Visual features of intermediate complexity and their use in classification

Shimon Ullman, Michel Vidal-Naquet and Erez Sali

Department of Computer Science and Applied Mathematics, The Weizmann Institute of Science, PO Box 26, Rehovot 76100, Israel
Correspondence should be addressed to S.U. (shimon.ullman@weizmann.ac.il)

Published online: 10 June 2002, doi:10.1038/nn870

The human visual system analyzes shapes and objects in a series of stages in which stimulus features of increasing complexity are extracted and analyzed. The first stages use simple local features, and the image is subsequently represented in terms of larger and more complex features. These include features of intermediate complexity and partial object views. The nature and use of these higher-order representations remains an open question in the study of visual processing by the primate cortex. Here we show that intermediate complexity (IC) features are optimal for the basic visual task of classification. Moderately complex features are more informative for classification than very simple or very complex ones, and so they emerge naturally by the simple coding principle of information maximization with respect to a class of images. Our findings suggest a specific role for IC features in visual processing and a principle for their extraction.

A fundamental question in the study of visual processing is the problem of 'feature selection': which features of an image are extracted and represented by the visual cortex? Several brain areas are involved in visual object processing, and different features are represented at different stages. In the earliest processing stages, which involve the retina, lateral geniculate nucleus (LGN) and primary visual cortex (V1), the image is represented by simple local features such as center–surround receptive fields and oriented lines and edges. This encoding can arise from the computational principles of decorrelation and redundancy reduction[1–3] or from faithful reconstruction of the input using sparse encoding[4–6]. After this early processing, moderately complex features are represented in areas V4 and the adjacent region TEO, and finally, partial or complete object views are represented in anterior regions of inferotemporal (IT) cortex[7–11].

Here we show, by computational analysis and simulations, that features of intermediate complexity and partial object views are optimal for visual object classification. These features were automatically selected when the system was set to maximize the information delivered with respect to a class of images, and thus serve as basic building blocks for representing the class. The simulations show that IC features are more informative than very simple or very complex ones, and that during visual classification, the extracted features have the capacity to generalize broadly to new exemplars within the class.

It has previously been proposed that complex objects are represented in the visual cortex in terms of simpler elements such as wavelets or Gabor basis functions[4,5], both of which have been used in object recognition models[12]. These 'building blocks' are universal in the sense that they are equally applicable to all natural images. Alternatively, we propose that the visual system encodes features of intermediate complexity that are class-specific, that is, selected for encoding images within a class of related images. These features are used after the encoding of simple features in V1 but before the encoding of complex object views in anterior IT cortex, and they are specifically selected to support visual classification—one of the basic tasks of visual perception. Here we present examples of such features, the coding principle used to extract them (maximizing information for classification), their advantages and their biological implications.

Results

Features extracted by maximizing information

From a training set of 138 roughly-frontal face images and 40 side-view images of cars (examples in **Fig. 1b**), we extracted sets of IC features or 'fragments' that are optimal building blocks for encoding those images (**Fig. 1a, c** and **d**). The fragments were extracted on the basis of maximizing the information delivered about the set of faces (or cars) using a search procedure. The search stored a large number (>10,000) of sub-images (see Methods), measured the information delivered with respect to the training set for each sub-image and extracted the most informative fragments.

The amount of information delivered by a candidate fragment about the class of images was calculated using the mutual information equation:

$$I(C, F) = H(C) - H(C|F) \quad (1)$$

In this equation, $I(C, F)$ is the mutual information between the fragment F and the class C of images, and H denotes entropy. This is a natural measure of information conveyed by F about C, as it measures how the uncertainty about the presence of the class C in the image is reduced by the possible presence of the fragment F in the image. In simulations, we found that simplified

