Psychol Res (1992) 54: 187–193

Psychological Research
Psychologische Forschung

The efficiency of holistic template matching in the recognition of unconstrained handwritten digits

Axel Larsen[1] and Claus Bundesen[2]

[1] Institute of Clinical Psychology, Copenhagen University, Njalsgade 90, DK-2300 Copenhagen S, Denmark
[2] Psychological Laboratory, Copenhagen University, Njalsgade 90, DK-2300 Copenhagen S, Denmark

Received September 3, 1991/Accepted February 9, 1992

Summary. Psychological evidence suggests that simple visual patterns can be recognized by the use of internal representations as holistic templates, but the efficiency of holistic template matching in the recognition of real-life patterns, such as handwritten characters, has been doubted. To clarify this issue, we measured the efficiency of holistic template matching in machine recognition of totally unconstrained handwritten digits. Our learning and recognition algorithm was simple; no previous knowledge of 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 .88, respectively, when about 5, 10, or 50 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 the discounting of classifications based on weak evidence. Apparently, in high reliability recognition, holistic template matching can be used as a first operation by which recognition is achieved for most of the handwritten digits that are seen in real life.

Introduction

Psychological evidence suggests that visual pattern recognition can be achieved by the use of mental images as holistic templates – that is, by the superimposing of mental images upon visual impressions of stimulus patterns and the determination of degrees of match by a process akin to template matching in machine vision (see, e.g., Shepard & Cooper, 1982). Mental images can be regarded as representations in visual short-term memory. They can be transformed, and the transformation of mental images (short-term templates) appears to be one way of achieving recognition regardless of stimulus position, size (cf. Bundesen & Larsen, 1975; Jolicoeur & Besner, 1987; Larsen & Bundesen, 1978), and orientation (cf. Cooper & Shepard, 1973; Koriat & Norman, 1988; Shepard & Metzler, 1971; for combined transformations of size and orientation, see Bundesen, Larsen, & Farrell, 1981; Larsen, 1985).

There are other routes to visual-pattern recognition. In most cases, presumably, recognition is achieved by comparison of stimulus patterns with visual representations in long-term, rather than in short-term, memory. As is indicated below, the nature of the process by which a stimulus is compared with a representation in long-term memory is controversial.

On the one hand, theoretical simplicity favors the view that the process of comparing a stimulus with a representation in long-term memory is similar to the process of comparing a stimulus with a representation in short-term memory (i.e., a mental image). Furthermore, extant data from psychological experiments are consistent with the hypothesis that recognition of simple patterns such as letters and digits can be achieved by use of visual long-term representations as holistic templates. Two types of evidence are particularly suggestive.

1. A number of reaction-time studies of the effects of size (e.g., Cave & Kosslyn, 1989; Larsen & Bundesen, 1978) and orientation (e.g., Cooper, 1975; Jolicoeur, 1985) on the recognition process supports the notion that "visual pattern recognition is based on position-wise comparison of stimulus patterns with memory representations" (Larsen & Bundesen, 1978, p. 19). Holistic template matching is the most elementary way of making *position-wise* comparisons. (See Ullman, 1989, for a recent review of the empirical data and for an alignment approach to object recognition which elaborates the notion of position-wise comparison. For alternative approaches, see Biederman, 1987, and Gibson, 1969.)

2. Template-matching models of character recognition have had much success in accounting for variations in legibility across character sets, and some success in accounting for visual-confusion matrices (cf. Holbrook, 1975; Loomis, 1990). (See Gervais, Harvey, & Roberts, 1984, for a closely related model based on the analysis of

Offprint requests to: A. Larsen

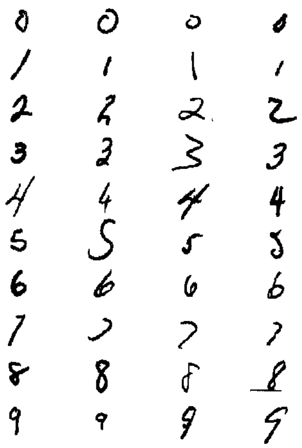

Fig. 1. Examples of input patterns.

spatial frequencies. For alternative models based on feature analysis, see Geyer & DeWald, 1973; Gibson, 1969, pp. 86–91; Keren & Baggen, 1981; for a more comprehensive review, see Townsend & Landon, 1983.)

On the other hand, the data available from psychological experiments are suggestive rather than conclusive, and the efficiency of holistic template matching has been questioned on computational grounds. In particular, serious doubts have been raised about the efficiency of holistic template matching in the recognition of unconstrained real-life patterns (see, e. g., Corcoran, 1971; Hummel & Biederman, in press; Humphreys & Bruce, 1989; Lindsay & Norman, 1972; Neisser, 1967; Reed, 1973). To provide evidence on this issue, we measured the efficiency of holistic template matching in machine recognition of totally unconstrained handwritten digits.

