Spectral Vs Spatial Coding for Non-Linear Neural Networks

Martial Mermillod
University of Clermont-Ferrand 2 – CNRS UMR 6024
34, av Carnot, 63037 Clermont-Ferrand
mermillod@srvpsy.univ-bpclermont.fr
Serban C. Musca
Laboratoire de Psychologie et NeuroCognition – UMR 5105
Universite Pierre Mendes France, BP 47, 38040 Grenoble Cedex 09
smusca@upmf-grenoble.fr
Thierry Atzeni
Laboratoire de Psychologie et NeuroCognition – UMR 5105
Universite Pierre Mendes France, BP 47, 38040 Grenoble Cedex 09
atzeni@free.fr
David Alleysson
Laboratoire de Psychologie et NeuroCognition – UMR 5105
Universite Pierre Mendes France, BP 47, 38040 Grenoble Cedex 09
David.Alleysson@upmf-grenoble.fr
Julien Barra
Laboratoire de Psychologie et NeuroCognition – UMR 5105
Universite Pierre Mendes France, BP 47, 38040 Grenoble Cedex 09
Julien.barra@upmf-grenoble.fr
Richard Palluel
Laboratoire de Psychologie et NeuroCognition – UMR 5105
Universite Pierre Mendes France, BP 47, 38040 Grenoble Cedex 09
richard.palluel@upmf-grenoble.fr
Christian Marendaz
Laboratoire de Psychologie et NeuroCognition – UMR 5105
Universite Pierre Mendes France, BP 47, 38040 Grenoble Cedex 09
Christian.Marendaz@upmf-grenoble.fr

Abstract

V1 receptive fields show different sensitivities to different scales suggesting spatial frequency coding of visual information. The purpose of the present paper is to determine if such spectral decomposition at a perceptual level would be efficient for non-linear categorization purpose. Specifically, we compare the advantage of providing an artificial neural network with spectral information vs. spatial information in a visual image classification task. In order to keep the same amount of information in the spatial and spectral domains, we applied a subsampling of original images while having the same amount of data in the frequency domain after applying a low-pass cut-off. Results show that a standard back-propagation neural network aimed at classifying visual images yields better performances when provided with the spectral information than spatial information. This result suggests that artificial and possibly biological neural systems would be better of using spectral rather than spatial information in the visual classification of natural scene images.

1. Introduction

Using directly pixels in images as input of a connectionist neural network is often impossible due to the too large amount of data. Indeed, a nxn image needs a n^2 number of neurons in the input layer. Many different methods are used in connectionist modeling to compress visual information. For example, the use of Principal Component Analysis (PCA) permits one to describe a pattern of data in an orthogonal space. Therefore, each stimulus in the original space can be specified in this new space by the value of its projection along the first axis describing the larger variability of the inputs. Applying this method on a gender categorization task of human faces, Abdi, Valentin, Edelman, O'Toole [1] showed that reliable gender categorization can be obtained using a posteriori features automatically derived from the statistical structure. They reported a correct classification rate of 91.8 % when associated with a multi-layer perceptron (MLP) and 90.6 % when associated with a Radial Basis Function network. Similarly, Cottrell [2] proposed using the hidden layer of a back-propagation auto-associator to compress visual information. This process, generally considered as a non-linear version of PCA, results in a compression of the input layer called "holons".

Another method consists of using local receptive fields that handle different parts of the input layer (the image), and associating their centers with category nodes. These types of neural networks, such as Radial Basis Function (RBF) networks [13] seem to be reliable for identification and classification purposes [5, 6,]. Nonetheless, Jones & Palmer [11] have shown that Gabor functions are better at simulating receptive fields of biological visual neurons than Gaussian functions. Therefore, Wiskott [14]; Wiskott, Fellous, Krüger & Von der Malsburg [15]; Dailey, Cottrell, Padgett & Ralph [4] used convolution of Gabor wavelets at specific locations of the original image in the spatial domain in order to compress visual images. However, this last technique requires an additional (and not biologically motivated) step, generally PCA, to compress visual information for MLP.

