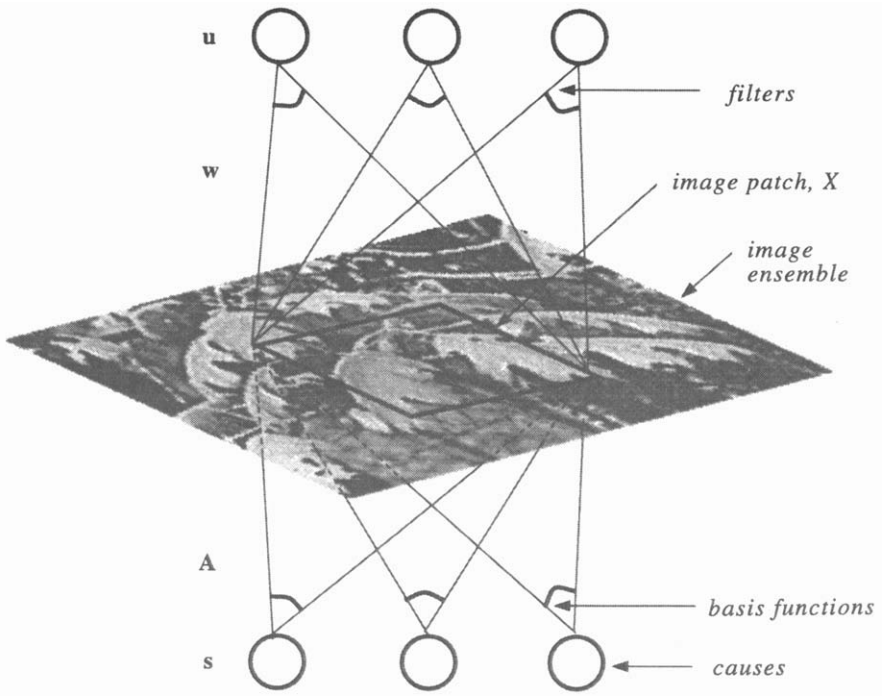

FIGURE 1. The Blind Linear Image Synthesis model (Olshausen & Field, 1996). Each patch, **x**, of an image is viewed as a linear combination of several (here three) underlying basis functions, given by the matrix **A**, each associated with an element of an underlying vector of "causes", **s**. In this paper, causes are viewed as statistically independent "image sources". The causes are recovered (in a vector **u**) by a matrix of filters, **W**, more loosely "receptive fields", which attempt to invert the unknown mixing of unknown basis functions constituting image formation.

demonstrated the ability of this nonlinear information maximization process (Bell & Sejnowski, 1995a) to find statistically independent components to solve the problem of separating mixed audio sources (Jutten & Hérault, 1991). This "Independent Components Analysis" (ICA) problem (Comon, 1994) is equivalent to Barlow's redundancy reduction problem, therefore, if Barlow's reasoning is correct, we would expect the ICA solution to yield localized edge detectors.

That it does so is the primary result of this paper. The secondary result is that the outputs of the resulting filters are indeed, more sparsely distributed than those of other decorrelating filters, thus supporting some of the arguments of Field (1994), and helping to explain the results of Olshausen's network from an information-theoretic point of view.

We will return to the issues of sparseness, noise and higher-order statistics in the Discussion. First, we describe more concretely the filter-learning problem. An earlier account of the application of these techniques to natural sounds appears in Bell & Sejnowski (1996).

BLIND SEPARATION OF NATURAL IMAGES

The starting point is that of Olshausen & Field (1996), depicted in Fig. 1. A perceptual system is exposed to a series of small image patches, drawn from one or more larger images. Imagine that each image patch, represented by the vector **x**, has been formed by the linear combination of N basis functions. The basis functions form the columns of a fixed matrix, **A**. The weighting of this linear combination (which varies with each image) is given by a vector, **s**. Each component of this vector has its own associated basis function, and represents an underlying "cause" of the image. The "linear image synthesis" model is therefore given by:

$$\mathbf{x} = \mathbf{A}\mathbf{s}. \tag{1}$$

which is the matrix version of the set of equations $x_i = \sum_{j=1}^{N} a_{ij} s_j$, where each x_i represents a pixel in an image, and contains contributions from each one of a set of N image "sources", s_j, linearly weighted by a coefficient, a_{ij}.

The goal of a perceptual system, in this simplified framework, is to linearly transform the images, **x**, with a matrix of filters, **W**, so that the resulting vector:

$$\mathbf{u} = \mathbf{W}\mathbf{x} \tag{2}$$

recovers the underlying causes, **s**, possibly in a different order, and rescaled. Representing an arbitrary permutation matrix (all zero except for a single "one" in each row and each column) by **P**, and an arbitrary scaling matrix (non-zero entries only on the diagonal) by **S**, such a system has converged when:

$$\mathbf{u} = \mathbf{W}\mathbf{A}\mathbf{s} = \mathbf{P}\mathbf{S}\mathbf{s}. \tag{3}$$

The scaling and permuting of the causes are arbitrary, unknowable factors, so we will consider the causes to be defined such that $\mathbf{PS} = \mathbf{I}$ (the identity matrix). Then the basis functions (columns of **A**) and the filters which