

ELASTIC MATCHING AND PATTERN RECOGNITION IN NEURAL NETWORKS

Elie BIENENSTOCK and René DOURSAT

E. S. P. C. I.
Laboratoire d'Electronique
10, rue Vauquelin
75005 PARIS.
FRANCE

ABSTRACT

Some problems in pattern recognition may be phrased as *graph-matching* tasks. Indeed, relational descriptions provide a powerful theoretical tool for solving recognition problems which require invariances with respect to various transformation of the image. We propose a notion of graph-matching, actually a form of *elastic matching*, based on the definition of a *cost-function* H , the minima of which provide the solution to the matching problem. In the neural network perspective, this leads to considering systems where the state is a *connectivity state* $J = (J_{ij})$ rather than a vector of neuron activities $x = (x_i)$, and computation is a J -dynamics instead of an x -dynamics. This *dynamical connectionist* approach has first been advocated as a theory of brain function in 1981 by Christoph von der Malsburg [14]. In this approach, matching one graph with another consists in finding a connectivity state or map J which satisfies at best many local requirements. This map is precisely the one which minimizes the cost H . We illustrate the approach by an application to character recognition. The problem is to locate and identify characters on a page, in spite of shape variation and in the absence of any positional information. The labelled elastic matching approach is shown to provide a reliable solution to this problem.

1. INTRODUCTION

In spite of the considerable progress of the neurosciences in the last decades, it is probably fair to say that some fundamental issues concerning the workings of the brain remain unsettled. One such issue is that of the neural code. In the peripheral nervous system, that is, in its sensory and motor ends, there is overwhelming evidence that stimulus intensity and motor command are encoded in the activity rates of specific neurons. In association areas however, the nature of the code is much less obvious. On the one hand, many researchers hold that there too, information is encoded in average activity rates of neurons. In this view, individual spikes are totally irrelevant, just like single electrons in an electric wire. It should be noted that electrophysiological experiments are by and large carried out in this conceptual framework, focusing on average neuronal firing rates. Thus for instance, the notion of a receptive field is

strongly linked to the technique of single cell recording, whereby the average activity of a single cell is considered in isolation from the activity of its fellow cells in cortex.

On the other hand, a few authors [1,14], see also [21], have proposed that the accurate timing of single spikes may carry essential information. These authors believe that synchrony of spikes and more generally precise temporal relationships between spikes in different fibres may be crucial elements in the functioning of the cortex. In this view, single cell activity, particularly in association areas, loses a good deal of its relevance to brain function, as information is encoded mainly in temporal relationships between the spikes carried by different fibres. Note that, by the same token, the very notion of a receptive field becomes somewhat questionable. More generally, the technique of single cell recording appears as a rather narrow bottleneck for information to pass from the nervous system to the observer. However, avoiding this bottleneck is not easy: experiments designed to study directly the role of temporal patterns in multi-neuron activity face serious methodological problems.

Another central issue is the role of synapses. There is of course good agreement that chemical synapses are the substrate of memory: they are believed to be the main sites undergoing modification during learning. Synaptic plasticity underlying long term memory is by its very nature a long lasting, essentially irreversible process [10]. Yet synapses could, in addition, be the site of a much faster type of plasticity. Fast synaptic plasticity, which remains speculative at the present time, would consist of transient, reversible changes, which would significantly affect the efficacy of a synapse for periods of time ranging from milliseconds to seconds. In this view, synapses would play a much more active role than the one traditionally assigned to them: the fast varying synaptic weights would provide a medium for storing and manipulating information on the short term. Note again that the existence of fast synaptic plasticity as well as its possible role in brain function are not easily amenable to direct experimental investigation.

However, if we switch to an abstract mode of reasoning about the brain, and consider these issues on the *computational-representational* level, it would seem that using temporal relationships and fast modifiable synapses could be extremely valuable in nearly all cognitive tasks performed by cortex. Indeed, what our brains seem to be mostly busy at is the manipulation of *relational information*. For instance, when analyzing a visual scene or a speech signal, we focus on various types of spatial, temporal, or spatio-temporal *relationships* between visual primitives, objects, parts of objects, acoustic-phonetic features etc. Spike synchronicity as well as dynamical synapses are by essence relational variables. They therefore provide a natural medium for representing and processing relational information. They also provide a straightforward solution to the so-called binding problem, regarded by many authors as the main theoretical difficulty with conventional theories of brain function.

