Independent Component Analysis of Natural Image Sequences Yields Spatio-Temporal Filters Similar to Simple Cells in Primary Visual Cortex

J. H. van Hateren; D. L. Ruderman

Proceedings: Biological Sciences, Vol. 265, No. 1412. (Dec. 7, 1998), pp. 2315-2320.

Stable URL:
http://links.jstor.org/sici?sici=0962-8452%2819981207%29265%3A1412%3C2315%3AICAONI%3E2.0.CO%3B2-5

Proceedings: Biological Sciences is currently published by The Royal Society.

THE ROYAL SOCIETY

Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex

J. H. van Hateren[1*] **and D. L. Ruderman**[2]

[1]*Department of Neurobiophysics, University of Groningen, Nijenborgh 4, NL-9747 AG Groningen, The Netherlands*
[2]*Sloan Center for Theoretical Neurobiology, Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA*

Simple cells in the primary visual cortex process incoming visual information with receptive fields localized in space and time, bandpass in spatial and temporal frequency, tuned in orientation, and commonly selective for the direction of movement. It is shown that performing independent component analysis (ICA) on video sequences of natural scenes produces results with qualitatively similar spatio-temporal properties. Whereas the independent components of video resemble moving edges or bars, the independent component filters, i.e. the analogues of receptive fields, resemble moving sinusoids windowed by steady Gaussian envelopes. Contrary to earlier ICA results on static images, which gave only filters at the finest possible spatial scale, the spatio-temporal analysis yields filters at a range of spatial and temporal scales. Filters centred at low spatial frequencies are generally tuned to faster movement than those at high spatial frequencies.

Keywords: natural scenes; independent component analysis; information theory; simple cells; visual cortex

1. INTRODUCTION

It has been argued (Barlow 1972, 1989; Field 1987; Zetzsche *et al.* 1990) that the visual cortex uses a factorial code to represent the visual environment. Such a code uses components in the visual scene that occur independently of each other for large ensembles of scenes, and thus are more easily processed when coinciding in a particular scene. Recently, evidence has accumulated that a factorial code may indeed capture part of the function of the primary visual cortex: research on sparse coding (Field 1994; Olshausen & Field 1996, 1997; Harpur 1997) and independent coding (Bell & Sejnowski 1997*a*,*b*; van Hateren & van der Schaaf 1998) of natural images shows the emergence of spatial receptive fields with properties similar to those of simple cells in the primary visual cortex of mammals. These studies are, however, limited to the space domain, and the question is how well this line of research predicts the full spatio-temporal receptive fields of simple cells.

Here we investigate this question by applying independent component analysis (ICA) to video sequences of natural scenes. The ICA technique (Comon 1994) assumes that each particular signal of an ensemble of signals is a superposition of elementary components that occur independently of each other. The technique is different from principal components analysis (PCA) because it also imposes higher-order independence, not just up to the second order (decorrelation) as in PCA. For images, PCA generally leads to global components (cf. Fourier components), whereas ICA leads to localized components (cf. wavelets).

Algorithms performing linear ICA (Bell & Sejnowski 1995; Hyvärinen & Oja 1997) try to find those components ('ICs') that are as independent of each other as possible by a linear transformation of the signals. For example, if the signal is an image patch $I(x, y)$, it can then be represented as a sum of the ICs

$$I(x, y) = \sum_i a_i C_i(x, y), \tag{1}$$

where a_i is the amplitude of independent component $C_i(x, y)$. This amplitude is extracted by a corresponding independent component filter $F_i(x, y)$:

$$a_i = \sum_{x,y} F_i(x, y) I(x, y). \tag{2}$$

If all image pixels $I(x, y)$ are concatenated, and collected into a single vector $\boldsymbol{I}$, and all amplitudes a_i into a vector $\boldsymbol{a}$, then equation (1) can be written as

$$\boldsymbol{I} = M_{\mathrm{IC}} \boldsymbol{a}, \tag{3}$$

with M_{IC} a matrix containing the independent components $C_i(x, y)$ as its columns. Similarly, equation (2) can be written as

$$\boldsymbol{a} = M_{\mathrm{ICF}} \boldsymbol{I}, \tag{4}$$

with M_{ICF} a matrix containing the independent component filters $F_i(x, y)$ as its rows. From equations (3) and (4) we see that the matrices M_{IC} and M_{ICF} are each other's

*Author for correspondence (hateren@bcn.rug.nl).

