SIMILITUDE-INVARIANT PATTERN RECOGNITION USING PARALLEL DISTRIBUTED PROCESSING.

K.Prazdny

Artificial Intelligence Center, FMC Corporation
P.O. Box 580, Santa Clara, CA 95052

ABSTRACT.

Translation-, rotation-, and scale-invariant recognition of multiple, superimposed, partially specified or occluded objects can be accomplished in a fast, simple, distributed and parallel fashion using localizable features with intrinsic orientation. All known objects are recognized, localized, and segmented simultaneously. The method is robust and efficient.

INTRODUCTION.

Classification of signals on the basis of their shape is a fundamental problem in biological and machine information processing. Interactive activation models view this process as a continuous transaction between the incoming information and stored knowledge. For example, McClelland and Rumelhart (1981) developed an interactive activation model of perceiving letters within the context of words. When a stimulus (a particular spatial distribution of features) is presented to the system, the relevant letter units are activated. The letter units then activate word units that in turn reinforce the activation of the appropriate letter units. In this model, the connections between the letter and word units are fixed. In a more recent model (McClelland, 1986) the connections are "programmable" in the sense that they can be switched on and off by multiplicative signals from the "knowledge" modules. The model uses both the bottom-up and the top-down processing simultaneously to home in on a consistent interpretation of the stimulation. An important feature of such models is parallelism: processing occurs both between and within levels simultaneously. This is one of the few known methods permitting large scale exploitation of mutual, often only partially valid constraints in a distributed and parallel fashion.

Such connectionistic approach is powerful but unfortunately it also has some drawbacks. The sheer number of connections and units necessary to implement a full scale "programmable" similitude-invariant (i.e., translation-, rotation-, and scale-invariant) pattern recognizer is simply enormous. Another problem is that these models usually implicitly assume that the segmentation (i.e. the knowledge of what features belong to the same object) has already been achieved. For example, most word recognition systems of the connectionist variety (e.g. McClelland, 1986) use the spatial locality constraint: a feature at a position x cannot possibly interact with a feature at y because they are separated by more than z units (i.e. they cannot belong to the same letter). In other words, not only the segmentation but also the scale is given (e.g. Hinton & Lang, 1985). In general, however, different objects may be superimposed or occlude each other, and spatially widely separated features may cooperate to define an object. The problem, then, is not the recognition of an isolated object (for which many techniques are available [e.g. Hu, 1962; Casasent & Psaltis, 1976]) but recognition of multiple objects without a prior segmentation or knowledge of which objects are present in the image.

Sometimes, segmentation can be achieved on the basis of "peripheral" information without the involvement of pattern specific evidence. For example, a selective attention mechanism may (at least partially) extract one voice from the jumble of noise and other voices at a party (the cocktail party effect [Cherry, 1953]) based on stimulus onset synchrony or other distinguishing components (e.g. pitch) of one sound. Similarly, in motion and stereopsis, objects can be separated from the background and each other using only motion or disparity information without being recognized first (Julesz, 1971; Prazdny, 1985). Often, such peripheral segmentation based on similarity of a local signal quality is not sufficient and more central evidence based on the geometrical/topological disposition of the individual features has to be invoked. In short, in many (and perhaps most) situations, segmentation, recognition and localization of objects occur simultaneously as the perceptual system tries to "explain" the world.

This paper presents a simple but surprizingly powerful technique for position-, rotation-, and scale-invariant pattern recognition of two-dimensional objects. It is based on parallel distributed processing and uses message passing instead of fixed connections as the primary means of communication. In general, message passing is more economical, versatile, and powerful than "programming" through conjunctive connections.

FEATURES, CODING, AND MODELS.

People and animals (Hollard & Delius, 1982; Rock, 1973) can recognize objects from a variety of viewpoints even when the objects were never seen from those viewpoints previously. The problem addressed in such task can be stated simply: given a set of model objects, each specified independently and in isolation, locate (possibly) occluded, overlapping and/or partially specified instances of the models in an image (described as a "bag" of initially a priori unrelated features) (FIGURE 1).

