A Connectionist Learning Model for 3–Dimensional Mental Rotation, Zoom, and Pan

Bartlett W. Mel

Artificial Intelligence Group
Coordinated Sciences Laboratory
University of Illinois

Abstract

A connectionist architecture is applied to the problem of 3–D visual representation. The **VI**sual **P**erception System (VIPS) is organized as a flat, retinotopically mapped array of 16K simple processors, each of which is driven by a coarsely–tuned binocular feature detector. By moving through its environment and observing how the visual field changes from state to state for various kinds of motion, VIPS learns to run internal simulations of 3–D visual experiences, e.g. mental rotations of unfamiliar objects. Unlike traditional approaches to visual representation, VIPS learns to perform 3–D visual transformations purely from visual–motor experience, without actually constructing an explicit 3–D model of the visual scene. Instead, the third dimension is represented *implicitly* in the knowledge as to how the pattern of activation on its flat sheet of binocularly–driven processors will shift about as VIPS moves, or only imagines moving, through space. VIPS is argued to be more compatible with a variety of phenomena from the psychology of 3–D perception than previous vision systems, particularly with respect to development, plasticity, and stability of perception, as well as the analogical, linear–time mental rotation phenomena.

1. Introduction

A fascinating puzzle in the study of visual perception lies in the mechanisms by which a compelling impression of 3–dimensionality is achieved via two, 2–dimensional sensory ports. Within the field of computer vision, the most common approach to the problem of 3–D vision has been to develop sophisticated algorithms that use the 2–dimensional retinal images to construct explicit, 3–dimensional models of the visual scene [e.g. Marr & Nishihara, 1978].

The **VI**sual **P**erception System (VIPS) described herein is a connectionist architecture that embodies a new and very different approach to 3–D visual representation. Inspired by the architecture of the brain, which consists to a remarkable degree of 2–dimensional, topographic maps of the sensory and motor modalities, VIPS demonstrates how a flat, retinotopically mapped sheet of simple processors, each driven by a coarsely coded binocular feature detector, can to a large degree achieve the *effect* of a 3–D representation without actually constructing an explicit 3–D model of the visual scene. Furthermore, VIPS is argued to be more compatible with a variety of phenomena from the psychology of 3–D perception than previous vision systems, particularly with respect

This work was originally supported by Thinking Machines Corporation, Boston MA, and subsequently by a Hewlett–Packard/AEA Fellowship.

to development, plasticity, and stability of perception, as well as the analogical, linear-time mental rotation phenomena.

Two tenets illustrate the nature of the VIPS approach to 3–D vision. The first holds that the faculty for running *internal simulations* of visual events, or *envisionment*, is of central importance to an intelligent vision system. More concretely, VIPS learns to drive state sequences on its internal, binocular visual map that approximate the state sequences on the *same* map that would be driven *externally*, by the retinae, were VIPS actually moving through its environment and observing the visual changes. A 3–dimensional visual experience is therefore represented as a *sequence* of states on an internal binocular visual map, where each state codes for the instantaneous 3–D surface layout in the visual field. VIPS' representational repertoire consists of any mostly–continuous transformation of the binocular visual array, that is, any transformation in which *most* of the low–level features in the visual field follow smooth trajectories from state to state in a motion sequence. This is to say that in each state, only a small fraction of the features in the visual field should appear or disappear abruptly at an occluding contour. Included in this class of transformations are *zoom*, *pan*, and *rotation* of the visual field along or about any axis in 3 dimensions, as well as arbitrary combinations of these, and optionally, with different sub–portions of the visual field undergoing different transformations. Visual changes brought about by autonomously moving bodies fall within the representational power of VIPS, but cannot yet be learned or predicted. This limitation is discussed in the concluding section with respect to future stages in VIPS' design.

The second tenet holds that no *a priori* knowledge of 3–D visual transformations need be built in. Rather, VIPS is based on the philosophy that an extremely rich source of knowledge about the 3–D world is available to an intelligent vision system that can act on the environment through its "musculature", and observe and record the resulting changes in its sensory stream. In just this way, VIPS learns the relationship between its own state of motion, and the resulting, highly predictable way in which the visual scene changes through time. In lay terms, VIPS embodies a "learn–by–doing" approach to 3–D vision, as dependent on its own state of activity as on the content of the incoming sensory stream.

While VIPS and its constituent elements are being proposed elsewhere as an abstract model for a visual/motor association cortex in the brain [Mel, 1986a], it will be described here only in its capacity as a connectionist architecture applied to the problem of 3–D visual representation, with particular attention to its compatibility with a variety of phenomena in the psychology of 3–D visual perception.

