Representation and Recognition of Biological Motion [2]

Nigel H. Goddard [1]
Hughes Artificial Intelligence Center
23901 Calabasas Road Calabasas, CA 91302
and
Division of Biology
California Institute of Technology

I. Introduction

The human visual system has a remarkable ability to discriminate between different types of movement. The classic illustration of this ability is Johannson's Moving Light Display (MLD) [Johansson, 1973]. Reflective pads were placed at the joints of an actor dressed in black, and the actor illuminated. Films were taken of the actor walking, jumping and making various other movements against a black backdrop. When these films were shown to subjects, they all recognized the display to be of a person walking, jumping, *etc.*, but reported single frames to be meaningless patterns of dots. A presentation time of no more than 200 msec was sufficient for all subjects to make the correct discrimination. In forced choice experiments, all subjects accurately identified 6 human and 3 puppet generated patterns with a presentation time of 400 msec. Further experiments [Kozlowski and Cutting, 1977, Cutting and Kozlowski, 1977] demonstrated the sensitivity of this faculty: subjects could determine the actor's gender, and could even identify the actor if (s)he was known to the subject.

This paper describes the early stages of an attempt to produce a computational account of this capability, consistent with the psychological and neurophysiological literature. The following section discusses data on the human visual system which may shed light on MLD processing. Section 3 introduces Feldman's Four Frames computational architecture [Feldman, 1985] for the visual system, outlines the low level processing we believe occurs, and develops and motivates our target representation the *scenario*. Section 4 describes the processing architecture which activates scenarios using the output of the low level system, and details its implementation in a connectionist network.

II. MLD Processing in the Human Visual System

There are two obvious ways the motion information available in the Johannson experiments could be used to generate the percepts of *person* and *walking*, or the single percept of *walking person*. The first method would be to use the motion information to index directly into memory, implying a memory representation rich in temporal information. This method places motion information in a central position *vis-a-vis* the recognition process. The second method would use the motion information to reconstruct various static qualities of the scene object (such as structure), and use those static qualities to index into memory and recognize the object. Having recognized the object, the motion of various key parts of the object could be used to discriminate between gaits. In this second method, the motion information is used in two ways: to recover static qualities; and to disambiguate a small number of gaits.

In this paper we address the first method, but do not rule out the second. Such a motion-specific process and memory structure must play a role in MLD experiments. Johansson's subjects could distinguish gait with a presentation time of less than a quarter of a cycle of the periodic motion (i.e. less than a quarter of a step in walking or running) [Johansson, 1976]. This implies that phasal relationships between joints throughout a cycle of the gait must be represented in memory. Recognition must be based on invariants and the absolute dot motions in an MLD are not invariant with respect to scale or rotation in the image plane. Something like the Johannson's Visual Vector Analysis [Johansson, 1973] must be taking place, with the movement of dots treated as relative to that of other dots. Relative speed of rotation about a joint of the limbs connected at the joint is invariant with respect to scale and rotation in the image plane, and may be a good candidate for the recognition process. However the fact that upside-down MLDs are not recognized as such [Sumi, 1984], while upside down moving stick figures are easily recognized (our own informal observation) implies that the motion invariants used in recognition cannot be computed by the visual system for upside down MLDs. This implicates top-down feedback in the computation of the invariants, under the assumption that the gait is represented in memory for the object in its normal orientation. The memory representation then provides no help in computation of invariants for moving objects in an unfamiliar orientation.

Further evidence for memory structures devoted to representation of sequence and time is provide in [Freyd, 1983]. She presented single frames of a motion sequence to subjects, and then tested their memory for other frames

[1] correspondence should be sent to the first address

[2] An early version of this paper is to appear in *Proceedings of DARPA Image Understanding Workshop 1988* under the title "Recognizing Animal Motion".

from the same sequence. Subjects found it harder to distinguish frames *later* in the sequence from the stimulus frame than they did to distinguish frames *earlier* in the sequence from the stimulus frame. [O'Connell and Gerard, 1985] found that children develop the ability to reproduce familiar sequences earlier than the ability to reproduce the same events prevented in an unfamiliar sequence, implying an early development of representation of sequence. [Runeson and Frykholm, 1983] argue persuasively that representation of body motion is couched in terms of causal factors as well as descriptive components (i.e. force and mass as well as velocity). MLD's of actors lifting boxes were presented to subjects who had no difficulty discriminating the box's weight qualitatively. If the actor attempted to deceive, not only was the deception detected but both real and intended weights were discriminable. Although we will ignore causaul factors in our model, they will have to be integrated eventually.