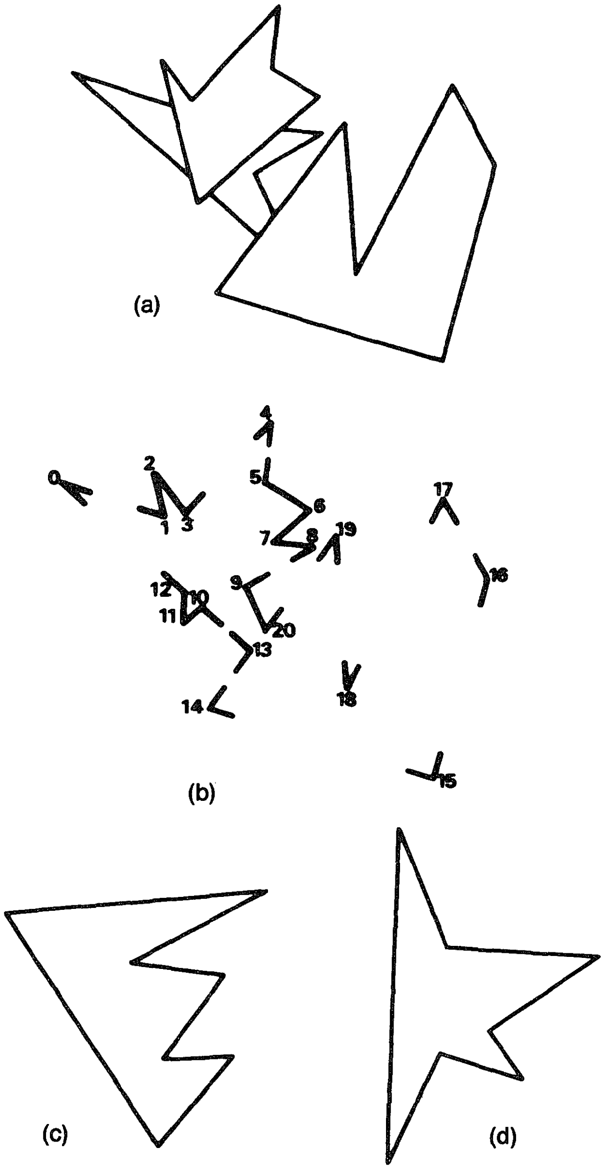

FIGURE 1.

An image [set of overlapping polygons] (a), and the actual input to the program [a set of features] (b). Can you find the two model shapes (c) and (d) in the input image?

There are three important issues involved in such pattern recognition task: (i) the nature of primitives or features used, (ii) the nature and the encoding of information in the model, and (iii) the use of the model in the actual recognition process. Most objects can be characterized by a set of features or landmarks. The landmark encoding of a shape does not, of course, contain as much information about shape as the whole bounding contour but it simplifies the recognition process and captures the essence of shape (Attneave, 1954).

It is difficult to impose a global reference frame on a collection of (retinotopic) features when we do not know which features belong to a single object. Current approaches to this problem (Ballard, 1981; Hinton & Lang, 1985) rely on combinatorial optimization methods. The problem is simplified enormously if one can impose a unique *local* reference frame on each feature separately. The fact that a well defined coordinate frame can be imposed on a locally detected feature or landmark (in general, any orientable point of interest that has some important distinguishing shape attribute) makes it possible to specify the positions, directions and orientations of *all other* features of the same object with respect to the given feature-based frame. Similar coding techniques were used by Baird (1978) in one of the first successful practical applications of computer vision, and by Bolles (1982). Such relational coding means that the learned pattern is be stored and represented in a highly distributed, robust and redundant fashion: a (correct) correspondence between any image and model feature can be used to re-construct (given the overall scale) all other features in the object. This is in many ways similar to the holographic information storage where any piece of the interference plate can be used (given a suitable reference illumination) to reconstruct the whole original image.

