Feature Analysis

Michael Morgan
City University
Applied Vision Research Centre

Address for correspondence:

M.J. Morgan
Applied Vision Research Centre
City University
Northampton Square
London EC1V 0HB

Email: m.morgan@city.ac.uk

Introduction

A popular idea is that natural images can be decomposed into constituent ***objects***, which are in turn composed of ***features.*** The space of all possible images is vast, but natural images occupy only a small corner of this space, and images of significant objects like animals or plants occupy a still smaller region. The visual brain has evolved to analyze only the interesting regions of image space.

A feature description of an image reduces the number of dimensions required to describe the image. An image is a two-dimensional (N by N) array of pointwise (or pixel-wise) intensity values. If the number of possible pixel values is p, then the number of possible images is a set $\aleph$, of size pN^2. To distinguish all possible images having **N** by **N** pixels, we need a space of N^2 dimensions, which is too large in practice to search for a particular image.

The core idea behind feature analysis is that in real images, objects can be recognized in a space $\Re$ with a much smaller number of dimensions (a smaller dimensionality) than $\aleph$. The space $\Re$ is a ***feature space*** and its dimensions are the features. A simple example of a feature space is colour space, where all possible colours can be specified in a 3 dimensional space, with axes L-M, L+M-S and L+M+S, and L,M and S are the photon catches of the long, medium and short wavelength receptors respectively. The reason why a three-dimensional space suffices to distinguish the very much higher dimensional space of surface reflectance spectra is that there is huge ***redundancy*** in natural spectra. The reflectance at a given wavelength is highly correlated with reflectance at nearby wavelengths. We seek similar redundancies in space that will allow dimensional reduction of images.

Features are not necessarily localized.
Note that in this very general framework there is no implication that features are spatially localized. Features could, for example, be Fourier components. The global Fourier transform has the same dimensionality as the original image and is thus not, according to the present definition a feature space. But if we throw away Fourier components that are unimportant in distinguishing objects, or if we quantise phase, dimensional reduction has been achieved and we have a feature space. The familiar example of JPEG compression involves a feature space.

To distinguish between spatially-localized and non-localized features we shall follow physicists in calling the former ***particles*** and the latter ***waves***. Wavelets (see Olshausen & Field, 1986)) are hybrids that are waves within a region of the image, but otherwise particles. Another important distinction is between particles that have ***place tokens*** and those that do not. Although all particles have places in the image it does not follow that these places will be represented by tokens in feature space. It is entirely feasible to describe some images as a set of particles, of unknown position. Something like this happens in many description of texture. A very active source of debate in visual

psychophysics has been the extent to which place tokens (sometimes called local signs) are used by the visual system. For example, if the distance between two points A,B is seen as greater than between two other points C,D does this imply that there are place tokens for A,B,C,D, or is some other mechanism (Morgan & Watt, 1997)?

The feature concept has proved useful in an impressive variety of different contexts.

Features in Ethology: Lorenz and Tinbergen's concept of the *innate releasing mechanism* (IRM) with its *releasing stimulus* foreshadowed much later work in behaviour and physiology. The red spot at the base of the Herring-gull beak, which the young attack to get food from the parent; and the silhouette of the hawk/goose which elicits fear when moved only in the 'hawk' direction, are classical features to have entered folk psychology.

Features in Physiology: The classical paper 'What the frog's eye tells the frog's brain' (Lettvin, Maturana, McCulloch, & Pitts, 1959) popularised the idea that special low-level sensory analysers might exist for the purpose of responding to simple input features, the canonical example being the response of 'bug detecting' retinal ganglion cells to a small, moving spot. Hubel & Wiesel (1997) described 'bar' and 'edge detectors' in the visual cortex of cat and introduced an influential feature analysis scheme in which hierarchies of mechanisms would combine elementary features into ever increasingly complex objects. The hierarchical scheme, although not without its critics, was supported by the discovery of neurones in inferotemporal cortex (IT) responding selectively to images of complex objects such as a face or hand. Further studies of IT with simpler shapes found a columnar organisation of IT with cells having similar response properties organised in repeating columns (Tanaka, 1996). The proposal that a 'ventral pathway' leading to IT is responsible for object recognition gains support from lesioning and functional brain imaging studies, although the idea of a single area for feature analysis is almost certainly too simple (Logothetis & Sheinberg, 1996).