Whatever the interpretation of the psychological results, it is clear there must be a memory structure rich in information about change in the environment. Brain damaged patients provide neuroanatomical evidence for a separate recognition process based on motion alone. Lesions to the temporal lobe can lead to the inability to identify faces, while leaving intact the ability to identify from body motion; and lesions to parietal cortex can impair recognition from body motion while leaving object recognition unimpaired [Damasio, 1988]. [Perret *et al.*, 1985] found cells in the superior temporal sulcus of the macaque monkey which responded selectively to differing body motions in view, into view and out of view. [Chitty *et al.*, 1987 (in submission)] found cells in the same area that responded selectively to MLD displays, to MLD displays with limb segments suggested by contour; also some cells selective for static form which responded to MLD stimuli, implying computation of form from motion. All these results indicate representation and recognition of sequence is explicitly performed in the visual system.

III. Four Frames

We take as a basis for our computational model the Four Frames architecture [Feldman, 1985] for the visual system, and its extension to deal with kinematics [Feldman, 1988]. Using this computational framework, we may address the question of how MLD image sequences are processed into percepts. Of the four frames (*retinotopic, stable-feature, world-knowledge* and *environmental*), we shall focus on the interaction between the stable-feature frame and the world-knowledge formulary.

A. Retinotopic to Stable-Feature

The first frame, the *retinotopic frame*, is "intended to model the view of the world that changes with each eye movement" [Feldman, 1985, page 265]. In our case the information in the retinotopic frame at any instant is a representation of the dot pattern image on the retina at that instant, modulated by the receptor properties and their time characteristics. The most salient information here will be retinal smear due to the motion of the dots. The second frame, the *stable-feature frame*, computes intrinsic features of the scene being viewed, which do not change with eye-movement. For our purposes, the first important feature encoded in the stable-feature frame will be the motion of the dots. Whether via the short range process or apparent motion, the stable feature frame will provide at each instant the velocity and position of each dot in the moving pattern. Considerable work has been done on the detailing the kind of computations necessary for transformating motion information from the retinotopic to stable-feature frame [Hildreth, 1983]. We assume this computation is performed along the lines suggested in [Feldman, 1988] and [Olson, 1988].

However, as stated in the previous section, the position and motion parameters for each dot in the MLD images are not suitable data to use in indexing. The stable-feature frame must also compute the invariants in terms of which object motion is represented in the next frame, the world-knowledge formulary. We suggested above that for biological motion relative speed of rotation of the two limb segments about their common joint might be such an invariant. In fact we also need to know direction of rotation, and will need some relative positional information, such as angle formed at the joint. In the following subsection we develop a representation for biological motion, the *scenario*, based on relative angular velocity and relative angular position of limb segments. We assume that the invariants used to represent scenarios are computed in the stable feature frame from the position and velocity parameters of the MLD dots, using top-down feedback. The nature of the feedback and the details of the computation are hard unsolved problems, but we would assume at least that the feedback is from the same type of memory used to hold scenarios. The recognition problem is then to index into scenario memory from the invariant values computed in the stable-feature frame.

B. Representing Biological Motion in the World-Knowledge Formulary

Under the Four Frames analysis, scenario memory is contained in the *world-knowledge formulary*. This frame is "the observer's general knowledge of the world, including items not dealing with either vision or space" [Feldman, 1985, page 266]. Knowledge of types of movements of types of objects is general knowledge of the world.

Our scenario representation will be couched in terms of visual *events*. Informally an event is any significant change in one of the invariants specifying the object and motion. A *form event* could be arrival at colinearity of various object features, a *color event* a change in color, and a *motion event* a change in speed or direction of movement.