This idea has been put forward by Christoph von der Malsburg in his Correlation Theory of Brain Function [14], and further developed in an approach termed *dynamical connectionism*, or *graph coding* [4,15,16]. The main tenet in this theory is the use of dynamical links. Formally, one is led to describe the state of the brain as a modifiable connectivity graph – possibly a labelled graph – rather than a neuronal activity state. If one uses the statistical mechanics analogy between brain states and thermodynamical systems, the relevant thermodynamical variables in the theory are the exchange interactions J_{ij} rather than the spins x_i . Accordingly, the dynamics of the system will depend on the definition of an appropriate energy function (or cost-function) $H(J)$ rather than $H(x)$.

Connectionist models have been applied in recent years to a variety of "computational paradigms", such as error correction, interpolation, extrapolation,

approximation, classification [11,19,20]... In order for these to make sense, the state space has to be equipped with a metric. For instance, error correction in a Hopfield type model [11] means that initial states whose *Hamming distance* to a stored memory is small are attracted to that memory state. Similarly, in a layered feed-forward network, the metric underlying the computation, e.g. classification, is most often the mean-square distance [20]. In this paper, we shall define a notion of distance based on the operation of *elastic matching* between two images. This operation can be performed by minimizing a cost-function (sometimes termed energy function) defined over the space of all links between the two images, i.e., a function of the form $H(\mathbf{J})$ rather than $H(\mathbf{s})$. The distance we shall use is the minimum value of this energy function, and we shall call it the "elastic distance" between the two images. Computing this distance will be equivalent to minimizing the energy function $H(\mathbf{J})$. We shall argue that using this notion of distance, based on comparing relationships between elements in one image to the relationships between corresponding elements in the other image, is a very natural way to obtain generalization with respect to moderate distortions.

The main goal of the present paper is to show that this notion of elastic distance is indeed well suited to shape recognition. From the neurobiological standpoint, this paper is meant to provide support to the dynamical connectionist approach to brain function through an abstract argument on the "computational-representational" level. In short, this argument runs as follows: it might be easier to understand how humans solve hard tasks of shape recognition if one is willing to consider the possibility that the brain manipulates links rather than lists. For a more complete discussion, both on the abstract computational level and regarding the "implementation" in the nervous system, the reader is referred to previous papers [2,3,4,14,15,16].

In the context of machine vision on the other hand, elastic matching for shape recognition has been suggested and studied by several authors, see in particular [5]. However, it is often considered impractical in real applications. The results presented in this and subsequent papers provide further support to the elastic matching approach to shape recognition, from the practical computational point of view. Indeed, using an energy function which lends itself to fast minimization, we shall demonstrate that satisfactory results may be obtained with a reasonable amount of computation in a task involving the recognition of distorted shapes.

Note that an alternative neural network approach to graph-matching is proposed in this volume by G. Dreyfus and A. Zippelius, using a representation scheme where the presence of edges is encoded in the activity levels of neurons (the usual connectionist coding scheme). The model of Dreyfus and Zippelius addresses the general problem of isomorphism of abstract graphs.

The structure of the present paper is as follows. In Section 2, we formulate the elastic matching problem as a constrained optimization task, and define the "elastic distance" between two images. In Section 3, we present a fast algorithm for solving this optimization problem. Section 4 is a brief description of an application to character recognition.

2. ELASTIC MATCHING

In this section, we give a brief outline of the *dynamical connectionist* approach to the general problem of graph matching, and we describe in more detail the *elastic matching* formulation that will be used in the present paper. Let G and G' be two abstract graphs. The general problem of matching G with G' consists in finding a (one-to-one) map f , from the set of nodes of G to the set of nodes of G' , which preserves the graph structure: any two nodes linked by an arc in G should be mapped to two nodes linked by an arc in G' (and vice-versa). When dealing with labelled graphs, one may

impose further label-related constraints, which ensure that the mapping f also preserve the labels.

The labelled graphs we shall deal with in the present paper are subgraphs of the 4-nearest-neighbor square lattice, obtained as follows. Consider a binary-valued image X defined on a square array S : $X(s) = 0$ or 1 , for s in S . We shall also say that the value of pixel s is white ($X(s) = 0$) or black ($X(s) = 1$). A labelled graph G is obtained from X by first selecting from S all nodes which are either black or within distance 1 of a black node, and by then connecting, within the set of selected nodes, every node s to those of its four nearest neighbors which are also in that set (see Fig.1). This choice is somewhat arbitrary: one may for instance use a richer connectivity, e.g. where each node s is connected to its eight nearest neighbors; one may also include more "white nodes" than the first neighbors of black nodes. It is necessary however, for reasons which will become clear later, to include both black, or figure nodes, and white, or background nodes. In the following, we shall identify the nodes of graph G to lattice points in S , and use the same notation s, t, \dots for the two. In particular, it should be understood that in the expression $\|s - t\|$, s and t are viewed as lattice points, and the distance is the euclidean distance in the plane.