Gabor [8] has shown that his function efficiently describes the informative content of a signal in the frequency domain while losing a minimum of information in the spatial domain. Thus, recent articles have proposed applying Gabor filtering in the frequency domain for visual categorization purposes [9, 10, 12]. This method is a common tool to simulate spatial invariance properties observed at the level of the striate cortex (which is not the case for visual pathways from the retina to the lateral geniculate nucleus). Unfortunately, the spectral analysis of visual information in these previous papers was not compared to a direct spatial analysis of visual images in identification and categorization tasks. Accordingly, the aim of the present paper is to test the reliability of neural

computation of the Gabor technique compared to the raw code provided by natural images. More recent spiking neuron networks are using the raw code of pixel output to perform fast and reliable categorization purpose (Thorpe, Delorme & VanRullen, 2001 [16]; VanRullen & Thorpe, 2002 [17]). The aim of the present paper is not the comparison of these different families of neural networks but, instead, a test of different perceptual information pre-processing widely used in connectionist modeling for a *same* PDP classifier. In other words, our current goal is to show that a biologically inspired spectral integration of visual data could outperform spatial coding for a same non-linear classifier. In order to conduct a fair comparison of these inputs one must keep the same amount of data in both simulation conditions. Therefore, visual compression was done in the spatial domain by means of a regular subsampling of the low-pass filtered original image that was identical to the low-pass filtering applied before computing the spectral domain.

2. Gabor filters and stimuli pre-processing

In order to avoid a boundary effect in subsequent Fourier transformation, a Hanning window was first applied. Boundary effects could result in a bias toward an over-representation of cardinal orientations, and the Hanning window is a common tool to eliminate this bias. The following formula describes the Hanning window applied to each image:

$$w(i) = 0.5 + 0.5 \times \cos(\frac{2\pi i}{N})$$

Then we applied the following spectral or spatial processes to the images.

2.1. Spectral analysis

The purpose of this section is the use of Gabor wavelet filtering to describe a visual stimulus in terms of orientation and spatial frequency information that is similar to V1 cortical cell neurons. Computationally, this spectral analysis is widely used because it relies on well-known mathematical signal decomposition tools, such as discrete Fourier transform (DFT). A low pass filter was then used to remove high spatial frequency information. This step was done in order to retain the same amount of information in the Fourier domain as in the spatial domain, for the subsampled images. Subsampling of visual images is possible because high spatial frequency information is removed. Therefore, we used the same low pass filter to reproduce the removal of high spatial frequency information in the spectral domain.

Images were 256x256 pixels in size. This defines spatial frequencies up to 128 cycle/images. The low pass filtering suppressed all the frequencies to 16 cycle/images in order to be able to represent images in an array of size 32x32. The filter we used applied in the Fourier domain; it is a step function putting all frequency values greater than 16 cycle/images to zero as shown in Figure 1.

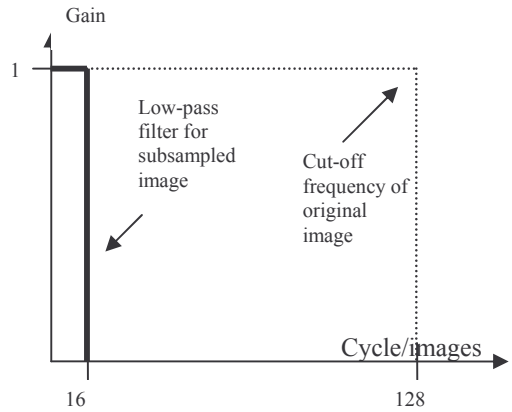

Figure 1. Spectral cut-off and spatial subsampling.

Then, we applied Gabor receptive fields in the spectral domain. A Gabor function has the shape of a sine wave modulated by a Gaussian filter. However, it is also possible to perform the filtering in the frequency domain by multiplying the spatial frequency information by the kernel of the Gabor function:

$$G(x, y, f_c, \theta) = \frac{1}{2\pi\sigma_r\sigma_t} e^{-\frac{(x.u)^2}{2\sigma_r^2}} e^{-\frac{(x.u_\perp)^2}{2\sigma_t^2}} e^{j2\pi \underline{x} f_c}$$

With: $\left| \begin{array}{l} \underline{x} = [x, y]^t, f_c = [f_0 \cos\theta, -f_0 \sin\theta]^t \\ u = [\cos\theta, \sin\theta]^t, u_\perp = [\sin\theta, \cos\theta]^t \end{array} \right.$

A Gabor filter is constructed with a Gaussian modulated by a complex exponential: parameters σ_r and σ_t of the Gaussian determine the spatial extent of the filter (figure 1). The vector f_c with module f_0 and direction θ describes this sine wave.