To make our task tractable, we make the following assumptions. The motion to be recognized is that of an articulated stick figure with bright spots at the joints, moving parallel to the image plane and viewed orthogonally. If the trunk of the figure is moving, we assume the imaging system is tracking the center of rotation of the trunk, so that in the image the trunk is undergoing pure rotation. The limbs are rotating about an end of the trunk, and so on. Thus we have a movement which can be completely described by the length of each stick in the figure, and the change in angle at each joint over time. We shall assume that the change in joint angle over time is piecewise linear, i.e. a sequence of segments of constant angular velocity. This treatment is similar to that of [Johansson, 1973] and recalls Cutting's hierarchy of "centers of moment" [Cutting, 1981]. As in the related Tinker Toy recognition project [Cooper and Hollbach, 1987], we have assumed a principal views treatment and features which are invariant to scale[3]. As viewpoint changes, so that motion is no longer parallel to the image plane, these angular position and velocity cues vary little and can be considered invariant for the purposes of indexing.

We have now delimited the class of movements in such a way that a complete representation is possible. Each joint undergoes a sequence of constant angular velocity changes, which for biological movements such as walking is periodic. The set of sequences *together with information co-ordinating them* describe the motion completely: sufficiently to unambiguously regenerate it. Such a set of sequences of events constitute a *scenario*.

The fundamental motion *event* under these assumptions is a change in angular velocity. The choice of angular velocity as the basis of motion representation is due to data suggesting that we have velocity information availible for a variety of tasks, but not any higher derivative such as acceleration [Jagacinski *et al.*, 1983, Todd, 1981, Runeson, 1975]. There are cells sensitive to rotation [Saito *et al.*, 1986, Sakata *et al.*, 1985], although not sufficiently highly tuned for particular velocities. However the output of several broadly tuned cells can be combined to achieve finer tuning. [Perret *et al.*, 1985] found cells sensitive to velocity change, which would be required to detect our events. But these neurophysiological data are more an indication of what is possible in the visual system rather than definitive evidence for a particular computational or representational scheme.

A simple graphical representation follows from the specification of a scenario given above. We represent each event, or point (in time) when the angular velocity changes, by a graph node. The nodes are labeled with the new angular velocity and the absolute angle of the joint at that time. Directed edges between nodes represent sequence, each edge being labeled with the time between the two nodes. Each sequence is represented by such a graph. The graph is cyclical if the sequence is periodic. The graphs for the sequences are linked with directed edges that specify the *co-ordination* between the sequences, using labels on

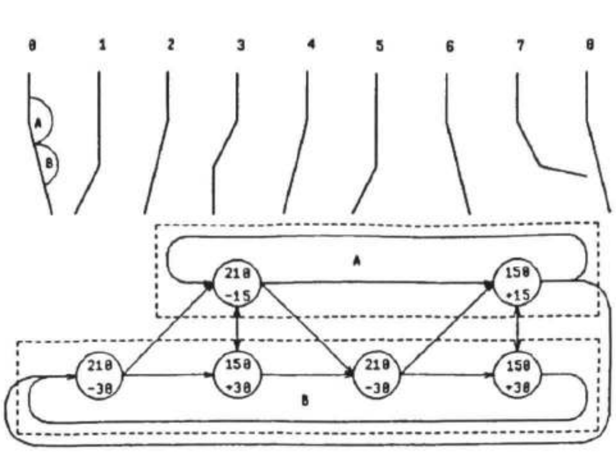

Figure 1: **abstract scenario and graph**

the edges as before.