Received 4 August 1998 *Accepted* 12 August 1998

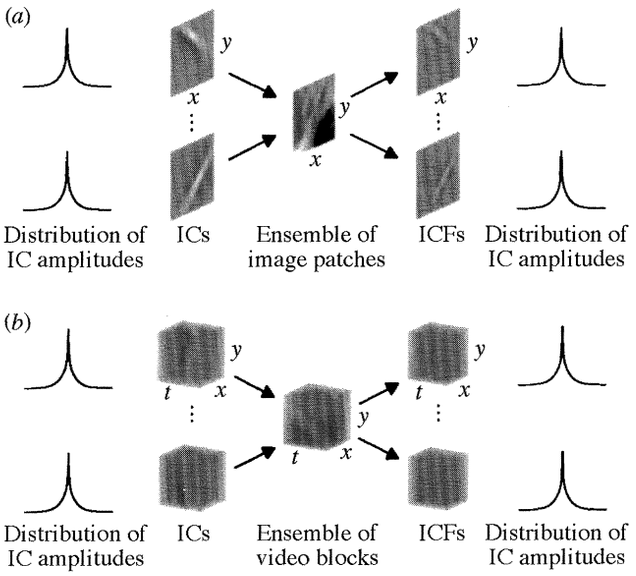

Figure 1. Independent component analysis (ICA) of images and video. (*a*) ICA performed on an ensemble of image patches yields both a set of independent components (ICs) and a set of independent component filters (ICFs). The ICs are the building blocks of image patches, and occur according to a sparse distribution of amplitudes (symbolized by the line plots, showing a high peak at amplitude zero, and long tails). The amplitude of an IC in a patch can be determined by applying the corresponding ICF to the patch. (*b*) ICA performed on video blocks produces spatio-temporal ICs and ICFs, denoted in the figure by a stack of 12 consecutive time-frames (denoted by t), each consisting of 12×12 pixels (denoted by x and y).

inverse, thus directly relating the ICs to the independent component filters (ICFs). Whereas the ICs can be regarded as the basic building blocks of a signal (as in equation 1), the ICFs are used for analysing a signal by determining how strongly each building block is present (as in equation 2).

Figure 1(*a*) illustrates this for an ensemble of image patches (centre). Each of these patches can be decomposed (as in equation 1) into a set of ICs, two of which are shown in the figure (centre left). The amplitudes with which a particular IC is present in the patches of the ensemble are maximally independent of those of the other ICs. For natural images, this requirement produces a sparse distribution of amplitudes for each IC (figure 1*a*, left), where the probability of small amplitudes is high (the central peak), but large amplitudes occur as well (the tails). For extracting the amplitude of an IC, the corresponding ICF (centre right) is applied (as in equation 2) to an image patch. Applying a particular ICF to the entire ensemble again produces the original distribution of IC amplitudes (figure 1*a*, right). Note that an ICF differs somewhat from the IC for which it gives the amplitude. This is because the ICs overlap strongly (they are not orthogonal) and the ICFs solve this by concentrating their power in those areas in spatio-temporal frequency space where the overlap of ICs is minimal and the amplitude of an IC can be obtained without interference from others.

