Emergence of simple-cell receptive field properties
of each coefficient's activity passed through a nonlinear function $S(x)$:

$$[\text{sparseness of } a_i] = -\sum_i S\left(\frac{a_i}{\sigma}\right) \tag{4}$$

where $\sigma$ is a scaling constant. The choices for $S(x)$ that we have experimented with include $-e^{-x^2}$, $\log(1+x^2)$ and $|x|$, and all yield qualitatively similar results (described below). The reasoning behind these choices is that they will favour among activity states with equal variance those with the fewest number of non-zero coefficients. This is illustrated in geometric terms in Fig. 2.

Learning is accomplished by minimizing the total cost functional, $E$ (equation (2)). For each image presentation, $E$ is minimized with respect to the $a$. The $\phi$ then evolve by gradient.

The results of these tests (Fig. 3) confirm that the algorithm is indeed capable of discovering sparse structure in input data, even when the sparse components are non-orthogonal. The result of training the system on $16 \times 16$ image patches extracted from natural scenes is shown in Fig. 4a. The vast majority of basis functions are well localized within each array (with the exception of the low-frequency functions). Moreover, the functions are oriented and selective to different spatial scales. This result should not come as a surprise, because it simply reflects the fact that natural images contain localized, oriented structures with limited phase alignment across spatial frequency. The functions $\phi$, shown are the feedforward weights that, in addition to other terms, contribute to the value of each $a_i$ (refer to term $b$, in equation (5)). To establish the correspondence to physiologically measured receptive fields.
FIG. 4 Results from training a system of 192 basis functions on 16 × 16-pixel image patches extracted from natural scenes. The scenes were ten 512 × 512 images of natural surroundings in the American northwest, preprocessed by filtering with the zero-phase whitening/lowpass filter $R(f) = e^{-\pi f / f_s}$, $f_s = 200$ cycles/picture (see also ref. 9). Whitening counteracts the fact that the mean-square error (or m.s.e.) preferentially weights low frequencies for natural scenes, whereas the attenuation at high spatial-frequencies