General method

Input patterns. The input data consisted of 4,000 handwritten characters, namely, 400 tokens of each of the 10 types of digit. The characters were taken from ZIP codes collected by the US Postal Services from dead-letter envelopes. The number of authors is not known, but can be assumed to approach the actual number of samples. The material had previously been used by Lam and Suen (1988) and was provided by Ching Suen in digitized and binarized form. Typical samples are shown in Figure 1.

Algorithm. In each experimental run through the input data, the input patterns (handwritten characters) were presented one by one to the recognition system. The order of presentation was random, and no pattern was presented more than once. Unless otherwise noted, each input pattern was processed as follows:

1. The *centroid* ("center of gravity"), C, 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*, s, 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).

3. The character was assigned an intrinsic *orientation*. Let $P_1, P_2 \ldots P_n$ be those pixels that belonged to the character and were located at the distance s from the centroid C, and let vector $\mathbf{V}$ be the sum of vectors $\overline{CP_i}$, $i = 1, 2 \ldots n$. If $\mathbf{V}$ was a nonzero vector, then the intrinsic orientation of the character was determined as the direction of $\mathbf{V}$. If $\mathbf{V}$ was the zero vector, then an integer i, $1 \leq i \leq n$, was chosen at random, and the intrinsic orientation of the character was determined as the direction of $\overline{CP_i}$.

4. The input pattern was *normalized* with respect to the position, the size, and the orientation 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, the y axis was aligned with the intrinsic orientation of the character, and the units of length along the x and y axes equaled the size of the character.

5. 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.

6. The smoothed input pattern was compared with every template in a library of stored templates, and the input was *classified* as a token of the same type (Digit Type 0, 1 . . . or 9) as the template that gave the highest degree of match. The degree of match between the input and a template was defined as the product-moment correlation between the two patterns.

The library of stored templates was empty when the first input pattern was processed, and the classification of the first input pattern was made at random.

7. If the classification of the input was correct (according to the specifications provided by Ching Suen), processing ended (i. e., no learning occurred). If the classification of the input pattern was in error, the smoothed input pattern was *copied* into the library of stored templates with a label that indicated the correct classification (Digit Type 0, 1 . . . or 9). After this step, the recognition system was ready for a new input pattern to be processed.

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

The input characters varied widely in size, covering from about $7 \times 12\ px$ up to $53 \times 53\ px$. After normalization, each character was represented in a format such that the greatest distance from the centroid pixel to any pixel that belonged to the character equaled $15\ px$ (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. As is described in the following sections, the standard deviation σ of the Gaussian filter was varied between experimental conditions. Let d be the greatest distance from the center of the Gaussian filter to a pixel at which the quantized filter coefficient was different from 0. For σ equal to 0.5, 1, 1.5, 2, 3, and $4\ px$, d equaled 1, 2, 3, 3, 4, and $4\ px$ respectively.

The smoothed input pattern had a value of 0 at any pixel located at a distance greater than $15 + d\ px$ from the centroid. Every comparison between the smoothed input pattern and a pattern in the library of stored templates was made by computation of the Pearson product-moment correlation coefficient between the two patterns across all pixels located at or within the distance of $15 + d\ px$ from the centroid.

Experiment 1

Method

In Experiment 1 each input character was presented in the (approximately upright) orientation in which it was written on the envelope, and the algorithm was simplified by the omission of normalization with respect to orientation (i. e., the omission of Step 3 and the modification of Step 4 so that the y axis was aligned with the upward direction on the envelope). The standard deviation σ of the Gaussian filter was varied between experimental conditions. In one condition – referred to as the condition in which $\sigma = 0$ – Step 5 of the algorithm was simply omitted. In the other six conditions, σ was 0.5, 1, 1.5, 2, 3, and $4\ px$ respectively, where $1\ px = 1/15$ unit of length (i. e., 1/15 of the greatest distance from the centroid of the normalized character to any pixel that belonged to the character).

Each of the seven experimental conditions comprised 500 independent runs (i. e., replications): 50 long runs (full replications) and 450 short runs (partial replications). For every new run, the recognition

