Network: Computation in Neural Systems **7** (1996) 1–5. Printed in the UK

VIEWPOINT

The vocabulary of shape: principal shapes for probing perception and neural response

Joseph J Atick, Paul A Griffin and A Norman Redlich

Computational Neuroscience Laboratory, The Rockefeller University, 1230 York Avenue, New York, NY 10021-6399, USA

Received 3 October 1995

Abstract. Humans perceive shape rapidly and effortlessly but have great difficulties describing what they perceive. This suggests that the representation of shape in the brain is abstract and very unlike that used in conscious thought. Here we explore the proposal that this representation is matched to the statistical properties of objects in the environment. From an ensemble of several hundred laser-scanned three-dimensional (3D) human heads we extract the principal components which provide a compact basis for head shape. We show that, with good accuracy, a given head can be represented by linear combinations of a few dozen primary shapes just as colours can be synthesized by combining the three principal colours. We suggest new perceptual adaptation experiments for testing the brain's shape representation system. The principal head shapes can also be used to probe response properties of 'face-cells' in the inferior temporal cortex.

The idea that the brain's 'visual vocabulary' is in some way matched to properties of its environment implies that one should be able to uncover this representation by measuring the statistical properties of natural visual signals (Barlow 1961, 1989, Marr 1970, Atick and Redlich 1990, Atick 1992, Ruderman and Bialek 1994). This general philosophy has so far proven valuable in understanding low-level vision, where it was shown that the details of retinal (Atick and Redlich 1992) and lateral geniculate nucleus (Dong and Atick 1995a) processing can be predicted from knowledge of the spatio-temporal power spectrum of natural time-varying images (Dong and Atick 1995b)—the simplest non-trivial statistical regularity. Here we suggest that this approach can also yield powerful tools for probing high-level vision.

The representation of visual signals formed by the photoreceptors is very high-dimensional and is not suitable for object recognition or other high-level vision tasks. We believe the brain must perform dimensional reduction to find a relatively compact representation of objects. As is well known, low-dimensional representations can be derived from knowledge of statistical regularities, using the tools of information theory (Shannon and Weaver 1949). Here we analyse an important class of 3D objects, human heads, and derive from their regularities a minimal 'vocabulary' for head shape. Our hypothesis is that the brain is using such visual vocabulary.

We begin with a set of several hundred laser-scanned three-dimensional heads, with each head represented as a surface in cylindrical coordinates $r(\theta, \ell)$; ℓ is the height coordinate, θ is the angular variable, and $r(\theta, \ell)$ is the radial coordinate of points on the surface. In this database an entire head is sampled at 512×256 angular $\times$ height points; thus over 100 000 real values are required to specify a particular person's head shape. Of course,

0954-898X/96/010001+05$19.50

these parameters are not independent within the class of heads. To find the independent variables we apply principal component analysis (Jolliffe 1986).

Principal components can be thought of as a basis set $\{\Psi_i(\theta, \ell)\}$, just like sinusoidal gratings or Walsh patterns, but, unlike these generic bases, principal components are derived from the data. For details of how they can be derived, we refer the reader to the literature (Jolliffe 1986, Sirovich and Kirby 1987). In terms of this basis, any given head is expanded as:

$$r(\theta, \ell) = r_0(\theta, \ell) + \sum_{i=1}^{N} a_i \Psi_i(\theta, \ell)$$

where $r_0(\theta, \ell)$, the 'mean-head,' is the average over head surfaces used to derive the basis. In figure 1 we show r_0 and the 15 most significant eigenmodes $\{\Psi_i\}$. The claim is that any given head shape—not just those in the set used to derive $\{\Psi_i\}$—can be reconstructed by specifying the N coefficients $\{a_i\}$. How many coefficients are required depends on how accurately we wish to construct the shape. This is very easy to determine since eigenmodes are hierarchical in importance so one simply keeps the top N modes which achieve the desired accuracy. We have done this for 150 out-of-sample head surfaces (i.e. heads that were not used in the derivation of the basis) and we find that the reconstruction error (defined as $< |\Delta r/r| >$ averaged over all points and over the 150 examples, with $\Delta r \equiv r^{\mathrm{actual}} - r^{\mathrm{reconstructed}}$) decays exponentially with N and is less than 1% when $N = 40$. Thus head shape can be represented by specifying as few as three or four dozen coefficients a_i, a compression of several thousandfold.

We should point out that principal components were used earlier to derive a representation of *images* of human faces (Sirovich and Kirby 1987). There it was shown that eigenmodes (so-called eigenfaces) provided an excellent low-dimensional characterization of face images. Here, of course, we compute eigenmodes for surfaces and not images, so these functions have a different interpretation and utility. In analogy with eigenfaces, one may use the term eigenheads to refer to these modes.