Each individual image was transferred into the Fourier domain and filtered by a set of Gabor filters. The filtered outputs were used to determine energy coefficients by coding the local energy spectra. We applied a bank of 56 Gabor wavelets corresponding to seven different spatial frequency bands and eight different orientations (0, $\pi/8$, $2\pi/8$, $3\pi/8$, $4\pi/8$, $5\pi/8$, $6\pi/8$, $7\pi/8$). The distance between two consecutive centers was one octave and spatial extent was increasing by one octave per spatial frequency channel [7].

2.2. Spatial analysis

Spatial analysis consisted of applying a subsampling in the spatial domain that keeps the same amount of information as the low pass cutt-off applied in the Fourier domain (figure 2). In other words, it is possible to convert the subsampled image in sine wave decomposition in the Fourier domain and then return to the spatial domain without any loss of information. This enables retaining the same amount of information in the spectral and spatial domains for further neural computation processes. The subsampling was applied in the following way.

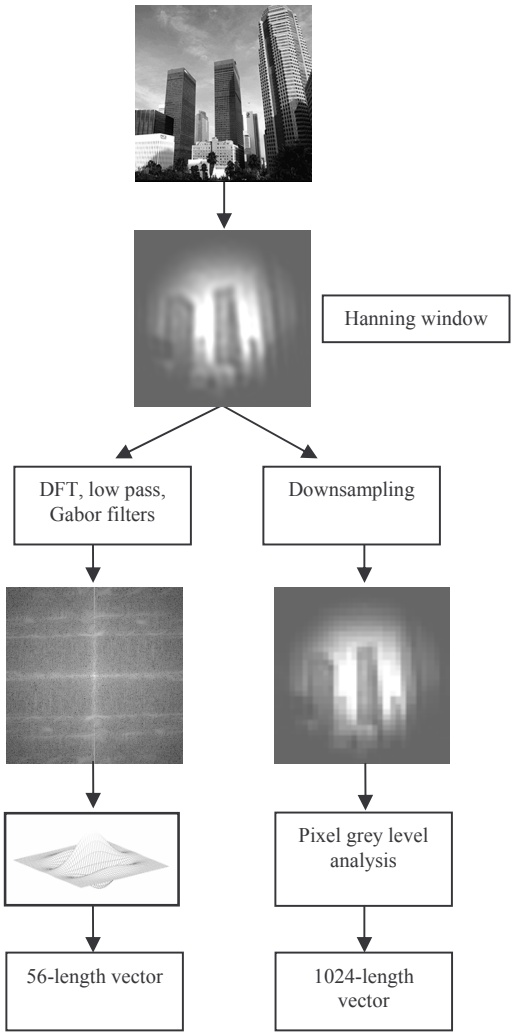

Figure 2. Spectral and spatial image processing.

3. Neural network processing

The connectionist network is a 3-layer back-propagation neural network. We used the standard hetero-association training algorithm, whose function is to associate each of the different category exemplars with a specific output vector coding for each category.

An important point of this paper is that neural network modeling was not proposed as a simulation of cortical processes proper, but rather as a tool allowing us to explore the statistical properties of the two input signals for non-linear dynamic systems.

4. Simulation 1: Testing spatial information in a connectionist network

4.1. Network

The connectionist architecture was the standard 1024-512-6 backpropagation hetero-associator described above.

4.2. Stimuli

Stimuli were composed of 72 exemplars from 6 categories of natural scenes that were subsampled. Stimuli description was detailed in section 2.2. Data were normalized between 0 and 1 across all category exemplars for neural network computation.

4.3. Procedure

The training phase was done on 36 training exemplars randomly chosen from each of the 6 natural scene categories (6 training exemplars per category). The training consisted of associating each exemplar with a specific vector category during a fixed number of 200 epochs. Output vectors were 000001 for the beach category, 000010 for the village category, 000100 for the mountain category, 001000 for the indoor category, 010000 for the city category and 100000 for the forest category. The learning rate was fixed to 0.1, momentum to 0.9 and the Fahlman offset to 0.1. The only difference between the two simulations was the nature of the input vectors: Subsampling of grey level pixels vs. spectral information provided by Gabor filters.

The test phase consisted of exposing the 6 remaining exemplars per category to the neural network. The observed output was gathered through the whole set of test vectors and a winner-take-all algorithm was then applied to the output nodes in order to determine which node provides the strongest activation associated with each input vector.

Fifty runs of this training/test procedure were conducted and a ratio of correct responses was computed for each run.