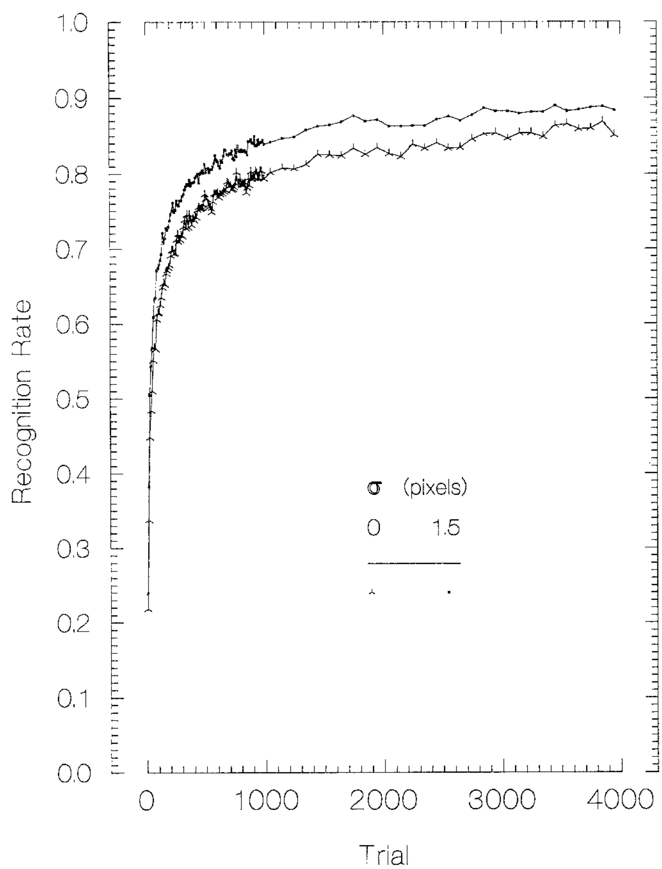

Fig. 2. Mean rate of recognition for successive blocks of trials in two conditions of Experiment 1. In one condition (upper curve), the Gaussian smoothing parameter σ was 1.5 *px*. In the other condition (lower curve), no smoothing was done.

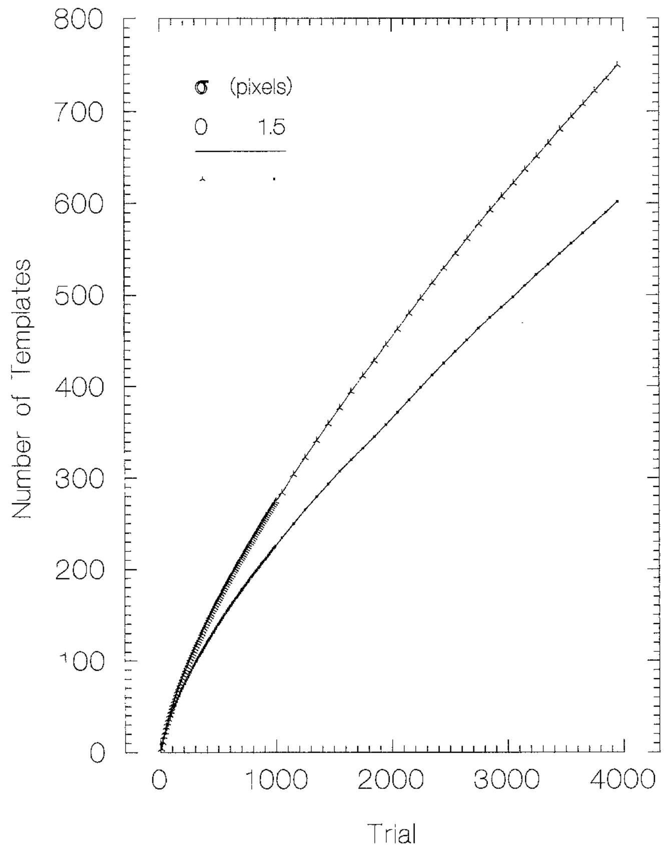

Fig. 3. Mean number of stored templates for successive blocks of trials in two conditions of Experiment 1. In one condition (lower curve), the Gaussian smoothing parameter σ was 1.5 *px*. In the other condition (upper curve), no smoothing was done.

system was initialized so that the library of stored templates was empty. During a long run, all of the 4,000 input patterns were presented to the system, one after another. The order in which the input patterns were presented over the 4,000 trials was random, and a new random order was prepared for each run. Short runs were identical with long runs, except that a short run ended when 1,000 input patterns had been presented (i. e., after 1,000 trials).

Results

Performance was better with Gaussian smoothing than without it, and the best performance was found with $\sigma = 1.5$ *px*. Figure 2 shows the rate of recognition (i. e., the relative frequency of correct classifications) for successive blocks of trials in the two conditions in which σ was 0 and 1.5 *px* respectively. The first 100 blocks of trials consisted of Trials 1–10, Trials 11–20 . . . and Trials 991–1,000 respectively. The succeeding 30 blocks of trials consisted of Trials 1,001–1,100, Trials 1,101–1,200 . . . and Trials 3,901–4,000. Thus, each data point in Figure 2 was based on 5,000 observations (classifications of input patterns).

By the nature of the algorithm, the recognition rate was at chance level (i. e., .10) on Trial 1 and low on the first block of trials. As can be seen in Figure 2, performance improved rapidly during the initial phases of training, but the rate of improvement decreased, and progress was slow later in training. Over the last five blocks of trials, the recognition rate averaged .86 for $\sigma = 0$ and .89 for $\sigma = 1.5$ *px*.