In our experiments, we use the simplest localizable and orientable features possible, the corners. A corner is just the intersection of two lines (based on linear approximation): its orientation is determined by the bisector of the enclosed angle. A model is a set of relations specifying, for all features of an object, the position, orientation, and a general appearance of all other distinguishing features of the object as viewed from the given feature. Thus, all features are coded relative to all other features! In our implementation, a model for an object consisting of N features contains 4 N×N arrays containing the relative descriptions. Each row, i, contains descriptions of other features with respect to the local coordinate frame imposed by the feature *i*. The arrays describe the type (e.g. obtuse corner darker "inside"), the value (e.g. the enclosed angle is about 50 degrees), the relative orientation (e.g. 60 degrees clockwise), and the relative direction (e.g. the feature is to be found along a radial oriented about 30 degrees counterclockwise) respectively. A model is thus basically a co-occurence matrix where each row specifies which features (belonging to the same object) occur (or are expected) where relative to the given feature (i.e. the matrix entry M[i,j] specifies some appearance parameter of the feature *j* relative to the coordinate frame imposed by the feature *i*). Presently, the entire feature vocabulary consists of just one type, a corner. Other localizable and orientable features can also be used. It is also possible to use other image entities as "confirmation" features, i.e. features that do not generate "expectations" because they are not localizable (e.g. line segments) or orientable (e.g. points) but can serve to confirm an expectation.

PATTERN RECOGNITION USING PARALLEL DISTRIBUTED PROCESSING.

Given such encoding, the recognition process is surprizingly simple mainly because the relational specification does not involve absolute locations or angles (i.e. the relational coding takes care of the position and rotation invariance). First, image features are extracted. Each feature is associated with a processor or unit that can send to, and accept messages from all other feature processors. Each processor, *i*, has also access to the knowledge sources (the model catalogue) and can determine a set of admissible correspondences$\{c_{ijk}\}$ between itself and the model features. A model is instantiated when a match c_{ijk} is found that maps an image feature (processor) *i* to a feature *j* of model *k*. Two features match if they are of the same type (e.g., both are a corner) and have approximately the same value (in our implementation, the image and model corner angles have to be within *x* degrees of each other, where *x* is a user defined parameter). Each putative correspondence c_{ijk} immediately allows the demployment of the knowledge because each model contains information about all other object features relative to the local reference frame of the model feature *j*. Thus, each c_{ijk} results in a set of messages being broadcast (FIGURE 2a). Each direction can be thought of as priming or gating signals along which information is propagated that can be "digested" only by specific "receptors" within its path. Each message stipulates exactly which feature with what orientation is expected at each direction; features outside the "attention" beams are ignored (this is in many ways analogous to an expectation driven parsing). This is a very general way to obtain *programmable* "connections" not available in the contemporary connectionist schemes. Connections are not fixed but rather generated "on the fly". It is similar in many ways to the earlier proposals of Waltz (1978).

A message is accepted by a processor if the feature type, value, and direction specified by the message agree with the processor's feature specification. When a message is accepted it is in effect a vote for a particular correspondence c_{abc} of the feature processor *a* that accepted the message. It can be thought of as arguing: "the sender of this message says that if you accept this then you are feature *b* of model *c*". Accepting a message also immediately defines the scale of the object relative to the model: the scale is simply the ratio of the original model distance to the signalled distance between the sender and the receiver.

Each processor *i* monitors the support for all feasible correspondences c_{ijk}. After accepting all relevant messages and updating its state accordingly it chooses that c^*_{ijk} (and the associated scale s^*_{ijk}) with the largest support as the "correct" correspondence. All subsequent messages are sent only on the basis of this best "hypothesized" correspondence c^*_{ijk} (this is a "winner-takes-all" computation). The knowledge of the "correct" scale s^*_{ijk} enables the transmitting processor to considerably narrow the message target area: instead of the "attention beams" small regions around the predicted locations are used (FIGURE 2b). That is, the first iteration uses the "attention beams" for message propagation while all subsequent iteration use the more refined estimates based on the knowledge of s^*_{ijk} and the model distance, d^*_{iak} between *i* and *a* under *k*. The radius of the message area is proportional to the distance between the centre of the target area and the sender. This is like sending a message with a delayed "fuse" that "explodes" in a "fireball" some time after its launch as opposed to one

