Proc. Int'l Conf. on Artificial Neural Networks, ICANN'02,
ed. José R. Dorronsoro, Springer Verlag, pp. 81-86

Applying Slow Feature Analysis to Image Sequences Yields a Rich Repertoire of Complex Cell Properties

Pietro Berkes and Laurenz Wiskott

Institute for Theoretical Biology,
Humboldt University Berlin,
Invalidenstraße 43, D - 10 115 Berlin, Germany,
{p.berkes,l.wiskott}@biologie.hu-berlin.de,
http://itb.biologie.hu-berlin.de/

Abstract. We apply Slow Feature Analysis (SFA) to image sequences generated from natural images using a range of spatial transformations. An analysis of the resulting receptive fields shows that they have a rich spectrum of invariances and share many properties with complex and hypercomplex cells of the primary visual cortex. Furthermore, the dependence of the solutions on the statistics of the transformations is investigated.

1 Introduction

In the past years there has been an increasing interest in understanding the computational principles of sensory coding in the cortex. One of the proposed principles is known as *temporal coherence* or *temporal smoothness* [1–4]. It is based on the assumption that the *sources* of the sensory input (e.g. objects in vision) vary on a slower time scale than the sensory signals themselves, which are highly sensitive even to small transformations of the sources (e.g. rotation or translation). By extracting slow features from the raw input one can recover information about the sources independently of these irrelevant transformations. We focus here on vision and apply Slow Feature Analysis (SFA) [4, 5] to image sequences for a comparison with receptive fields of cells in the primary visual cortex V1.

2 Methods

The problem of extracting slow signals from time sequences can be formally stated as follows: given an input signal $\mathbf{x}(t) = (x_1(t) \dots x_N(t))$, $t \in [t_0, t_1]$ and a set of real-valued functions $\mathcal{F}$, find a function $\mathbf{g}(\mathbf{x}) = (g_1(\mathbf{x}), \dots, g_M(\mathbf{x})), g_j \in \mathcal{F}$ so that for the output signals $y_j(t) := g_j(\mathbf{x}(t))$

$$\Delta(y_j) := \langle \dot{y_j}^2 \rangle \quad \text{is minimal} \tag{1}$$

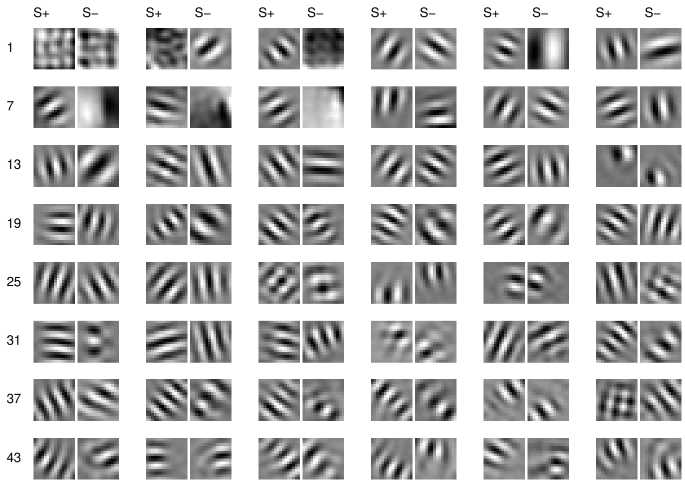

Fig. 1. Optimal excitatory and inhibitory stimuli for Run #1.

$$\text{under the constraints:} \quad \langle y_j \rangle = 0 \quad \text{(zero mean)}\,, \tag{2}$$

$$\langle y_j^2 \rangle = 1 \quad \text{(unit variance)}\,, \tag{3}$$

$$\forall i < j, \quad \langle y_i y_j \rangle = 0 \quad \text{(decorrelation)}\,, \tag{4}$$

with $\langle . \rangle$ indicating time averaging. Here we choose $\mathcal{F}$ to be the set of all polynomials of degree 2 and use Slow Feature Analysis (SFA) [4, 5] to find the optimal input-output functions $g_j(\mathbf{x})$.