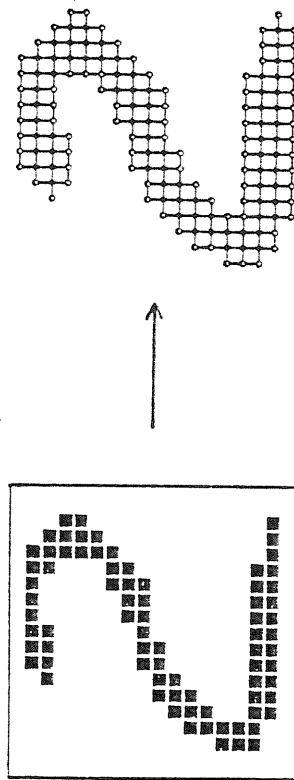


Fig.1. An example of graph-representation: from the image to the labelled graph.

Consider now two labelled graphs G and G' , obtained from two images X and X' on the lattice S as explained above. If the two images are identical, or if one is a mere shift of the other, there exists a mapping f from G to G' which preserves perfectly both the graph structure (it is an isomorphism of graphs), and the labelling. In this case, the graph-matching problem admits a perfect solution. However, in general, one might wish to declare two images X and X' "similar" even when there is no perfect match between them. In particular, if X' is a mildly distorted version of X , it would seem reasonable to call them indeed similar. It is precisely this notion of similarity that we wish to formalize, and actually quantify, with the help of the best possible map f from G to G' , where "best map" is given a precise mathematical definition.

There are several ways to define the notion of a best possible map. We now outline one such possibility. First, we define the *state space* as the set of all correspondences, or many-to-many mappings, between the nodes of G and the nodes of G' : A correspondence may be viewed as a collection of "active links": such a state $\mathbf{J} = (J_{ss'})_{s \in G, s' \in G'}$ is defined by specifying, for each s in G and s' in G' , whether link $J_{ss'}$ is active ($J_{ss'} = 1$) or inactive ($J_{ss'} = 0$). One then defines a function $H(\mathbf{J})$ over this state space, with the property that $H(\mathbf{J})$ is small when \mathbf{J} is a good match. Specifically, H is a sum of many small contributions, each of which is the embodiment of a constraint, of one of three types: (i) link preservation; (ii) label preservation and (iii) a constraint which for each node s prohibits two many links $J_{ss'}$ originating from s to be simultaneously active. All of these are *weak* constraints, which can always be violated, but each violation incurs a penalty, the sum of which adds up to the cost function $H(\mathbf{J})$. If the three terms of H are given appropriate forms and appropriate relative weights

those pairs of labelled graphs (G,G') for which there exists a J of low H are precisely those which are created from pairs of qualitatively similar images (X,X'). This approach to graph matching has been studied in previous papers, and the reader is referred to references [2,3,4] for further details.

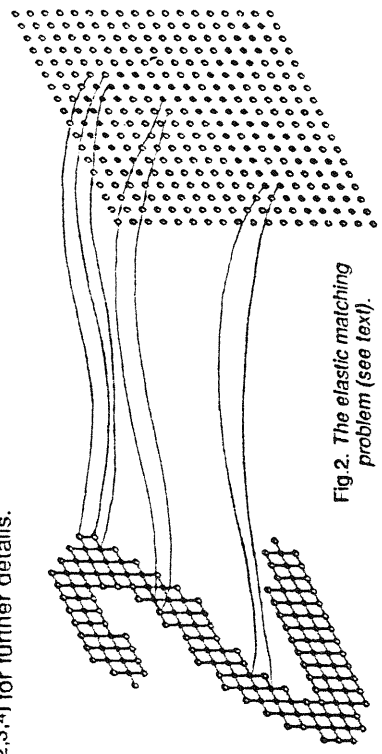


Fig.2. The elastic matching problem (see text).

In this paper, we shall use a slightly different notion of "best map", one which is tailor-made for problems of shape identification, where recognition should be robust against mild "rubber-sheet" deformations. First, we consider only proper mappings f from G to G' : each s in G maps into a single $s' = f(s)$ in G' . The state space is thus the set $\{f\}$ of all maps from G to G' , a subset of the set $\{J\}$ of correspondences between G and G' . Another difference is that we exploit the 2D-structure of our graphs to view f as a spatial transformation from a 2D-lattice (actually a subset thereof) to another 2D-lattice, rather than from an abstract graph to another graph (see Fig.2). We define a cost-function $H(f)$ along the same principles as outlined above, i.e., H is the sum of local contributions, each of which embodies a particular weak constraint. However, rather than using only the abstract graph structure in G and G' , we also use, in the definition of H , the euclidean distance in the 2D-lattice S . Specifically:

$$H(f) = H_1(f) + \beta H_2(f), \text{ with}$$

$$H_1(f) = \sum_{\langle s,t \rangle \in G} \|f(s) - f(t)\| - (s-t) \|^2 \text{ and}$$

$$H_2(f) = \sum_{s \in S} (1 - \delta(X(s), X'(f(s)))).$$

The first term, $H_1(f)$, measures the extent to which f differs from a rigid shift, in the following sense. For each pair of nodes s and t which are connected in G , i.e., such that the euclidean distance $\|s - t\|$ is exactly 1, the vector $s - t$ in the lattice S is compared to the transformed vector $f(s) - f(t)$. The local contribution to the cost, that is, the penalty incurred by f when modifying the position of s and t relative to each other, is the square of the norm of the difference between these two vectors, denoted d in Fig.3. This quadratic distance cost is reminiscent of an elastic energy term, and one may indeed view $H_1(f)$ as an approximation of the energy associated to a deformation of a rubber-like sheet.

Another, and perhaps more relevant analogy in this context, is a neurobiological one. Note that keeping $H_1(f)$ small means enforcing the principle that neighbors should stay neighbors through the transformation f , in much the same way as fibres originating from nearby locations in the retina project to nearby locations in the various brain structures. This principle results in an organized, topology preserving map, termed a retinotopic, and more generally topographic map, in the neurobiological literature. The fact that projections from one brain structure to another are very often organized in

topographic maps is a quite striking phenomenon. In the case of sensory systems, it clearly subserves the function of providing the brain with a central representation of the external space. During the course of ontogenetic developments of the nervous system, Nature is almost certainly using several different mechanisms to enforce the principle of neighbor-to-neighbor connectivity, and a variety of mathematical models have been proposed in this context, see e.g. [13,17,22]. Our cost-function $H_1(f)$ is related to these models, although they need not use the particular quadratic form used here.

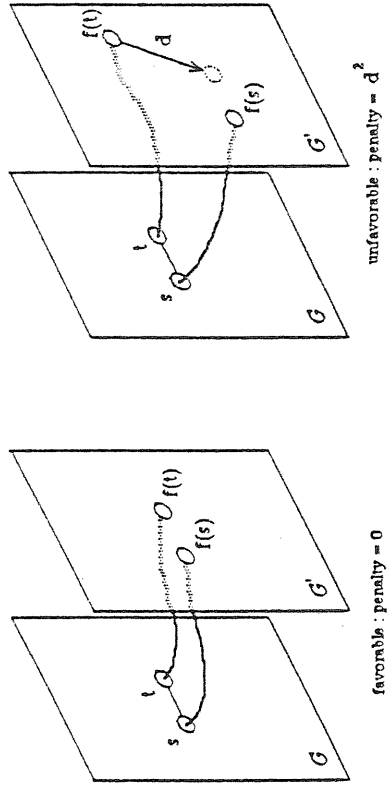


Fig.3. The principle underlying elastic matching as well as retinotopy is that neighbors should be connected to neighbors (see text).

The second term $H_2(f)$ in the cost-function is made of penalties incurred for each s such that the labels of s and $f(s)$ are different, i.e., $\delta(X(s), X'(f(s))) = 0$. The positive parameter β is used to adjust the relative weight of the two terms. In the following, we shall use an infinite β , which turns the weak constraint into a strict one: black (resp. white) nodes map only into black (resp. white) nodes. Minimizing H is then a constrained optimization problem. In some cases however, it proves useful to "soften" this constraint.

To summarize, when given two images X and X' on the square lattice S , minimizing the cost-function $H_{X,X'}(f)$ over all mappings f means trying to fulfill two types of requirements in the same time: preserve the pixel-values, and minimize local distortions. We now use this cost-function H to define a measure of dissimilarity $\mu(X,X')$ as follows:

$$\mu(X,X') = \min_f H_{X,X'}(f), \text{ and}$$

$$\mu^*(X,X') = \max \{ \mu(X,X'), \mu(X',X) \}.$$

Note that the measure of dissimilarity $\mu^*(\dots)$ is symmetric, but does not satisfy the triangle inequality. It is therefore not a metric in the strict sense. We shall nevertheless refer to $\mu^*(X,X')$ as the "elastic distance" between images X and X' .