2. The Organization of VIPS

2.1. Connectivity

Structurally, VIPS consists of 16K simple processors, or Contextrons, organized in a flat, retinotopically mapped grid, with each processor receiving its dominant input, and therefore its "meaning", from a coarsely tuned binocular feature detector that responds optimally to a short, physical micro–feature at one of 4 orientations and 4 depths in the visual field. This type of visual feature detector is purposely analogous in design to the binocular cells in the mammalian visual cortex [e.g. Hubel & Wiesel, 1979], where the depth is given by the horizontal disparity between the Contextron's left and right–eye receptive fields. In addition to this powerful retinally–derived input, a Contextron receives a weighted, excitatory feedback connection from each of its neighbors within some fixed radius, allowing the *current* state of activity in its neighborhood to influence its own *next* state. Finally, each Contextron receives a motion–context input from VIPS' controlling "motor center".

2.2. The Contextron

The Contextron is the simple connectionist processor out of which VIPS is built, essentially performing the function of a multiplexer (fig. 1). In each state, a context field selects some subset of a Contextron's

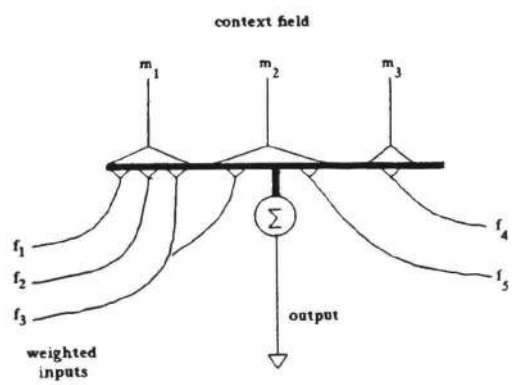


Figure 1. The Contextron.

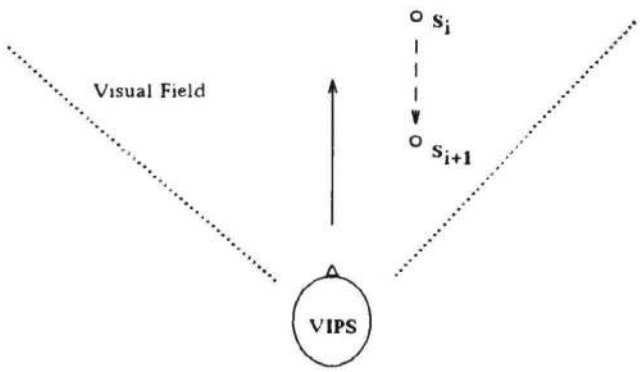
weighted inputs to be gated to the output[1]. Its function as a multiplexer is most easily understood for the degenerate case in which the context field enables only a *single* weighted input. In this case, the Contextron simply gates the single input directly to the output through a weighting factor. In general, however, a context field will usually enable not one, but some small number of weighted inputs. Crucially, each of the enabled inputs in a given context state is signalling the presence of the *same* physical event E, where the weight for each input is a measure of its historical reliability in reporting E. The Contextron can therefore be thought of as taking a weighted "poll" of its selected inputs in order to determine the probability that event E occured. In VIPS, the subset of weighted inputs that is enabled in a given context state will tend to originate from a localized *cluster* of the Contextron's neighbors, and will be seen to signal the imminent incursion of a physical micro-feature into its own receptive field.

2.3. Theory of Operation

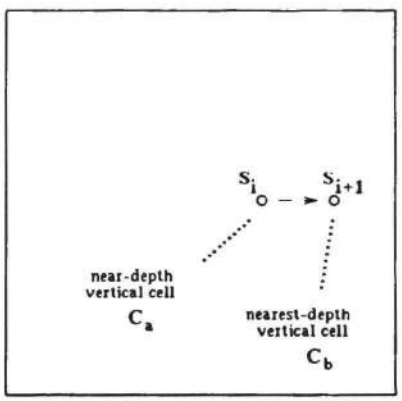
The operation of VIPS depends most directly on the assumption that within a given motion-context, such as "move forward", *most* of the physical features in the visual field will follow smooth, predictable trajectories relative to the moving observer. This is identical to the continuity of flow assumption detailed by Marr [1982].

Consider F, a short, vertically oriented feature in the near-right visual field, and C_a, the Contextron that is most strongly activated by F. If VIPS is moving smoothly forward in space, then the same physical feature that is exciting C_a in state i will move deterministically into a new position relative to VIPS, depending only on the rate of forward motion (fig. 2a). In state $i+1$ therefore, the *same* physical feature will be somewhat less distant and further to the right in the visual field, and will therefore excite a different Contextron, C_b (fig. 2b). Assuming the feature has not had time to travel far across the retina in a single state in the motion sequence, then C_b will be a relatively close neighbor to C_a, within the radius of the feedback paths. Under the continuity-of-optic-flow assumption, it is clear that in the context of forward motion, C_a is an excellent predictor of the next state of C_b. This fact is reflected by a heavily weighted feedback connection from C_a to C_b, for the context of forward motion (fig. 2c). As a useful abstraction, C_a may be given the special title *Probabilistic Physical Predecessor* (PPP) of

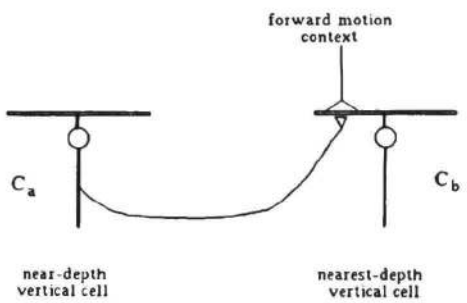
[1] In the current implementation, the output is computed as a simple weighted sum of the selected inputs, but the particular transfer function is not of immediate importance to the model.