Fig. 1:

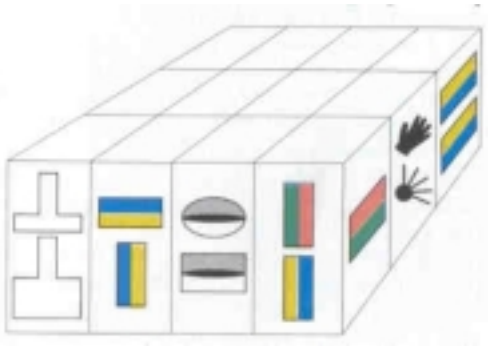

Fig. 1 legend: Columnar organsiation of inferotemporal cortex, based on work of Fujita et al (see review by Tanaka, 1996). Columns of cells selective for similar complex shapes are interspesed with columns of cells unresponsive to these stimuli but selective to different complex shapes. Reproduced with permission from Stryker, M.P (1992) Nature (Lond), 360, 301).

Concept learning: Most dogs bark, and have hair, tails and ears. But not all dogs bark, and Wittgenstein famously pointed out that some concepts like a 'game' have *no necessary* features. His challenge to the feature concept was met by philosophers and animal psychologists by the 'polymorphous concept' (Watanabe, Lea, & Dittrich, 1993), an **n**-dimensional feature space in which instances of a concept occupy a sub-space without sharp boundaries. Considerable research effort has been devoted to investigating the abilities of animals to learn both natural and artificially polymorphous concepts, a key issue being whether a *linear* model of feature combination will serve.

Cognitive Psychology. The idea that certain 'features' can be analysed at a pre-conscious level has proved fertile. Using the technique of 'visual search', (Treisman, 1988) suggested that only certain elementary features, and not their combinations, could serve as pre-conscious markers (Fig. 2). This idea proved especially popular when linked, on rather slender evidence, with the discovery of specialised pre-striate cortical areas in monkey devoted to the analysis of colour and motion. However, the simple dichotomy between a fast, parallel search for features, and a slow, serial search for combinations of features, has come to be questioned (see legend to Fig. 2).

Fig. 2

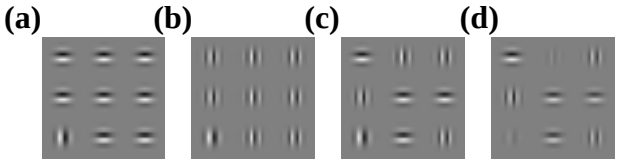

Fig. 2 legend: Searching for a single 'odd man out' is easy when the target has a very different orientation from the background elements (a) or when it has a different spatial frequency (b). Search times do not increase with the number of background elements ('parallel search'), provided the orientation difference is sufficiently large. With smaller orientation differences (< 10 deg) search times do increase with the number of background elements, indicating a 'serial search'. It might be thought that orientation and spatial frequency are easy search features because they are represented in primary visual cortex. However, the conjunction of a particular spatial frequency and orientation is much harder to find (c), despite the fact that many V1 neurones are jointly tuned to orientation and frequency. If contrast is randomised the search becomes harder still (d). The most powerful generalisation about search is that it becomes harder when the number of different background elements increases. A standard ('back pocket') texture segmentation mechanism that responds to local contrasts in orientation, frequency, contrast and colour, can explain most of these findings. Figures by J.A. Solomon.

Image compression The earliest pictures to be sent across the Transatlantic telegraph took more than a week to transmit. Engineers soon reduced this time to three hours by encoding the image more economically. Image compression techniques are divided into those that preserve all the information in the original ('error free'), and the 'lossy' which try to transmit only the important features (Gonzalez & Woods, 1993). An early pioneer of speeding up telegraph transmission by a 'lossy' feature decomposition was Francis Galton, who proposed as set of features for transmitting face profiles over the telegraph using only four telegraphic 'words'. Appropriately for the man who invented the fingerprint, he saw this primarily as a forensic aid in sending profiles of wanted criminals around the world. Galton's feature space lays stress on five *'cardinal points'*. These are

the notch between the brow and the nose, the tip of the nose, the notch between the nose and the upper lip, the parting of the lips and the tip of the chin (Fig. 3).