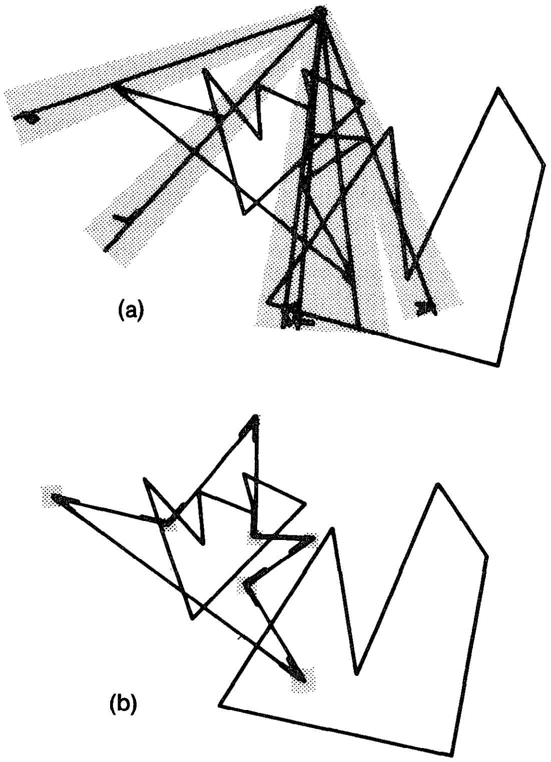

FIGURE 2.

A. Messages generated by the correct feature-to-model correspondence. Each admissible correspondence, c_{ijk}, between an image feature *i* and a feature *j* of model *k* generates a set of messages along directions determined by the local reference frame imposed by the image feature *i* and the relative directions of all other features of the object *k* with respect to the feature *j*. Each message in effect says: "if the image feature *i* (oriented along angle *a*) is the feature *j* of model *k* then the feature *x* of the same model (which lies at an angle β relative to *j*) has to lie somewhere along *a*+β." Each message, denoted by a small angle in the figure, can be caught only by a processor whose feature is sufficiently similar (in type, value, and orientation) to the message specification. An "attention" beam (instead of a single radial) centered at the predicted direction is used to overcome various forms of noise (measurement errors and deformations).

B. When the scale is known the message delivery area can be considerably narrowed to a region around the predicted feature location. The message specification (the type/value pair and the feature direction) is denoted by the thick lines.

that is continuously active. Each processor thus simultaneously and continuously sends and receives messages, and updates its "beliefs" about the "correct" image-to-model feature correspondence based on the information supplied by the other features. After only a few such iterations a global order emerges due to mutual "coercion"; the "conspiracy of partial evidence" drives the system to a stable state (i.e. the state where, for all i, $c^*_{ijk}(t) = c^*_{ijk}(t-1)$) (FIGURE 3). In our experiments we have found that in general, no more than three iterations are required; mostly, the first pass already results in a correct assignment. The mutual support the features belonging to the same object provide to each other is also reminiscent of the concept of synergistic reverberations.

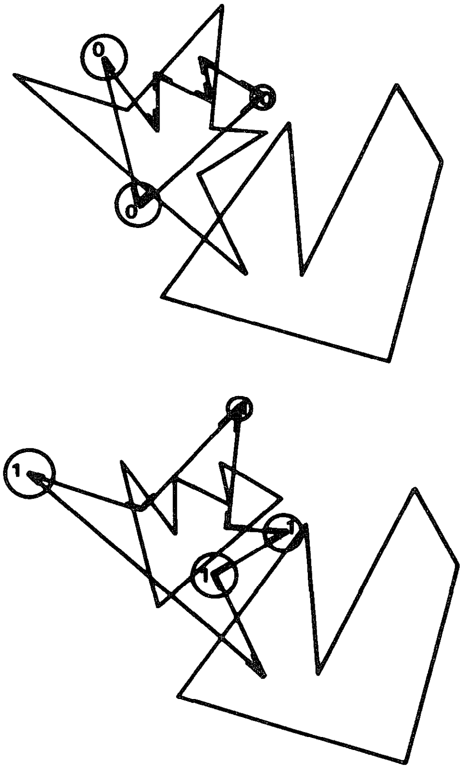

FIGURE 3.

The final state of the system. The models (thick lines denote model features) are superimposed on the image. Numbers identify the models (10 models were in the catalogue in this experiment). Circles around the features identify the features of the same object. The lower right polygon (image features 14-19) has not been recognized, segmented or localized since no model in the catalogue was supported by the spatial distribution of those image features.