(a) Visual change with forward motion.



(b) Internal binocular state.



(c) Resulting feedback connection.

Figure 2. (a) As VIPS moves forward, a vertical feature in its near-right visual field moves closer and further to the right. (b) This same vertical feature stimulates a vertically oriented *near*-cell in state i, followed by a vertically oriented *nearest*-cell in state $i+1$. (c) Since C_a tends to predict the activity of C_b in the context of forward motion, a strong feedback connection will develop for that context.

C_k for this "forward" motion–context, though because the feature detectors driving VIPS are coarsely tuned, a small *population* of Contextrons clustered around the PPP will also develop connections to C_k. In the context of *reverse* motion, C_k should receive strongly weighted connections from the opposite neighborhood, that is one that codes for a physical feature even further to the right in the visual field and still nearer.

For each motion–context, some *different* subset of C_k's neighbors will tend to predict, one state in advance, the onset of its own activation. This reflects the determinism of the trajectories of physical features moving in 3–D space, when the motion state of the observer is known.

2.4. The Contextron Learning Rule

We have seen the desired final distribution of weights for a Contextron, in which each motion–context enables only the weighted input connections from those neighbors that tend to predict the Contextron's own activation, one state in advance. In order to justify this final distribution of weights, we describe here the Contextron learning rule, a variant of the Perceptron learning rule [Rosenblatt, 1958], and tracing its roots to a learning rule proposed by Hebb [1949] for neurons:

> "If a neuron, A, is near enough to another, B, to have any possibility of firing it, and it does take part in firing it on one occasion...the probability is increased that when A fires next B will fire as a result."

Within VIPS, the weight modification rule takes the following form: whenever the feedback input from a neighbor is repeatedly in temporal agreement with the *retinal* input to a Contextron, which acts as a "teacher", then the weight for this neighbor's feedback input is selectively increased for the motion–context currently in effect. Feedback inputs that are *uncorrelated* in their activity with the retinal input are punished, down to a minimum weight of zero. The details of the weight modification algorithm are ommitted here, as they are essentially identical to iterative weight modification rules described elsewhere [e.g. Rosenblatt, 1958].

The physical significance of this rule is as follows: if the current–state activity of some neighbor, as conveyed by its radially projecting feedback path, is repeatedly "felt" to arrive at the same time as the *retinal* stimulation to a Contextron, then the retinal input becomes, in a sense, informationally redundant. Ultimately therefore, a Contextron is able to compute its next state solely on the basis of the current state of its neighbors, the primitive capacity that enables VIPS to internally simulate visual events without benefit of the retina. No negative weights are needed on the assumption that visual micro–features will not in general act consistently as *negative* evidence for other visual micro–features, from state to state in a motion sequence.

The motivation behind the continuity–of–optic–flow assumption can now be made more clear. VIPS can represent exactly those global visual transformations that are both smooth enough and slow enough that each Contextron can know (with high probability) its next state as a function of current state of activity in its fixed local neighborhood.

3. Current Status of VIPS Implementation

3.1. A Simple Visual/Motor Environment

A standard graphics package is used to generate left– and right–eye views of simple 3–D objects on the 32 by 32 bit retinae (fig. 3). VIPS "moves" by issuing a motor command, which has the dual function of setting up the motion–context input to the field of Contextrons, as well as stimulating its virtual "musculature", having the desired side–effect of producing motion–related changes to the binocular visual display. Thus if VIPS were to issue the command "circle object to left", it would "see" the binocularly depicted 3–D object rotating to the right in depth, in 10° intervals about a vertical axis in the center of the visual field.

3.2. Visual–Motor Learning Phase

Since each VIPS motion–context enables a *logically* distinct (though perhaps *physically* overlapping) set of feedback connections for each Contextron, the single type of circular motion described above has been implemented as a test of system principles[2].

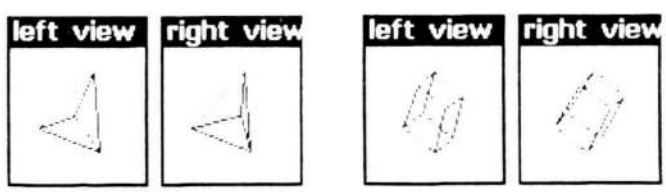


Figure 3. Examples of binocularly displayed VIPS objects.

