A template-matching pandemonium recognizes unconstrained handwritten characters with high accuracy

AXEL LARSEN and CLAUS BUNDESEN
University of Copenhagen, Copenhagen, Denmark

Psychological data suggest that internal representations such as mental images can be used as templates in visual pattern recognition. But computational studies suggest that traditional template matching is insufficient for high-accuracy recognition of real-life patterns such as handwritten characters. Here we explore a model for visual pattern recognition that combines a template-matching and a feature-analysis approach: Character classification is based on weighted evidence from a number of analyzers (demons), each of which computes the degree of match between the input character and a stored template (a copy of a previously presented character). The template-matching pandemonium was trained to recognize totally unconstrained handwritten digits. With a mean of 37 templates per type of digit, the system has attained a recognition rate of 95.3%, which falls short of human performance by only 2%–3%.

Analysis of an input pattern by template matching consists in superimposing a stored pattern (template) on the input and determining the degree of match (overlap or correlation) between the input and the template. In traditional recognition by template matching, the input is first compared against a number of templates and then classified as a member of the same category as the best matching template. To compensate for irrelevant variations in the spatial position, size, and orientation of the input, the input pattern and the template can be aligned by being shifted, size-scaled, and rotated before their degree of match is determined.

Psychological studies suggest that simple visual patterns such as letters or digits can be recognized by use of internal representations as holistic templates. Mental images form one type of representations that seem to be used as templates. They can be transformed, and transformation of mental images appears to be one way of achieving recognition regardless of stimulus position, size (see Bundesen & Larsen, 1975; Jolicoeur & Besner, 1987; Larsen & Bundesen, 1978), and orientation (see Shepard & Cooper, 1982; Shepard & Metzler, 1971; also see Bundesen, Larsen, & Farrell, 1981; Larsen, 1985).

Most cases of visual recognition are presumably based on comparing input patterns against long-term memory representations instead of short-term representations such as mental images. The nature of visual long-term memory representations is controversial, but simplicity favors the view that visual representations in long-term memory are similar in format to visual representations in short-term memory (template-like mental images).

Suggestive empirical evidence that visual long-term representations may be used as templates in recognition of simple visual patterns has come from two sources. First, a number of reaction time studies have shown systematic decrements in recognition speed and accuracy caused by irrelevant variations in visual size (see Cave & Kosslyn, 1989; Larsen & Bundesen, 1978) and orientation (see Cooper, 1975; Jolicoeur, 1985, 1990; Jolicoeur & Landau, 1984). The results support the notion that "visual pattern recognition is based on position-wise comparison of stimulus patterns with memory representations" (Larsen & Bundesen, 1978, p. 19), and template matching is the most elementary way of making *position-wise* comparisons (also see Ullman, 1989). Second, template-matching models have yielded good fits to observed visual-confusion matrices (see Gervais, Harvey, & Roberts, 1984; Holbrook, 1975) and excellent fits to observed variations in legibility across character sets (see Loomis, 1990).

The computational efficiency of template matching has been seriously questioned. For several decades, common wisdom has held that template matching is insufficient for recognition of unconstrained real-life patterns (see, e.g., Eysenck & Keane, 1990; Hummel & Biederman, 1992; Humphreys & Bruce, 1989; Lindsay & Norman, 1972; Neisser, 1967; Reed, 1973). Recognition of unconstrained handwritten characters is probably the most

frequent textbook example of a task on which template matching should fail.

We recently tested the efficiency of traditional template matching in machine recognition of totally unconstrained handwritten digits (Larsen & Bundesen, 1992). Our learning and recognition algorithm was simple; no previous knowledge concerning handwritten digits was presupposed, and preprocessing was limited to Gaussian smoothing and normalization with respect to position, size, and orientation. For patterns presented in a known orientation, recognition rates were 69%, 77%, and 89%, respectively, when about 5, 10, and 60 templates had been learned for each type of digit. For patterns presented in unknown orientations, recognition rates were slightly lower. High levels of reliability could be attained by omitting classifications based on weak evidence. However, at the end of training, the effect of further increase in the number of templates was extremely small, and recognition rates substantially higher than 90% seemed practically impossible to obtain. For comparison, human subjects tested with a random sample of the handwritten digits (presented in upright orientation) achieved a mean recognition rate of 97%.

In traditional recognition by template matching, the classification of an input depends solely on the type of the best matching template. This is wasteful of information. The method fails to utilize the diagnostic power of templates for any given type of digit in discriminating between digits of other types. For example, the degree of match with a template for a digit of Type 0 yields information about the likelihood that the input belongs to Type 6 rather than Type 7, but this information is not utilized. When the template-matching approach is combined with a feature-analysis (*pandemonium*; Selfridge, 1959; Selfridge & Neisser, 1960) approach to recognition, the degree of match with any given template may be treated as a particular feature of the input, and the feature may be used as positive or negative evidence for any classification. We explored such a system. It encodes new patterns as templates, uses the templates for feature analysis, strengthens the role of useful templates, and weakens the role of useless ones.