Figure 1 shows an example abstract scenario and its associated graph. At the top of the figure we show two connected pendula, rotating at joints A and B. The topmost stick is stationary. Over the course of eight time steps, the pendula undergo the motion depicted from left to right. Inter-step motion is of constant angular velocity. Thus it is easily seen that the upper pendulum (A) oscillates with period 8, and the lower pendulum (B) oscillates with period 4. For each pendulum, a cycle of this artificial oscillation consists of two constant angular velocity segments, one clockwise and one counter-clockwise. In the graph we represent changes of angular velocity by nodes. Thus there are two nodes for joint A, shown in the upper dotted box, corresponding to the two changes of direction of rotation during the 8 step cycle. Similarly for joint B there are four nodes corresponding to the four changes of direction during two cycles of the 4 step period. The nodes are labeled with the angle at the joint and the new angular velocity at the joint. For example, the leftmost node in the dotted box for A specifies that this event occurs when the joint angle is 210 degrees and the new angular velocity is -15 degrees/step. Within each dotted box are the sequence links indicating the order of the nodes. Links between the dotted boxes indicate coordination of the joint sequences. The critical *simulataneity* links are shown by the two lines with arrows at both ends.

IV. Connectionist Network Implementation

The preceding section described the information availible in the stable-feature frame and the world-knowledge formulary. For our implementation, the stable-feature frame provides the input representation, and the world-knowledge formulary the memory representation. Recognition is the process of using the stable-feature frame information to activate structures in the world-knowledge formulary.

[3] In fact cells selective for faces by principle view have been found in the macaque monkey [Perret *et al.*, 1987].

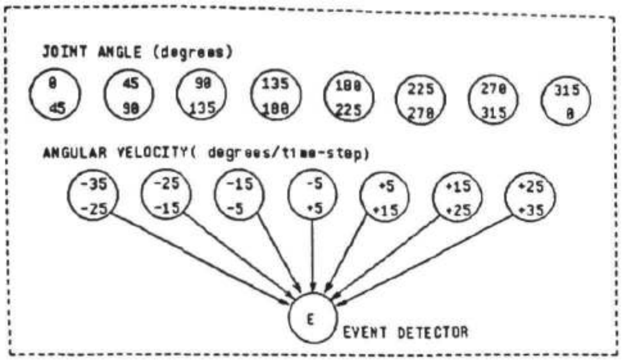

Figure 2: **input module**

A. Input and Scenario Representation

The stable-feature frame will compute the invariants used for recognition. The invariants we have chosen are the angular position and velocity at each joint. Using the unit value principle [Ballard, 1986] we have as input representation a number of *input modules*, each providing the angle and angular velocity for one joint.

Figure 2 shows an input module. Each input module consists of a set of *angle*, a set of *angular velocity* units, and a single unit to detect events (change in angular velocity). Each angle and velocity unit responds to a range of values, shown by the two numbers in each unit; for example, the first angle unit responds to joint angles from 0 to 45 degrees. This figure is an example of an input module for illustrative purposes. In the implementation, we actually use many more units to represent the 360 degree range. The pattern of activity of these units over time describes the kinematics of the joint. We assume that the stable feature frame has segmented the image information into information for each joint, and activates the appropriate number of input modules.

How should we represent scenarios? The graphical representation developed above is naturally implemented as a connectionist network[4]. Each graph node is represented by a unit, and each directed labeled edge by a link with an associated time-delay. The units have a site for *priming* activation which arrives along these delay links, and another site for input from the lower levels of the visual system. Initially all units receive a small amount of priming activation. Units expect activation to arrive at both sites simultaneously. If priming activation or input activation arrives, but not both, then the events in the image are not corresponding to the scenario represented. If the image events do indeed correspond to the scenario represented, then priming activation should flow through the network, building up as it does so. For periodic motions this activation should saturate quickly. Figure 1 is easily re-interpretable as a connectionist network, the graph nodes becoming scenario event units.

[4] In our networks units have one or more *sites* at which links arrive, and where input activation is processed. This enables differential treatment of inputs.