3. MINIMIZING THE COST-FUNCTION

Given two binary images X and X' , computing $\mu^*(X,X')$ as defined in Section 2 means finding the map f which realizes the minimum of $H_{X,X'}(f)$. Clearly, finding the absolute minimum of H may be quite difficult. One may use stochastic relaxation techniques such as simulated annealing [6,12], but these may be too expensive in many practical applications. We have therefore experimented with a simple iterative improvement method, i.e., in the statistical mechanics terminology, a 0-temperature dynamics: we visit the nodes s in G in a given sequence (generally a fixed random

sequence), and for each $f(s)$ to be updated, we compute the best possible value, given the rest of the map f . Note that due to the form of H , this best value is a function of the value of f at neighbors of s only. Specifically, if we ignore the second term of H for the moment, it is easily seen that the best $f(s)$ is given by:

$$u = 1/n_G(s) \sum_{t \text{ such that } \langle s,t \rangle \in G} (f(t) + (s-t))$$

where $n_G(s)$ is the number of neighbors of s in G , i.e., the number of sites t such that $\langle s,t \rangle \in G$. Remember that this number is always 4 for black nodes, and is generally less than 4 for white nodes. In particular, if $n_G(s) = 4$, i.e., if all four nearest neighbors of s are within G , the best $f(s)$ is simply the center of mass of the images of these four neighbors under f , for the sum of the four vectors $s-t$ then adds up to 0.

Now when taking into account both terms of H , updating $f(s)$ may be more complicated. As mentioned above, we shall only consider here the case $\beta = +\infty$, which implements a strict constraint on labels. In this case, to update $f(s)$, we visit all lattice points around the lattice point v closest to u , ordered in increasing distance from v , until we hit a point w of appropriate color, i.e., such that $X(w) = X(s)$. This point w , which in many cases is actually v , is the updated value of $f(s)$.

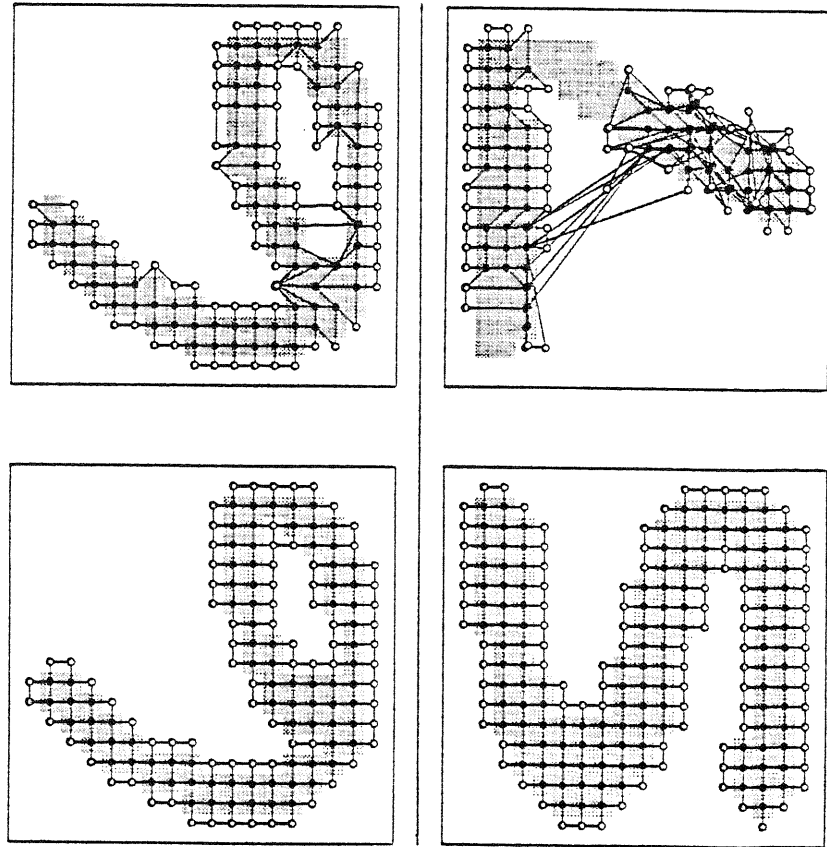


Fig. 4. The solution to the elastic matching problem found by the fast sub-optimal algorithm, in two examples. Graph G is shown on the left; the image of graph G under the mapping f is shown on the right. Note that distinct nodes in G may be mapped to the same node; whenever this happens, the different image nodes are shown with a slight offset.

This is of course a suboptimal procedure: it is not guaranteed to yield a map f which realizes the absolute minimum of H . Its main advantage is its quick convergence: the system stabilizes in a few full iterations (in a full iteration, each s in G is visited once). Not too surprisingly, the behavior of this suboptimal algorithm was found to critically depend on the initialization: when properly "seeded" (see below), the process reliably converges to the absolute minimum, or to a local minimum very close to the absolute minimum. On the other hand, the system can easily get trapped in a poor local minimum, i.e., a "high-energy valley", when given a wrong start. We used the following simple seeding process: the images X and X' are first registered using a fast coarse-grained correlation technique (see Section 4). A small subset K of nodes of G is then selected at random (typically, K contains 10 nodes, of black color only), and for each s in K , $f(s)$ is set to be the pixel s' in G' closest to s (using euclidean distance and under the given registration of the two images) within all pixels in G' of same color as s . After f has been seeded in this way, the value of $f(s)$ for s in $G-K$ is computed using the updating procedure described above, with the obvious modification that only those neighbors t of s with $f(t)$ already defined enter the summation.