Figure 3 shows the mean number of templates in the library of stored templates for successive blocks of trials with $\sigma = 0$ and $\sigma = 1.5$ *px* respectively. On Trial 1 the number of templates was 0. On the last block of trials the mean number of templates was 749 for $\sigma = 0$ and 602 for $\sigma = 1.5$ *px*. The relationship between Figures 2 and 3 is clarified if it is noted that the number of templates in the library of stored templates on a given trial of a certain run equaled the total number of errors made on previous trials of the same run.

Figure 4 shows the mean rate of recognition compared with the mean number of templates per type of digit for successive blocks of trials in each of the seven experimental conditions. Results for $\sigma \leq 1.5$ *px* are shown in the left panel, and results for $\sigma \geq 2$ *px* in the right panel. With a mean of only 5 templates per type of digit, recognition rates were approximately .61, .63, .66, .69, .69, .67, and .67 for σ = 0, 0.5, 1, 1.5, 2, 3, and 4 *px* respectively. With a mean of 10 templates per type of digit, the recognition rates had increased to .70, .73, .75, .77, .77, .75, and .75 respectively. With a mean of 50 templates per type of digit, the corresponding recognition rates were .84, .85, .88, .88, .87, .86, and .86.

At the expense of getting omission errors (*rejections*), the reliability of the recognition responses (i. e., the relative

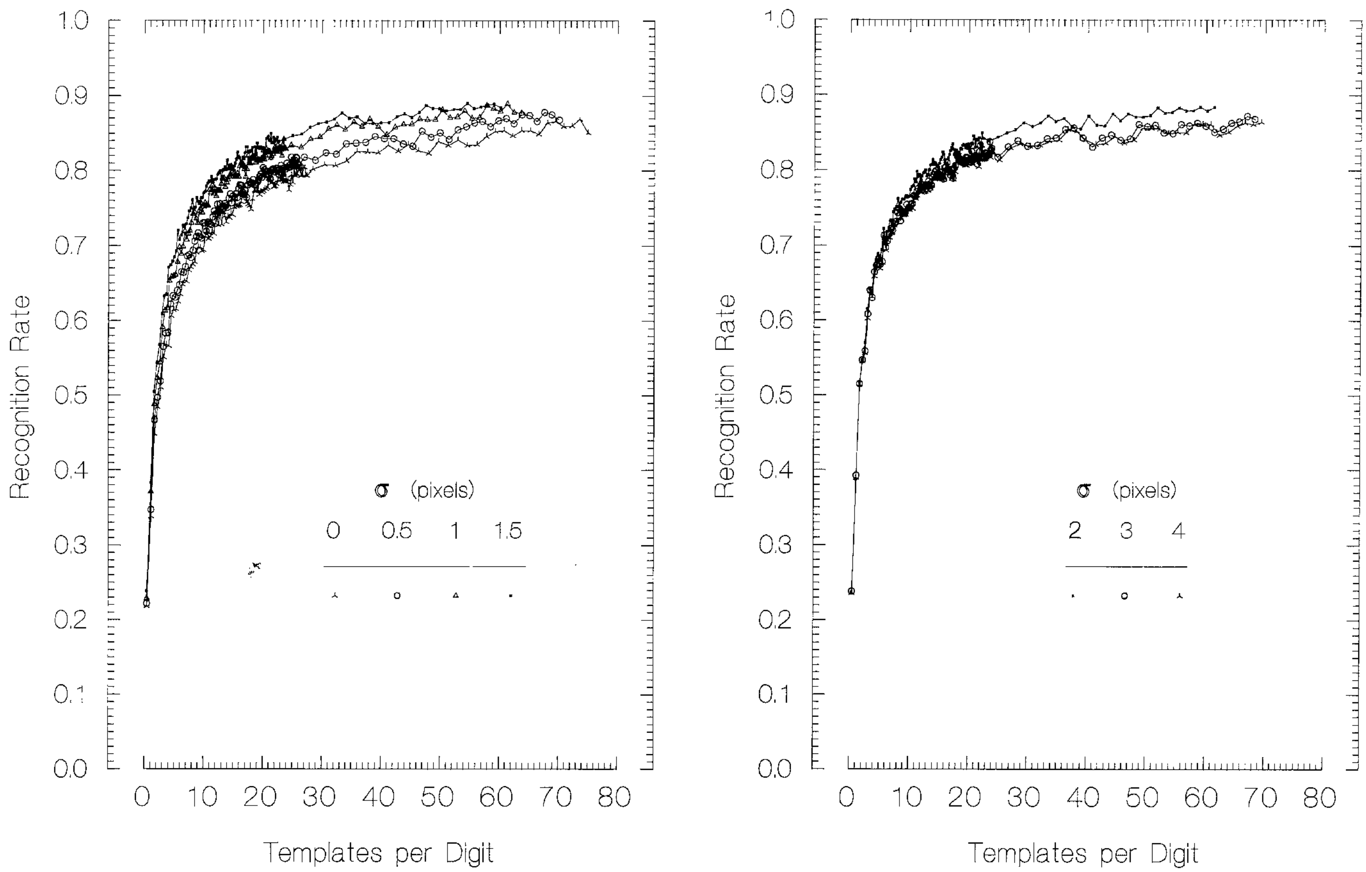

Fig. 4. Mean rate of recognition compared with mean number of templates per type of digit for successive blocks of trials in the seven conditions of Experiment 1. The Gaussian smoothing parameter σ was varied between conditions. Left panel: Results for σ = 0, 0.5, 1, and 1.5 *px* respectively; performance improves as σ increases. Right panel: Results for σ = 2, 3, and 4 *px* respectively; performance degrades as σ increases.