4.4. Results

Results show weak classification performance on the Forest category (35.3 % of correct categorization) and average performance for the City (57.7 %), Indoor (65.3 %), Mountain (69 %), Village (52.3 %) and Beach (64.7 %) categories. Surprisingly, the neural network often confused Mountain exemplars (31.3 %) with other category exemplars, especially the Forest and Village exemplars. An exhaustive confusion matrix is reported in Table 1.

	NEURAL NETWORK CLASSIFICATION					
INPUT VECTOR	Forest	City	Indoor	Mount	Village	Beach
Forest	35.3	5.0	7.3	31.3	13.7	7.3
City	7.3	57.7	9.0	15.7	6.7	3.7
Indoor	3.7	1.3	65.3	14.7	6.3	8.7
Mountain	8.3	2.7	2.0	69.0	7.3	10.7
Village	7.0	6.7	5.7	20.0	52.3	8.3
Beach	4.7	3.3	0.3	12.7	14.3	64.7

Table 1: Confusion matrix produced by the neural network after training on spatial information provided by subsampled data.

4.5. Discussion

Neural network performance was fair or average in almost all categories (except for the Forest category, where a significantly lower performance was observed). For most categories, the neural network seems to be able to use the spatial information provided by direct grey level pixels to generalize categorization to new exemplars after training on a random set of exemplars.

5. Simulation 2: Testing spectral information in a connectionist network

5.1. Network

The connectionist architecture was a 56-28-6 backpropagation neural network corresponding, in terms of hidden unit compression, to the neural network used in simulation 1. All parameters were kept identical to simulation 1.

5.2. Stimuli

Stimuli were the same 72 exemplars from the 6 nature scenes categories used in simulation 1. However, the stimuli were filtered by means of a low-pass cut-off applied in the Fourier domain, in order to keep the same amount of information as for the subsampled stimuli. Subsequently, a bank of Gabor filters was applied in the frequency domain. Each filter provided an average energy value related to one spatial frequency and one orientation channel resulting in a 56-length vector for neural network processing.

5.3. Procedure

The training/test procedure was identical to that used in simulation 1.

5.4. Results

Reliable categorization performances were obtained for each test condition. The Forest category produced 89.9 % correct categorization, the city category 88.3 %, the Indoor category 88.3 %, the Mountain category 91 %, the Village category 68 %, and the Beach category 79 %. The confusion matrix is reported in Table 2. For each category, spectral information provided by Gabor wavelet filtering provided a slightly better categorization performance than the spatial information provided by pixel grey levels. The effect is significant for the Forest category ($F(1, 98) = 136.2$, $p<0.001$), the City category ($F(1, 98) = 77.6$, $p<0.001$), the Indoor category ($F(1, 98) = 41.6$, $p<0.001$), the Mountain category ($F(1, 98) = 19.7$, $p<0.001$), the Village category ($F(1, 98) = 12.2$, $p<0.001$) and the Beach category ($F(1, 98) =14.5$, $p<0.001$).

	NEURAL NETWORK CLASSIFICATION					
INPUT VECTOR	Forest	City	Indoor	Mount	Village	Beach
Forest	89.0	0.0	3.0	1.0	4.3	2.7
City	1.7	88.3	3.7	1.0	4.0	1.3
Indoor	0.0	4.7	88.3	4.0	1.7	1.3
Mountain	1.3	0.7	1.7	91.0	4.0	1.3
Village	20.7	3.0	3.7	1.7	68.0	3.0
Beach	1.0	0.0	3.0	11.7	5.3	79.0

Table 2: Confusion matrix for the neural network after training on spectral information provided by Gabor wavelet filtering.

5.5. Discussion

Globally, these results imply that the spectral information coded by Gabor filters in the Fourier domain seems to permit better categorization performance of new category exemplars. This effect was obtained on a standard non-linear classifier based on parallel and distributed processing.

6. Conclusion

Among the different techniques used to compress visual information for subsequent connectionist processing, spectral decomposition of visual images seems to be a

promising tool. Gabor filtering of spectral information combines a biologically inspired method for signal analysis with an efficient coding of visual inputs for categorization tasks. Spectral analysis is less sensitive to translation within a same image or among different or similar details related to different exemplars of the same category. For example, if the horizontal lines characterizing a beach appear at different places in the beach images, spectral analysis will be less sensitive to these changes than spatial analysis.