Using this method, it was found that stopping after a single pass through all nodes of G , i.e., before the system really settles in a minimum-energy state, is enough to yield a good estimate of the elastic distance $\mu(X,X')$. This easily computed estimate was reliable enough for solving the task described in the next section. Fig. 4 shows this sub-optimal map f for two pairs of images: the map between the two "6"'s has a "cost" of 67, whereas the map from the "5" to the "7" has a cost of 956. These are the estimated elastic distances in the two cases.

4. APPLICATION

In this section, we illustrate the elastic matching approach and the notion of elastic distance defined in Section 2, by an application to a problem of shape recognition, where the task is to both locate and identify occurrences of numerals on a large page. For each of the 10 numerals, a single template is stored: we attempt only to recognize those shapes which are moderate deformations of one of the 10 templates. Yet the problem is a non-trivial one, for it combines two difficulties: shape deformation and a total lack of positional information.

An example of a page to be analyzed is shown in Fig. 5: it contains several occurrences of numerals, as well as various pieces of clutter, in random positions. The search will be done independently for each numeral. Thus, an occurrence of any numeral except the one we are currently looking for may actually be regarded as clutter. Since it is not known where the numerals will occur, every location in the image should be considered a legitimate candidate. However, applying our graph-matching method at all positions in the image would be impractical, even when using the fast optimization method described in the previous section. We therefore divide the search in two steps. Step 1 consists in screening the image to isolate locations which we deem reasonable candidates for further inspection. Step 2 consists of a more careful examination of those "sub-images" which have been isolated in Step 1. Such coarse-to-fine strategies are commonly used in computer vision, and they also convey the flavor of natural perception.

Clearly, what is required from the screening mechanism is speed — the computation is to be performed at every location in the image — and that it be forgiving: one is willing to tolerate a relatively high rate of false positive errors (wrongly deciding that a shape is present) to eliminate the risk of even a single error of the false negative type (failing to identify as such the occurrence of a shape). We shall only briefly describe the screening method that we have actually employed, for it is rather straightforward, and does not bear direct relevance to the topic of the present paper.

We first use spatial averaging to reduce both the image and the template by a factor of k (typically, k is 4, 9 or 16). We then slide the reduced template over the reduced image, and compute the correlation coefficient between the two at every location. We thereby obtain a 2D-surface, of height between -1 and +1. We are clearly interested in the hills of this landscape. We therefore isolate all local maxima of the correlation function (where "local" refers to position in the image), and from these we keep only those peaks which "culminate" at 0.5 or more. Fig. 5 shows the outcome of this screening process for the numeral "2". Every candidate occurrence of a "2" which has been isolated is indicated by a number positioned at the top left corner of the candidate sub-image. This number indicates the rank of the candidate on the correlation scale. For instance, the sub-image marked 1 is actually a perfect replica of the template: the correlation coefficient at this point is 1.0, which is of course the absolute maximum over the entire image.

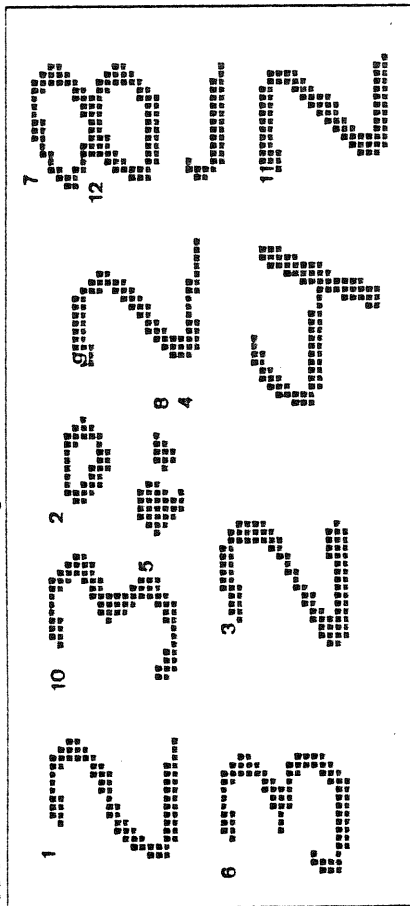


Fig. 5. The page to be analyzed, and the outcome of the screening process (see text).