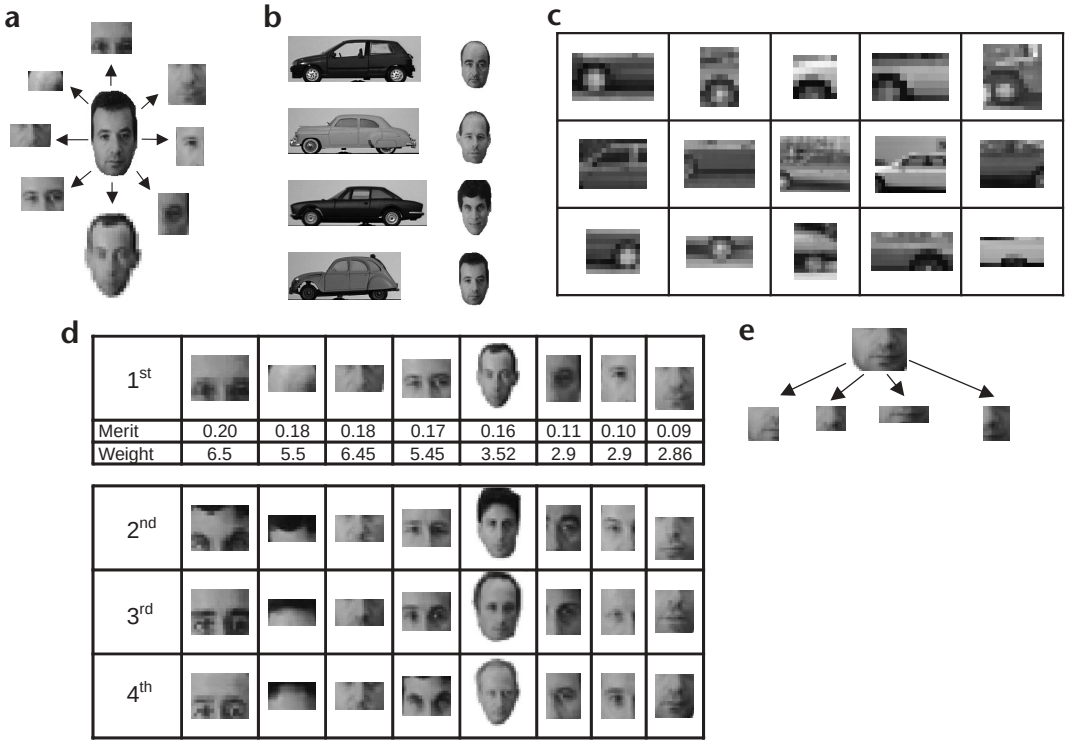

Fig. 1. Intermediate complexity visual features were chosen by maximizing delivered information with respect to a class of objects. (**a**) The best eight face features found, arranged around a face from the learning set (in decreasing order of mutual information from the top, moving counter-clockwise). (**b**) Examples of faces and cars in the training set. (**c**) Selected car fragments. (**d**) Additional fragments organized by type; first row same as (**a**) with merit (mutual information in bits) and weight ($\log_2$ of the likelihood ratio, equation 2) shown. Each column corresponds to a single type, rows arranged in decreasing levels of merit. (**e**) One fragment represented in terms of simpler sub-fragments.

approximations to the mutual information measure, which may be easier to implement biologically, were also effective.

The most informative face fragments extracted by this procedure were selected successively (**Fig. 1a**). After finding the fragment with the highest mutual information score, the search identified the fragment that delivered the maximal amount of additional information, and so on. The i+1 fragment was selected to increase the mutual information of the fragment set by maximizing the minimal addition in mutual information with respect to each of the first i fragments. (A more comprehensive criterion would be to select the i+1 fragment that maximizes the additional mutual information with respect to the joint distribution of all i previously selected fragments together, but this is a more demanding computation that requires substantially more training data.) After extracting the first eight fragments, the search extracted additional fragments at the same locations, arranged by location and decreasing mutual information (**Fig. 1d**).