[2] Pragmatically, the number of possible motion–contexts for each Contextron is limited by the number of resolvable PPP's for that Contextron, a point that is developed further in [Mel, 1986b].

During this phase, each Contextron learns to anticipate retinal stimulation on the basis of the current state of its relevant neighbors. In actively circling a number of randomly constructed and situated parallelograms, the pattern of weighted input connections to each Contextron is sharpened in the following way: input weights from neighbors that only spuriously concur with a Contextron's retinal input will eventually decay to zero, while inputs from neighbors that reflect real, physical predecessors in a motion sequence will become strengthed over time. In this way, each Contextron in VIPS begins to preferentially develop feedback connections from its PPP. Figure 4 shows the pattern of input weights to a particular far-depth/oblique cell that developed through training with 60 total views (e.g. 5 objects through 12 rotation steps each). The figure shows that this particular Contextron developed heavily weighted input connections exclusively from its neighbors of similar orientation, lying slightly to the right and slightly less distant on the average, and lined up in a curious, obliquely-oriented macro-pattern around the PPP. This macro-pattern was unexpected but has the following explanation: whenever the PPP is active, so will a continuous line of *other* Contextrons of similar orientation and depth be active in general, since the objects seen by VIPS tend to be composed of lines much longer than the receptive field of single Contextrons. These may be referred to as the physical correlates of the PPP. An interesting consequence of this elongated pattern of feedback connections is that a Contextron's PPP need not *itself* be present in the current state in order for the Contextron to fire in the next state--instead, VIPS has a tendency to "see" an absent feature if its presence is strongly implied by its physical correlates. The possible relationship of this effect to the perception of *subjective contours* has yet to be investigated.

3.2.1. Envisionment Phase

The second phase of VIPS' operation is the phase of *internal simulation* or *envisionment*, and runs concurrently with phase 1--but only becomes accurate after sufficient phase 1 training. In phase 2, let us assume the array of Contextron's in VIPS is excited into some initial state of activation by the retina, when confronted by a novel 3-D object viewed from an arbitrary perspective. This internal visual state of the Contextron array may be thought of as a "mental image". (Graphic representations of several mental images can be seen in figure 6, each actually a 16K element vector of activation levels over the field of Contextrons, where the intensity at each pixel represents the combined activation levels from cells of 4 orientations at 4

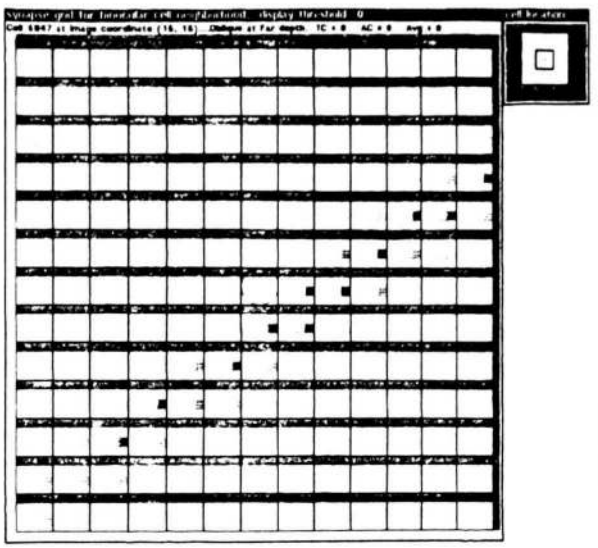


Figure 4. One Contextron's pattern of weighted inputs from its neighbors.

depths with receptive fields centered on that pixel.) By issuing a motion-context to its Contextron's and temporarily inhibiting the retinal pathway, VIPS can transform (e.g. rotate, zoom, or pan) this mental image through time in an *approximation* to the internal state sequence that would be driven by the retinae, were VIPS *actually* moving through its environment and "seeing" the changes. This internal simulation can be successful since each Contextron has learned during phase 1 to compute its *probable* next state of activation solely on the basis of the current state of its neighbors--a level of activation that is normally confirmed by the retinae during a *real* motion sequence. An internal simulation will continue, state by state, until the random error introduced at each state transition renders the mental image unrecognizable.

It was first desired to establish VIPS' capacity to represent 3-D visual transformations under optimal conditions. To this end, VIPS was trained from scratch with a *single* visual-motor sequence, i.e. circling a wire-frame rectangle to the left through 70°, and was simply asked to reproduce the sequence internally. In this configuration, VIPS simply acted as a sensory "tape-recorder", and performed with a high degree of accuracy. Figure 5 depicts both the ideal, retinally driven sequence and the subsequent, internally simulated sequence, with a correlation coefficient reflecting the increasing difference between the two sequences. After 7 steps in the sequence (70 degrees), VIPS maintained a correlation of 0.74 to the ideal retinally produced state. For purposes of comparison, two adjacent states in a visual sequence are typically

correlated at only 0.25, this being a measure of the difference between two views of the same object separated by 10 degrees. This first test therefore indicates that VIPS *is* in fact capable of accurately representing 3–D transformations.