Compared to RBF networks, combining Gabor filtering with MLP has the advantage of optimizing receptive fields within both Fourier and spatial domains. In addition, this combination also permits parallel and distributed processing, allowing non-linear dynamic classification to occur.

In summary, this work shows the advantage of using the spectral domain over the use of a raw coding of the images, in a categorization by a connectionist network. Therefore, we infer that the perceptual system may have many advantages in processing spatial frequency information. In this paper, we used Gabor filters as a computational shortcut to perform this scale processing but [17] have shown that spiking neuron networks would be able to transform spatial pixel coding into scale decomposition for cognitive processes.

Finally, there are new types of connectionist networks that use Gabor patches directly in the spatial domain by means of a sliding window across the entire image. This technique requires an additional PCA at the output of these filters allowing compression of visual information for MLP (Dailey & Cottrell [3]; Dailey, Cottrell, Padgett & Ralph [4]). The next step that must be taken is to compare spectral Gabor filtering with spatial Gabor filtering (in addition with PCA) in the domain of connectionist modeling.

Acknowledgments

This work was supported by a post-doctoral grant from the Fyssen Foundation and the French CNRS to MM and the French CNRS to CM, DA, SCM, JB, TA & RP.

References

[1] Abdi, H., Valentin, D., Edelman, B.E., O'Toole, A.J. (1995). More about the difference between men and women: Evidence from linear neural networks and the principal component approach. Perception, 24, 539-562.

[2] Cottrell, G.W. (1990). Extracting features from faces using compression networks: Face, identity, emotion, and gender recognition using holons. In Proceedings of the 1990 Connectionist Models Summer School, (eds. D. Touretsky, J. Elman, T. Sejnowski & G. Hinton) Kaufman, 328-337.

[3] Dailey, M. N. & Cottrell, G. W. (1999). Organization of Face and Object Recognition in Modular Neural Networks. Neural Networks, 12(7-8), 1053-1074.

[4] Dailey, M. N., Cottrell, G. W., Padgett, C., & Ralph A. (2002). EMPATH: A neural network that categorizes facial expressions. Journal of Cognitive Neuroscience 14(8), 1158-1173.

[5] Duvdevani-Bar, S., & Edelman S. (1999). Visual recognition and categorization on the basis of similarities to multiple class prototypes. International Journal of Computer Vision, 33(3), 201-228.

[6] Edelman, S. & Duvdevani-Bar S. (1997). A model of visual recognition and categorization, Proceedings of Royal Society, London, 352, 1191-1202.

[7] Field, D.J. & Brady N. (1997). Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. Vision Research, 37, 3367-3383.

[8] Gabor, D. (1946). Theory of communication. J. Inst. Electr. Eng, 93, 429-457.

[9] Guyader, N., Chauvin, A., Peyrin, C., Hérault, J., & Marendaz, C. (2004). Image phase or amplitude? Rapid scene categorization is an amplitude based process. C. R. Biologies 327, 313-318.

[10] Hérault, J., Oliva, A., & Guerin-Dugue, A. (1997). Scene Categorisation by Curvilinear Component Analysis of Low Frequency Spectra. Proceedings of the 5th European Symposium on Artificial Neural Network, Bruges, Belgium, pp. 91-96.

[11] Jones, J.P. & Palmer L.A. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. Journal of Neurophysiology, 58(6), 1187-1211.

[12] Oliva, A., Torralba, A. B., Guérin-Dugué, A., & Hérault, J. (1999). Global semantic classification of scenes using power spectrum templates. Proceedings of the Challenge of Image Retrieval (CIR99), Newcastle.

[13] Poggio, T. & Edelman, S. (1991). A network that learns to recognize three-dimensional objects. Nature, 343, 263-266.

[14] Wiskott L. (1997). Phantom Faces for Face Analysis. Pattern Recognition 30(6), 837-846.

[15] Wiskott, L., Fellous, J.M., Krüger, N. & Von der Malsburg C. (1999). Face Recognition by Elastic Bunch Graph Matching. In Intelligent Biometric Techniques in Fingerprint and Face Recognition, eds. L.C. Jain et al., CRC Press, 11, 355-396.

[16] Thorpe S.J., Delorme A. & VanRullen R. (2001). Spike-based strategies for rapid processing. Neural Networks, 14(6-7), 715-726.

[17] VanRullen R. & Thorpe S.J. (2002). Surfing a spike wave down the ventral stream. Vision Research, 42(23), 2593-2615.