ESANN'2005 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 27-29 April 2005, d-side publi., ISBN 2-930307-05-6.

New evidences for sparse coding strategy employed in visual neurons: from the image processing and nonlinear approximation viewpoint

Tan Shan and Licheng Jiao

Institute of Intelligent Information Processing,
Xidian University, Xi'an, 710071, China

Abstract. '*Sparse coding*' is a ubiquitous strategy employed in the sensory information processing system of mammals. Some work has focused on the validation of this strategy through finding the sparse component of sensory input, and then illustrating a fact that the resulting basis functions or corresponding filter response have the visually similar receptive field to those found in primary visual cortex (V1). In this review, we show that several newly proposed systems in the area of image processing and nonlinear approximation provide new evidences for the 'sparse coding' strategy along a contrary line. Inspired by the property of receptive field of neuron in V1, the bases functions of these systems are constructed with special structures, namely, band-pass, being localized and multi-orientation. Interestingly, these systems can sparsely represent the special classes of images dominated with edges.

1 Introduction

Understanding the mechanism of brain to process sensory information in mammals is a primary but challenging goal of neuroscience.

1.1 'Sparse coding' strategy in sensory information processing system

Some researches indicated that there exists a ubiquitous strategy employed in the sensory information processing system of mammals. This strategy, referred to as '*sparse coding*', represents information only using a relatively small number of simultaneously active neurons out of a large population [1]. In 1954, Attneave hypothesized that the goal of visual perception is to produce an efficient representation of the incoming signal [2]; In 1961, Barlow suggested that the neurons involved in sensory information processing should encode as much information as possible in order to most effectively utilize the available computing resource [3]. Early work on associative memory models also shown that sparse representations are most effective for storing patterns, as they maximize memory capacity because of the fact that there are fewer collisions between patterns [4]. Later work has similarly showed that sparse representations would be advantageous for learning associations in neural networks, as they enable associations to be formed effectively using local learning rules, such as Hebbian learning [5].

1.2 **Evidences for 'sparse coding' strategy: the receptive field of learning basis function derived using sparseness criterion resemble those of neurons in V1**

The spatial receptive fields of neurons in mammalian striate cortex have been reasonably well described physiologically and can be characterized as being localized, oriented, and band-pass [6] and hence well suited to the structure of images that falls upon the retina when viewing the natural world [1]. It is such special structure that makes the '*sparse coding*' possible.

Some work dedicated to the validation of the '*sparse coding*' strategy and showed that when the receptive fields of an entire population of neurons are optimized to produce sparse representations, the set of receptive fields that emerge resemble those of neurons in V1. These work, with the purpose to find the sparse representation of natural images, as well known to researcher in neuroscience, is viewed as evidences of the 'sparse coding' strategy employed in the sensory system in mammals. In these work, the key finding was obtained by Olshausen and Field [7]. They created a model of images based on a linear superposition of basis functions and adapted these functions so as to maximize the sparsity of the representation while preserving information in the images. The set of functions that emerges after training on hundreds of thousands of image patches randomly extracted from natural scenes, starting from completely random initial conditions, strongly resemble the spatial receptive field properties of neurons in V1, i.e. they are spatially localized, oriented, and band-pass in different spatial frequency bands. Example basis functions derived using sparseness criterion in [7] are shown in Fig.1. In addition, Hateren and Schaaf compared properties of the receptive fields of neurons in macaque cortex with the properties of independent component filters generated by independent component analysis (ICA) on a large set of natural images. Their results showed that the two kinds of receptive field properties match well, according to the histograms of spatial frequency bandwidth, orientation tuning bandwidth, aspect ratio and length of the receptive fields [8]. Hateren and Ruderman showed that performing independent component analysis (ICA) on video sequences of natural scenes produces results with qualitatively similar spatio-temporal properties [9]. In [10], Hyvarinen and Hoyer demonstrated that the principle of independence maximization, which is similar with that used in Olshausen and Field's experiment in [7], could explain the emergence of phase- and shift-invariant features, similar to those found in complex cells.