frequency of correct responses among all responses) could be increased by the discounting of classifications based on weak evidence. Suppose the highest degree of match on a given trial (i.e., with a given input pattern) was obtained with a template of type i (i = 0, 1 . . . or 9), and let r_1 be this degree of match. Let r_2 be the highest degree of match, on the same trial, found with templates that were not of type i. Rejection criteria based on the value of $r_1^n - r_2^n$ were tested for various values of n. (In general, $0 < r_2 < r_1 \leq 1$. For $n = 1$, responses were omitted when $r_1 - r_2 < c$, where c was a constant such that $0 < c < 1$. In the limit as $n \rightarrow \infty$, the effect of omitting responses when $r_1^n - r_2^n < c^n$, where $0 < c < 1$, was identical with the effect of omitting responses when $r_1 < c$, regardless of r_2. Thus, omitting responses when, for instance, $r_1^7 - r_2^7$ was below a given threshold was a way of compromising between omitting responses when $r_1 - r_2$ was below a given threshold and omitting responses when r_1 was below a given threshold.)

By analysis of all classifications made in the last five blocks of trials with σ = 1.5 *px* (a total of 25,000 classifications) and by use of a rejection criterion based on the value of $r_1^7 - r_2^7$, a reliability of .95 was attained with a rejection rate of .24. Reliabilities of .96 and .97 were attained with rejection rates of .33 and .46 respectively.

Experiment 2

Method

Experiment 2 tested the effect of presenting the input patterns to the recognition system in orientations that varied at random. The method was the same as in Experiment 1, with the exceptions noted below.

To take account of the fact that orientation information is critical in discriminating between digits of Types 6 and 9, the 400 digits of Type 9 were removed from the input data. The remaining 3,600 input patterns were used in three experimental conditions. In each condition the standard deviation σ of the Gaussian filter was set to a value of 1.5 *px*.

In Condition 1 (upright orientation), each input character was presented in the (approximately upright) orientation in which it was written on the envelope. This condition was identical with the condition of σ = 1.5 *px* in Experiment 1, except that no digits of Type 9 were used. In Conditions 2 and 3 the input patterns were presented to the recognition system in orientations that varied at random from trial to trial. Effectively, the input characters were treated as if they had been written in randomly determined orientations on the envelopes. In Condition 2 (random orientation without normalization), recognition was achieved by the simplified algorithm used in Experiment 1 – that is, there was no normalization with respect to orientation. In Condition 3 (random orientation with normalization), the full algorithm was used, including normalization with respect to orientation.

Results

Figure 5 shows the mean rate of recognition compared with the mean number of templates per type of digit for succes-

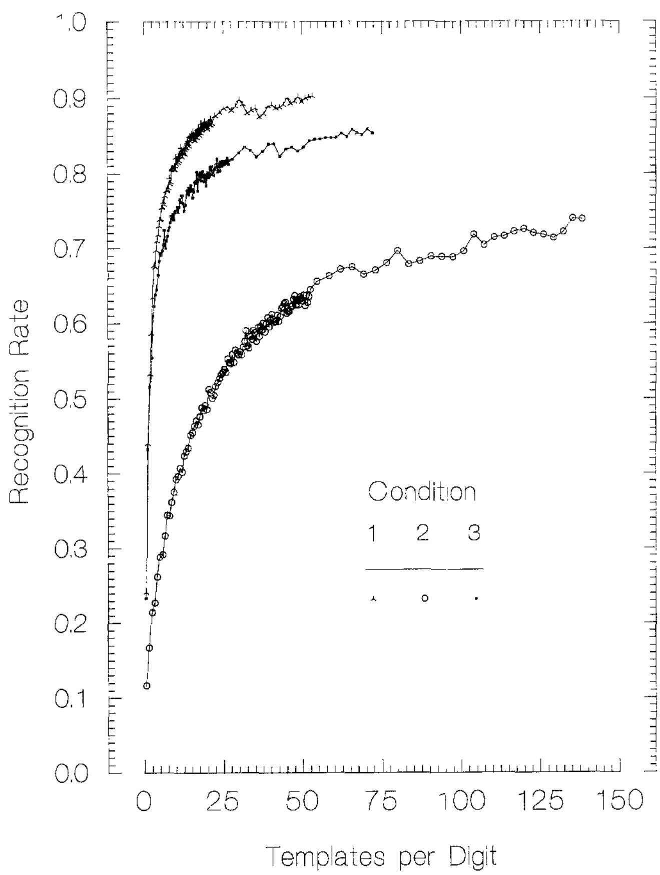

Fig. 5. Mean rate of recognition compared with mean number of templates per type of digit for successive blocks of trials in the three conditions of Experiment 2. Results for Conditions 1 (upright orientation), 2 (random orientation without normalization), and 3 (random orientation with normalization) are shown by the top, bottom, and middle curves respectively.