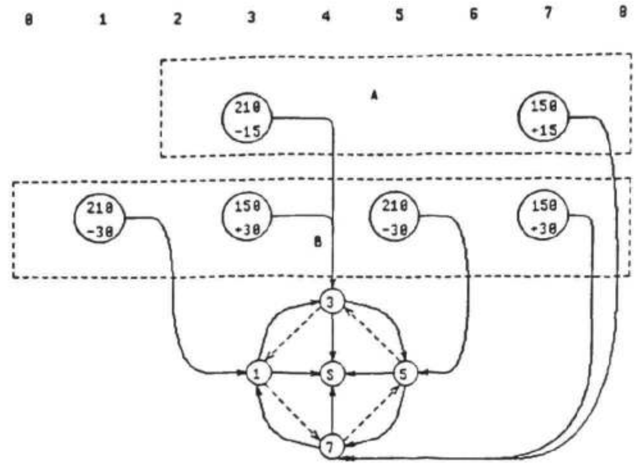

Figure 3: **evaluation network**

These scenario event units are complex. The unit has four sites: one receives inhibition from other event units in the same sequence; one receives priming from event units that represent simultaneous events in other sequences; another receives priming from preceding events; and the last receives input originating in the input modules. The sites compute relatively complex functions of their inputs, such as sum of squares and exponential decay. The unit output is a multiplicative function of the total priming (less inhibition) and the input activation.

A scenario is recognized when activation flows around the network representing it. It is a simple matter to attach a network to the scenario network to detect when and how strongly activation is flowing through the scenario network. The output of this *evaluation* network is a measure of how similar the input is to this particular scenario. Figure 3 illustrates the evaluation network. The dotted boxes from Figure 1 are reproduced, together with the time step count (0 to 8). The evaluation network is the bundle of five units at the bottom of the diagram. The central unit is the summator, which computes the final evaluation. The four peripheral units represent the four time steps at which events take place for this scenario; they are labelled in the diagram with the time steps they represent. If the scenario event units are activated in the correct order, then they will activate these evaluation units in the correct order. The solid links represent the correct order of activation. These have an associated delay equal to the time step difference between the source and destination units. The dotted links are reverese direction inhibitory links, with no delay associated. The links ensure that the evaluation units will not become significantly active unless they are being activated in the correct order. The central summation unit checks that only one is active at a time.

B. Recognition

Assuming the input described above, i.e. at each time step, for each joint in the image, a readout of the angle and angular velocity at that joint, how do we index into scenario memory? We would like the indexing algorithm to be tolerant of missing data points (for example, due to occlusion), and to incrementally converge on the correct scenario as more and more data arrives. At the same time we must avoid exponential growth in the number of units and links required as the number of scenario memories increases. We would also like to be able to take advantage of evidence based on structural or other static qualities of the object if it is available.

Our input modules detect *changes* in angular velocity, thus discretizing the input into a set of sequences of events at which angular velocities change for each joint. These sequences are exactly analogous to the sequences of events represented by the nodes in the scenario graph. For a given scenario we must match the input sequences against the stored sequences to determine which input sequence corresponds to which stored joint sequence. Not only must a mapping from input to scenario be established, the *co-ordination* between input sequences must match the co-ordination between joint sequences in the scenario. We must perform this match for each scenario memory. If we assume a solution to the first problem (matching a particular scenario against the input), then we can achieve recognition time independent of the number of scenarios stored in memory at a cost of linear increase in the number of units and links: we match against all scenario memories in parallel. This is trivial to do in a connectionist network; we simply duplicate the matching machinery for each scenario.

C. The Correspondence Problem

Solving the correspondence problem is harder. We cannot wait until we have all the data before attempting to match. This must also be an incremental process over time. Our approach is to attempt to match all input sequences against all stored sequences in parallel. Again it is trivial to achieve parallel matching in a connectionist network, if one is willing to pay the price in terms of the number of units and links required.