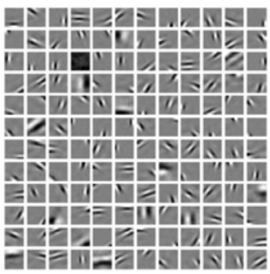

Fig. 1: Example basis functions derived using sparseness criterion in [7]

2 New evidence for 'sparse coding' strategy: new systems having basis functions with similar structure to receptive field of neurons in V1 can provide sparse representation for some special classes of images

Mathematically, the question to what degree a transform can efficiently represent signals or functions can be quantitatively characterized using the theory of nonlinear approximation. The higher the nonlinear approximation ability is, the less transform coefficients are needed to represent a signal within a given error level. The error level, in the theory of nonlinear approximation, commonly expressed as $\varepsilon[M] = \|f - f_M\|^2$, here f_M denotes the reconstruction using the *m*-term transform coefficients with largest amplitude. To some degree, the sparse representation of a transform to signals or functions, roughly, is the same as the 'sparse coding' strategy employed in V1 when one consider that V1 functions as if it takes a transform that maps the sensory input into the combination of states of neurons: active or not. Hence, it is not surprising that the advances in the theory of nonlinear approximation should shed light on the understanding the mechanism of sensory information processing system.

Wavelet analysis has achieved tremendous success in many fields of contemporary science and technology, especially in signal and image processing application. The success of wavelet mainly arises from its optimal nonlinear approximation ability to broad function classes, for example, functions smooth away from point singularity. In neuroscience, it is well known that there were claims that the human visual system acted, in early stages, by wavelet analysis [11]. Unfortunately, from the nonlinear approximation viewpoint, the 2-D separable wavelet analysis cannot provide sparse representation for bivariate functions with straight or curved singularity (corresponding to those images with straight and curved edges). The sensory system has long been assumed that neurons are adapted, at evolutionary, developmental, and behavioral timescales, to the signals to which they are exposed. Some work has shown that it is the edges that dominate the scenes human senses [12]. As a result, we suggest that wavelet be not an accurate enough and ultimate model to describe the function of neurons in V1.

As shown in Fig.2, the basis functions of 2-D separable wavelet system are 'isotropic', hence only have few orientations. The support of wavelet basis functions is multi-resolution, being localized but has not multi-orientation, which is obviously different from the receptive field of neurons in V1. It is the shortness of orientation selectiveness that makes the nonlinear approximation ability of 2-D separable wavelet system is contaminated by straight and curvilinear singularity contained in functions. In other words, wavelet cannot provide sparse representation for images with straight and curved edges. Precisely, let $x \in R^2$, $\theta^0 \in [0, 2\pi)$, and we consider function of the form $g(x_1, x_2; \theta^0, t^0) = 1_{\{x_1 \cos\theta^0 + x_2 \sin\theta^0 > t^0\}} e^{-x_1^2 - x_2^2}$, here, g is Gaussian function with singularity along line $t^0 = x_1 \cos(\theta^0) + x_2 \sin(\theta^0)$ and smooth elsewhere, then the error of the *m*-term nonlinear approximation of wavelet to g is $\|f - f_M^W\|^2 \sim O(M^{-1}), M \to \infty$, which is far from the optimal nonlinear approximation rate.

Recently, several new systems for function analysis and image processing have been proposed that provide more orientation selectiveness than separable wavelet system hence can efficiently deal with straight and curved edges in images.

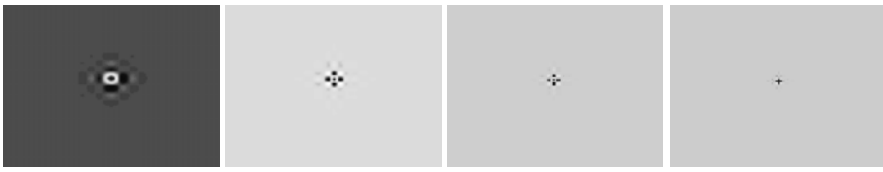

Fig. 2: Example basis functions of 2-D separable wavelet, from left to right: the scale varies from coarse to fine