Fig. 3:

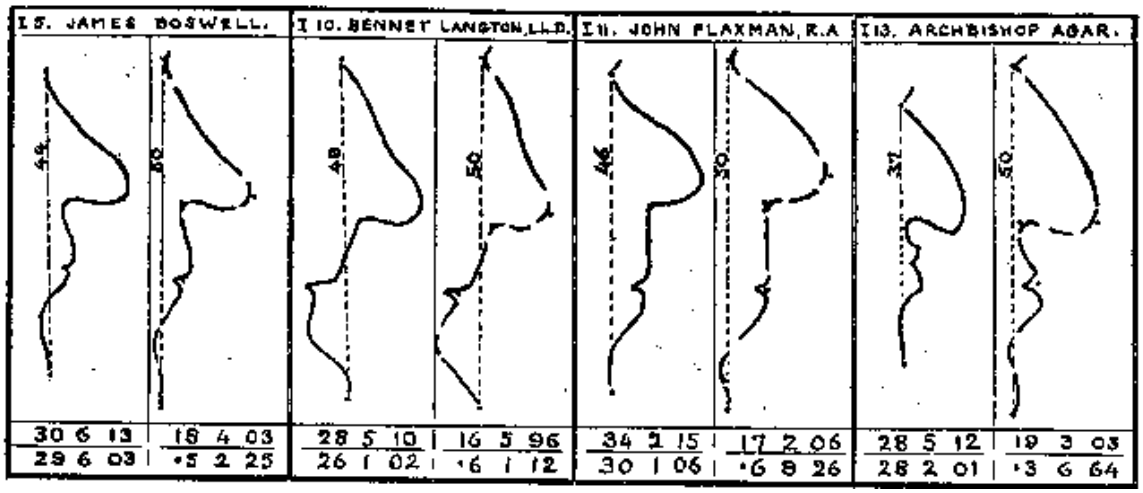

Fig. 3 legend: Face profiles described (left member of each pair) and reconstructed (right member of each pair) from 20 telegraphic numbers arranged in 5 groups of 4 (bottom). The code uses the relations between 5 'cardinal points' on the profile. Galton was perhaps the first to use the term 'cardinal points' to describe a multi-dimensional feature space for object recognition. From Galton, F. (1910) Nature (Lond), March 31, 127-130.

The remainder of this review will illustrate the concept of feature spaces, and describe key issues, some resolved and others not.

Principal Component Analysis of faces.

Galton was the first to treat faces as mathematical objects and to add them photographically. The *average face* was held to be especially beautiful, possibly because smallpox scars and other blemishes were removed. Galton also presented the average faces of groups such as rapists, clergymen and athletes. Any particular face could then be correlated with each average in turn and described by a vector **V**. This vector will tell us the extent of resemblance of that face to the mean rapist, the mean clergyman, the mean athlete, and so on. The average images are a *basis set* for describing all faces. Dimensional reduction has been achieved, so long as the dimension of **V** is less than that of the original image space, N^2. Whether the clergyman-athlete-rapist space is a good one is another matter. It is probably not, because the dimensions are correlated. The aim of most feature analysis, including PCA is to ensure that the dimensions of the feature space are uncorrelated, or in other words, that the axes in the space are *orthogonal.*

The idea behind PCA for faces (or Karhunen-Loeve expansion) is to find a set of features called *eigenvectors* which span the sub-space of images in which faces lie (Turk & Pentland, 1991). Each eigenvector has dimensions N^2 and is a linear combination of a set of training faces, each of dimension N by N pixels. Equivalently, each face in the training set is a linear combination of the N^2 eigenvectors. To achieve dimensional

reduction only the most important eigenvectors are chosen as a basis set. These are the vectors that correlate most highly with the members of the training set. The most important, in this sense, is the average image, as defined by Galton.

Since the eigenvectors resemble faces when they are represented as 2-D images, they can be called *eigenfaces*. Examples are shown in Fig. 4. Once the eigenfaces have been created, each face in the training set can be described by a set of numbers representing its correlation to each of the eigenfaces in turn. If 7 eigenfaces are chosen to span the face space, each face will be described by a vector of 7 numbers, instead of by its pixel values. A huge dimensional reduction has been achieved. The problem of recognising a face is now a simple one of pattern recognition: finding the vector in memory that it most closely resembles.

PCA has been used for face detection, face recognition and sex classification. Effective 'caricatures' can be derived by exaggerating the differences of a face from the average. An intriguing experiment by Leopold, O'Toole, & Blanz (2001) suggests that the brain may use a feature-space to identify faces. Observers were trained to discriminate 'Adam', who was described by a vector in face-space from 'Anti-Adam', who had the directly opposite vector. After adapting to 'Adam' for some minutes, observers were more likely to classify the average face (mid-way between Adam and Anti-Adam) as Anti-Adam than as Adam. The inference is that there exist feature detectors that are tuned in face-space, and that identification depends on a population code comprising these detectors.