Figure 4 gives a schematic outline of the functional architecture we adopt to solve the correspondence problem. This diagram shows four functional units, depicted by the bold boxes. The input modules and evaluation network, at the top of the diagram, were detailed above. In this figure we show three input modules labelled #X, #Y and #Z. On the left is an example scenario graph network, with three event sequences labelled #A, #B and #C. Each dotted box in the scenario graph represent the event sequence for one joint sequence, as in Figure 1. There is one scenario graph network for each scenario in memory. In the middle of Figure 4 is the grid of binding networks (the dotted boxes). There is one such grid for each scenario graph in

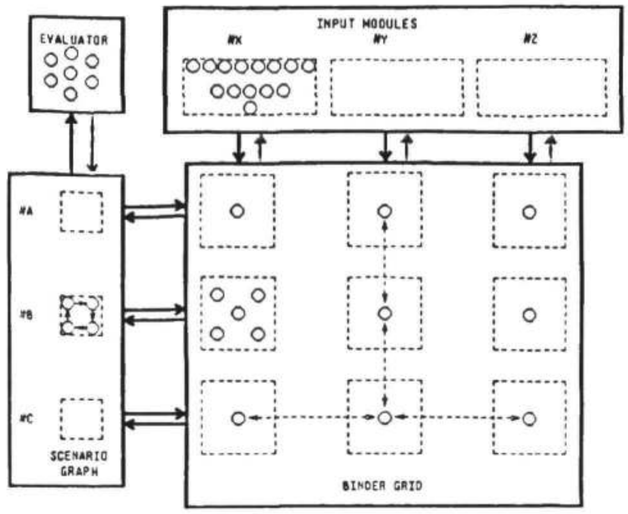

Figure 4: **functional architechture**

memory. A binder grid is always composed of n by n binding networks, where n is the number of event sequences in the scenario which the binder grid is associated with. The heavy arrows between the functional units indicate activation flow, the striped arrows indicating top-down feedback not yet implemented. Each input module sends activation to the column of binding networks below it in the diagram. Each row of binding networks sends activation to the scenario event sequence which is directly to its left, and receives feedback from that event sequence. The scenario event sequences send activation to the evaluation network as shown in Figure 3.

The function of the binder grid is to establish a one to one mapping, or correspondence, between the active input modules and the event sequences in the scenario graph. The input is a time-varying pattern of activation over the set of input modules. There is one input module for each joint in the scene. The the time-varying pattern at an input module should match the expected pattern represented in one of the scenario sequences. The binder grid compares in parallel the sequence of events at each input module with the events in each scenario event sequence. This is achieved by having a separate binding network for each input-module/scenario-sequence pair. Thus the top left dotted box in the binder grid in Figure 4 represents the binding network that is attempting to match the events arriving at input module #X with the events represented in scenario sequence #A. Binding networks with competing interpretations are arranged to inhibit each other, so that if a consistent interpretation can be found there will be exactly one binding network active in each row and column of the grid.

If the match between the input events is close to the sequences in a scenario, a good binding will be found, activation will flow around the scenario network, and the evaluation network will become active. If there is a partial match, the activation in the scenario network will be

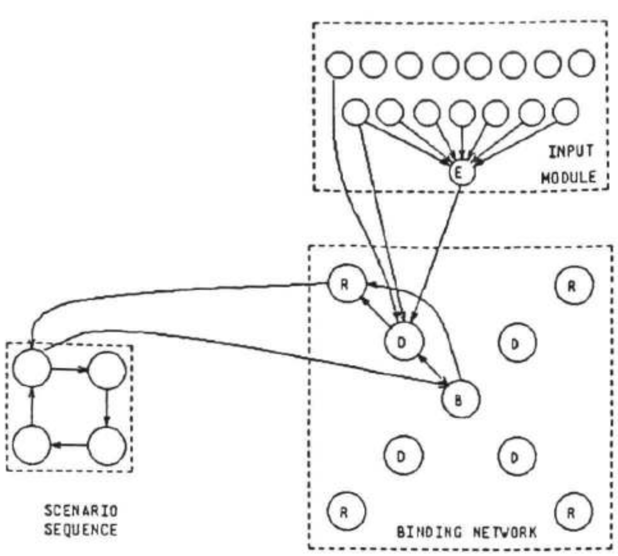

Figure 5: **binding network details**

lower, and less consistent in time, so that the evaluation network may become somewhat active. The best match is found by reading off the activity of the summator units in the evaluation network for each scenario.