In [13], Candès developed a new system, ridgelet analysis, and showed how they can be applied to solve important problems such as constructing neural networks, approximating and estimating multivariate functions by linear combinations of ridge functions. In a following paper [14], Donoho constructed orthonormal ridgelet, which provide an orthonormal basis for $L^2(R^2)$. In paper [15], it was shown that both ridgelet analysis and orthonormal ridgelet are optimal to represent functions that are smooth away from straight singularity. For example, for the function mentioned above, g, the number of orthonormal ridgelet coefficients with amplitude exceeding $1/M$ grows with *M* more slowly than any fractional power of *M*, namely, $\#\{\alpha_\lambda^R \,\|\, |\alpha_\lambda^R| > 1/M\} \sim O(M^{-s})$ for $\forall s \in Z^+$, where α_λ^R is the orthonormal ridgelet expansion coefficient of g with index λ. By comparison, when decomposing function g into wavelet series, we only have $\#\{\alpha_\lambda^W \,\|\, |\alpha_\lambda^W| > 1/M\} \sim O(M^{-1})$. It is exciting that the much higher approximation rate of g is achieved by orthonormal ridgelet. In fact, orthonormal ridgelet can optimally represent functions smooth away from straight singularity in the sense that nor orthonormal system achieves higher approximation rate. The key idea of orthonormal ridgelet is that it first transforms the straight singularity in spatial domain into point singularity in Radon domain, then deal with the resulting point singularity using wavelet system. As a result, in effect, it has 'anisotropic' basis functions as shown in Fig.3. As a generalization version of orthonormal ridgelet, ridgelet frame and dual ridgelet frame was proposed in paper [16][17], both of which also can effectively deal with straight edges in images. Though the multi-resolution and localization property, rigorously, are not introduced into these systems, they can provide the sparsest representation for images with straight edges yet. We suggest that the multi-orientation property, maybe, plays a more important role than others, i.e. band-pass and localization property, in the 'sparse coding' strategy in V1, considering that edges are dominating features in natural images.

Curvelet, which was derived from ridgelet analysis and can efficiently deal with smooth images with smooth edges, is a kind of multi-resolution representation with several features that set it apart from existing representations such as wavelets, multiwavelet, steerable pyramids, and so on [18]. Besides band-pass and localized, the basis functions exhibit very high direction sensitivity and are highly anisotropic, as shown in Fig.4.

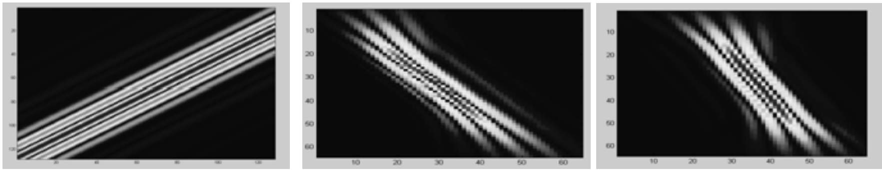

Fig. 3: From left to right: example basis function of ridgelet analysis, orthonormal ridgelet and ridgelet frame

Fig. 4: Example basis functions of Curvelet (the first two) and Countourlet (the last one)

Suppose we have an object supported in $[0,1]^2$ which has a discontinuity across a nice curve Γ, and is otherwise smooth. The error of *m*-term nonlinear approximation using curvelet can achieve the nearly optimal approximation rate: $\left\| f - f_M^C \right\|^2 \sim O(M^{-2}(\log M)), M \to \infty$. Whereas using a wavelet representation, the error of *n*-term nonlinear approximation only satisfies $\left\| f - f_M^W \right\|^2 \sim O(M^{-1}), M \to \infty$.