For time-varying images (video sequences) the analysis is very similar (figure 1*b*), where the ensemble of image patches is replaced by an ensemble of video blocks, each consisting of a stacked array of time frames. Equations (3) and (4) still apply, where the vector $\boldsymbol{I}$ is now the concatenation of all voxels $I(x, y, t)$ of a video block. The ICs (centre left) are the spatio-temporal constituents of video blocks, and the ICFs are again used for extracting their amplitudes. If analysing the underlying structure of time-varying images is a primary task of simple cells in the visual cortex, they perform an operation similar to that of the ICFs: their receptive fields weight the stimulus such that it yields the strength of each component present. Therefore, the ICFs (and not the ICs) can be compared with simple cells. This comparison should not be taken too far, however. First, the linear IC model for natural scenes can only be approximate (Bell & Sejnowski 1997*a*,*b*; van Hateren & van der Schaaf 1998), because objects in scenes do not superimpose linearly, but by occlusion (Ruderman 1997). Thus the ICA technique merely serves as a heuristic for maximizing independence. Nevertheless, it appears that even linear ICA is dominated by the occurrence of occlusion: in preliminary ICA experiments on constructed static images we found that both occlusion and independent statistics (variances) of the occluding surfaces are necessary to obtain edge detectors. A second reason to be careful when comparing ICFs with simple cells is that responses of simple cells are not expected to be completely independent of each other, because they appear to achieve a strongly overcomplete representation. This may be partly related to the fact that single neurons are noisy, and have a limited information capacity. Some form of (local) ensemble coding may therefore be necessary. Thus the ICFs should be viewed only as generic receptive fields, realizing a code that presumably underlies the actual neural implementation.

2. METHODS

Broadcasts from Dutch, British and German television were recorded on S-VHS tape and digitized with a JVC BR-S622E VCR and a Data Translation DT3851-8 frame grabber. Half-frames were spatially aligned by interpolation as in Dong & Atick (1995). Final resolution was reduced, by block-averaging, to 128×128 pixels at 25 frames s^{-1}. The total database consisted of 216 video clips of 192 s each, on a wide range of subjects, including sports, films, and wildlife (more than half of the data).

The ICA algorithm was the fixed-point algorithm of Hyvärinen & Oja (1997) with serial deflation scheme, using function g_2 as in Hyvärinen (1997). It was implemented on a Cray J932, and performed on 524 288 different $12 \times 12 \times 12$ video sequences, drawn randomly from the database, with intensities logarithmically transformed (van Hateren & van der Schaaf 1998). Each $12 \times 12 \times 12$ video sequence consisted of a stack of 12 time-frames of 12×12 image patches. This was extracted from the larger videos by fixing a random starting time and patch position, and then extracting the patches at that position from 12 consecutive time frames. The ICA yielded approximately 15 ICs per day, and this made it impracticable to calculate the entire basis set of 1728 ICs. Instead, the analysis was performed on the first 288 ICs recovered. With different random selections of video sequences, or with slight scaling of the data ($2 \times 2 \times 2$ block averaging), similar results were obtained. ICs and ICFs, as shown in figure 2, were normalized to their maximum.