MODEL

Our pandemonium model of human character recognition contains a number of feature demons (analyzers), each of which stores a particular template. The template is a copy of a previously presented character. When a character is presented to the system for recognition, the character is first normalized in spatial position and size and smoothed by convolution with a circularly symmetric two-dimensional Gaussian filter. Next, each feature demon determines the degree of match between the character and the template stored with the demon. The degree of match is a measure of the maximum correlation that can be found between the two patterns by permitting some displacement between their centroids.

Above the level of feature demons is a number of cognitive demons, one for each type of character. Each cognitive demon is connected to every feature demon, and the net input to a cognitive demon is a weighted sum of the degrees of match determined by the feature demons. The activation of a cognitive demon increases with the net input to the unit.

On top of the processing hierarchy is a decision demon. It classifies the input character as belonging to the type that corresponds to the cognitive demon with the highest activation.

SIMULATIONS

General Method

Input patterns. The input patterns presented to the pandemonium consisted of 6,000 totally unconstrained handwritten digits (600 tokens of each of the 10 types of digit). The digits were taken from zip codes collected by the U.S. Postal Service from dead letter envelopes (see Larsen & Bundesen, 1992). The number of writers is not known, but can be assumed to approach the actual number of samples. The material was provided in digitized and binarized form. Typical samples are shown in Figure 1.

Recognition algorithm. When an input pattern was presented to the system, it was processed as follows.

1. The *centroid* ("center of gravity") of the character (defined as a figure consisting of pixels with a value of 1 on a ground of pixels with a value of 0) was found.

2. The *size* of the character was determined as the greatest distance from the centroid to any pixel that belonged to the character (i.e., any pixel with a value of 1).
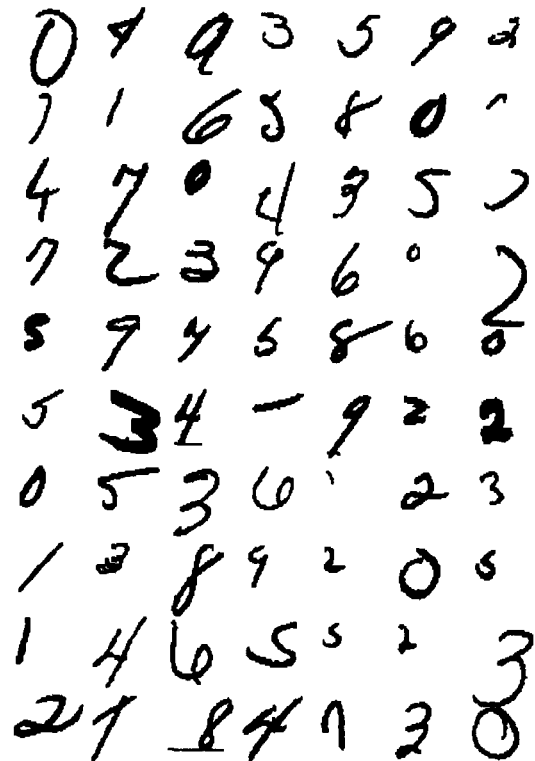
Figure 1. Examples of input patterns.

3. The input pattern was *normalized* with respect to the position and the size of the character. Effectively, an object-centered Cartesian xy coordinate system was imposed on the input pattern so that the origin of the coordinate system coincided with the centroid of the character and the units of length along the x and y axes equaled the size of the character.[1]

4. The normalized input pattern was *smoothed* by convolution with a two-dimensional Gaussian filter,

$$G(x,y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right),$$

with standard deviation σ.

5. The smoothed input pattern was *compared* with every template stored in memory. When the input was compared with a particular template, the two patterns were first aligned so that their centroid pixels coincided. Then the template was shifted relative to the input by a certain number of pixels along the x axis and a certain number of pixels along the y axis. All shifts of up to ±0.2 units of length (20% of the size of a character) along the x axis and ±0.2 units along the y axis were tested. For each shift, the product moment correlation between the two patterns was determined. Letting r be the highest product moment correlation obtained between the two patterns, the degree of match between the input and the template was defined as r^7 (see Larsen & Bundesen, 1992).[2] The degree of match was stored as the level of activation of a unit (*template node*) associated with the template.

6. The input was *classified* as a digit of Type 0, 1, ..., or 9. Each digit type was represented by a unit (*class node*), the net input of which was a weighted sum of the degrees of match computed at Step 5 (the levels of activation of all template nodes). The activation of the class node equaled the hyperbolic tangent of its net input.[3] The input pattern was classified as a token of the digit type represented by the class node with the highest activation.