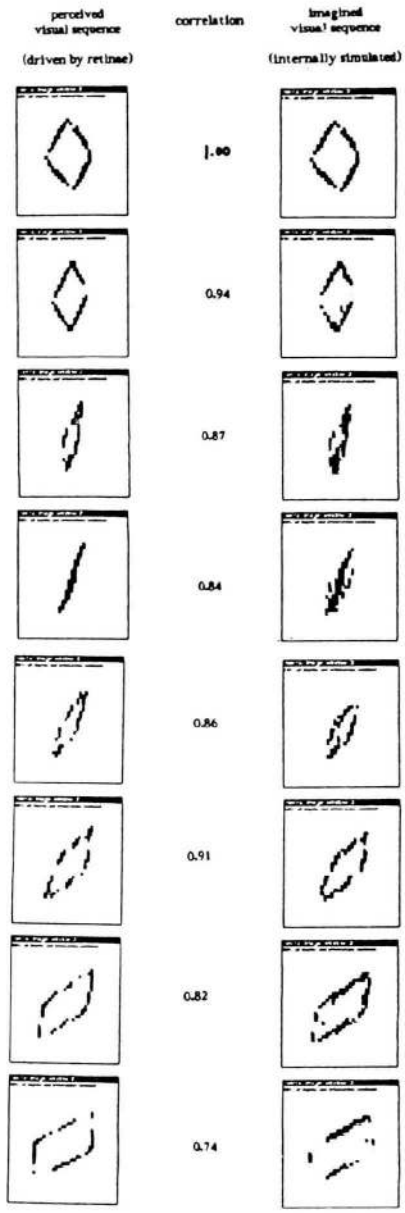


Figure 5.

The results of a more comprehensive training run in the same motion–context appear in Figure 6, which shows the gradual improvement in VIPS' ability to mentally rotate a novel object (a wire–frame pyramid) through one step (i.e. 10°). The left–hand column represents the same, ideal, retinally produced coding in each case, while the right–hand column represents the internally transformed coding after one step. The absolute accuracy of the simulations is still unspectacular, but improving as the implementation is refined.

4. VIPS: Psychological Issues

The idea that visual perception is fundamentally an interactive visual–motor process extending through time, is a rather old one. Whereas the prevailing philosophy in the fields of artificial intelligence and computer vision have focused on the algorithmic extraction of invariant features for the recognition or classification of static visual images, many authors in the psychology of visual perception have emphasized the dynamic character of vision [Miles, 1931; Koffka, 1935; Gibson et al., 1959; Wallach & O'Connell, 1953; Ullman, 1979]. Gibson [1979] argues forcefully that optical motion is the true substance of visual perception, that vision fundamentally involves the extraction of meaning from a continuously changing retinal image, and that optical rest is but a limited special case.

Others have stressed more explicitly that visual–motor *interactions* are the essential ingredient in the perception of space and in the development of such perception [Berkeley, 1709; Washburn, 1916; Helmholtz, 1925; Piaget, 1956; Held & Hein, 1963; Gyr et al., 1979]. Piaget [1952] gives a detailed account for the development of a child's conception of space in terms of the coordination of early motor and visual schemas, an accurate if metaphorical description of VIPS' learning phase. He further describes the child's transition from the *sensory–motor* to the *pre–operational* period as the time when the child learns to take his motor schemas "underground", allowing "abbreviated" movement to drive visual imagery. Again, an apt metaphor for VIPS' envisionment phase of operation.

Metaphor aside, three general areas from the experimental psychology of visual perception seem to support the VIPS account for 3–D visual representation.

4.1. Perceptual Stability

The intimate link between visual and motor processes in perception is perhaps most evident in the long tradition of work in *perceptual stability* [e.g. Held,