Inspecting Fig. 5 shows that the screening process has the required properties: it correctly identifies all occurrences of the numeral "2" in the image, in spite of various deformations. This forgivingness is the result of both the spatial averaging and the use of a low "detection threshold", namely 0.5. As expected, Step 1 also produces a number of "false positives": it isolates as candidates for further examination occurrences of other numerals as well as pieces of clutter which happen to correlate favorably with the template.

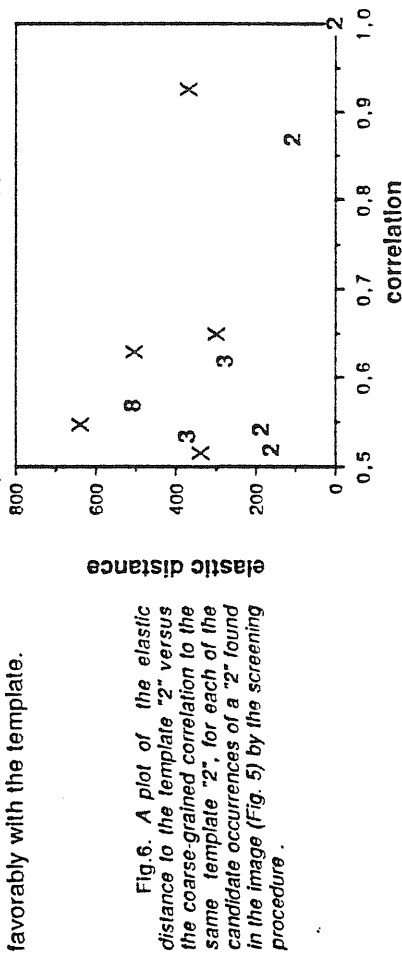


Fig. 6. A plot of the elastic distance to the template "2" versus the coarse-grained correlation to the same template "2", for each of the candidate occurrences of a "2" found in the image (Fig. 5) by the screening procedure.

It is then the task of Step 2 to eliminate those false negatives. Note that at the end of Step 1, we have a list of candidates as well as a rough positioning of the template X on each candidate sub-image X' (positioning is inaccurate due to the deformation of the shape and the spatial averaging employed in the screening process). We are then

ready to apply the method described in Section 3 to compute an estimate of the elastic distance $\mu^*(X, X')$. The result of this computation is shown in Fig. 6: each of the 12 symbols in this diagram corresponds to one of the 12 Step-1 candidates of Fig. 5. The symbol used identifies each case either as the occurrence of a numeral (that particular numeral is then used) or as clutter (the symbol X is used).

It is clear from Fig. 6 that all occurrences of the numeral "2" in the image can be isolated from the rest of the 12 Step-1 candidates on the basis of their μ^* -distance to the template "2". Thus, the elastic distance μ^* defined in Section 2 can be used to effectively recognize as such deformed versions of a given shape, while rejecting the "wrong" candidates, although some of these may correlate quite favorably with the template.

5. DISCUSSION

Several recent works have shown that connectionist models can be successfully used in paradigms such as associative memory or classification. However, some tasks which our brains perform effortlessly seem to resist this traditional neural network approach. In particular, pattern recognition problems often require invariances which cannot be easily obtained in conventional neural network architecture.

We claim that his limitation is intimately linked to the format of knowledge representation, as this format determines the metric underlying the computations performed by the network. Indeed, information is generally represented as a list of low-level attributes, or features. The corresponding metric, e.g. Hamming or euclidean distance in feature space, is most often inadequate to invariant pattern recognition tasks. In a dynamical connectionist model however, information is represented directly in a relational format, i.e., as a configuration of links (rather than a list of features). Accordingly, the metric underlying the computations in such a model may be fundamentally different from the conventional distance in feature space. The present paper shows indeed that the notion of "elastic distance" (actually not a metric in the strict mathematical sense), related to the task of matching relational structures to one another, stands in good agreement with our intuitive notion of similarity of shape.

From the neurobiological standpoint, the approach advocated here is intimately linked to the hypothesis of fast synaptic plasticity [14,16], and to the notion that the fine temporal structure of firing in a family of neurons may be extensively used to carry and process information in the brain [1,14,16]. Another important implication of this approach is the notion that the computations performed in a neural network engaged in a perceptual task may be far more complex than the one-pass propagation of activity through a layered network of fixed connectivity.

From the artificial vision perspective on the other hand, the contribution of this paper is in the notion of the "elastic distance" between shapes construed as the outcome of a constrained optimization problem, along with a fast algorithm for solving this problem. We shall present elsewhere the results of an extensive series of experiments in statistical pattern recognition using a large data-base, showing that the distance μ^* , and actually the fast computed approximation to μ^* discussed in Section 3, yield classification performances which compare favorably with alternative methods.