sive blocks of trials in each of the three experimental conditions. As one would expect, performance in Condition 1 (baseline condition with 9 types of digit in upright orientation) was slightly better than performance in Experiment 1 (which used 10 types of digit in upright orientation). With a mean of only 5 templates per type of digit, the recognition rate was .74. With a mean of 10 templates per type of digit, the recognition rate had increased to .82. Near the end of training, with a mean of 50 templates per type of digit, the recognition rate had reached a level of .90.

In relation to the baseline established in Condition 1, performance deteriorated dramatically when the input patterns were presented to the recognition system in orientations that varied at random and when there was no normalization with respect to orientation. In this case (Condition 2), recognition rates with about 5, 10, and 50 templates per type of digit were only .29, .39, and .63 respectively. Over the last two blocks of trials, the mean number of templates per type of digit was 137, but the recognition rate was no higher than that attained with only 5 templates per type of digit in Condition 1 (viz., .74).

When the full algorithm (including normalization with respect to orientation) was applied to input patterns presented in random orientations (Condition 3), performance was inferior to that in Condition 1, but greatly superior to that in Condition 2. With means of 5, 10, and 50 templates per type of digit, recognition rates were .69, .75, and .83. Over the last two blocks of trials, the mean number of templates per type of digit was 71 and the recognition rate was .86.

As in Experiment 1, the reliability of the recognition responses could be increased if classifications based on weak evidence were discounted. By analysis of all classifications made in the last five blocks of trials in Condition 3 (a total of 25,000 classifications) and by use of a rejection criterion based on the value of $r_1^7 - r_2^7$, reliabilities of .94 and .96 were attained with rejection rates of .30 and .43 respectively.

General discussion

The efficiency of holistic template matching

The purpose of Experiments 1 and 2 was to obtain solid evidence as to the efficiency of holistic template matching in the recognition of real-life patterns such as totally unconstrained handwritten digits. The learning and recognition algorithm used in the experiments was very simple, and no previous knowledge concerning the nature of the stimulus material (handwritten digits) was built into the algorithm. The results can be summarized as follows.

In Experiment 1 the input characters were presented in the (approximately upright) orientation in which they were written on the envelopes, and the characters were normalized only in position and size. With slight Gaussian smoothing ($\sigma = 1/10$ of the greatest distance from the centroid of the normalized character to any pixel of the character) and with means of 5, 10, and 50 templates per type of digit, recognition rates were .69, .77, and .88 respectively. Near the end of training, the recognition rate was fairly high (.89), but the rate of improvement with increase in the number of templates was very low. The reliability of recognition responses could be increased by classifications based on weak evidence being discounted; reliabilities of .95 and .97 were attained with rejection rates of .24 and .46 respectively.

In Experiment 2 performance deteriorated severely when the input characters were presented in orientations that varied at random and when there was no normalization with respect to orientation. When the characters were normalized in intrinsic orientation, the recognition rates approached those attained in Experiment 1. Near the end of the training, the recognition rate was .86, but the rate of improvement was low. Reliabilities of .94 and .96 were attained with rejection rates of .30 and .43 respectively.

Experiments 1 and 2 provide well-founded lower bounds on the efficiency of holistic template matching in recognition of totally unconstrained handwritten digits. Implications for upper bounds are less certain. Our algorithm is attractive for its simplicity and generality, but it is possible that performance could be improved at the expense of making the algorithm less simple and less general. For example, performance might possibly be improved by the use of a more complex and intelligent learning algorithm and by a smart smoothing routine, designed specifically for handwritten digits (see, e.g., Brown, Fay, &

Walker, 1988). Prima facie, however, our experiments suggest that, by holistic template matching with a reasonably small number of templates, a high level of reliability in the recognition of totally unconstrained handwritten digits can be attained only with a fairly high rate of rejection.

Our algorithm is less efficient than the most successful of the recognition algorithms designed specifically for handwritten digits. For example, by the use of a hybrid algorithm based on both structural classification and relaxation matching, Lam and Suen (1988) obtained reliabilities of nearly .97 and .98 with rejection rates of about .01 and .05 respectively for digits presented in upright orientation. (For other recent machine algorithms, see Baptista & Kulkarni, 1988; Huang & Chuang, 1986; LeCun et al., 1989; Shridhar & Badreldin, 1986; and Suen, 1990.)