Like in electrophysiological studies of neurons in V1, we are interested in characterizing the receptive fields (RFs) of the single components g_j being extracted. We do that by determining the input vector with norm r that maximizes and the one that minimizes the output signal y_j, yielding the optimal excitatory stimulus (S^+) and the optimal inhibitory stimulus (S^-) (cf. Fig. 1). We choose r to be the mean norm of the training input vectors, since we want the optimal stimuli to be representative of the typical input.

Of interest are also the invariances learned by the system, which correspond to the directions in which a variation of S^+ has the least effect on the output. We extract them by computing the Hesse matrix of the function $g_j(\mathbf{x})$ restricted to the sphere of radius r in $\mathbf{x} = S^+$and then choosing the directions corresponding to the smallest eigenvalues. For visualization we move the S^+ vector on the sphere of points with norm r in the direction of the invariance, thereby producing image sequences as those shown in Figure 2 [1].

[1] Animations corresponding to the image sequences shown in Figure 2 can be found at `http://itb.biologie.hu-berlin.de/~berkes/ICANN02/results.html`

Training data were taken from 36 natural images from van Hateren's natural stimuli collection and preprocessed as suggested in the original paper [6]. The end resolution was 2 minutes of arc per pixel. We produced input sequences by choosing an initial position in a randomly chosen image, cutting a 16×16 pixels patch (ca. 0.5×0.5 degrees of arc) and moving it around according to different transformations: translation, rotation and zoom. The transformations were performed simultaneously, so that each frame differed from the previous one by position, orientation, and magnification. Patches were computed by bilinear interpolation. In the default settings, the translation speed v was chosen uniformly between 1 and 5 pixel/frame, the rotation speed ω between 0 and 0.1 rad/frame and the magnification factor z between 0.98 and 1.02 per frame. The parameters were varied every 30 frames, for a total of 150,000 frames per simulation.

A run with these settings requires the computation of a covariance matrix having $\mathcal{O}(N^4)$ elements, where N is the input dimension. We thus performed a standard preprocessing step using PCA to reduce the dimensionality of the input patches to 50 components.

3 Results

3.1 Receptive field analysis

Figure 1 shows the $\mathrm{S}^+/\mathrm{S}^-$ pairs of the first 48 components for a typical run in decreasing order of temporal slowness. Notice that these optimal stimuli cannot be interpreted as linear filters and only give a first hint at the response properties of the RFs. Additional information can be extracted from their invariances; examples of the main types of invariances found are shown in Figure 2. The analysis of the units led to the following observations:

Gabor-like optimal stimulus The optimal excitatory stimulus for most of the components looked like an oriented Gabor wavelet and tended to be localized, i.e. it did not fill the entire patch. This property is linked to rotation and zoom, as localized receptive fields are less sensitive to them (since they are less localized in the Fourier space). If rotation and zoom are absent, localization disappears (cf. Sec. 3.2).

Phase invariance The first invariance for almost all analyzed units was *phase shift* invariance (Fig. 2a). In fact, the response of the units never dropped by more than 20% changing the phase of S^+. As a consequence the units responded well to an oriented edge but were not selective for its exact position, and thus matched the properties of *complex cells* in V1 [7]. Some units showed in addition to the phase insensitive response a phase sensitive response at a lower frequency and different orientation. These cells thus showed complex cell as well as *simple cell* behavior, which might be difficult to detect in an experimental situation, since the optimal simple and complex cell responses have different frequency *and* orientation. The simple cell component, being linear, was relatively stronger for stimuli of low contrast. The clear dichotomy between simple and complex cells in V1 has already been questioned in the experimental literature (see [8] for a discussion).