The eigenheads have many applications. For example, we have used them to solve the problem of shape-from-shading (Atick *et al* 1995). In this problem one attempts to recover the 3D shape of an object from a 2D image, which is a very difficult mathematical problem and is known to be ill-posed in many instances. The difficulties encountered can be attributed to the fact that most previous shape-from-shading algorithms attempt to solve the problem in a general fashion for any smooth surface, regardless of whether it is a head or a lunar surface. An alternative approach is to accept that objects belong to classes, and within each class to use a low-dimensional parameterization of shape as we do here. In this way, the problem of shape-from-shading becomes equivalent to estimating a small number of coefficients $\{a_i\}$ given an image and the knowledge of what class the object belongs to. Recently we showed that this approach leads to excellent reconstruction of 3D surfaces of human heads from single 2D real face images†.

We also would like to propose using the eigenheads as a tool for mapping the response properties of neurons in the temporal lobe. It is well known that neurons in the inferior temporal (IT) cortex exhibit strong object specificity. For example, so-called face-cells respond selectively to human faces and heads (Desimone 1991, Gross 1992, Rolls 1992). What is not clear yet is what properties of faces these neurons are partial to. The proposal is to use the eigenheads as stimuli to systematically measure the response of face cells. One

† For other approaches to shape-from-shading see Horn and Brooks (1988) and references therein, Pentland (1984) and Lehky and Sejnowski (1988).