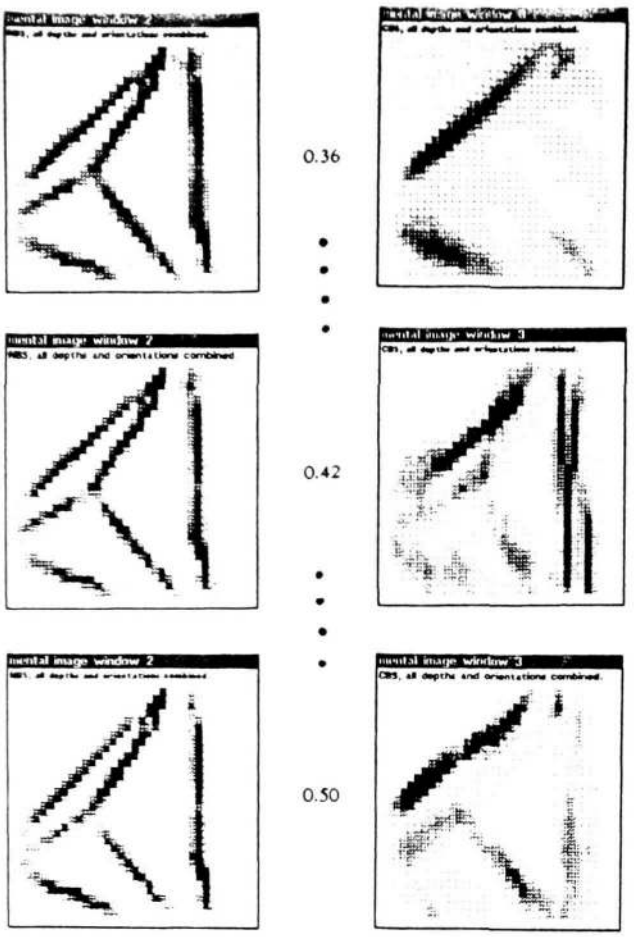


Figure 6. Gradual improvement in 10° mental rotation with training.

1965; Epstein, 1977; Wallach, 1985]. The question addressed by these workers is as follows: Given that the retinal images change with great rapidity during normal visual behavior, what accounts for the subjective stability of perception? The clearest answer to emerge from a large body of experimental work holds that the brain allows active movements of the eyes, head, and body to generate precise expectations of change in the visual field. To the extent that the internal predictions *concur* in real-time with the actual changes in the visual field, the visual environment is subjectively perceived as *stable*, while a lack of concurrence between internal prediction and perception is an indication of movement or change in the environment. In one of its modes of operation, VIPS can be described as just such an "on-line" prediction mechanism, where its internal motion-context field indirectly causes VIPS to "move", while simultaneously maintaining each Contextron in the appropriate context in which to compute its own probable next state of activation. VIPS has yet to be put to the task of identifying parts of its internal visual map for which the locally predicted next-state of each cell is *not* in agreement with the incoming retinal signal, though it is ideally suited to perform this kind of computation. The identification and analysis of autonomously moving bodies in the visual field is an area of particular interest for future work and is discussed briefly in the concluding section.

4.2. Development and Plasticity

Interestingly, the brain's predictive mechanisms for perceptual stability are *plastic*. That is, if the relationship between movements of the self and the associated patterns of visual change is suddenly altered, artificially or otherwise, the predictive mechanisms will at first make incorrect predictions, generally by reporting a stationary environment to move in disconcerting ways. As the subject of such a sudden alteration continues to move about however, he gradually learns the *new* visual-motor relationship, until his internal predictions coincide once again with the stream of retinal input, at which time the visual environment is reported to be stable once again. This illustrates the fact that the stability of perception is much less a function of the absolute rate-of-change in the visual field, than it is of the degree of correspondence between predicted visual change and that which is actually reported by the retinae. Stratton [1896] was the first to experiment with this effect, demonstrating that perceptual stability could be *reattained* in less than a week, even when the visual field was turned upside down with inverting lenses.

Entirely in keeping with the notion that the development of 3-D perception involves the learning of visual-motor relationships, neither development in the young nor such plasticity effects in the adult are observed, in general, in the absence of *active* motor participation on the part of the perceiver. A startling example of this fact is witnessed in an experiment of Held & Hein [1963]. Two cats were yoked together on opposite sides of a circular treadmill, one of the cats walking under its own power, the other being passively carried through the same circular trajectory in a basket (fig. 7). Thus, while both cats shared essentially identical visual experience, only the active cat had the opportunity to learn a consistent relationship between its own activity and the resulting changes in its visual array. On subsequent testing, the active cat appeared normal, while the passive cat was determined to be deficient in a variety of 3-D spatial tasks. A second developmental study with related results is the celebrated visual-cliff study of Gibson & Walk [1960], in which the initial hypothesis and ultimate conclusion both stated that infants of a variety of species begin to

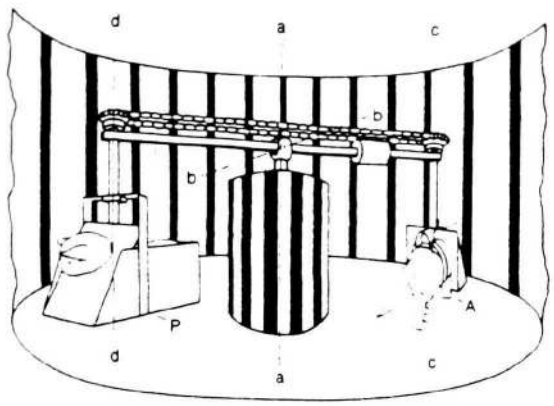


Figure 7. Active vs. passive visual change during development. (From "Movement-produced stimulation in the development of visually guided behavior", by R. Held & A. Hein. In *Journal of Comparative and Physiological Psychology*, 1963, *56*, 872-876. Copyright 1963 by the American Psychological Association.

appreciate the behavioral significance of the third dimension about the time they begin to move under their own power.

