Spatiotemporal Receptive Fields Maximizing Temporal Coherence in Natural Image Sequences

Jarmo Hurri and Jaakko Väyrynen and Aapo Hyvärinen

Neural Networks Research Centre
Helsinki University of Technology, P.O.Box 9800, 02015 HUT, Finland

Article Summary

1 Introduction

The relationship between the structure and functionality of the visual system and the properties of natural visual stimuli is an active research topic in computational visual neuroscience. It has previously been shown that maximization of temporal coherence of activity levels is one computational principle which leads to the emergence of simple-cell-like spatial receptive fields from natural image sequences [2].

In this paper we extend previous results by examining the case of *spatiotemporal* receptive fields. We show that application of the same principle of temporal coherence in the spatiotemporal case yields receptive fields which are not only localized, oriented and multiscale, as in the spatial case, but also have temporal characteristic similar to spatiotemporal simple-cell receptive fields. Quantitative measurements of the temporal properties of the resulting receptive fields are also provided, and these are compared against similar results obtained with independent component analysis.

2 Definition of temporal coherence

A vectorization of spatiotemporal filtering can be done by scanning the frames of an image sequence one by one column-wise into a vector. Let a vectorized sequence of 8 frames of size 11×11 pixels, taken from natural video at time

Email addresses: jarmo.hurri@hut.fi (Jarmo Hurri), jaakko.j.vayrynen@hut.fi (Jaakko Väyrynen), aapo.hyvarinen@hut.fi (Aapo Hyvärinen).

t, be denoted by the 968-dimensional ($= 8 \times 11^2$) vector $\mathbf{x}(t)$. Let $\mathbf{y}(t) = [y_1(t) \cdots y_K(t)]^T$ represent the outputs of K simple cells. We use the standard linear receptive field model

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t), \tag{1}$$

where the set of filters (vectors) $\mathbf{w}_1, ..., \mathbf{w}_K$ corresponds to the receptive fields of simple cells, and $\mathbf{W} = [\mathbf{w}_1 \cdots \mathbf{w}_K]^T$ denotes a matrix with all the filters as rows.

Temporal response strength correlation [2], the objective function, is defined by

$$f(\mathbf{W}) = \sum_{k=1}^{K} \mathrm{E}_t \left\{ g(y_k(t)) g(y_k(t - \Delta t)) \right\}, \tag{2}$$

where the nonlinearity g is strictly convex, even (rectifying), and differentiable, and Δt denotes a delay in time. Examples of choices for the nonlinearity are $g_1(x) = x^2$ and $g_2(x) = \ln \cosh x$. A set of filters which has a large temporal response strength correlation is such that the same filters *often respond strongly at consecutive time points*, outputting large (either positive or negative) values. This means that the same filters will respond strongly over short periods of time, thereby expressing temporal coherence of a population code. To keep the outputs of the filters bounded we enforce the unit variance constraint on each of the output signals $y_k(t)$, and to keep the filters from converging to the same solution we force their outputs to be uncorrelated. A gradient projection method can be used to maximize (2) under these constraints. The initial value of $\mathbf{W}$ is selected randomly.

3 Data collection and preprocessing

The data used in the experiments was sampled from the database of natural image sequences described in [3]. The sampled data consisted of 120,000 image sequence blocks of size $11 \times 11 \times 9$, where the first two dimensions denote spatial size and the last dimension denotes length in time. Each sample of length 9 was divided into two overlapping samples of length 8 to yield $\mathbf{x}(t)$ and $\mathbf{x}(t - \Delta t)$. The sampling frequency of the data was 25 Hz, so the duration of $\mathbf{x}(t)$ and $\mathbf{x}(t - \Delta t)$, and the spatiotemporal filter, was 320 ms, and Δt was 40 ms. Preprocessing consisted of removal of local mean (DC component of the spatiotemporal block) and dimensionality reduction by 50% to 484 (this retains 95% of original signal energy).

4 Results

Some of the resulting $11 \times 11 \times 8$ spatiotemporal filters maximizing objective function (2) are shown in Figure 1. As can be seen, the emerged receptive fields

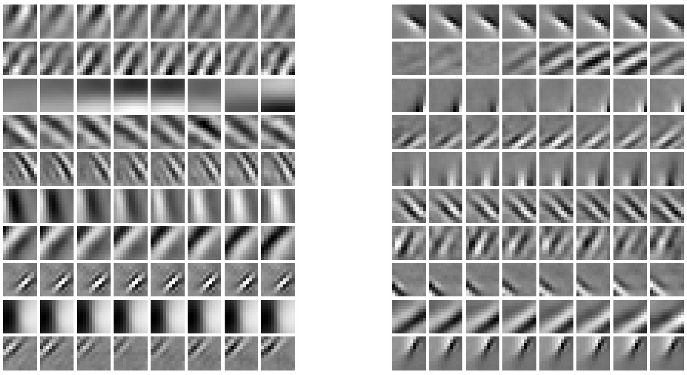

Fig. 1. A subset of 20 spatiotemporal receptive fields obtained by maximizing temporal coherence of activity levels in natural image sequences (10 receptive fields in the image on the left and 10 on the right). Each of the 20 rows corresponds to one spatiotemporal receptive field, and each frame in the row corresponds to the spatial receptive field profile of the spatiotemporal filter at a certain time instance.

share the primary spatial properties of simple cells in that they are localized, oriented, and have multiple scales. In addition to these spatial properties, the receptive fields also have physiologically relevant temporal properties. Some of the cells are space-time separable [1], while others are inseparable. [1] Some of the separable receptive fields have constant time profiles, some of them have changing time profiles. Also, different space-time inseparable receptive fields respond to different velocities.

Measurements of mammalian primary visual cortex indicate that the spatiotemporal receptive fields of different simple cells also have different temporal characteristics [1], as in our results. The final version of this paper contains quantitative measurements of these properties from our results, and comparisons against physiological measurements and similar results obtained with independent component analysis.

References

[1] G. C. DeAngelis, I. Ohzawa, and R. D. Freeman. Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *Journal of Neurophysiology*, 69(4):1091–1117, 1993.

[2] J. Hurri and A. Hyvärinen. Simple-cell-like receptive fields maximize temporal coherence in natural video. *Neural Computation*, 15(3), 2003.

[3] J. H. van Hateren and D. L. Ruderman. Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proceedings of the Royal Society of London B*, 265(1412):2315–2320, 1998.

[1] A space-time separable receptive field can be expressed as a product of a one-dimensional temporal profile and a two-dimensional spatial profile.