Fig. 4

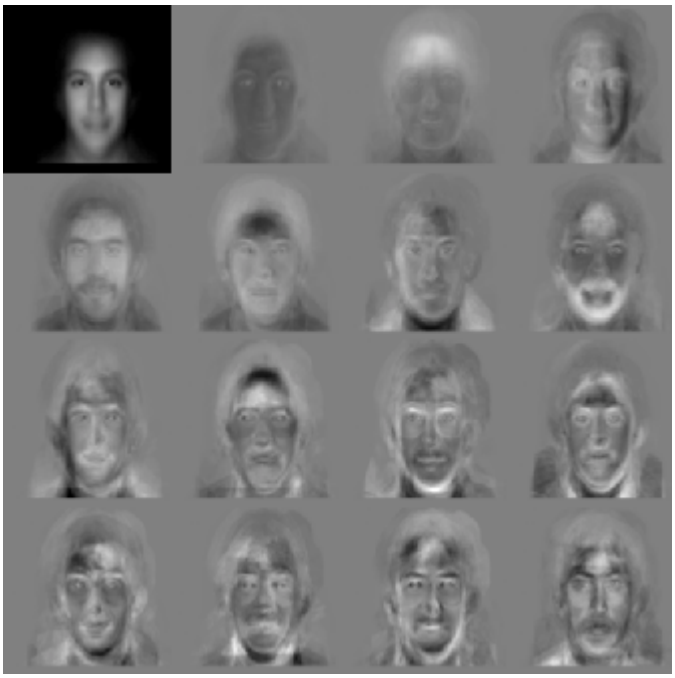

Fig. 4 legend: Eigenfaces for face recognition. Reproduced with permission from the interactive MIT Media Lab Web site:
http://www-white.media.mit.edu/vismod/demos/facerec/basic.html

'Fourier freaks and feature creatures'

In broad terms, we now see that the code for early vision will be cracked when we find the basis set used by the brain for describing the image. One such basis set is the Fourier Transform: sinusoids of differing frequency, orientation and phase are the eigenfunctions of linear systems (Turk & Pentland, 1991). Following the application of certain key ideas in linear systems theory to vision (Robson, 1980), there was much debate about whether feature analysis or Fourier analysis was the preferred vehicle for understanding biological vision. This entertaining but ultimately fruitless debate is now essentially dead and buried. It made sense only when the 'Fourier freaks' maintained that objects were recognised exclusively from their global amplitude spectrum. Since no one will now admit to having thought such a thing, it is pointless to provide historical detail.

Contrary to the view that the Fourier amplitude spectrum carries the important information in natural images, it is now recognised that the amplitude spectrum of most images is very similar (Field, 1987). The interesting features of images such as the boundaries of objects are represented in the relative *phases* of Fourier components. An edge or a bar in the image is a place where Fourier components undergo constructive interference. If the global amplitude spectra of two different images such as faces are interchanged, the hybrid images look like the images from which their phase spectra are derived (Fig. 5). However, this is true only for the global Fourier Transform. If the image is decomposed into a number of overlapping patches, and if each Patch is transformed, it is the amplitude rather than the phase information in the patches which determines the appearance of the image, if the patch size is sufficiently small (Fig. 5). This is self-evidently true when the patch size is a single pixel, because the transform contains only the DC level, and no phase information. But the limit is reached before a single pixel. Further work is needed to determine how the limiting patch size varies in different images, and whether it is determined by cycles/image or cycles/deg of visual angle.

Fig. 5

Fig 5 legend: The Figure shows images of two political theorists (rightmost panels), one of whose ideas have been recently discredited. Each thinker contributes his/her phase from the Fourier Transform to each

of the images on the left; the amplitude information comes from the other face. The Fourier transform is not global, but is rather derived from overlapping patches, of size of which decreases (64, 32, 16, 8 and 4 pixels) from left to right. When the patch size is large, appearance of the image is dominated by phase information; when it is small, appearance is dominated by the amplitude of the Fourier components. At intermediate patch sizes, the appearance is composite. (Reproduced with permission from Morgan et al, 1991).