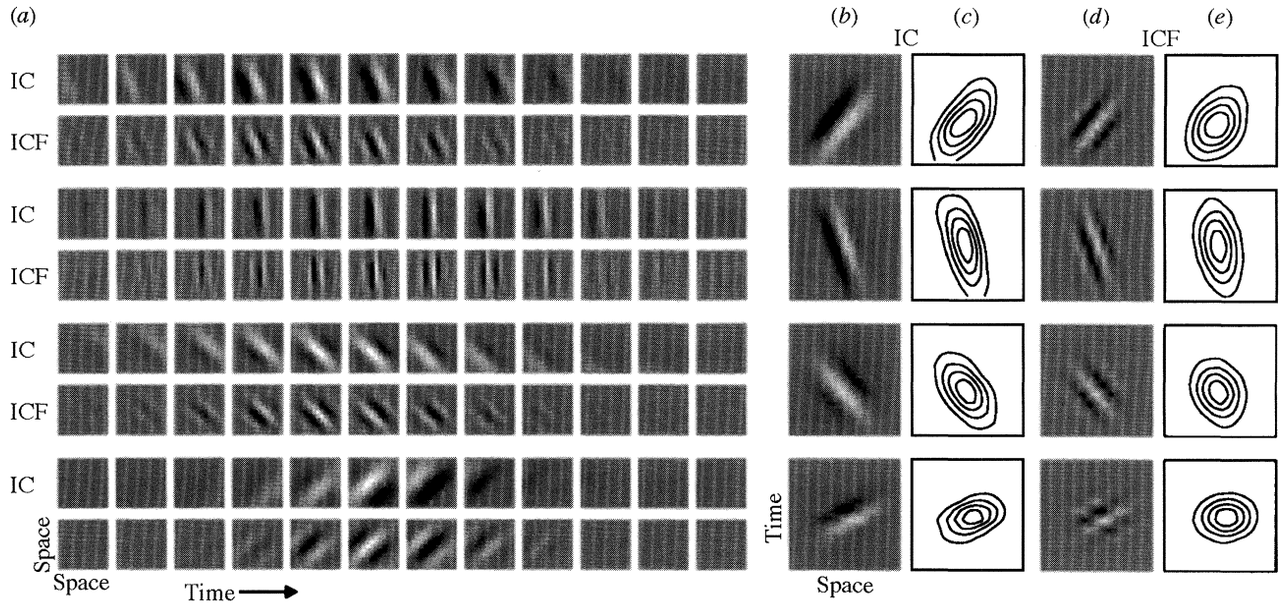

Figure 2. Four examples of an IC and the corresponding ICF. (*a*) Odd rows: ICs, shown as 12 consecutive time-frames each of 12×12 spatial pixels. Grey denotes zero, black negative, and white positive values. Even rows: corresponding ICFs. (*b*) Space–time diagrams of the ICs shown in (*a*), constructed by projecting the IC in each time-frame onto the direction of propagation; the result was slightly smoothed for presentation purposes, just enough to blur the edges of the constituent pixels. (*c*) Envelopes of (*b*), calculated as the magnitude of the vector sum of each spatial profile of (*b*) and its Hilbert transform (DeAngelis *et al.* 1993). Contours are drawn at 0.2, 0.4, 0.6, 0.8 and 1.0 of the maximum. (*d*) Space–time diagrams of the ICFs shown in (*a*). (*e*) Envelopes of (*d*).

As only a limited set of ICs was calculated for the $12 \times 12 \times 12$ video data, it is possible that, given the particular ICA algorithm used, the properties of this set are not representative of the complete set. As a control, we therefore computed a complete set of ICs for $8 \times 8 \times 8$ video data. The results are qualitatively consistent with those of $12 \times 12 \times 12$ video, and there was no significant change in IC properties depending on the order of recovery. This was also observed in earlier work on spatial ICA (van Hateren & van der Schaaf 1998). As the $8 \times 8 \times 8$ ICA is more strongly influenced by boundary effects, and spans only a very limited range of spatial and temporal frequencies, further analysis is limited to the results for $12 \times 12 \times 12$ video.

3. RESULTS

ICA on digitized video of natural scenes yields ICs that resemble edges or bars, moving with a fixed velocity perpendicularly to their main axis of orientation (figure 2*a*, odd rows). The corresponding ICFs (even rows) move with similar velocity, but at higher spatial and temporal frequencies. Both ICs and ICFs are confined to a particular region in space and time, but their behaviour within these spatial and temporal envelopes is different. Whereas the ICs move as a whole (identical group and phase velocity), the ICFs consist of an undulation moving within a relatively steady envelope (group velocity small compared with phase velocity). The spatio-temporal behaviour of ICs and ICFs can be summarized by integrating them along their main axis of spatial orientation, i.e. by projecting them onto the direction of propagation (Adelson & Bergen 1985; DeAngelis *et al.* 1993). The resulting space–time diagrams (figure 2*b*,*d*) show how the spatial profile develops as a function of time. The identical slopes in these diagrams indicate an identical velocity of the undulations of corresponding ICs and ICFs. The envelopes of the space–time diagrams of the ICFs (figure 2*e*), however, have a position that is more steady as a function of time than those of the ICs (figure 2*c*), consistent with the steady ICF envelope noted above.

The spatio-temporal properties of the ICFs closely resemble those of receptive fields measured in simple cells (DeAngelis *et al.* 1993). Both are confined to a limited region in space and time and both resemble an undulation travelling through a steady envelope. Moreover, both are found for a range of velocities, spatial frequencies, and spatial scales. However, there are also differences. First, different ICFs are centred at different positions and times. For visual space, this is similar to how the visual cortex is organized: the finite extent of (angular) visual space can be covered with a finite number of neurons at different positions. For visual time, however, it implies coding with many cells, each only responding if a stimulus occurs at a specific time. As time is unbounded, this would be an unwieldy processing strategy. In practice, ICFs that are spatially identical but centred at different times can be combined into a single continuously acting filter (with a proper delay to comply with causality). A second difference with the receptive fields of simple cells is that ICFs are more symmetrical in time and more narrowly tuned in temporal frequency (DeAngelis *et al.* 1993).