Superiority of IC features

The most informative fragments were typically fragments of intermediate size (**Fig. 2**). The amount of mutual information, defined as the fragment's 'merit' (equation 1), was calculated for fragments of different size centered at the same image location. The merit typically peaked at an intermediate size (median 11%, s.d. 16% of object size). The superiority of intermediate-size fragments can be explained as the interplay of two factors: specificity and relative frequency. A large face fragment can provide reliable indication of the presence of a face in an image, although the likelihood of encountering such a fragment in a novel face image is low. Consequently, the information carried by such a fragment with respect to the class is limited. A smaller fragment has a higher likelihood of appearing in different face images, but the likelihood of its presence in non-face images is also higher. The optimal fragments we found are considerably more complex than V1-like receptive fields, but still correspond to local image structures rather than to the global shape templates that are used in some current visual recognition models[13,14].

We also found a superiority effect with respect to changes in resolution. A common approach in computer vision is to process images at multiple resolutions[15], and it has been suggested that the mammalian visual system also performs multi-resolution processing using multiple receptive field sizes[16]. A plot of the merit of a face fragment as a function of image resolution (**Fig. 2d**, see Methods) shows that the mutual information peaked at intermediate resolution. This could explain why intermediate resolution face templates are computationally useful for face detection[17]. For car images, high-merit fragments were typically of low resolution. On the whole, the highly informative features were of intermediate complexity: intermediate size at high resolution and larger size at intermediate resolution.

The features were selected by this procedure to support generalization and classification rather than economic reconstruction of the image[5,18]. An important difference between classification and reconstruction schemes is that in classification, fragments are determined by images within as well as outside the class, resulting in features that are more informative than those selected for effi-

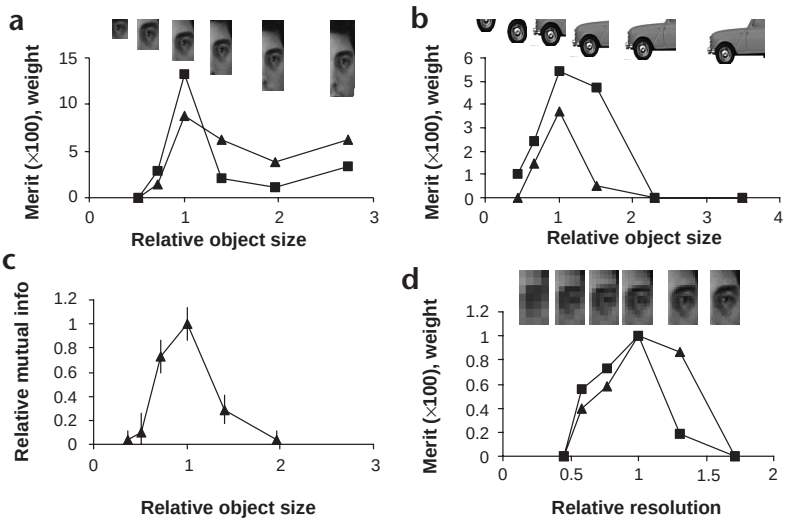

Fig. 2. Superiority of intermediate fragments. Mutual information (merit, ▲) and weight (■) as a function of fragment size (**a**–**c**) and resolution (**d**). (**a**,**b**) Examples of size effect on mutual information and weight for two fragments. Horizontal axis, relative size in terms of fragment area (the size of maximum mutual information is defined as 1). Vertical axis, merit × 100 (equation 1) and weight (equation 2). (**c**) Average size effect on mutual information (n = 15). Horizontal axis as in (**a** and **b**). Vertical axis, relative mutual information ± s.d. (the maximal mutual information is defined as 1). (**d**) Effect of image resolution on mutual information and weight. The decrease in information was significant for a 15% resolution change ($P < 0.05$). Horizontal axis, relative resolution in pixels across the fragment (the resolution of maximum mutual information is defined as 1). Vertical axis as in (**a**).

cient reconstruction. In addition, representations resulting from classification schemes contain overlapping fragments at multiple scales that may be redundant for reconstructing the training set.

Informative features are useful for classification

We tested how useful the intermediate features were for classification, compared with more local and more global features. The main difficulty in classification arises from the variability in shape within a natural class of objects[19]. We compared the matching of a novel image using composition of fragments to that using full-face shapes (**Fig. 3**). The full-face condition searched the database for the most similar stored face. The fragments condition matched the novel face with fragments extracted from the same image set as the full-face condition. The fragment-based approximation was markedly better (as judged by eight observers), showing that a modest number of appropriate building blocks can be used in different combinations to deal with shape variation within a class of face images.