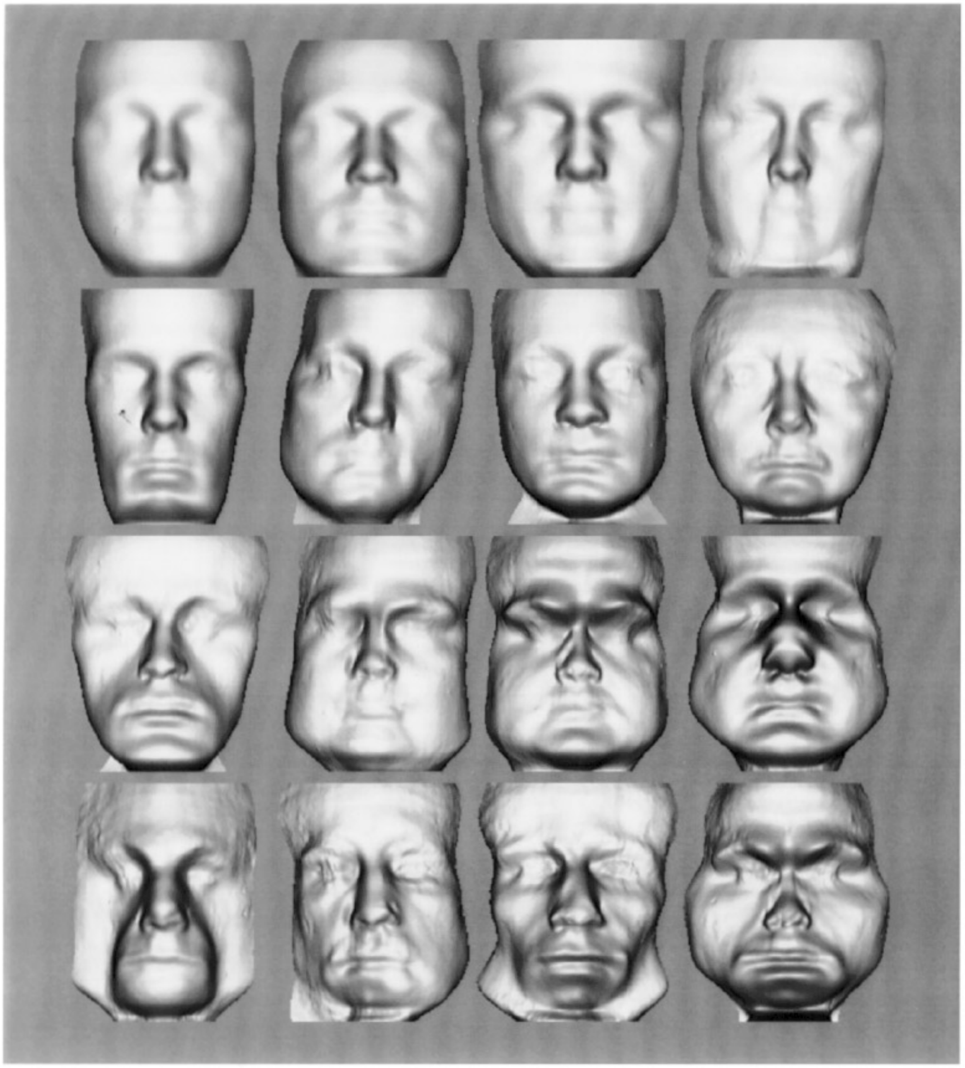

Figure 1. The mean-head r_0 (upper left-most corner) and the 15 most significant eigenheads, Ψ_i. In order to display them, the modes Ψ_i have been added to r_0 and then rendered; what is displayed is $r = r_0 + 10 * \Psi_i$. The mean-head and the modes were derived from 200 three-dimensional laser-scanned human head surfaces made available to us by the Human Engineering Division of the Wright Patterson Airforce Base. The modes form a basis in the sense that, by taking linear combinations of these and adding them to r_0, we can generate the surface of any human head—not just those used to derive the representation—up to a given accuracy. From 150 out-of-sample heads, we find that 40 modes are enough to represent a head with 1% accuracy. Of course, since the database used to extract the Ψ_i had only adult males, these modes will not be suitable for representing women's and children's heads. Eigenmodes for those can be derived using the same techniques from the appropriate databases. We suggest that the brain uses this minimal 'vocabulary' to represent human heads. Finally we should point out that there are higher-order regularities in shape-space which constrain what principal shapes can be put together to make a human-looking head. This 'grammar of shape' can be systematically studied; we hope to report on this in a future publication.