The consensus view now is that Fourier analysis in the visual system is limited to a local or *patchwise* Fourier analysis, and that this is performed by neurones in V1 with localised, oriented and spatial-frequency tuned receptive fields (Robson, 1980). The idea of patchwise spatial frequency analysis fits in well with the architectural division of V1 into *hypercolumns*, each containing a full range of orientations and spatial frequencies, with a scatter of their receptive fields within a region of the image (Hubel & Wiesel, 1977). According to this model, the receptive fields of simple cells in V1 provide a *basis set* for describing local properties of the image, comparable to the wavelet transform in image processing (Olshausen & Field, 1986). Putting this simply, the idea is that an object can be recognised locally in an image by a series of numbers representing its effect on the activity of a population of detectors tuned individually in orientation and spatial frequency. Dimensional reduction is achieved because pointwise intensity values have been discarded and replaced by a more economical code. Just how many types of receptive field are needed to provide a satisfactory basis set for describing natural images is a question of equal interest to psychophysics and image processing.

PCA is far from being the only way to find a suitable basis set for natural images. Receptive fields like those in V1 emerge naturally as a basis set from a learning algorithm that seeks a sparse code for natural images (Olshausen & Field, 1986). The aim of sparse coding is to have each image activate the smallest possible number of members of the basis set. In physiological terms, the aim would be to have as many neurones as possible not activated at all by the image. Using this approach, Olshausen & Field derived a basis set having an impressive similarity to the receptive fields of V1 neurones (Fig. 6).

The 'Primal Sketch'

The ability of line drawings to convey shape is very strong evidence that the feature space for object recognition may be of drastically reduced dimensionality, compared to the space of all possible images. Just a few lines drawn on a flat surface can suggest the face of a well-known person; or the idea 'no skateboarding allowed'. Impressed by the effectiveness of cartoons David Marr (1992) proposed that the earliest stages of vision transform the continuous gray-level image into a neural cartoon, which he called 'The Primal Sketch'. Since cartoons emphasise primarily the outlines of shapes, the main objective of the primal sketch is to find *edges* in the image, edges being the loci of points on the 3-D object where an object occludes itself or objects further away (edge of a cube), or where there is a rapid change in the direction of the tangent plane to the object surface (outline of the nose on a face) , or where there is some abrupt change in surface reflectance (boundary of the iris). It is a remarkable fact that much of the edge structure of an object in the image depends upon its 3-D structure: remarkable because we

recognise objects from a variety of viewpoints, or from sketches made from a variety of viewpoint.

Fig. 6

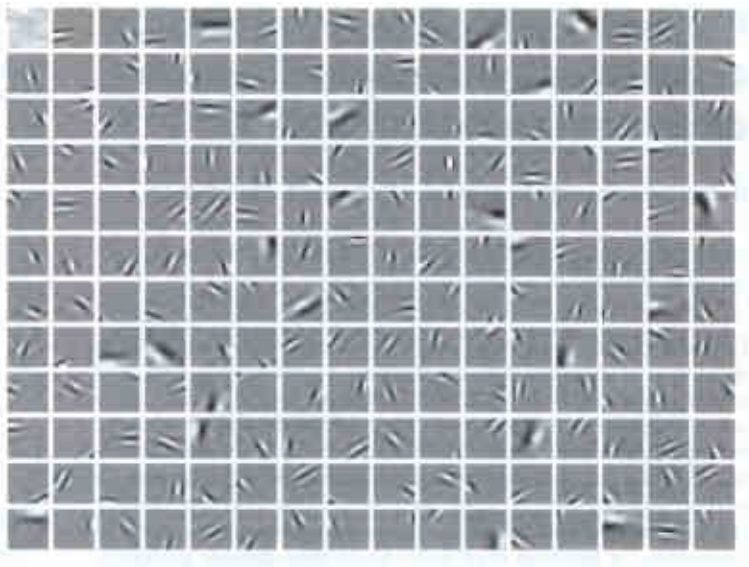

Fig. 6 legend A basis set derived from ten 512 by 512 natural images of the American northwest, by a training algorithm aimed at maximizing the sparseness of the representation. The basis set bears a striking similarity to the receptive fields of V1 neurones, suggesting that they too form an efficient basis set for describing natural images. Reproduced with permission from Olshausen & Field (1996).