We tested the generalization capacity of informative features by using them to classify novel images that were substantially different from the training set. For the purpose of classification, the features were organized by type: different fragments covering the same region of the face, such as the hairline region, were grouped together into the same fragment-type group (**Fig. 1d**). We used this organization because, in classifying an image, the best-matching fragments from each type are selected first, and then combined to produce the final decision. The features of a common type can be represented in terms of simpler fragments (**Fig. 1e**), leading to a hierarchical representation in which an intermediate fragment is defined by the conjunction of lower level features. In this hierarchical representation, an intermediate feature cannot be simply characterized by a preferred sub-image. It is defined instead by the presence and arrangement of its own constituents, similar to the 'critical features'[10] of intermediate- complexity IT units.

Once a fragment F was detected in the image, the strength of the evidence it supplied for the presence of an object of class C was measured by the likelihood ratio (R) of fragment F being found inside and outside of class C:

$$R(F) = \frac{P(F \mid C)}{P(F \mid \bar{C})} \qquad (2)$$

This ratio is commonly used in signal detection, and it is an optimal detection criterion for the presence of an object from C given the fragment F. We used $w = \log_2(R(F))$ as the 'weight' of the fragments. Note that the merit and weight of a feature are two different criteria, as a fragment can have a high weight but still have a low merit. It is efficient for the visual cortex to use features with high merit and to use them according to their weights, which can be implemented by synaptic strengths. To classify an image as a face or non-face, the following formula was used (see Methods for more details):

$$\sum_k w_{i_k} \max_{i_k}(F_{i_k}) > \theta \qquad (3)$$

where F_{i_k} stands for the i^{th} fragment of type k. This means that for each type (such as 'hairline region'), the maximal response of all fragments of this type is selected. The maximum operation is taken also over different retinal positions within a region 25% of object size. We used a scheme similar to a biological model incorporating the maximum operation[20]. The fragments detected from each type were then combined by summing their weights and comparing the sum to a threshold. The threshold was set to the lowest value that gave no false classification on a collection of test images that contained no faces. Once the threshold was fixed, 200 novel face and 200 novel non-face images were tested yielding 97% detection and 2.1% false detection, showing that a biologically plausible combination of informative features is competitive with current classification systems[21] (see **Fig. 4** for examples). Of particular interest is the ability to generalize to novel exemplars, including face paintings that are markedly different from the training set. To show the location of the detected object, each fragment was used to estimate the location of the object center (**Fig. 4**, white boxes). To detect objects at different scales, the input image was re-sampled and analyzed at multiple (2–4) scales[15].

We compared the classification results obtained using optimal fragments with those obtained using fragments that were selected in a similar manner but with a fixed size. These fixed-size fragments were either smaller (one-third the size, 4% of average face area) or larger (33% of face area) than the optimal fragments. They were spaced over a regular 5 × 6 grid, covering the same total area as the optimal fragments, and applied to a new image set. The IC fragments were significantly better (95.6% detection, 0% false alarms) than both the small (97% detection, 30.4% false alarms)

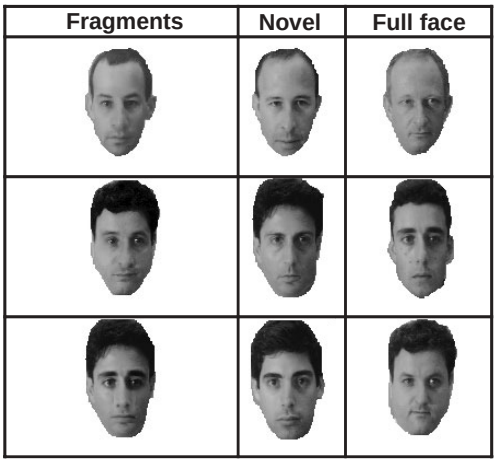

Fig. 3. Approximating novel faces by fragments. Novel faces (middle column) were approximated by full-face images from a data set (right) and by fragments extracted from the same set (left). For the full-face condition, a novel face image was matched against the existing images in the data set, and the best matching face was selected. In the fragments condition, the novel face was matched using the best matching fragment of each type (borders between fragments were blurred to create a smoother image). The fragments improved the compensation for intra-class shape variability.