VIPS displays both of these effects of development and plasticity: motor activity must be correlated with visual change for successful development of 3-D appropriate behavior, and if the visual-motor relationship is ever altered, the simple weight modification rule will adaptively maintain VIPS' predictive function in as close agreement with the retinal stream as possible. From a systems point of view, this state of affairs is extremely advantageous, since the system designer need not anticipate a *particular* set of visual-motor transformations by building in the appropriate set of special-purpose algorithms [e.g. Hinton, 1981; Kosslyn & Schwartz, 1977; Funt, 1983]. Instead, only the *continuity-of-optic-flow* condition on the visual transformations need be assumed *a priori*, and to the extent that the visual world is well behaved in this respect, VIPS can learn (and relearn) to reliably predict 3-D, movement-induced visual change.

4.3. Mental Rotation

A extensive and elegant body of work on the chronometry of mental rotation, zoom, pan, folding, and apparent motion [Shepard & Metzler, 1971; Robins & Shepard, 1977; Shepard, 1978; Kosslyn, 1980; Shepard & Cooper, 1982; Finke & Pinker, 1983] elucidates a further quality of the internal representations of 3-D objects in human subjects with which VIPS is compatible. The most striking characteristic common to all of these mental transformations has been that the time necessary to carry out the mental transformation grows with its spatial extent, often linearly. Thus, for example, it takes twice as long to mentally rotate an object through 60° as 30°. While this result may seem intuitive, it is very troublesome for those theories of vision that argue for an internal, *view-independent* representation of a 3-D object, for which no mental rotation should be necessary [e.g. Marr, 1982]. On the basis of the chronometry and other data, Shepard [1979] draws the following conclusion on the nature of the internal visual representations:

> "The mental transformation is carried out over a path that is the internal analog of the corresponding physical transformation of the external object...By analogical or analog process I mean just this: a process in which the intermediate internal states have a natural one-to-one correspondence to appropriate intermediate states in the external world."

VIPS has exactly these analogical properties, where the intermediate states during an internal simulation correspond to intermediate visual states, and the time necessary to carry out a transformation is indeed a linear function of its extent. Kosslyn & Schwartz' influential model [1977] captures this same property for a variety of 2-D image plane transformations, but gives no account for the development or plasticity effects, and more importantly, does not deal with 3-D visual transformations. Funt [1983] models stepwise 3-D mental rotations with a built-in spherical shell of processors, but points out that the model cannot account for stepwise transformations of any other kind.

5. Conclusions and Future Directions

The VIPS architecture has been proposed as a more "natural" approach to the representation of the 3-D visual world and its transformations. Starting only with a set of simple, built-in binocular feature detectors driving a regular grid of radially interconnected Contextrons, VIPS learns to envision the visual consequences of an arbitrary motion in its repertoire. VIPS' knowledge of the third dimension is not embodied in explicit 3-D models of the visual scene, as is the common *modus operandi* in computer vision systems. Instead, the knowledge lies in the pattern of modifiable weights that determine how the pattern of activation

over its internal binocular map of the 3–D visual world shifts about during each kind of motion.

Beyond simply improving the performance of the current implementation, there are three directions currently planned for VIPS that reflect its significant limitations.

Firstly, VIPS has a disturbing "out–of–sight, out–of–mind" quality, with no representation of surfaces hidden from view. This limitation highlights an important future direction for this project. Currently, the initial visual state that begins every VIPS simulation is restricted to come from the retinae. If instead, VIPS were configured as one of several distinct fields of Contextrons representing a variety of modalities, then a visual state could be induced in VIPS via a parallel projection from one of the other *internal* fields, supplanting the function of the retinae. The induced visual state could, among other things, be a "remembered" view of the backside of an object. Clearly, however, the complex and varied processes underlying full object–recognition remain far from specified within this model, and are topics for future work.

Secondly, VIPS cannot currently "understand" autonomously produced motion in its visual field. On the other hand, VIPS is capable of *representing* any smooth transformations of the binocular visual field, including those produced by autonomously moving bodies. One possible solution to this problem is a process by which VIPS can "fiddle" with its internal motion contexts until its Contextrons are predicting, as well as is possible, the autonomously produced visual changes.

Most importantly however, to be truly useful, VIPS must be embedded in a larger system that has some form of representation of plans and goals. Currently, VIPS is *capable* of running internal simulations of 3–D visual transformations, but embodies no knowledge as to *why* a particular simulation might be useful for some particular goal, such as bringing a novel view of a 3–D object into registration with another for comparison. VIPS is therefore a useful tool to a system that wishes to run particular internal simulations in service of its goals. This remains a rich and unexplored area for further research.