Assume that each image feature matches, on the average, n features of m models, and that each model has, on the average, p features. Then, each feature processor generates, on the average, mnp messages. Only a small fraction of such messages is actually accepted by the image feature processors. The computations inside each processor are very simple. The feasibility of the algorithm is thus mainly determined by the speed and costs associated with message broadcasting.

SEGMENTATION.

The parallel distributed approach described above is fast and robust: if there are features that are mutually consistent under some mapping to the same model, they will support each other and survive. The approach leaves one additional problem to solve, however: while we know which image feature correponds to which feature o which model, we do not explicitly know which features belong the the same object! In effect we have solved for the location and identity of objects without segmentation. One simple way to perform such labeling is through support coincidence. Every correspondence or explanation c^*_{ijk} generates a set of locations where it expects to find other image features corresponding to the same object. All other correspondences c^*_{opk} ($o \neq i$, $p \neq j$) generating the same set of expectations are thus instances of the same object. Note that for this computation, it is immaterial whether there are any image features found at the predicted locations! This "labelling" strategy is fast and convenient and does not have the problems of most methods based on consensus. For example, the accumulator peak coincidence problem of the generalized Hough transform (Ballard, 1981) (when two shapes, each specified with respect to a reference point are positioned so that their reference points coincide) is handled in a natural way. Observe also that multiple instances of the same model are handled in the same way as instances of different models.

DISCUSSION AND CONCLUSION.

Experiments with the algorithm (implemented as a simulation on a Symbolics 3670 Lisp machine) revealed that it is fast, robust and efficient. It is robust because any mutually consistent set of features will reinforce each other and survive the "winner-takes-all" competition. It is efficient because of the automatic narrowing of the focus of attention during the search. The system performance is, of course, limited by the accuraccy of its input, mainly by the reliability of the image feature direction measurements.

In conclusion, a position-, rotation-, and scale-invariant pattern recognition can be performed fast and in a simple parallel distributed fashion using features only slightly more complicated than points or edge segments. The algorithm's power derives from relational coding, distribution of processing into many interacting but independent modules performing identical computations, and from the use of directed message passing. This approach is well suited for a fine-grained massively parallel architecture, e.g. the Connection Machine (Hillis, 1985).

REFERENCES.

Attneave F., "Some informational aspects of perception", *Psychological Review*, 61, 1954

Baird M.L., "SIGHT-1": a computer system for automated IC chip manufacture", *IEEE Trans.Systems, Man, Cyber.*, 8, 133-139, 1978

Ballard D., "Generalizing Hough transform to detect arbitrary shapes", *Pattern Recognition*, 13, 11-122, 1981

Bolles R., "Robust feature matching through maximal cliques", *SPIE Technical Symposium on Imaging and Assembly*, 1979 (December)

Casasent D. Psaltis D., "Position, rotation, and scale invariant optical correlation", *Applied Optics*, 15, 1795-1799, 1976

Cherry E.C, "Some experiments on the recognition of speech", *Journal Acoustical Soc. America, 25,* 975 979, 1953

Hillis D. *The connection machine*, MIT Press, 1985

Hinton G. Lang K.J., "Shape recognition and illusory conjunctions", *Proceedings IJCAI*, 252-259, 1985

Hollard V.D. Delius J.D., "Rotational invariance in visual pattern recognition by pigeons and humans", *Science*, 218, 804-806, 1982

Hu M.K., "Visual pattern recogntion by moment invariants", *IEEE Trans.Inf.Theory*, IT-8, 179-187, 1962

Julesz B., *Foundations of cyclopean perception*, University of Chicago Press, 1971

McClelland J.M., "The programmable blackboard model of reading", in: *Parallel distributed processing*, McClelland J.M. and Rumelhart D.E. (eds), MIT Press, 1986

McClelland J.M. Rumelhart D.E., "An interactive activation model of context effects in letter perception, *Psychological Review*, 88, 375-497, 1981

Prazdny K., "Detection of binocular disparities", *Biological Cybernetics*, 52, 387-395, 1985

Rock I., *Orientation and form*, Academic Press, 1973

Waltz D.L., *A Parallel model for low-level vision*, in "Computer Vision Systems", Hanson A.R. Rieseman E.M. (ed), 175-186, Academic Press, 1978