and large (39% detection, 0% false alarms) fragments. Thus, intermediate fragments selected for informativeness achieved markedly better classification than larger or smaller ones. The selected features were also more informative than were wavelet features of different orientations and scales. We further examined the use of informative fragments within a back-propagation neural network, with the fragments replacing the first-layer features selected by the network itself. Classification results improved markedly, indicating that incorporating informative fragments in standard network models can enhance their classification performance.

Image rearrangement and spatial relations

A representation using multiple-scale, overlapping fragments was useful for enforcing the correct overall arrangement of the features. As there was no explicit representation of the exact location of the fragments or of their spatial relationships, the scheme might have confused a given shape with a shape constructed from the same fragments arranged in a different configuration. However, owing to the use of overlapping fragments at multiple scales, correct configurations were preferred by our model. The model's response to faces at different levels of rearrangement (**Fig. 5**) shows that with increased degree of rearrangement, fragments were increasingly lost and the overall response (equation 3) decreased. In contrast, uniform displacement of the entire figure by a similar amount had a smaller effect on the response. Similar effects have previously been shown[20,22,23] for the identification (rather than classification) of simple shapes.

Physiological studies show a gradual decrease in response as a function of rearrangement in macaque IT neurons[24], as well as in human visual cortex (mapped by functional magnetic resonance imaging)[25]. Information regarding the relative arrangement of fragments is captured because the fragments themselves have a 'jigsaw puzzle' property[22]: their shapes determine their possible interactions, and their assembly is often unique[23]. This is a useful property of IC features that is not included in schemes that use universal V1-type features directly for recognition and classification[12,21].

Discussion

Our results have two main implications for the problem of feature selection in visual processing. First, they show that visual features of intermediate complexity emerge naturally from a coding principle of maximizing the delivered information with respect to a class of objects. Regardless of the particular mechanism used to extract the features, our results explain the relative advantage of IC features for visual classification. Second, they show that visual features based on combinations of object fragments provide a rich set of potential features from which informative ones can be effectively selected. This stands in contrast to back-propagation and other network models for feature selection. Such models typically start from randomly selected features and then seek to improve them locally by small changes, and therefore perform a

