural images," *Network: Computation In Neural Systems* **5**, 517–548 (Nov. 1994).




Downloaded From: http://spiedigitallibrary.org/ on 04/20/2013 Terms of Use: http://spiedl.org/terms

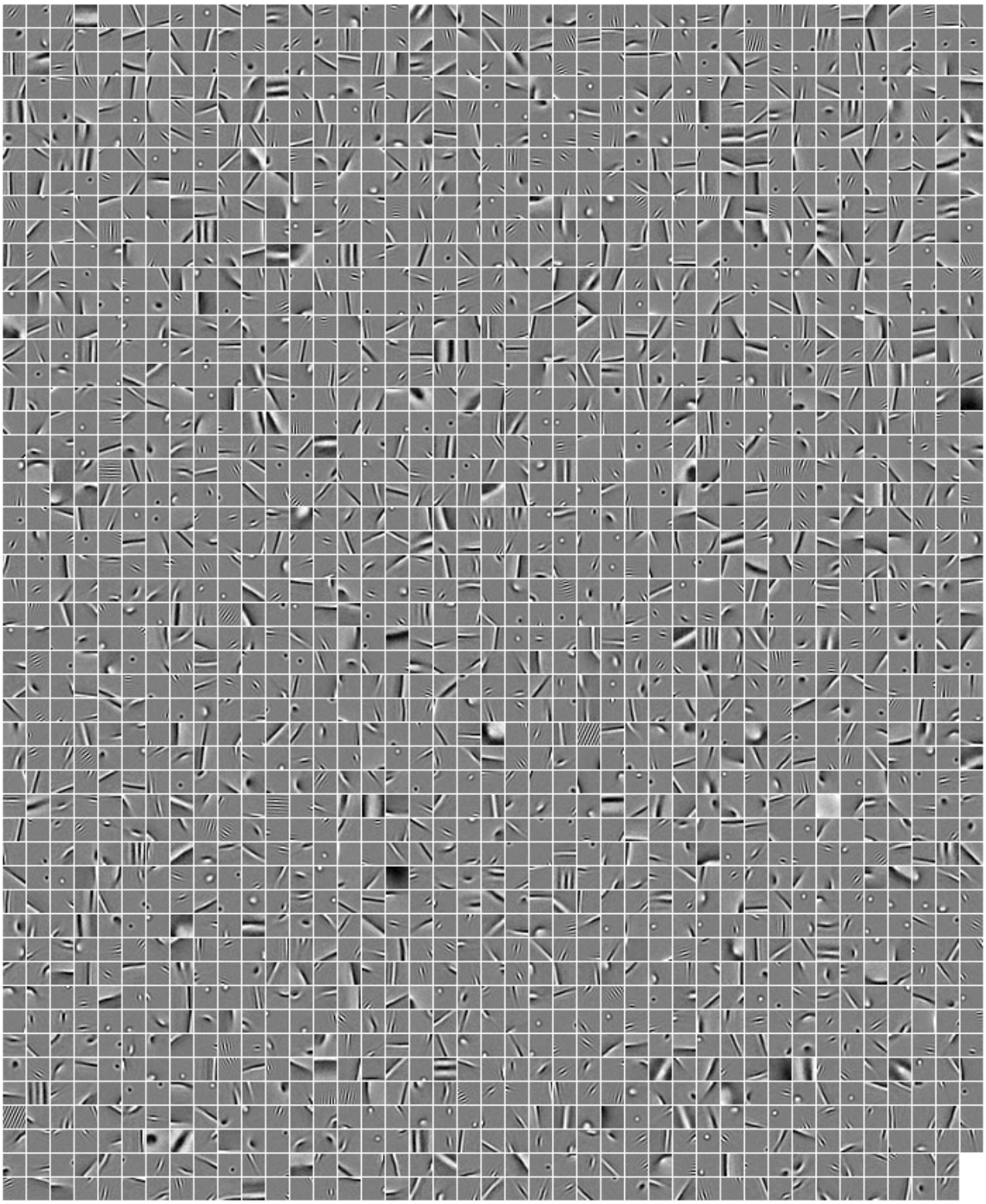


Figure 7. The full set of 2048 bases of the 10x overcomplete dictionary



Downloaded From: http://spiedigitallibrary.org/ on 04/20/2013 Terms of Use: http://spiedl.org/terms