Figure 3(*a*,*b*) shows an IC and ICF in frequency space, with the surface showing points where the amplitude is at half maximum. The ICF is shifted towards higher spatial and temporal frequencies than the IC (median ratios 1.76

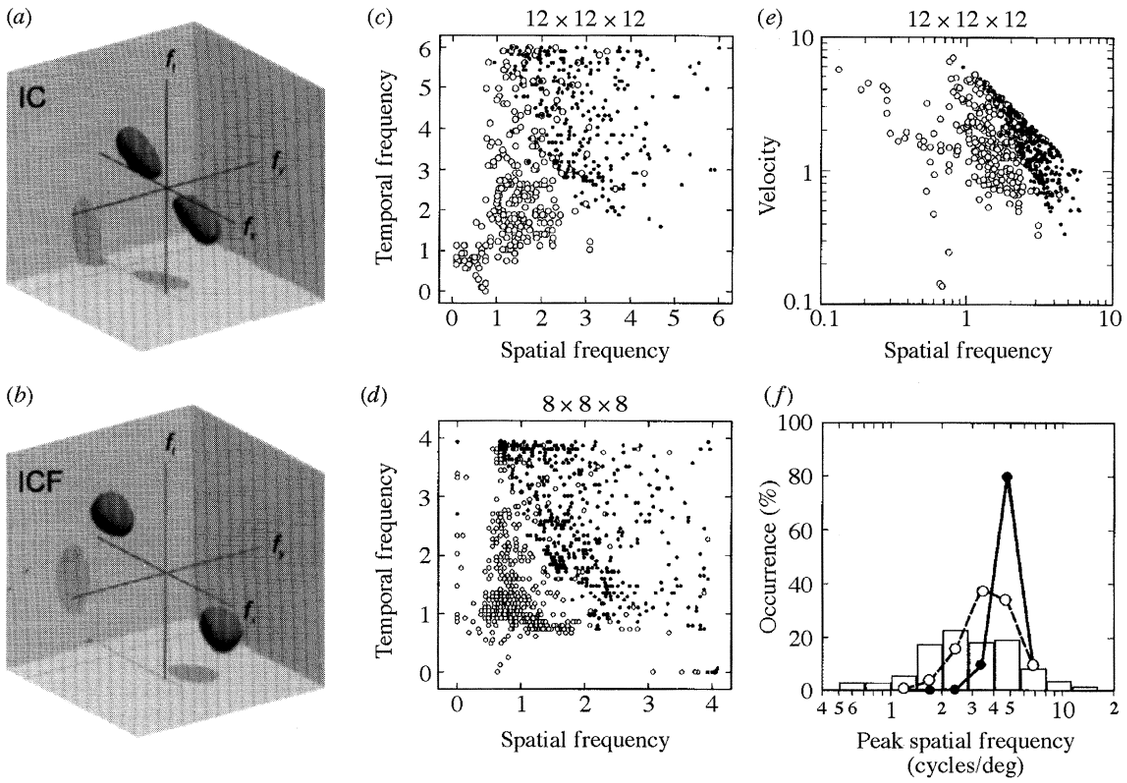

Figure 3. Spatial and temporal frequency analysis of ICs and ICFs. (*a*) Amplitude spectrum of an IC, showing the surface at half maximum; f_x and f_y are spatial frequencies (cycles per 12 pixels), and f_t temporal frequency (cycles per 12 video frames). (*b*) Amplitude spectrum of the ICF corresponding to (*a*). (*c*) Positions of the peaks in frequency space of 288 ICs (open circles) and ICFs (dots). Spatial frequency (f_s) is the magnitude of the vector sum of f_x and f_y. Frequency resolution was enhanced by performing a $128 \times 128 \times 128$-point FFT with zero-padding of the $12 \times 12 \times 12$ data. (*d*) As (*c*), all 512 ICs and ICFs of $8 \times 8 \times 8$ video data. (*e*) Peaks in spatial frequency–velocity ($= f_t/f_s$) space. Details as in (*c*). (*f*) Distribution of the spatial frequencies at the peak in the amplitude spectrum of ICFs for single video frames (solid line), $12 \times 12 \times 12$ video sequences (dashed line), and measurements of simple cells in the foveal area of macaque primary visual cortex (histogram, from DeValois *et al.* 1982). The ICF histograms have an arbitrary position along the spatial frequency axis, because the video data do not have a fixed spatial scale. Thus only the relative widths of the distributions can be compared. Note that for the $12 \times 12 \times 12$ video, a total ratio of only six is possible for the peak spatial frequency; thus the width of the calculated histogram (dashed line) may be underestimated.