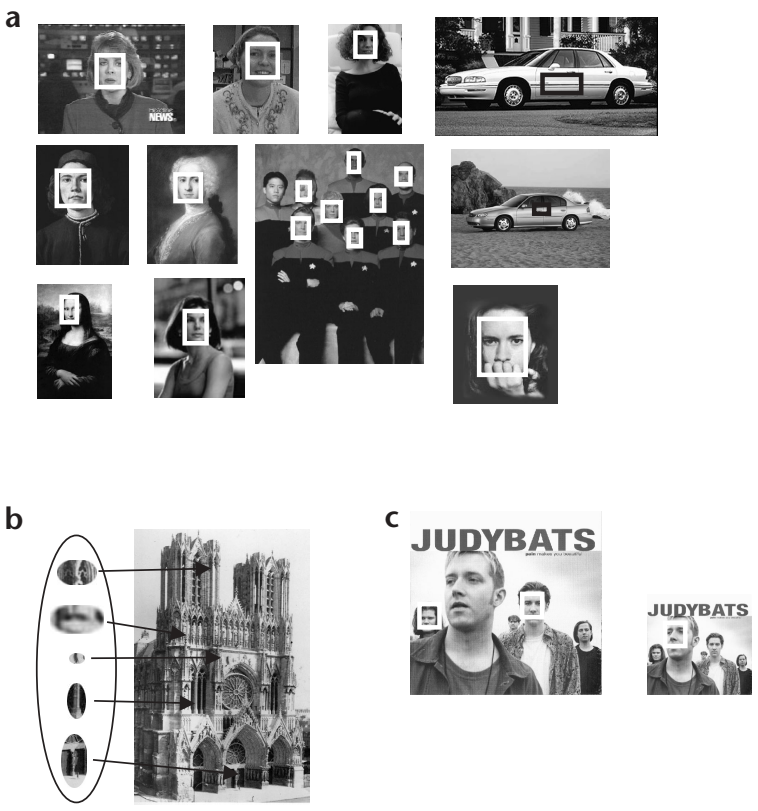

Fig. 4. Face and car detection examples showing broad generalization. Fragments in Fig. 1 and equation (3) were used. (**a**) Detected faces and cars marked by outline squares. (Cars were detected in reduced-resolution images.) (**b**) Individual face-like features can appear occasionally in non-face images, but the conjunction of a sufficient number to exceed threshold is highly unlikely. (**c**) Images were tested at several scales by re-sampling the input image.

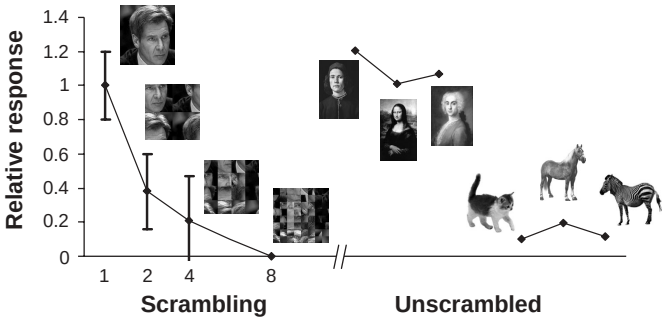

Fig. 5. Detection response (equation 3) decreases with degree of image scrambling. The original image was cut into 2×2, 4×4 or 8×8 sub-images, then scrambled. Left, average response (n = 10) for different degrees of scrambling. Example of one scrambled image shown above the curve. Vertical axis, the response for the original image is defined as 1. Horizontal axis, level of scrambling (8 denotes 8×8 sub-images). For comparison, response is shown for three face (middle) and non-face images (right).

local search in the very large space of all possible shapes. In contrast, the fragment-based scheme performs a more global search in a restricted space of features, composed of combinations of common object parts. This results in more informative features, probably because the unconstrained search converges to a local optimum that is lower in information than the features obtained by fragment selection.

Unlike in many previously described schemes[3–6], the features that emerged here are not universal, but shaped by visual experience with particular classes of objects[7,10,11]. For the task of visual object classification, these features are more informative than simple generic features used by some recognition models, and also more informative than global complex features selected by other schemes[13,14]. As a result, the IC fragment–based features provide an efficient basis for classification and generalization. In IT cortex, many units encode partial rather than complete object views. It would be of interest to compare empirically the features preferred by IT with fragments selected by the model, and also to test whether such features can be shaped, as predicted, by classification experience. With respect to resolution, the model suggests that if cortical cells are tested, the response of some units with large receptive fields will saturate at intermediate rather than full resolution.

In view of these findings and the fundamental role of classification and generalization in vision, the visual system is likely to use a similar coding principle that favors informative features, selected for their use in classification. Although specific neural implementations of the fragment selection process are beyond the scope of this discussion, they may be based on neural connections that are facilitated by the co-occurrence of a feature and a class, and depressed by the occurrence of one without the other. Finally, in the overall hierarchy of features used by the visual system, this work suggests a distinct role for visual features of intermediate complexity. Local V1-like features provide an efficient encoding of natural images in general, global object views are useful for the identification of specific objects under familiar viewing conditions[26], and intermediate complexity features are optimally suited to support generalization and classification.

Methods

Image sets. Training images for fragment extraction were 138 face images and 40 car images. Faces were roughly rescaled to 40 columns horizontally. Non-class images were a random collection including landscapes, fruits and toys with a similar gray-level range. Examples of the images as well as informative fragments extracted from them can be viewed at http://www.wisdom.weizmann.ac.il/~michel/fbc/fbc.html. The site also contains additional details on the fragment extraction computation and comparisons with alternative features.