Our algorithm is also less efficient than normal literate human beings. In an experiment described in the Appendix, we presented subjects with a random sample of the digits that were used as input patterns in Experiments 1 and 2. In Condition 1 each digit was presented in the (approximately upright) orientation in which it had originally been written on an envelope. In Condition 2 the digits were presented in orientations that varied at random from trial to trial, but digits of Type 9 were excluded. The mean rate of recognition was .97 in Condition 1 and .96 in Condition 2.

Template matching in human recognition

Psychological evidence suggests that stimulus patterns can be recognized by the use of visual short-term representations in the form of mental images as holistic templates. By analogy, it is tempting to conjecture that the recognition of simple patterns, such as letters and digits, can be achieved by the use of visual long-term representations as holistic templates. Data from psychological experiments are consistent with this conjecture, but serious doubts have been raised about the efficiency of holistic template matching in the recognition of unconstrained real-life patterns.

Experiments 1 and 2 provide evidence that most handwritten digits seen in real life can be recognized by holistic template matching with a reasonably small number of templates per type of digit and with simple routines for normalization with respect to position, size, and orientation. Moreover, high levels of reliability can be attained by the discounting of classifications based on weak evidence. Thus, in high reliability recognition, holistic template matching can serve as a first-phase operation in which recognition is achieved for the bulk of the input patterns.

Considering the temporal properties of the visual system in humans, holistic template matching seems fit for a first attempt at recognition. Much evidence suggests that the visual impression of a stimulus develops over time, so that global (low-spatial-frequency) information becomes available before local (high-spatial-frequency) information. When a stimulus is presented very briefly, only its low-frequency components appear to be represented in the visual impression (cf. Petersik, 1978). As the duration of exposure is increased, the impression gains in clarity (i.e., high-frequency components are added), so that gross figure-ground differentiation and vague contour resolve into a well-defined figure with sharp contour (cf. Eriksen & Schultz, 1978; Flavell & Draguns, 1957; Lupker, 1979; Wever, 1927). Reaction-time studies corroborate this interpretation. For example, Breitmeyer (1975) found that simple reaction time to sinusoidal gratings increased by 40–80 ms over spatial frequencies ranging from 0.5 to 11 cycles/degree, even when the gratings were matched for subjective contrast. Similarly, holding eccentricity constant, Navon and Norman (1983) found that choice-reaction time to letter stimuli increased as the size of the letters decreased over a wide range of visual angles and retinal positions (see Kimchi, in press, for a review of related results).

Being based on the global shape of an object, holistic template matching is relatively insensitive to low-pass filtering (blurring) of the stimulus. Hence, given that global (low-spatial-frequency) information is available before local (high-spatial-frequency) information, holistic template matching seems fit for a first attempt at recognition. There is evidence from the analysis of asymmetries in alphabetic-confusion matrices that suggests a two-phase account of character recognition: a first phase based on global information and a later phase based on local information (see Dawson & Harshman, 1986). Against this background, we offer the following speculations on the role of template matching in human recognition of handwritten characters.

Consider a subject who is ready to process a handwritten digit (see, e.g., Bundesen, 1990, for an account of perceptual readiness). In normal circumstances, we propose, recognition by holistic template matching is attempted first (that is, the stimulus is compared as a whole against a number of visual long-term representations of digits). If strong evidence is obtained in favor of a particular classification, the digit is recognized as a whole with no attention to individual strokes and with no conscious effort. Introspectively, recognition appears to be immediate. If only weak evidence is obtained in favor of any particular classification, the level of perceptual analysis is shifted, and recognition of the components of the stimulus (e.g., individual strokes) is attempted. By this procedure, recognition of the digit may be mediated by the recognition of its components. However, for most of the handwritten digits that are seen in real life, the first attempt at recognition should succeed. In other words, holistic template matching should suffice.

Acknowledgements. This work was supported by grants from the Danish Ministry of Education and Research and the International Human Frontier Science Program Organization. Thanks are due to Michael Dawson and an anonymous reviewer for constructive comments on the draft.

References

Baptista, G., & Kulkarni, K. M. (1988). A high accuracy algorithm for recognition of handwritten numerals. *Pattern Recognition, 21*, 287–291.

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

Breitmeyer, B. G. (1975). Simple reaction time as a measure of the temporal response properties of transient and sustained channels. *Vision Research, 15*, 1411–1412.