If we consider an edge like the outline of the nostril on a face, we shall find in its image a sudden change in luminance, which is not predictable from the gradual changes around it. A discontinuity is conveniently found by a local maximum or minimum in the first spatial derivative of the luminance profile, or equivalently, a *zero-crossing* in the second derivative. Ernst Mach (followed by William McDougall in his quaint 'drainage' theory) was the first to conjecture that the visual system uses second derivatives to find features in the luminance profile, his evidence being the appearance of 'Mach Bands' on the inflexion points in luminance ramps. Marr & Hildreth (see Marr, 1982) proposed that the receptive fields of retinal ganglion cells and simple cells in V1 make them nearly ideal second-derivative operators (Laplacians of Gaussians) and that their function is to produce the primal sketch. Different sizes of receptive field produce cartoons at different *spatial scales*, corresponding to the different frequencies in the Fourier Transform (qv), but agreeing on the position of zero-crossings at the most significant points in the image. This recalls the fact that an edge or a bar in the image is a place where Fourier components undergo constructive interference. The Primal Sketch neatly uses wavelets to locate edges and turns them into particles.

A wide variety of phenomena have been used to investigate the nature of the 'spatial primitives' or features in human vision. These include Mach Bands, the Chevreul illusion, the apparent location of bars and edges in gratings and plaids with different spatial

frequency components, and edge blur discrimination. Various primitives have been considered such as zero-crossings, zero-bounded regions and local energy maxima. Although it is now possible to predict the apparent location of edges and bars in images with a fair degree of accuracy, there is no consensus as yet about the nature or existence of primitives, or about the way they are combined over spatial scale (Morgan & Watt, 1997).

Summary: Features are useful for describing natural images because the latter have massive informational redundancy. Image space itself is too vast to search directly. Feature analysis depends on the proposition that the search for particular objects can be concentrated in a sub-space of image space : *the feature space*. Biologists expect that there will be special sensory mechanisms for searching just the right sub-space for a particular task. Ethologists and animal learning theorists concur. Useful hints about likely feature spaces may be obtained from engineers working on image compression. Although in the past feature analysis was contrasted with Fourier analysis, the modern synthesis is that a patchwise Fourier analysis by localised receptive fields in primary visual cortex (V1) provides the primitive basis set for the feature space of vision. These form the basis set for the elaboration of neurones responding selectively to geometrical features in area TE of the inferotemporal cortex, and these in turn from the basis for object recognition in different but overlapping areas of IT.

References

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A, 4*(12), 2379-2394.

Gonzalez, R., & Woods, R. (1993). *Digital Image Processing*. Reading, Mass: Addison-Wesley.

Hubel, D. H., & Wiesel, T. N. (1977). Functional architecture of the macaque monkey visual cortex. Ferrier Lecture. *Proceedings of the Royal Society of London (Biology), 198*, 1-59.

Leopold, D., O'Toole, A., T, & Blanz, V. (2001). Prototype-references shape encoding revealed by high-level aftereffects. *Nature Neuroscience, 4*, 89-94.

Lettvin, J. Y., Maturana, R. R., McCulloch, W. S., & Pitts, W. H. (1959). What the frog's eye tells the frog's brain. *Proc. Inst. Rad. Eng., 47*, 1940-1951.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual Object Recognition. *Annual Review of Neuroscience, 19*, 577-621.

Marr, D. (1982). *Vision*. San Francisco: WH Freeman & Co.

Morgan, M. J., Ross, J., & Hayes, A. (1991). The relative importance of local phase and local amplitude in patchwise image reconstruction. *Biological Cybernetics, 65*, 113-119.

Morgan, M. J., & Watt, R. J. (1997). The combination of filters in early spatial vision: a retrospective analysis of the MIRAGE model. *Perception, 26*, 1073-1088.

Olshausen, B., & Field, D. (1986). Emergence of simple-cell receptive field properties by learning a spare code for natural images. *Nature, 381*, 607-609.

Robson, J. G. (1980). Neural Images: The Physiological Basis of Spatial Vision. In C. S. Harris (Ed.), *Visual Coding and Adaptability* (pp. 177-214). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience, 19*, 109-139.

Treisman, A. M. (1988). Features and objects: The 14th Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology A, 40*, 201-237.

Turk, M., & Pentland, A. (1991). Face recognition using Eigenfaces. *IEEE*, 586-591.

Watanabe, S., Lea, S., & Dittrich, W. (1993). What can we learn from experiments in pigeon concept formation? In H. Zeigler & H.-J. Bischof (Eds.), *Vision, Brain and Behaviour in Birds*. Cambridge, Mass: MIT Press.

Word count: 4011 including title page, references and Figure legends