Procedure. The search for informative fragments examined candidate fragments at multiple locations and sizes, where a fragment is a sub-image of some size $p \times q$ taken from one of the images. These sub-images were searched for in every database image and their mutual information was computed. The fragment-to-image comparison used a weighted sum of gray-level gradient and orientation differences. We also tested normalized cross-correlation and the ordinal measure (in ref. 27) and obtained similar classification results.

To compute the mutual information $I(C,F)$ in equation (1), we measured the frequency of detecting a given fragment F in the database of images that contain or do not contain objects in the class, and assumed $P(C) = 0.05$. The entropy $H(x)$ of a random variable x is given by $-\Sigma P(x)\log(P(x))$. Here C and F are binary variables: $F = 1$ if the fragment is found in the image and 0 otherwise; $C = 1$ if the image belongs to the class in question and 0 otherwise. $I(C,F)$ is then given by:

$$I(C,F) = -P(C)Log(P(C)) - P(\bar{C})Log(P(\bar{C}) + P(F)((P(C\,|\,F)\,Log(P(C\,|\,F)) + P(\bar{C}\,|\,F)Log\,P(\bar{C}\,|\,F))) + P(\bar{F})((P(C\,|\,\bar{F})\,Log(P(C\,|\,\bar{F})) + P(\bar{C}\,|\,\bar{F})Log(P(\bar{C}\,|\,\bar{F}))) \quad (4)$$

The measured frequencies depend on the detection threshold used; for each fragment, the threshold that maximized the mutual information was selected.

To make the search more efficient, the algorithm was divided into two stages. The first stage identified the approximate location and size of candidate fragments, by searching over a restricted range of sizes and locations (steps of three pixels). The second stage made additional comparisons centered around the locations and sizes of the fragments identified in the first stage. We used a range of sizes from one-half the area of F_i to twice the area of F_i, in integral number of pixels. The process was repeated for 20 image resolutions spaced linearly over a total factor of ten. Image resolution was reduced by convolving the image with a smoothing kernel (cubic spline in Matlab, Mathworks, Natick, Massachusetts) that reduced the high-frequency cutoff of the image, and then by re-sampling the image with a smaller number of pixels. For car fragments, low-resolution fragments were typically superior and all were selected in low (20×40) resolution images. The most informative fragments were selected as described in the text. After the first eight face fragments, additional fragments were searched by types. For each type (for example, hairline region) the general location in the images was marked manually and the search proceeded in the marked regions. In total, we used 48 face and 28 car fragments (we also tested twice and three-times as many fragments, which resulted in only a small improvement in performance). For each fragment, merit was computed by mutual information (equation 1). The weight of fragment F_i was defined by the log likelihood ratio (equation 2). For $F_i = 1$ (fragment detected), $w_i(1) = \log_2 [P(F_i = 1/C)/ P(F_i = 1/ \bar{C})]$, similarly for $w_i(0)$. We used equivalently the weight $w_i(1) - w_i(0)$ when $F_i = 1$, and 0 otherwise. We used $P(C) = 0.05$, but found that fragment selection is insensitive to this value.

Classification experiments were done on a new set of 600 images, 200 for each of three classes (face, car and non-class). To detect a face, for example, all fragments were searched at each image location. Detected fragments were combined within a 40-pixel search window using equation (3). This combination rule assumes conditional independence between fragments given the class variable, and often gives good results[28]. We also

applied a more complex combination rule that takes into account pairwise correlations between fragments to a test set of car images, which resulted in a small increase in overall performance, and combination by a back-propagation network which resulted in an additional performance increment.

Acknowledgments

We thank J. Golberger, M. Bar and N. Rubin for helpful discussions. Supported by Grant 99-28 CN-QUA.05 from the James S. McDonnell Foundation and by the Moross Laboratory at the Weizmann Institute. Face images for testing were in part from the to Carnegie Mellon University (CMU) face images database http://www.cs.cmu.edu/~har/faces.html#upright.

Competing interests statement

The authors declare that they have no competing financial interests.