Minh N. Do and Vetterli developed a new 'true' two-dimensional representation for images that can capture the intrinsic geometrical structure of pictorial information [19]. The new system, called countourlet transform, is implemented using a double filter bank structure—pyramidal directional filter bank, which combined the Laplacian pyramid with a directional filter bank. The countourlet transform provides a flexible multi-resolution, local and directional expansion for images. Several basis functions of countourlet transform are shown in Fig.4. The countourlet transform also provides a sparse representation for bivariate piecewise smooth signals, namely, in such case, the error of *m*-term nonlinear approximation using *countourlet* can achieve the nearly optimal approximation rate: $\left\| f - f_M^C \right\|^2 \sim O(M^{-2}(\log M^3)), M \to \infty$.

3 Discussion

In this review, we have shown that several newly-proposed function analysis or image representation systems, whose basis functions imitated the response properties of neurons in V1, namely, band-pass (multi-resolution), localized and multi-orientation, interestingly, can provide sparse representation for some special classes of images dominated by edges. We believe that they are better models for functions of V1 than wavelet analysis and can be viewed as a new kind of evidences for the 'sparse coding' strategy along a line contrary to earlier work focused on this issue.

References

[1] B. A. Olshausen and D. J. Field, Sparse coding of sensory inputs, *Current Opinion in Neurobiology*, 14: 481-487, 2004.

[2] F. Attneave, Some informational aspects of visual perception. *Psychol. Rev.* 61:183–193, 1954.

[3] H. B. Barlow, Possible principles underlying the transformation of sensory messages. In W. A. Rosenblith, editor, *Sensory Communication*, pages 217–234, Cambridge, MA: MIT Press, 1961.

[4] Willshaw DJ, Buneman OP, Longuet-Higgins HC: Nonholographic associative memory. *Nature*, 222:960-962, 1969.

[5] Palm G: On associative memory. *Biol Cybern*, 36:19-31,1980.

[6] B. A. Olshausen and D. J. Field, Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Res.*, 37: 3311-3325, 1997.

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

[8] J. H. Hateren and A. Schaaf, Independent component filters of natural images compared with simple cells in primary visual cortex. Proc R Soc Lond B Biol Sci, 265:359-366,1998.

[9] J. H. Hateren and D. L. Ruderman, Independent component analysis of natural image sequences yields spatiotemporal filters similar to simple cells in primary visual cortex. Proc R Soc Lond B Biol Sci, 265:2315-2320, 1998.

[10] A. Hyvarinen and P. O. Hoyer, Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. *Neural Comput,* 12:1705-1720, 2000.

[11] D. J. Field, Relations between the statistics of natural images and the response properties of cortical cells. *J Opt Soc Am A* 4: 2379-2394,1987.

[12] E. P. Simoncelli and B. A. Olshausen, Natural image statistics and neural representation, *Annu. Rev. Neurosci*., 24: 1193-1216, 2001.

[13] E. J. Candès, Harmonic analysis of neural netwoks, *Appl. Comput. Harmon. Anal*., 6: 197–218, 1999.

[14] D. L. Donoho, Orthonormal ridgelet and straight singularities, *SIAM J. Math Anal*., 31: 1062–1099, 2000.

[15] E. J. Candès, On the representation of mutilated sobolev functions, *SIAM J. Math. Anal.*, l: 2495-2509, 1999.

[16] Tanshan, L. C. Jiao, and X. C. Feng, Ridgelet frame. In Campilho, Aurélio; Kamel, Mohamed, editors, *Proc. Int. Conf. Image Analysis and Recognition*, Lecture Notes in Computer Science, pages 479-486, Springer-Verlag, 2004.

[17] Tanshan, X. R. Zhang, and L. C. Jiao, Dual ridgelet frame constructed using biorthonormal wavelet basis. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, March 19-23, Philadelphia, PA, USA, 2005, to appear.

[18] E. J. Candès and D. L. Donoho, Curvelet—A surprisingly effective nonadaptive representation for objects with edges, In A. Cohen, C. Rabut, and L. L. Schumaker, editors, *Curve and Surface Fitting: Saint-Malo* Nashville, TN: Van-derbilt Univ. Press, 1999.

[19] M. N. Do and M. Vetterli, Contourlets, In J. Stoeckler, G. V. Welland, editors, *Beyond Wavelet*, Academic Press, 2002.