ACKNOWLEDGMENTS

This research was supported by a grant from the Program BRAIN (Basic Research in Adaptive Intelligence and Neurocomputing) of the E.E.C., under contract n°STJ2-0416-C, and by the French *Ministère de la Recherche et de l'Enseignement Supérieur*,

under contract n°87C0187. It is a pleasure to acknowledge useful discussions with several colleagues, in particular Donald Geman, Stuart Geman and Colin Windsor.

REFERENCES

- [1] Abeles M. (1982) *Local Cortical Circuits: An Electrophysiological Study*. V. Braitenberg, ed. Springer-Verlag, Berlin.
- [2] Bienenstock E. (to appear) *Connectionist Approaches to Vision*. In: *Models of Visual Perception: from Natural to Artificial*. M. Imbert Ed., Oxford Univ. Press.
- [3] Bienenstock E. (1988) *Neural-like Graph-matching Techniques for Image Processing*. In: *Organization of Neural Networks*. W. von Seelen, G. Shaw, and U.M. Leinhos Eds. VCH Verlagsgesellschaft, Weinheim, W.-Germany.
- [4] Bienenstock E. and von der Malsburg C. (1987) *A Neural Network for Invariant Pattern Recognition*. *Europhys. Lett.*, 4 (1), pp. 121-126.
- [5] Burr D.J. (1980) *Elastic Matching of Line Drawings*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-3 (6), Nov. 81, pp. 708-713.
- [6] Cerny V. (1982) *A Thermodynamical Approach to the Travelling Salesman Problem: an Efficient Simulation Algorithm*. *Journal of Optimization Theory and Applications*, 45 (1), Jan. 85, pp. 41-51.
- [7] Duda R.O. and Hart P.E. (1973) *Pattern Classification and Scene Analysis*. John Wiley & Sons, New York.
- [8] Geman S. and Geman D. (1984) *Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-6 (6), pp. 721-741.
- [9] Grenander U. (1976, 78, 81) *Lectures in Pattern Theory*. Volumes 1, 2 and 3. Springer-Verlag, New York.
- [10] Hebb D.O. (1949) *The Organization of Behavior*. Wiley, New York.
- [11] Hopfield J.J. (1982) *Neural Networks and Physical Systems with Emergent Collective Computational Abilities*. *Proc. Natl. Acad. Sc. USA*, 79, pp. 2554-2558.
- [12] Kirkpatrick S., Gelatt Jr. C.D. and Vecchi M.P. (1983) *Optimization by Simulated Annealing*. *Science*, 220, pp. 671-680.
- [13] Kohonen T. (1984) *Self-Organization and Associative Memory*. Springer-Verlag, New York.
- [14] von der Malsburg C. (1981) *The Correlation Theory of Brain Function*. Internal Report 81-2: Max-Planck Institute for Biophysical Chemistry, Dept of Neurobiology. Göttingen, W.-Germany.
- [15] von der Malsburg C. and Bienenstock E. (1987) *A Neural Network for the Retrieval of Superimposed Connection Patterns*. *Europhys. Lett.*, 3 (11), pp. 1243-1249.
- [16] von der Malsburg C. and Bienenstock E. (1986) *Statistical Coding and Short-Term Synaptic Plasticity: a Scheme for Knowledge Representation in the Brain*. In: *Disordered Systems and Biological Organization*. E. Bienenstock, F. Fogelman and G. Weisbuch Eds., pp. 247-272, Springer-Verlag, Berlin.
- [17] von der Malsburg C. and Willshaw D.J. (1977) *How to label nerve cells so that they can interconnect in an ordered fashion*. *Proc. Natl. Acad. Sc. USA*, 74 (11), pp. 5176-5178.
- [18] Metropolis N. and al. (1953) *Equation of State Calculations by Fast Computing Machines*. *J. Chem. Phys.*, 21, pp. 1087-1091.
- [19] Personnaz L., Guyon I. and Dreyfus G. (1986) *Collective Computational Properties of Neural Networks: New Learning Mechanisms*. *Ph. Rev. A*, 34 (5), pp. 4217-4228.
- [20] Rumelhart D.E., Hinton G.E. and Williams R.J. (1986) *Learning Representations by Back-Propagating Errors*. *Nature*, 323, pp. 533-536.
- [21] Sejnowski T.J. (1981) *Skeleton Filters in the Brain*. In: *Parallel Models of Associative Memory*. G.E. Hinton and J.A. Anderson, eds. Erlbaum Associates, Hillsdale, New Jersey.
- [22] Willshaw D.J. and von der Malsburg C. (1976) *How Patterned Neural Connections can be set up by Self-Organization*. *Proc. Roy. Soc. Lond.*, B194, pp. 431-445.