Implementation. The algorithm was written in C. It was executed on a computer system consisting of a Digital Equipment Corporation Micro-VAX 2 and a DEC-station 3100.

The input characters varied widely in size, about 7 $\times$ 12 pixels up to 53 $\times$ 53 pixels. After normalization, each character was represented in a format such that the greatest distance from the centroid (i.e., the centroid pixel) to any pixel that belonged to the character equaled 15 pixels (1 unit of length). Distances between pixels were measured from center to center.

Whereas the normalized character was represented as a figure of pixels with a value of 1 on a ground of pixels with a value of 0, the Gaussian filter coefficients were quantized on a 7-bit scale so that the sum of the coefficients was within the range of the scale. The standard deviation σ of the Gaussian filter equaled 1.5 pixels, and the greatest distance from the center of the Gaussian filter to a pixel at which the quantized filter coefficient was different from 0 equaled 3 pixels.

The smoothed input pattern had a value of 0 at any pixel farther than 18 pixels from the centroid. Each product moment correlation between the smoothed input pattern and a template stored in memory was made by computing the Pearson product moment correlation coefficient between the two patterns across all pixels located at or within a distance of 18 pixels from the centroid of one or the other pattern.

Experiments

The recognition system was trained on a fixed subset of the input patterns. The subset (training set) consisted of 4,000 digits (400 randomly selected tokens of each
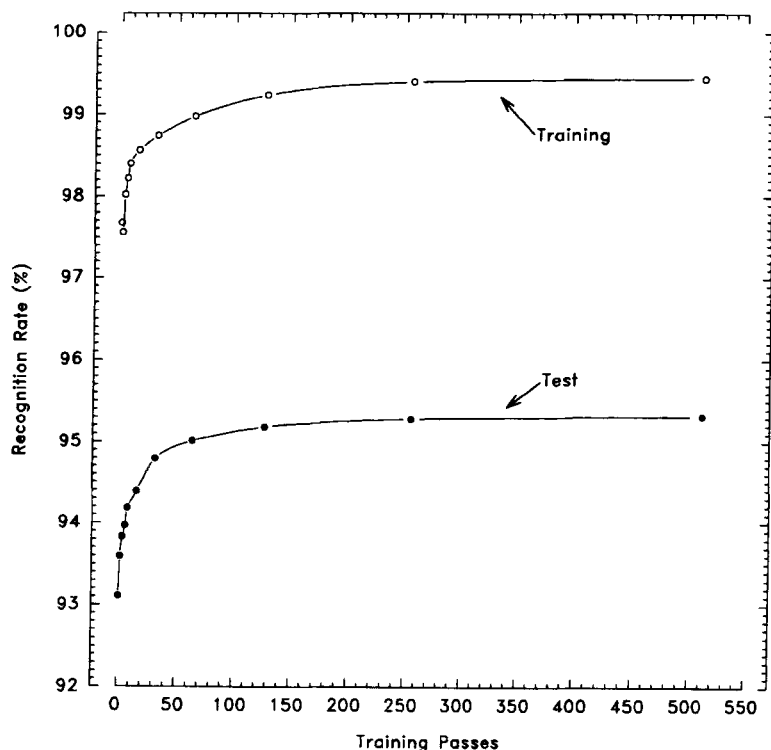

Figure 2. Mean recognition rates on training (open circles) and test sets (filled circles) as functions of the number of passes through the training set. (Results are based on five independent simulations.)

type). In each pass through the training set, the 4,000 digits were presented one at a time in random order.

At the beginning of training, only one template was stored in memory. This template was a smoothed normalized version of a randomly selected member of the training set. The weight on the connection from the template node to the class node that represented the character type of the template (the correct classification) was 1. Weights on connections to the other nine class nodes were -0.01.

During the first pass through the training set, learning occurred whenever an input was incorrectly classified. In this case, the smoothed normalized version of the incorrectly classified input was added to the set of templates stored in memory. The corresponding template node was connected to the correct class node with a weight of 1 and to the other nine class nodes with weights of -0.01. After the first pass, about 370 templates had been stored (mean of five independent simulations).

During later passes through the training set, no new templates were acquired, but weights on connections from template to class nodes were adjusted. After each presentation of a training pattern, weight adjustments were made by the delta rule (see Stone, 1986; Sutton & Barto, 1981; see also Donegan, Gluck, & Thompson, 1989). Specifically, the change in weight on the connection from template node i to the class node for digit type j equaled

$$0.025d_i(t - a_j).$$

Here, d_i is the degree of match signaled by the template node, and a_j is the activation of the class node. If the training pattern was of type j, then constant $t = 0.7$, else $t = -0.7$.

After each pass through the training set, performance was measured on a separate test set. The test set consisted of 2,000 digits (200 tokens of each type). No learning occurred during passes through the test set.