RECEIVED 6 MARCH; ACCEPTED 21 MAY 2002

1. Barlow, H. B. & Foldiak, P. in *The Computing Neuron* (eds. Durbin, R., Miall, C. and Mitchison, G.) 54–72 (Addison-Wesley, Reading, Massachusetts, 1989).
2. Atick, J. J. & Redlich, N. A. What does the retina know about natural scenes? *Neural Comput.* **4**, 196–210 (1992).
3. Bell, A. J. & Sejnowski, T. J. The 'independent components' of natural scenes are edge filters. *Vision Res.* **37**, 3327–3338 (1997).
4. Field, D. J. What is the goal of sensory coding? *Neural Comput.* **6**, 559–601 (1994).
5. Olshausen, B. & Field, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607–609 (1996).
6. Vinje, W. E. & Gallant, J. L. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* **287**, 1273–1276 (2000).
7. Rolls, E. T. Neural organization of higher visual functions. *Curr. Opin. Neurobiol.* **1**, 275–278 (1991).
8. Fujita, I., Tanaka, K., Ito, M. & Cheng, K. Columns for visual features of objects in monkey inferotemporal cortex. *Nature* **360**, 343–346 (1992).
9. Gallant, J. L., Braun, J. & Van Essen, D. C. Selectivity for polar, hyperbolic, and cartesian gratings in macaque visual cortex. *Science* **259**, 100–103 (1993).
10. Tanaka, K. Neuronal mechanisms of object recognition. *Science* **262**, 685–688 (1993).
11. Logothetis, N. K., Pauls, J., Bulthoff, H. H. & Poggio, T. Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.* **5**, 552–563 (1995).
12. Wiskott, L., Fellous, J. M., Krüger, N., & von der Malsburg, C. Face recognition by elastic bunch graph matching. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 775–779 (1997).
13. Turk, M. & Pentland, A. Eigenfaces for recognition. *J. Cogn. Neurosci.* **3**, 71–86 (1990).
14. Belhumeur, P. N., Hespanha, J. P. & Kriegman, D. J. Eigenfaces versus fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 711–720 (1997).
15. Ballard, D. H. & Brown, C. M. *Computer Vision* (Prentice-Hall, Inglewood Cliffs, New Jersey, 1982).
16. DeValois, R. L., Albrecht, D. G. & Thorell, L. G. Spatial frequency selectivity of cells in the macaque visual cortex. *Vision Res.* **22**, 545–559 (1982).
17. Brunelli, R. & Poggio, T. Face recognition: features versus templates. *IEEE Trans. Pattern Anal. Mach. Intell.* **15**, 1042–1052 (1993).
18. Lee, D. D. & Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* **401**, 788–791 (1999).
19. Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M. & Boyes-Braem, P. Basic objects in natural categories. *Cognit. Psychol.* **8**, 382–439 (1976).
20. Riesenhuber, M. & Poggio, T. Hierarchical models of object recognition in cortex. *Nat. Neurosci.* **2**, 1019–1025 (1999).
21. Yang, M. H., Kriegman, D. J. & Ahuja, N. Detecting faces in images: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 34–58 (2002).
22. Mel, B. W. SEEMORE: combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Comput.* **9**, 777–804 (1997).
23. Ullman, S. & Soloviev, S. Computation of pattern invariance in brain-like structures. *Neural Net.* **12**, 1021–1036 (1999).
24. Vogels, R. Effects of image scrambling on inferior temporal cortical responses. *Neuroreport* **10**, 1811–1816 (1999).
25. Grill-Spector, K., Kushnier, T., Hendler, T., Edelman, S., Itzchak, Y. & Malach, R. A sequence of object-processing stages revealed by fMRI in human occipital lobe. *Hum. Brain Mapp.* **6**, 316–328 (1998).
26. Logothetis, N. K. & Sheinberg, D. L. Visual object recognition. *Annu. Rev. Neurosci.* **19**, 577–621 (1996).
27. Bhat, D. & Nayar, K. S., Ordinal measures for image correspondence. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 415–423 (1998).
28. Friedman, N., Geiger, D. & Goldszmidt, M. Bayesian network classifiers. *Mach. Learn.* **29**, 131–163 (1997).