Brown, R. M., Fay, T. H., & Walker, C. L. (1988). Handprinted symbol recognition system. *Pattern Recognition, 21*, 91–118.
Bundesen, C. (1990). A theory of visual attention. *Psychological Review, 97*, 523–547.
Bundesen, C., & Larsen, A. (1975). Visual transformation of size. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 214–220.
Bundesen, C., Larsen, A., & Farrell, J. E. (1981). Mental transformations of size and orientation. In J. Long & A. 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., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual information processing* (pp. 75–176). New York: Academic Press.
Corcoran, D. W. J. (1971). *Pattern recognition*. Harmondsworth, England: Penguin.
Dawson, M. R. W., & Harshman, R. A. (1986). The multidimensional analysis of asymmetries in alphabetic confusion matrices: Evidence for global-to-local and local-to-global processing. *Perception & Psychophysics, 40*, 370–383.
Eriksen, C. W., & Schultz, D. W. (1978). Temporal factors in visual information processing: A tutorial review. In J. Requin (Ed.), *Attention and performance: VII* (pp. 3–23). Hillsdale, NJ: Erlbaum
Flavell, J. H., & Draguns, J. (1957). A microgenetic approach to perception and thought. *Psychological Bulletin, 54*, 197–217.
Gervais, M. J., Harvey, L. O., Jr., & Roberts, J. O. (1984). Identification confusions among letters of the alphabet. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 655–666.
Geyer, L. H., & DeWald, C. G. (1973). Feature lists and confusion matrices. *Perception & Psychophysics, 14*, 471–482.
Gibson, E. J. (1969). *Principles of perceptual learning and development.* New York: Appleton-Century-Crofts.
Holbrook, M. B. (1975). A comparison of methods for measuring the interletter similarity between capital letters. *Perception & Psychophysics, 17*, 532–536.
Huang, J. S., & Chuang, K. (1986). Heuristic approach to handwritten numeral recognition. *Pattern Recognition, 19*, 15–19.
Hummel, J. E., & Biederman, I. (in press). Dynamic binding in a neural network for shape recognition. *Psychological Review.*
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., & 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 and Performance, 13*, 478–487.
Keren, G., & Baggen, S. (1981). Recognition models of alphanumeric characters. *Perception & Psychophysics, 29*, 234–246.
Kimchi, R. (in press). The primacy of wholistic processing and the global/local paradigm: A critical review. *Psychological Bulletin.*
Koriat, A., & Norman, J. (1988). Frames and images: Sequential effects in mental rotation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 93–111.
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 and Performance, 4*, 1–20.
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 and Performance, 16*, 106–120.
Lupker, S. J. (1979). On the nature of perceptual information during letter perception. *Perception & Psychophysics, 25*, 303–312.
Navon, D., & Norman, J. (1983). Does global precedence really depend on visual angle? *Journal of Experimental Psychology: Human Perception and Performance, 9*, 955–965.
Neisser, U. (1967). *Cognitive psychology.* New York: Appleton-Century-Crofts.
Petersik, J. T. (1978). Possible role of transient and sustained visual mechanisms in the determination of similarity judgments. *Perceptual and Motor Skills, 47*, 683–698.
Reed, S. K. (1973). *Psychological processes in pattern recognition.* New York: Academic Press.
Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations.* Cambridge, MA: MIT Press.
Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701–703.
Shridhar, M., & Badreldin, A. (1986). Recognition of isolated and simply connected handwritten numerals. *Pattern Recognition, 19*, 1–12.
Suen, C. Y. (Ed.). (1990). *Frontiers in handwriting recognition.* Montreal: Concordia University, Centre for Pattern Recognition & Machine Intelligence.
Townsend, J. T., & Landon, D. E. (1983). Mathematical models of recognition and confusion in psychology. *Mathematical Social Sciences, 4*, 25–71.
Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition, 32*, 193–254.
Wever, E. G. (1927). Figure and ground in the visual perception of form. *American Journal of Psychology, 38*, 194–226.

Appendix

Measuring the efficiency of human recognition of unconstrained handwritten digits

A sample of the digits used as input patterns in Experiments 1 and 2 was presented to 6 subjects (4 students and 2 members of the technical staff at Copenhagen University). The sample consisted of 200 randomly selected tokens of each of the 10 types of digit. The digits were displayed one by one on a computer screen, and the subject was instructed to name the type of each digit as accurately as possible. Each digit was displayed until the subject gave his response.

The subjects participated individually in two experimental conditions spread over four sessions. In Condition 1 (upright orientation), the 2,000 digits were all presented in the orientations in which they were written on the envelopes. In Condition 2 (random orientation), digits of Type 9 were excluded, and the remaining 1,800 digits were presented in random orientations, varying unpredictably from trial to trial. Three subjects participated in 1,000 trials in Condition 1 in Session 1; 900 trials in Condition 2 in Session 2; 900 trials in Condition 2 in Session 3; and 1,000 trials in Condition 1 in Session 4. The other three subjects participated in 900 trials in Condition 2 in Session 1; 1,000 trials in Condition 1 in Session 2; 1,000 trials in Condition 1 in Session 3; and 900 trials in Condition 2 in Session 4.

The results showed considerable variation among subjects. In Condition 1 individual recognition rates for Subjects 1–6 were .98, .99, .90, .99, .98, and .99 respectively. In Condition 2 the corresponding recognition rates were .97, .98, .92, .96, .95, and .98 respectively. Averaged across subjects, the recognition rate was .97 in Condition 1 and .96 in Condition 2.