and 1.71, respectively, for all $12 \times 12 \times 12$ ICs and ICFs that were computed), but remains at the same orientation (correlation coefficient $r = 0.97$) and velocity ($r = 0.94$). Figure 3(*c*) shows the shift of frequencies of ICFs (dots) relative to ICs (open circles). As can be seen, the ICFs do not tile the spatio-temporal frequency space by more or less uniformly filling a sphere (as in the tiling recently proposed by Simoncelli & Heeger (1998)). Instead, they are concentrated mostly in an outer shell of such a sphere. This shift of frequency changes the correlation between spatial and temporal frequencies from slightly positive for the ICs ($r = 0.35$) to slightly negative for the ICFs ($r = -0.23$). Negative correlations ($r = -0.57$ and $r = -0.21$) were also measured in simple cells in the cat (DeAngelis *et al.* 1993; Baker 1990). Qualitatively similar results (figure 3*d*) were obtained for an $8 \times 8 \times 8$ video analysis, where all ICs were calculated rather than a subset as for the $12 \times 12 \times 12$ video. The frequency shift produces an increased correlation between spatial frequency and velocity for ICFs (figure 3*e*, $r = -0.42$ for ICs, and $r = -0.75$ for ICFs, with a slope of -1.28 for the logarithmic regression line; cf. $r = -0.84$ and a slope of -1.24 for cat simple cells (DeAngelis *et al.* 1993)). Thus ICFs centred at low spatial frequencies are generally tuned to faster movement than those at high spatial frequencies. As a result, there is a much higher proportion of filters at low spatial frequencies for moving images than for still images, which alleviates a lingering discrepancy between the properties of spatial ICFs and simple-cell receptive fields (van Hateren & van der Schaaf 1998). Figure 3(*f*) compares a histogram of the peak spatial frequency measured in macaque simple cells (DeValois *et al.* 1982) with ICFs of single video frames (solid line) and video sequences (dashed line) taken from the same data set. Although the histogram for the video sequences (dashed line) is still not as broad as the measurements, it should be remarked that its width is probably an underestimate, because for $12 \times 12 \times 12$ video only a total ratio of six is possible for the range of peak spatial frequencies (i.e. for horizontal and vertical orientations). The curve for single video frames is consistent with results for still images (van Hateren & van der Schaaf 1998), suggesting that it is not motion blur in the video that causes an increase in ICFs and ICs at low spatial frequencies, but rather space–time correlations introduced by the movement.

In control experiments with video computed by sliding a window at constant speed over still images, we found that video sequences with a single velocity produced ICs and ICFs moving at exactly that velocity. The results deviate from those for natural video in several ways: first, the ICs and ICFs have their main axis of orientation not exclusively perpendicularly to the direction of motion; second, ICFs are all at a single, high spatial frequency; and third, the ICs and the ICFs extend over all frames, i.e. they lack a temporal envelope. The first two results follow directly from ICA on still images, because spatio-temporal ICA of video with a single velocity is a degenerate case, which can be reformulated as a two-dimensional, spatial ICA. An ICA on sequences with a more natural movement distribution (continuously distributed random directions, and a velocity distribution $\sim 1/(v+v_0)^2$, with v angular velocity and v_0 a constant; see van Hateren (1993) and Dong (1997)) yielded perpendicular movement, and ICFs peaking at different spatial frequencies with different velocities. The ICs and ICFs are not as localized in space and time as those of natural video, however, thus it appears that local spatio-temporal structure (such as relative movement within a scene, acceleration, and scene cuts or saccades) is important.