Results based on five independent simulations are shown in Figure 2. Before weights were adjusted (i.e., after the first pass through the training set), the recognition rate averaged 97.7% on the training set and 93.1% on the test set. At the end of training, the recognition rates on the training and test sets were 99.4% and 95.3%, respectively.[4] For both sets, convergence to the higher level of performance was fast.

Figure 3 shows distributions of weights on connections from template to class nodes at the end of training (means of the five simulations). Let an *intrinsic connection* be a connection from a node for a template of a certain type of digit to the class node for this type of digit, and let an *extrinsic connection* be a connection from a node for a template of a certain type of digit to a class node for a different type of digit. The upper panel of Figure 3 shows the distribution of weights on intrinsic connections, and the lower panel shows the distribution of weights on extrinsic connections. The distribution of weights on intrinsic connections had a mean of 1.5 and a standard deviation of 0.7, whereas the distribution of weights on extrinsic connections had a mean of -0.3 and a standard deviation of 0.5. Thus, as would be expected,

almost all of the weights on intrinsic connections were positive. Most weights on extrinsic connections were negative, but a substantial proportion (21%) were positive. Many templates had heavy positive weights on their intrinsic connections, light positive weights on some of their extrinsic connections, and negative weights on the remaining ones.

The *reliability* of the recognition responses was defined as the relative frequency of correct responses among all responses. At the expense of getting omission errors (*rejections*), reliability of recognition responses could be increased by omitting responses that would have been based on weak evidence. During passes through the test set, we investigated the effect of rejecting a test pattern (omitting response to the pattern) if the difference in activation between the two most active class nodes was below a certain threshold. The rejection rate was varied by changing the threshold. The results are shown in Figure 4. As illustrated, reliabilities of 97.0% and 99.0% were
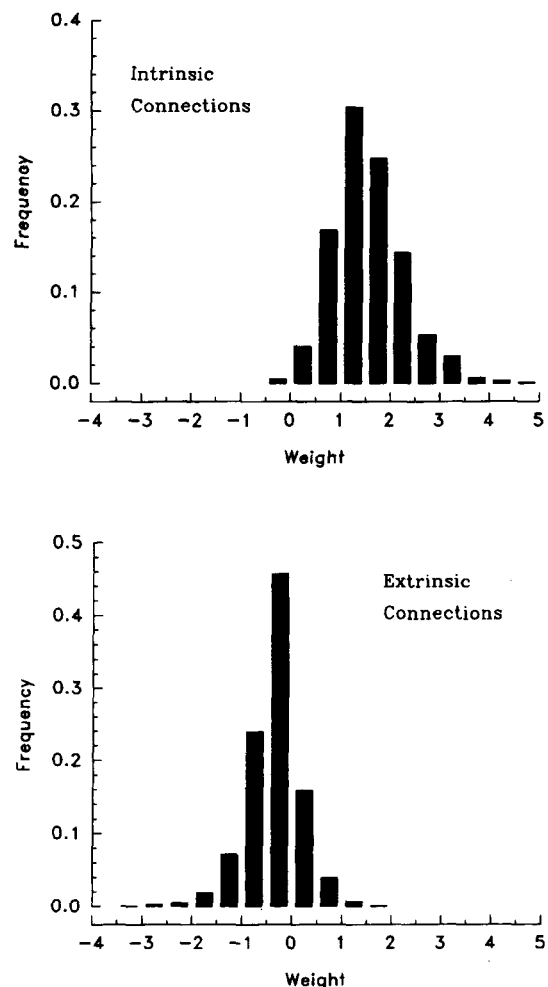

Figure 3. Distributions of weights on connections from template to class nodes at the end of training. Upper panel: Weights on intrinsic connections. Lower panel: Weights on extrinsic connections. Results are based on five independent simulations.

attained on the test set with rejection rates of 3.5% and 14.1%, respectively.

The number of templates stored in memory could be substantially reduced with only modest effects on performance. At the end of training, the strength of a template was calculated as the sum of the squares of the weights on the connections from the template's node to the class nodes. The weakest templates and their nodes were removed, the remaining weights were readjusted by further practice on the training set, and the recognition rate was measured on the test set. The curve shown in Figure 5 was generated by repeating this process many times. As can be seen, a recognition rate of 95% could be attained with a total of 278 templates or about 28 templates per type of digit. With a mean of 8 templates per type of digit, the recognition rate (89%) equaled the rate we have reported for traditional recognition by template matching with about 60 templates per type of digit (see Larsen & Bundesen, 1992).