Figure 5 shows the details for one binding network. At the top is an input module. It has a set of units tuned to a particular angular range and another set tuned to particular angular velocity. Unit E is an event detecting unit, connected to all the velocity tuned units in the module; it fires when the angular velocity changes, our definition of an event. At the left of Figure 5 is a sequence network from a scenario, similar to the dotted boxes in Figure 1. In the middle is a binding network.

A binding network performs two functions: it passes on input events from its input module to its sequence network; and it compares the events occurring in the input module with those occurring in the sequence network. Events are differentiated by the angle and the angular velocity at the joint, and occur when the angular velocity changes. For each event represented in the sequence network, there is a pair of units in the binding network, a *detector* unit, labeled D, and a *relay*, labeled R. In Figure 5 there are four such pairs, corresponding to the four event units in the scenario sequence. We show the links for one pair. The detector unit fires fires when the appropriate event occurs in the input module, i.e. when all three inputs from the input module are active. The relay unit passes on activation from the detector unit to the appropriate event unit in the scenario sequence network, modulated by the level of activation of the *binding* unit, labeled B. The detector unit also sends activation to the binding unit. This binding unit has a site for each event represented in the sequence network. Each site receives activation from a network event unit and from the corresponding detector unit. If a site receives input from the detector unit, it expects input soon after from the sequence network. Otherwise there is a mis-match occurring. The binding unit checks that the sites are receiving activation in this fashion, and if so increases its activation level. Otherwise its activation decreases.

All binding networks connected to the same input module have their binding units arranged in a mutually inhibitory network (see Figure 4). Similarly all binding networks connected to the same sequence network are arranged so that their binding units inhibit each other. Thus, even with locally ambiguous input, so long as globally the input is determinate, the correct scenario should be the most highly activated. If evidence for matching is available from other sources, for instance form or color matching, it can be used to influence the scenario match by providing input to the binding units in the binding networks.

D. Preliminary Qualitative Results

The architecture has been implemented using the Rochester Connectionist Simulator [Goddard *et al.*, 1988] for two scenarios - one abstract, and one corresponding to a running stick figure. Unsurprisingly, with such distinct choices, the network had no problem with discriminating inputs. The results depend on the actual parameters used in the activation functions and in the time-delayed links. As expected, presenting perfect input causes the scenario network to saturate quickly (within one cycle of the motion). Presenting imperfect input, e.g. with one of the input modules inactivated to simulate occlusion, caused the scenario network to activate more slowly. Recognition was fairly robust over quite large parameter variations. Overall it is clear that the architecture solves the problem, and moreover that it can be tuned along several dimensions: speed, sensitivity to missing data, sensitivity to incorrect data. Exactly how the network should be tuned is a matter for further research, and will require psychophysical experiments.

V. Conclusions

We have introduced a representation for articulated stick figure motion that is naturally implemented in a massively parallel network. A network architecture for indexing into this memory representation from biologically plausible input has been designed. The results of the preliminary implementation and tests are encouraging.

VI. Acknowledgements

Inspiration for this work came from Jerome Feldman at the Department of Computer Science, University of Rochester. Funding was generously provided by Hughes Aircraft Company. My colleagues at Hughes Artificial Intelligence Center critiqued a previous draft of this paper. Christof Koch at the California Institute of Technology, Division of Biology, gave financial support and help with the biological literature.

References

[Ballard, 1986] Dana H. Ballard. Cortical connections and parallel processing: structure and function. *The Behavioral and Brain Sciences*, 9(1):67–120, March 1986.

[Chitty *et al.*, 1987 (in submission)] Andrew J. Chitty, David I. Perret, Amanda J. Mistlin, and M. Harries. Visual cells sensitive to biological motion. 1987 (in submission).

[Cooper and Hollbach, 1987] Paul R. Cooper and Susan C. Hollbach. Parallel recognition of objects comprised of pure structure. In *Proceedings DARPA Image Understanding Workshop*, pages 381–391, February 1987.

[Cutting and Kozlowski, 1977] J.E. Cutting and L.T. Kozlowski. Recognizing friends by their walk: gait perception without familiarity cues. *Bull. Psychonometric Soc.*, 9(5):353–356, 1977.