Complex cells with phase invariant optimal responses to Gabor stimuli have been modeled earlier [3, 9]. In addition we found properties that have to our knowledge not yet been reproduced through unsupervised learning in a computational model:

End-inhibition Some S^+ only filled one half of the patch, while the missing half was filled by S^- (e.g. Units 28, 41, 44 in Fig. 1). These units responded to edges with a specific orientation in the S^+ half of the patch but failed to respond if the stimulus was extended into the S^- half. Complex cells with this behavior are called *end-inhibited* or *hypercomplex cells* [7].

Orientation tuning The optimal inhibitory stimulus was typically also a wavelet. Its orientation was often non-orthogonal to the preferred one (e.g. Units 14, 15, 24 in Fig. 1) resulting in sharpened or bimodal orientation tuning. This is often the case for cells in V1 and thought to contribute to the orientation selectivity of complex cells [10]. On the other side, some units showed invariance to orientation changes (Fig. 2e) and had a broad orientation tuning, a feature also found in V1 [10].

Frequency tuning Similar mechanisms can lead to sharp frequency tuning. In some units S^- had the same orientation and shape as S^+, but had a different frequency (e.g. Units 20, 36, 40 in Figure 1). Such units reacted to a change in frequency by an abrupt drop in their response, in contrast to other units which showed an invariance to frequency changes (Fig. 2d). This suggests that our results can account for the wide range of spatial frequency bandwidths found in complex cells [11].

Non-oriented receptive fields The first two units in all 6 runs with these settings responded to the mean and the squared mean pixel intensity, as was inferred directly from the learned functions. These units are comparable to the *tonic cells* described in [12]. A few other units responded to edges and had phase invariance but showed identical responses for all orientations. A small percentage of the neurons in V1 consists of *non-oriented complex cells* with these characteristics [10].

Additional invariances For each unit we found 4 to 7 highly significant invariances. These included the invariances mentioned above (phase shift, orientation change and frequency change) and also change in position, size, and curvature of the S^+ wavelet (Fig. 2b,c,f). Units corresponding to faster signals showed also more complex invariances, which were sometimes difficult to interpret. Some of the units were found to be responsive to corners or T-shaped stimuli (Fig. 2g-h).

Although the precise shape and order of the components can vary in different simulations, we observed a systematic relationship between the slowness of the output of a unit and its behavior. For example, slower functions have usually a non-structured or orthogonal S^-, while inhibition at non-orthogonal orientations was typical for units with a faster output. It is possible that this kind of dependence also holds for neurons in V1.

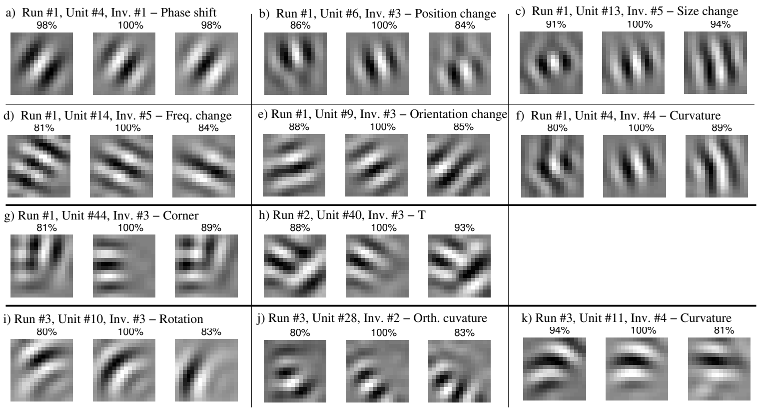

Fig. 2. Invariances. The central patch is S^+, while the other are produced by applying an invariance as described in Section 2. Each patch elicits in the considered unit the percent of the maximal output indicated over it.