DISCUSSION

Computational Efficiency

The recognition rate attained by the template-matching pandemonium model of human character recognition is remarkable. It falls short of human performance by only 2%–3% (see Larsen & Bundesen, 1992). Despite the extreme simplicity of the model, it appears to perform as well as the most complex and successful machine algorithms designed specifically for recognition of hand-

Figure 5. Recognition rate as a function of the mean number of templates per type of digit.
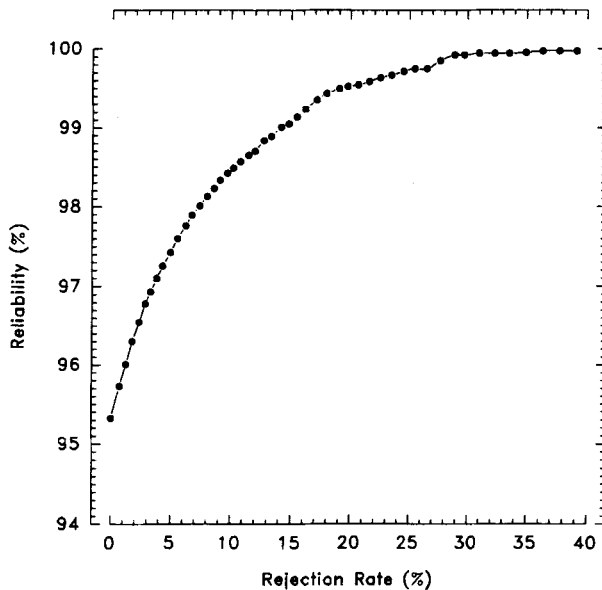

Figure 4. Reliability of recognition responses as a function of rejection rate. (The reliability is the relative frequency of correct responses among all responses. A test pattern was rejected [response was omitted] if the difference in activation between the two most active class nodes was below a certain threshold. The rejection rate was varied by changing the threshold. Results are based on five independent simulations.)
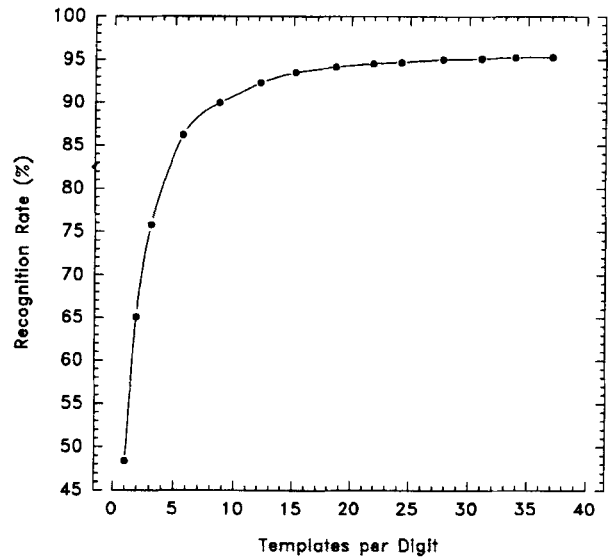
written digits (see, e.g., Lam & Suen, 1988; LeCun et al., 1989; Suen, 1990).

It is instructive to trace the steps by which the recognition rate was improved when our system for recognition by traditional template matching (Larsen & Bundesen, 1992) was developed into the template-matching pandemonium. In the system for recognition by traditional template matching, the input was always classified as a member of the same category as the best matching template. Various degrees of Gaussian smoothing were tested. Without any Gaussian smoothing, the recognition rate was 86% at the end of training (i.e., when about 75 templates had been stored for each type of digit). The highest recognition rate was obtained when the standard deviation σ of the Gaussian filter equaled 10% of the size of a normalized character. In this case the recognition rate reached a value of 89% (with about 60 templates per type of digit).

A small further increase in recognition rate was obtained by improving the normalization with respect to position. By tolerating minor displacements between centroids of templates and input patterns (see Step 5 of the current recognition algorithm), the recognition rate was increased by about 0.5%.

Dramatic improvements in performance were found when the system was rewired so that classification was based on evidence summed across many templates rather than evidence provided by the best matching template. When all weights on intrinsic connections from template to class nodes were equally great, and weights on extrinsic connections were close to zero, the recognition rate averaged 93% (with about 37 templates per type of digit). Adjusting the weights through learning yielded a further increment of several percentiles up to the final recognition rate of 95.3%.

Feature Analysis by Template Matching

Our template-matching pandemonium model for character recognition combines a template-matching with a feature-analysis approach to pattern recognition. In the combined approach, input patterns are analyzed on a number of form dimensions (dimensions of variation in form). Each form dimension is defined by a template. The value of the input pattern on the dimension equals the degree of match between the input and the template. As the value on the dimension (the degree of similarity to the template) is a visual feature of the input, the process of template matching is a process of feature analysis. Recognition of the input pattern is based on all of the visual features that are extracted by template matching. Each feature is used as positive or negative evidence for each of the possible classifications.