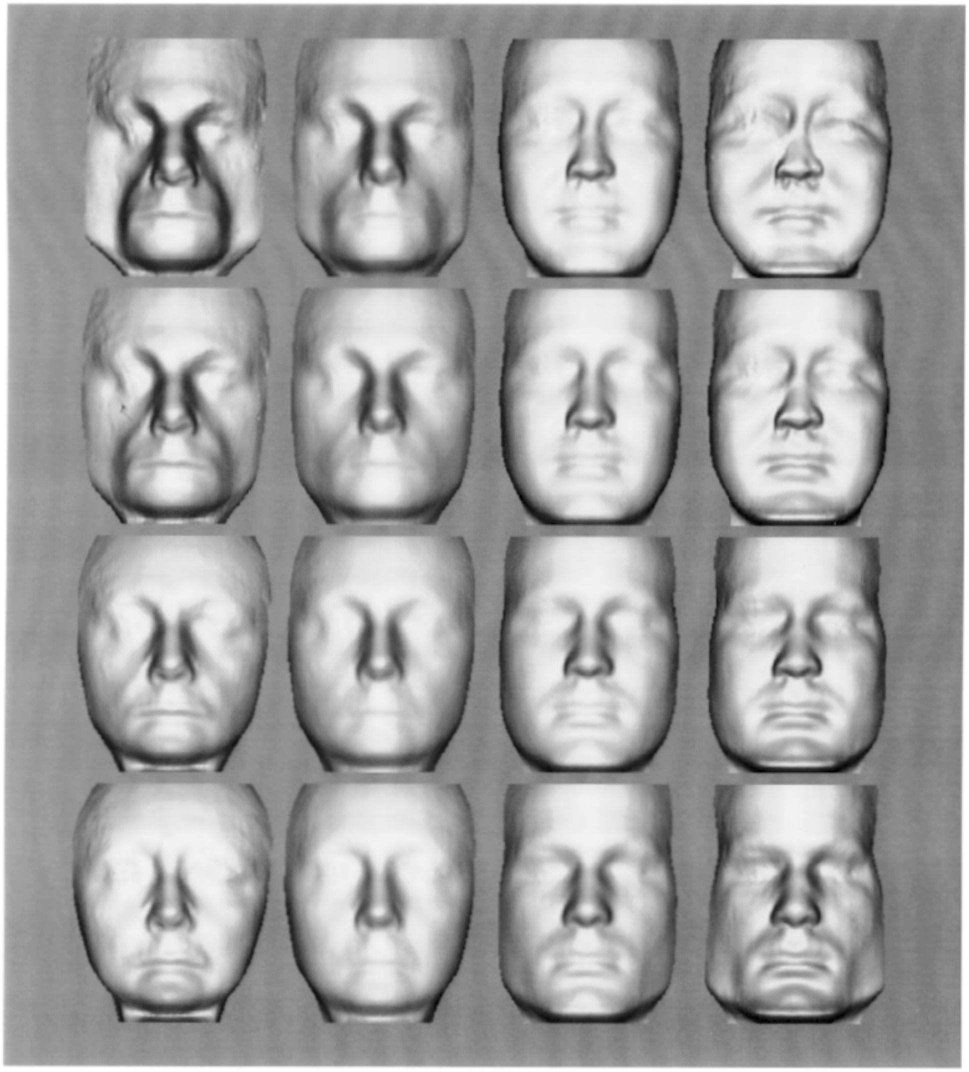

Figure 2. Modulation (from left to right) of head-shape about the mean-head along four different directions (rows) in shape space. The first and last rows are along the direction of the twelfth and seventh modes, respectively, while the second and third rows are along two intermediate directions in the plane of these modes. More precisely, the shapes from left to right are generated by the surfaces $r(\theta, \ell) = r_0(\theta, \ell) + a(\cos(\phi)\Psi_{12} + \sin(\phi)\Psi_7)$ with $a = +8, +4, -4, -8$, while the angle ϕ is equal to $0, 35, 55, 90$ degrees as we go from row one to four, respectively. An interesting application is to use this type of shape-contrast modulation to explore perceptual adaptation to shape.

can even apply texture maps to these principal shapes or render them with different lighting to determine the relative effects of 3D shape, colour, and lighting on the neural response.

Finally, if indeed there are principal shapes or ‘shape channels’ in the brain, then one expects to be able to uncover their presence by selectively adapting them and measuring perceptual after-effects. One class of experiments that we have in mind is the analogue for shape space of the experiments of Webster and Mollon (1994) in colour space. These experiments demonstrate colour contrast adaptation by adapting subjects to stimuli which are temporally modulated in colour about grey along a fixed direction in colour–luminance space. The perceptual changes measured are consistent with the theory that colour is represented by principal colour channels that alter their tuning functions by adapting to the statistical properties of input stimuli. By analogy, one can try to adapt subjects to stimuli which are facial shapes modulated in a fixed direction in shape space around the mean-head. In figure 2 we show four examples of 'shape-contrast" modulation along some selected directions.

Acknowledgments

We thank Kathleen Robinette, Jennifer Whitestone, Barbara McQuiston, and Glen Geisen for their cooperation in making the 'USAF database' available to us. This work is supported in part by a grant from the Office of Naval Research contract number N00014-95-1-0381.

References

Atick J J 1992 Could Information theory provide an ecological theory of sensory processing? *Network: Computation in Neural Systems* **3** 213–51

Atick J J and Redlich A N 1990 Towards a theory of early visual processing *Neural Comput.* **2** 308–20

——1992 What does the retina know about natural scenes? *Neural Comput.* **4** 196–210

Atick J J, Griffin P A and Redlich A N 1995 Statistical approach to shape from shading: Reconstruction of 3D face surfaces from single 2D images *Neural Comput.* submitted

Barlow H 1989 Unsupervised learning *Neural Comput.* **1** 295–311

——1961 Possible principles underlying the transformation of sensory messages *Sensory Communication* ed W Rosenblith (Cambridge, MA: MIT Press) pp 217–34

Desimone R 1991 Face selective cells in the temporal cortex of monkeys *J. Cogn. Neurosci.* **3** 1–8

Dong D W and Atick J J 1995a Temporal decorrelation: a theory of lagged and nonlagged responses in the lateral geniculate nucleus *Network: Computation in Neural Systems* **6** 159–78

——1995b The statistics of natural time-varying images *Network: Computation in Neural Systems* **6** 345–58

Gross C G 1992 Representation of visual stimuli in inferior temporal cortex *Phil. Trans. R. Soc. Lond.* B **335** 3–10

Horn B K P and Brooks M J 1989 *Shape from Shading* (Cambridge, MA: MIT Press)

Jolliffe I T 1986 *Principal Component Analysis*(Berlin: Springer Verlag)

Lehky S R and Sejnowski T J 1988 Network model of shape-from-shading: neural function arises from both receptive and projective fields *Nature* **333** 452–4

Marr D 1970 A theory for cerebral neocortex *Proc. R. Soc. Lond.* B **176** 161–234

Pentland A P 1990 Linear shape from shading *Int. J. Computer Vision* **4** 153–62

Rolls E T 1992 Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas *Phil. Trans. R. Soc. Lond.* B **335** 11–21

Ruderman D L and Bialek B 1994 Statistics of natural images: Scaling in the woods *Advances in Neural Information Processing Systems 6* ed J D Cowan, G Tesauro and J Alspector(San Mateo, CA: Morgan Kaufman)

Shannon C E and Weaver W 1949 *Mathematical Theory of Communication* (Urbana, IL: University of Illinois Press)

Sirovich L and Kirby M 1987 Low-dimensional procedure for the characterization of human faces *J. Opt. Soc. Am.* A **4** 519–24

Webster M A and Mollon J D 1994 The influence of contrast adaptation on colour appearance *Vision Res.* **34** 1993–2020