3.2 The role of transformations

The characteristics of optimal stimuli and invariances apparently depend more on the statistics of the transformations than on the statistics of the images. Results similar to the ones described above were found by replacing the natural images by a colored noise image with a natural power spectrum of $1/f^2$. However, when changing the transformation settings during training we found significant differences:

- If we increased the influence of rotation by choosing ω to lie between 0.15 and 0.25 rad/frame, many optimal stimuli assumed a curved shape and often included in their invariance spectra (in addition to the previously discussed transformations) rotation and curvature changes both parallel and orthogonal to the orientation of the wavelet (Fig. 2i-k). Wavelets of this shape have been proposed in context of object and face recognition as *banana wavelets.* [13].
- Units in simulations involving only translation did not show any localization and had fewer invariances.
- Optimal stimuli in simulations involving only rotations or only zooms did not look like Gabor wavelets anymore, but assumed a circular and a star-like shape, respectively.

4 Conclusion

We have shown that SFA applied to image sequences generated by translation, rotation, and zoom yields a rich repertoire of complex cell properties. We found receptive fields with optimal stimuli in the shape of Gabor wavelets and invariance to wavelet phase. These properties were found also in earlier modeling studies (e.g. [3, 9]). However, since SFA provides a more general functional architecture than these earlier models, we were able to reproduce additional complex cell properties, such as end-inhibition, inhibition at non-orthogonal orientations, inhibition at different frequencies, and non-oriented receptive fields. The units also showed additional invariances, such as invariances with respect to position, size, frequency, orientation, and/or curvature. Our experiments suggest that there could be a relation between the slowness of the output of complex cells and their behavior. They also suggest that some complex cells could exhibit a simple cell behavior at non-optimal frequency and orientation, particularly at low contrast. It is remarkable that the temporal smoothness principle is able to reproduce so many different properties found in the primary visual cortex, which indicates that it might be an important learning principle in cortex in general.

Supported by a grant from the Volkswagen Foundation.

References

1. Földiák, P.: Learning invariance from transformation sequences. Neural Computation **3** (1991) 194–200
2. Stone, J.V.: Learning perceptually salient visual parameters using spatiotemporal smoothness constraints. Neural Computation **8** (1996) 1463–1492
3. Kayser, C., Einhäuser, W., Dümmer, O., König, P., Körding, K.: Extracting slow subspaces from natural videos leads to complex cells. In: Artificial Neural Networks - ICANN 2001 Proceedings, Springer (2001) 1075–1080
4. Wiskott, L., Sejnowski, T.: Slow feature analysis: Unsupervised learning of invariances. Neural Computation **14** (2002) 715–770
5. Wiskott, L.: Learning invariance manifolds. In: Proc. Computational Neuroscience Meeting, CNS'98, Santa Barbara. (1999) Special issue of Neurocomputing, 26/27:925–932.
6. van Hateren J.H., van der Schaaf A.: Independent component filters of natural images compared with simple cells in primary visual cortex. Proc. R. Soc. Lond. B (1998) 359–366
7. Hubel, D., Wiesel, T.: Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. Journal of Physiology **160** (1962) 106–154
8. Mechler, F., Ringach, D.L.: On the classification of simple and complex cells. Vision Research (2002) 1017–1033
9. Hyvärinnen, A., Hoyer, P.: Emergence of phase and shift invariant features by decomposition of natural images into independent features subspaces. Neural Computation **12** (2000) 1705–1720
10. De Valois, R., Yund, E., Hepler, N.: The orientation and direction selectivity of cells in macaque visual cortex. Vision Res. **22** (1982) 531–44
11. De Valois, R., Albrecht, D., Thorell, L.: Spatial frequency selectivity of cells in macaque visual cortex. Vision Res. **22** (1982) 545–559
12. Schiller, P., Finlay, B., Volman, S.: Quantitative studies of single-cell properties in monkey striate cortex. I. Spatiotemporal organization of receptive fields. J. Neurophysiol. **39** (1976) 1288–1319
13. Krüger, N., Peters, G.: Object recognition with banana wavelets. In: Proceedings of the ESANN97. (1997)