References

Attneave, F. Representation of physical space. In A. W. Melton & E. Martin, (Eds.), *Coding Processes in human memory*. Washington, D.C., 1972.

Berkeley, *An essay towards a new theory of vision*. 1709 (see any modern edition).

Epstein, W. *Stability and Constancy in Visual Perception: Mechanisms and Processes*. New York: John Wiley & Sons, 1977.

Feldman, J. A. Four frames suffice: A provisional model of vision and space. *The Behavioral and Brain Sciences*, 1985, (in print).

Finke, R. A. & Pinker, S. Directional Scanning of Remembered Visual Patterns. *Journal of Experimental Psychology: Language, Memory, and Cognition*, 1983, *9(3)*, 398–410.

Funt, B. V. A parallel process model of mental rotation. *Cognitive Science*, 1983, *7(1)*, 67–93.

Gibson, E. J. & Walk, R. D. The visual cliff, *Scientific American*, 1960, *202*, 64–71.

Gibson, J. J. *The ecological approach to visual perception* Boston, MA: Houghton Mifflin, 1979.

Gibson, E.J., Gibson, J.J., Smith, O.W., & Flock, H. Motion parallax as a determinant of perceived depth. *J. Exp. Psychol.*, 1959, *8*, 40–51.

Gibson, E.J. & Walk, R.D. The visual cliff, *Scientific American*, 1960, *202*, 64–71.

Gyr, J., Willey, R., & Henry, A. Motor–Sensory feedback and geometry of visual space: an attempted replication. *Behavioral and Brain Sciences*, 1979, *2*, 59–94.

Hebb, D. O. *The organization of Behavior*. New York: Wiley, 1949.

Held, R. & Hein, A. Movement produced stimulation in the development of visually guided behavior. *Journal of Comparative and Physiological Psychology*, 1963, *56*, 872–876.

Held, R. Plasticity in sensory–motor systems. *Scientific American*, 1965, *213*, 84–94.

Helmholtz, H.v. *Handbook of physiological optics*, vol. 3. New York: Optical Society of America, 1925. (Translated by J.P.C. Southall).

Hinton, G. E. Shape representation in parallel systems. In *Proceedings of the 7th International Conference on Artificial Intelligence, 2,*, 1088–1096. Vancouver BC, Canada, 1981.

Hubel, D.H., & Wiesel, T. N. Brain mechanisms of vision. *Scientific American*, 1979, *241*, 150–162.

Koffka, K. *Principles of gestalt psychology*. New York: Harcourt, 1935.

Kosslyn, S. & Schwartz, S. Simulating visual imagery. *Cognitive Science*, 1977, *1*, 265–295.

Kosslyn, S.M. 1980 *Image and Mind*. Cambridge, MA: Harvard University Press, 1980.

Marr, D. *Vision*. San Francisco: Freeman Press, 1982.

Marr, D. & Nishihara, H. K. Representation and recognition of the spatial organization of three–dimensional shapes. *Proc. R. Soc. Lond.* B, 269–294, 1978.

Mel, B.W. Model for the structure and behavior of a hypothetical visual–motor association cortex. Working Paper #69, 1986a. Artificial Intelligence Group, Coordinated Sciences Lab,

University of Illinois, C-U.

Mel, B.W. A connectionist model for simple visual function. Working Paper #70, 1986b. Artificial Intelligence Group, Coordinated Sciences Lab, University of Illinois, C-U.

Miles, W.R. Movement in interpretations of the silhouette of a revolving fan. *Am. J. Psycholol.*, 1931, *43*, 392–404.

Minsky, M. A framework for representing knowledge. In *The psychology of computer vision*, P. H. Winston, (Ed.), New York: McGraw-Hill, 1975

Piaget, J., & Inhelder, B. *The child's conception of space.* New York: Humanities Press, 1956.

Robins, C. & Shepard, R. N. Spatio-temporal probing of apparent rotational movement. *Perception and Psychophysics*, 1977, *22*, 12–18.

Rosenblatt, F. *Principles of neurodynamics. Perceptrons and the theory of brain mechanisms.* Washington, D.C.: Spartan Books, 1961.

Shepard, R. N. The mental image. *American Psychologist*, 1978, *33*, 125–137.

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

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

Stratton, G.M. Vision without inversion of the retinal image. *Psychol. Rev.*, 1896, *3*, 611–617.

Ullman, S. *The interpretation of visual motion.* Cambridge: MIT Press, 1979.

Wallach, H. *Scientific American*, 1985.

Wallach, H. & O'Connell, D. N. The kinetic depth effect. *Journal of Experimental Psychology*, 1953, *45(4)*, 205–217.

Washburn, Margaret F. *Movement and Mental Imagery.* Boston & New York: Houghton Mifflin Company, 1916.