The templates in the current pandemonium are copies of previously presented instances of handwritten digits. Thus the long-term memory of the system contains pictorial representations of moderate complexity. The analysis of a new pattern is done by cross-correlating the input with each of these representations, so new patterns are analyzed by being filtered through a sample of previously experienced patterns. By the learning process governed by the delta rule, the role of useful feature analyzers (templates) is strengthened and the role of useless analyzers is weakened. If the weakest analyzers are purged (see Figure 5), feature analyzers are effectively selected on the basis of their diagnostic power with respect to the set of relevant pattern classifications. The feature analyzers that survive correspond to those copies of previously presented patterns that are highest in diagnostic power.

The template-matching pandemonium throws new light on the quest for the basic units or dimensions of visual pattern analysis. In the middle of the century, a strong plea was made for the development of a *perceptual psychophysics* (Gibson, 1950, 1959) that should include a *psychophysics of form* (Attneave & Arnoult, 1956). Quantitative studies of shape and pattern perception were initiated in attempts to create a psychological metric of visual form (see, e.g., Brown & Owen, 1967; Michels & Zusne, 1965). Ad hoc geometric measures such as number of sides in a figure, number of angles, moments of area, and moments of the perimeter were computed and correlated with behavioral measures (see, e.g., Zusne, 1970). However, the fundamental problem remained unsolved. The template-matching pandemonium model suggests that the quest for the basic units or dimensions of visual form was misguided. It suggests that there are very many dimensions of variation in visual form (one for each template), but no particularly basic ones. The important dimensions of variation in visual form are not mutually orthogonal, and few, if any, are universal.

Possible Role of Structural Descriptions

Three general approaches to visual pattern recognition are commonly considered: template matching, feature analysis, and structural description (see, e.g., Hummel & Biederman, 1992; Reed, 1973). In the structural-description approach, an object is represented by a structural description, that is, a symbolic representation of the geometric structure of the object. The structural description specifies the components of the object and ways in which the components are interrelated. The components may be elementary visual features (Sutherland, 1968) or three-dimensional primitives such as generalized cones (Marr, 1982; Marr & Nishihara, 1978) or geons (Biederman, 1987; Biederman & E. E. Cooper, 1992; see also related work by L. A. Cooper, Schacter, Ballesteros, & Moore, 1992).

The template-matching, feature-analysis, and structural-description approaches to pattern recognition are often contrasted, but they are not incompatible. As described in the previous section, our template-matching pandemonium model for character recognition combines a template-matching and a feature-analysis approach. The template-matching pandemonium might be included as a character recognition module in a model of word recognition based on structural descriptions (symbolic representations of spatial arrangements of letters). Such a model would represent a simple synthesis of template-matching, feature-analysis and structural-description approaches to visual pattern recognition.

Recognition of Three-Dimensional Objects

The template-matching pandemonium model can be extended to recognition of three-dimensional objects. A natural extension can be made by representing a three-dimensional object by a collection of two-dimensional perspective views obtained by inspecting the object from different viewpoints. This mode of representation is attractive because learning of a multiple-view representation seems much more easy than learning of a three-dimensional model of the object (a single viewpoint–invariant description of the three-dimensional geometric structure of the object; see Biederman, 1987; Lowe, 1987; Marr, 1982).

The efficiency of multiple-view representations of three-dimensional objects has been explored by Edelman and Weinshall (1991), who trained a simple two-layer network to recognize wire-frame objects from different viewpoints. The network learned to recognize 10 different objects, and the extent of generalization to novel views seemed comparable to that found in human subjects (see Rock & DiVita, 1987; Rock, Wheeler, & Tudor, 1989; Tarr & Pinker, 1989). At a more general level, Poggio and Edelman (1990) and Ullman and Basri (1991) have provided computational arguments that a three-dimensional object can be recognized from any viewpoint by use of a multiple-view representation based on a small number of views.

Neural Mechanisms of Visual Recognition

The template-matching pandemonium model seems generally consistent with electrophysiological findings on neural mechanisms of visual recognition. Shape-

based recognition is thought to be subserved by a visual pathway running from primary visual cortex (V1) via visual areas V2 and V4 to inferotemporal cortex (IT) (see Goodale & Milner, 1992; Mishkin, Ungerleider, & Macko, 1983; Ungerleider & Mishkin, 1982). In an extensive investigation of the way in which visual form is represented in the IT of the macaque monkey, Tanaka, Saito, Fukada, and Moriya (1991) determined the optimal stimulus of individual IT cells. In the anterior part of IT (i.e., TE), most cells required moderately complex features for their activation. Examples of features required for activation of individual cells in TE are the shape of an inverted T, a six-rayed star, a horizontally striped disk on top of a vertically striped one. These critical features are two-dimensional, and responses of cells were almost always selective for the orientation of stimuli. The selectivity to the optimal stimulus was fairly sharp, but not absolute.