4. DISCUSSION

Because the structure of the ICFs is determined by the statistics of video sequences, we can try to understand their form in terms of natural image composition. At a fundamental level, scenes consist largely of occluding objects undergoing various degrees of motion (Ruderman 1997). Several studies (Bell & Sejnowski 1997*a*,*b*; van Hateren & van der Schaaf 1998) have shown that the independent components of spatial images resemble edges. This suggests that moving edges have an analogous role for the spatio-temporal case. The velocity distribution in natural scenes is continuous, and this means that an ICF tuned to a particular velocity always has to deal with a range of velocities. For high spatial frequencies, even a small range of velocities will produce strongly varying phase relationships between frequency components in consecutive time frames, which makes it difficult to track fast-moving high spatial frequencies. Thus it can be understood that fast-moving ICFs are mainly limited to low spatial frequencies. Given the presence of moving object borders in natural scenes, it is also not unexpected that the ICFs reported here are producing sparse response distributions (i.e. distributions with both high central peaks and long tails). We suggest that the ICFs produce the high central peaks because they are local in space–time, and most of the time measure small pixel differences within a single object. They have long tails in the response distribution because they are able to momentarily track object borders: when the speed and orientation match, they measure large pixel differences between two (or a few) adjacent objects.

Although the properties of the ICFs found (see, for example, figure 2) are qualitatively similar to spatio-temporal receptive fields of cortical simple cells, the present analysis is too limited to make a full quantitative comparison possible. First, the spatial and temporal size of the video blocks used (12 in each dimension) is not large enough to avoid boundary effects completely. Although, for a purely spatial ICA, 12×12 image patches are already reasonably free from boundary effects (Olshausen & Field 1997; Bell & Sejnowski 1997*b*; cf. the 18×18 analysis in van Hateren & van der Schaaf (1998)), the movement in video blocks presumably increases boundary effects. This can only be tested with larger video blocks, but such testing is not possible with the present (already large) database and the available computational power. Furthermore, recent results on spatial ICA (van Hateren & van der Schaaf 1998) indicate that systematic variations in the statistics of different scenes are causing variability in ICF properties which should be taken into account when comparing these with cortical receptive fields. For the same reasons as outlined above, a similar analysis for spatio-temporal ICA is not yet possible.

In a recent study (DeValois & DeValois 1991), patterns called 'moving Gabors' were presented to human observers. These patterns consist of a moving sinusoidal grating windowed by a steady Gaussian envelope, and resemble both the ICFs reported here and the spatio-temporal receptive fields measured in simple cells. It was shown that this stimulus elicits a strong positional illusion: the envelope of the pattern appears to be displaced in the direction of movement of the sinusoid. This illusion can be understood from the results presented here. As already mentioned, the ICFs are different from the ICs, because ICs overlap strongly, and the ICFs need to extract the IC amplitude without interference from other ICs. A strong excitation of a particular ICF signifies the strong presence of the corresponding IC. However, the IC is not the most effective stimulus for the ICF: the strongest response will occur when a matched stimulus is given, i.e. a stimulus identical to the ICF itself (Watson *et al.* 1983; Watson & Turano 1995). Although neighbouring ICFs will also be excited to some extent, noise in the visual system will prevent a faithful representation of the ICF-shaped stimulus: we assume that noise in effect turns the set of ICs into an undercomplete basis for arbitrary video blocks, though still complete—or even overcomplete—for video blocks with natural spatio-temporal statistics. The response of the matching ICF and its neighbours is, however, not interpreted by the visual system as signifying the presence of an ICF-shaped stimulus, but rather the presence of an IC-shaped stimulus. Thus it causes a visual illusion. Fortunately, this will seldom interfere with normal vision, because natural scenes are not composed of ICF-shaped components (an unnaturally moving object with too many fringes), but rather of IC-shaped components (resembling moving edges or bars). Only when the visual system is deliberately presented with an ICF-shaped stimulus, such as a moving Gabor, is the illusion revealed: the steady envelope is misinterpreted as a moving one, which subsequently causes a positional misjudgement.