[Cutting, 1981] James E. Cutting. Six tenets for event perception. *Cognition*, 10:71–78, 1981.

[Damasio, 1988] Antonio R. Damasio. Seminar talk, California Institute of Technology, February 1988.

[Feldman, 1985] Jerome A. Feldman. Four frames suffice: a provisional model of vision and space. *The Behavioral and Brain Sciences*, 8:265–289, 1985.

[Feldman, 1988] Jerome A. Feldman. *Time, Space and Form in Vision*. Technical Report, Department of Computer Science, University of Rochester, 1988.

[Freyd, 1983] Jennifer J. Freyd. The mental representation of movement when static stimuli are viewed. *Perception and Psychophysics*, 33(6):575–581, 1983.

[Goddard *et al.*, 1988] Nigel H. Goddard, Mark A. Fanty, and Kenton Lynne. *The Rochester Connectionist Simulator*. Technical Report TR213, Department of Computer Science, University of Rochester, 1988.

[Hildreth, 1983] Ellen Catherine Hildreth. *The measurement of visual motion*. MIT Press, Cambridge, MA, 1983.

[Jagacinski *et al.*, 1983] Richard J. Jagacinski, Walter W. Johnson, and Richard A. Miller. Quantifying the cognitive trajectories of extrapolated movements. *Journal of Experimental Psychology: Human Perception and Performance*, 9(1):34–57, 1983.

[Johansson, 1973] Gunnar Johansson. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14:201–211, 1973.

[Johansson, 1976] Gunnar Johansson. Spatio-temporal differentiationn and integration in visual motion perception. *Psychological Research*, 38:379–393, 1976.

[Kozlowski and Cutting, 1977] L.T. Kozlowski and J.E. Cutting. Recognizing the sex of walker from dynamic point-light displays. *Perception and Psychophysics*, 21(6):575–580, 1977.

[O'Connell and Gerard, 1985] Barbara G. O'Connell and Anthony B. Gerard. Scripts and scraps: the development of sequential understanding. *Child Development*, 56:671–681, 1985.

[Olson, 1988] Thomas J. Olson. *Forthcoming dissertation*. PhD thesis, University of Rochester, 1988.

[Perret *et al.*, 1985] D.I. Perret, P.A.J. Smith, A.J. Mistlin, A.J. Chitty, A.S. Head, D.D. Potter, R. Broennimann, A.D. Milner, and M.A. Jeeves. Visual analysis of body movements by neurones in the temporal cortex of the macaque monkey: a preliminary report. *Behavioural Brain Research*, 16:153–170, 1985.

[Perret *et al.*, 1987] David I. Perret, Amanda J. Mistlin, and Andrew J. Chitty. Visual neurones responsive to faces. *Trends in Neuroscience*, 10(9):358–364, 1987.

[Runeson, 1975] Sverker Runeson. Visual prediction of collision with natural and nonnatural motion functions. *Perception and Psychophysics*, 18(4):261–266, 1975.

[Runeson and Frykholm, 1983] Sverker Runeson and Gunilla Frykholm. Kinematic specificattion of dynamics as an informational basis for person-an-action perception: expectation, gender recognition, and deceptive intention. *Journal of Experimental Psychology: General*, 112(4):585–615, 1983.

[Saito *et al.*, 1986] Hide-aki Saito, Masao Yukie, Keiji Tanaka, Kazuo Hikosaka, Yoshiro Fukada, and Eiichi Iwai. Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey. *Journal of Neuroscience*, 6(1):145–157, 1986.

[Sakata *et al.*, 1985] H. Sakata, H. Shibutani, K. Kawano, and T. L. Harrington. Neural mechanisms of space vision in the parietal cortex of the monkey. *Vision Research*, 25:453–463, 1985.

[Sumi, 1984] Shigemasa Sumi. Upside-down presentation of the johansson moving light-spot pattern. *Perception*, 13:283–286, 1984.

[Todd, 1981] James T. Todd. Visual information about moving objects. *Journal of Experimental Psychology: Human Perception and Performance*, 7(4):795–810, 1981.