Fujita, Tanaka, Ito, and Cheng (1992) found that cells located at nearby positions in TE had similar—but not identical—stimulus selectivity. Their results suggest that TE consists of columnar modules in which cells with overlapping but slightly different selectivity cluster together. According to Tanaka (1993), the selectivity and the columnar organization are not determined by genes or early development in infancy, but are subject to changes by perceptual learning in the adult. Extended discrimination training with particular shapes produces a marked increase in the number of cells that give a maximal response to some of the trained stimuli.

There are strong similarities between the foregoing description of processing in TE and the template-matching pandemonium. Just like the template-matching pandemonium, TE contains a lot of units working in parallel. Each unit has a critical feature of moderate complexity (much like a single handwritten digit), and each unit belongs to a cluster of units with overlapping but slightly different selectivity (like handwritten samples of a given type of digit). Units that are maximally sensitive to particular shapes are created by training with these shapes, but the critical features appear to be two-dimensional rather than three-dimensional. The two-dimensionality accords with the conjecture that three-dimensional objects are recognized by use of multiple-view representations in a template-matching pandemonium.

REFERENCES

ATTNEAVE, F., & ARNOULT, M. D. (1956). The quantitative study of shape and pattern perception. Psychological Bulletin, 53, 452-471.

BIEDERMAN, I. (1987). Recognition-by-components: A theory of human image understanding. Psychological Review, 94, 115-147.

BIEDERMAN, I., & COOPER, E. E. (1992). Size invariance in visual object priming. Journal of Experimental Psychology: Human Perception & Performance, 18, 121-133.

BROWN, D. R., & OWEN, D. H. (1967). The metrics of visual form: Methodological dyspepsia. Psychological Bulletin, 68, 243-259.

BUNDESEN, C., & LARSEN, A. (1975). Visual transformation of size. Journal of Experimental Psychology: Human Perception & Performance, 1, 214-220.

BUNDESEN, C., LARSEN, A., & FARRELL, J. E. (1981). Mental transformations of size and orientation. In J. Long & A. D. Baddeley (Eds.), Attention and performance IX (pp. 279-294). Hillsdale, NJ: Erlbaum.

CAVE, K. R., & KOSSLYN, S. M. (1989). Varieties of size-specific visual selection. Journal of Experimental Psychology: General, 118, 148-164.

COOPER, L. A. (1975). Mental rotation of random two-dimensional shapes. Cognitive Psychology, 7, 20-43.

COOPER, L. A., SCHACTER, D. L., BALLESTEROS, S., & MOORE, C. (1992). Priming and recognition of transformed three-dimensional objects: Effects of size and reflection. Journal of Experimental Psychology: Learning, Memory, & Cognition, 18, 43-57.

DONEGAN, N. H., GLUCK, M. A., & THOMPSON, R. F. (1989). Integrating behavioral and biological models of classical conditioning. In R. D. Hawkins & G. H. Bower (Eds.), Computational models of learning in simple neural systems: The psychology of learning and motivation (Vol. 23, pp. 109-156). New York: Academic Press.

EDELMAN, S., & WEINSHALL, D. (1991). A self-organizing multiple-view representation of 3D objects. Biological Cybernetics, 64, 209-219.

EYSENCK, M. W., & KEANE, M. T. (1990). Cognitive psychology: A student's handbook. Hillsdale, NJ: Erlbaum.

FUJITA, I., TANAKA, K., ITO, M., & CHENG, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. Nature, 360, 343-346.

GERVAIS, M. J., HARVEY, L. O., JR., & ROBERTS, J. O. (1984). Identification confusions among letters of the alphabet. Journal of Experimental Psychology: Human Perception & Performance, 10, 655-666.

GIBSON, J. J. (1950). The perception of the visual world. Boston: Houghton Mifflin.

GIBSON, J. J. (1959). Perception as a function of stimulation. In S. Koch (Ed.), Psychology: A study of a science (Vol. 1, pp. 456-501). New York: McGraw-Hill.

GOODALE, M. A., & MILNER, A. D. (1992). Separate visual pathways for perception and action. Trends in Neurosciences, 15, 20-25.

HOLBROOK, M. B. (1975). A comparison of methods for measuring the interletter similarity between capital letters. Perception & Psychophysics, 17, 532-536.

HUMMEL, J. E., & BIEDERMAN, I. (1992). Dynamic binding in a neural network for shape recognition. Psychological Review, 99, 480-517.

HUMPHREYS, G. W., & BRUCE, V. (1989). Visual cognition: Computational, experimental, and neuropsychological perspectives. Hillsdale, NJ: Erlbaum.

JOLICOEUR, P. (1985). The time to name disoriented natural objects. Memory & Cognition, 13, 289-303.

JOLICOEUR, P. (1990). Orientation congruency effects on the identification of disoriented shapes. Journal of Experimental Psychology: Human Perception & Performance, 16, 351-364.

JOLICOEUR, P., & BESNER, D. (1987). Additivity and interaction between size ratio and response category in the comparison of size-discrepant shapes. Journal of Experimental Psychology: Human Perception & Performance, 13, 478-487.