We thank Arjen van der Schaaf and Herman Snippe for comments. This research was supported by the Netherlands Organization for Scientific Research (to J.H.v.H.), the Alfred P. Sloan Foundation (to D.L.R.), and the Groningen Centre for High-Performance Computing.

REFERENCES

Adelson, E. H. & Bergen, J. R. 1985 Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am.* A **2**, 284–299.

Baker, C. L. Jr 1990 Spatial- and temporal-frequency selectivity as a basis for velocity preference in cat striate cortex neurons. *Vis. Neurosci.* **4**, 101–113.

Barlow, H. B. 1972 Single units and sensation: A neuron doctrine for perceptual psychology? *Perception* **1**, 371–394.

Barlow, H. B. 1989 Unsupervised learning. *Neur. Comput.* **1**, 295–311.

Bell, A. J. & Sejnowski, T. J. 1995 An information-maximization approach to blind separation and blind deconvolution. *Neur. Comput.* **7**, 1129–1159.

Bell, A. J. & Sejnowski, T. J. 1997*a* Edges are the 'independent components' of natural scenes. In *Advances in neural information processing systems*, vol. 9 (ed. M. C. Mozer, M. J. Jordan & T. Petsche), pp. 831–837. Cambridge, MA: MIT Press.

Bell, A. J. & Sejnowski, T. J. 1997*b* The 'independent components' of natural scenes are edge filters. *Vision Res.* **37**, 3327–3338.

Comon, P. 1994 Independent component analysis, a new concept? *Sign. Process.* **36**, 287–314.

DeAngelis, G. C., Ohzawa, I.. & Freeman, R. D. 1993 Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *J. Neurophysiol.* **69**, 1091–1117.

DeValois, R. L. & DeValois K. K. 1991 Vernier acuity with stationary moving Gabors. *Vision Res.* **31**, 1619–1626.

DeValois, R. L., Albrecht, D. G. & Thorell, L. G. 1982 Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res.* **22**, 545–559.

Dong, D. W. & Atick, J. J. 1995 Temporal decorrelation: a theory of lagged and nonlagged responses in the lateral geniculate nucleus. *Network* **6**, 159–178.

Dong, D. W. 1997 Spatiotemporal coupling and scaling of natural images and human visual sensitivities. In *Advances in neural information processing systems*, vol. 9 (ed. M. C. Mozer, M. J. Jordan & T. Petsche), pp. 859–865. Cambridge, MA: MIT Press.

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

Field, D. J. 1994 What is the goal of sensory coding? *Neur. Comput.* **6**, 559–601.

Harpur, G. F. 1997 *Low entropy coding with unsupervised neural networks*. PhD Thesis, University of Cambridge.

Hateren, J. H. van 1993 Spatiotemporal contrast sensitivity of early vision. *Vision Res.* **33**, 257–267.

Hateren, J. H. van & Schaaf, A. van der 1998 Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. R. Soc. Lond.* B **265**, 359–366.

Hyvärinen, A. 1997 A family of fixed-point algorithms for independent component analysis. In *Proc. Institute of Electronic Engineers Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'97)*, pp. 3917–3920. Munich, Germany: IEEE.

Hyvärinen, A. & Oja, E. 1997 A fast fixed-point algorithm for independent component analysis. *Neur. Comput.* **9**, 1483–1492.

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

Olshausen, B. A. & Field, D. J. 1997 Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision Res.* **37**, 3311–3325.

Ruderman, D. L. 1997 Origins of scaling in natural images. *Vision Res.* **37**, 3385–3398.

Simoncelli, E. P. & Heeger, D. J. 1998 A model of neuronal responses in visual area MT. *Vision Res.* **38**, 743–761.

Watson, A. B. & Turano, K. A. 1995 The optimal motion stimulus. *Vision Res.* **35**, 325–336.

Watson, A. B., Barlow, H. B. & Robson, J. G. 1983 What does the eye see best? *Nature* **302**, 419–422.

Zetzsche, C., Barth, E. & Wegmann, B. 1990 The importance of intrinsically two-dimensional image features in biological vision and picture coding. In *Digital images and human vision* (ed. A. Watson), pp. 109–138. Cambridge, MA: MIT Press.