JOLICOEUR, P., & LANDAU, M. J. (1984). Effects of orientation on the identification of simple visual patterns. Canadian Journal of Psychology, 38, 80-93.

LAM, L., & SUEN, C. Y. (1988). Structural classification and relaxation matching of totally unconstrained handwritten zip-code numbers. Pattern Recognition, 21, 19-31.

LARSEN, A. (1985). Pattern matching: Effects of size ratio, angular difference in orientation, and familiarity. Perception & Psychophysics, 38, 63-68.

LARSEN, A., & BUNDESEN, C. (1978). Size scaling in visual pattern recognition. Journal of Experimental Psychology: Human Perception & Performance, 4, 1-20.

LARSEN, A., & BUNDESEN, C. (1992). The efficiency of holistic template matching in the recognition of unconstrained handwritten digits. Psychological Research, 54, 187-193.

LARSEN, A., & BUNDESEN, C. (1993). An adaptive pandemonium of templates for visual pattern recognition. In C. Bundesen & A. Larsen (Eds.), Proceedings of the Sixth Conference of the European So-

ciety for Cognitive Psychology: Summaries (p. 5). Copenhagen: European Society for Cognitive Psychology.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1, 541-551.

Lindsay, P. H., & Norman, D. A. (1972). *Human information processing*. New York: Academic Press.

Loomis, J. M. (1990). A model of character recognition and legibility. *Journal of Experimental Psychology: Human Perception & Performance*, 16, 106-120.

Lowe, D. G. (1987). Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, 31, 355-395.

Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London: Series B*, 200, 269-294.

Michels, K. M., & Zusne, L. (1965). Metrics of visual form. *Psychological Bulletin*, 63, 74-86.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences*, 6, 414-417.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature*, 343, 263-266.

Reed, S. K. (1973). *Psychological processes in pattern recognition*. New York: Academic Press.

Rock, I., & DiVita, J. (1987). A case of viewer-centered object perception. *Cognitive Psychology*, 19, 280-293.

Rock, I., Wheeler, D., & Tudor, L. (1989). Can we imagine how objects look from other viewpoints? *Cognitive Psychology*, 21, 185-210.

Selfridge, O. G. (1959). Pandemonium: A paradigm for learning. In *Mechanisation of thought processes* (pp. 511-526). London: H.M.S.O.

Selfridge, O. G., & Neisser, U. (1960, August). Pattern recognition by machine. *Scientific American*, 203, 60-68.

Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.

Shepard, R. N., & Metzler, J. (1971, February 19). Mental rotation of three-dimensional objects. *Science*, 171, 701-703.

Stone, G. O. (1986). An analysis of the delta rule and the learning of statistical associations. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 444-459). Cambridge, MA: MIT Press.

Suen, C. Y. (Ed.) (1990). *Frontiers in handwriting recognition*. Montreal, Canada: Concordia University, Centre for Pattern Recognition & Machine Intelligence.

Sutherland, N. S. (1968). Outlines of a theory of visual pattern recognition in animals and man. *Proceedings of the Royal Society of London: Series B*, 171, 297-317.

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, 88, 135-170.

Tanaka, K. (1993, October 29). Neuronal mechanisms of object recognition. *Science*, 262, 685-688.

Tanaka, K., Saito, H.-A., Fukada, Y., & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology*, 66, 170-189.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, 21, 233-282.

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, 32, 193-254.

Ullman, S., & Basri, R. (1991). Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13, 992-1006.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549-586). Cambridge, MA: MIT Press.

Zusne, L. (1970). *Visual perception of form*. New York: Academic Press.

NOTES

1. Normalization with respect to orientation (see Larsen & Bundesen, 1992) was not invoked. Each digit was input in the (approximately upright) orientation in which it had originally been written on a letter envelope.

2. Use of r^7 (the 7th power of r) instead of r implies that variations in r have little effect unless $|r|$ is high. For example, an increment in r from 0 up to 0.5 makes a difference of 0.01 in r^7, but an increment in r from 0.5 up to 1 makes a difference of 0.99 in r^7.

3. The hyperbolic tangent of x is a sigmoid function defined as

$$\frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}.$$

It is a linear transform of the logistic function

$$\frac{1}{1 + \exp(-2x)},$$

and it squashes the continuum of real numbers into the open interval $(-1, 1)$.

4. In principle, any given pattern can be correctly classified by our recognition algorithm if that particular pattern is stored as a template with appropriate weights. If all patterns in the training set had been stored as templates, a recognition rate of 100% on the training set could have been obtained with weights of 1 and 0 on connections representing correct and incorrect classifications, respectively.

The recognition rate of 95.3% on the test set could be increased. It is a mean across five independent simulations. By running the five pandemonia simultaneously and letting them vote on a majority basis, the recognition rate rose to 95.9%.