CONSTRUCTIVE VISUAL IMAGERY AND PERCEPTION

Arthur M. Farley
University of Oregon
Eugene, OR 97403

Introduction

One active area of artificial Intelligence research is the Inquiry Into the nature of human cognition. One aspect of this Investigation is the attempt to embody theories of cognition In the form of programs which simulate both general characteristics and specific Instances of observed cognitive behavior. This paper Is a report of one such effort In this area of artificial intelligence viewed as theoretical psychology. It Is based upon research conducted and reported as a Ph.D. thesis In computer science at Carnegie-Melion University (1).

Classical or *atomistic* theory *proposed that* the visual perception of form Is an "unconscious conclusion" realized by "unconscious Inferences" which are based upon the values of the smallest discrlmlnable or homogeneous patches of the stimulus (2). The Gestalt theory of visual form perception developed as a reaction to the fallure of the Classical approach to adequately account for the effects of context upon the valuation (Interpretation) of any part (atom) of the stimulus. Gestalt theory defines several organizational principles (laws of proximity, continuity, symmetry, simplicity) which are applied to the whole perceptual field to produce the visual form perception (3). Gestalt theory, like atomistic theory, has proven to have Its weaknesses. The laws have proven difficult to specify In quantitative or operational terms. Perception of partial figure regions can have significant effects upon complete figural perceotlon (4,5). Most notably, Gestalt theory Ignores the fact that visual form perception normally Involves multiple fixations of the eye and so provides no means for the Integration of information from successive dlfferIng views.

The constructive theory of visual form perception has developed as an alternative to the Classical and Gestalt approaches. This theory proposes that an internal representation of the visual field Is constructed by the integration of a succession of views of (fixations upon) the environment. This representation Is both guide for and product of visual form perception. Hebb (6) began the modern psychological formulation of this theory, describing "cell assemblies" Joined together by (into) "phase sequences" as Its basic functional elements. Hochberg (7,8) has recently continued the investigation. He proposes "schematic maps" as the underlying structural organizations which make possible the selective attention to and the successive Integration of the visual environment.

The research which is reported here Is a further Investigation and specification of the constructive theory of visual form perception. More specifically, the goals have been: (1) to Investigate the nature of the processes and memories which are Involved In the fixation and Integration of successive views of the environment; (2) to Investigate the nature of the internal representation (symbolic visual Image) which Is capable of embodying the necessary partial and complete perceptions; (3) to specify the results of the Investigations in the form of an operational, computer-implemented visual Imagery and perception system (VIPS). VIPS is such a program which has been Implemented In LISP 1.6 on the CUM POP-10.

Motivation for the two Investigative goals Is abundant. The need Is best expressed by the following two statements:

> "There Is little evidence to guide our thinking on how these integrations and constructions take place. In fact,..., little attention has been paid to how such processes occur at all." (9,p174)

> "What we need is a set of operations for deflning and studying the kind of visual storage that will build up the structures of perceived forms out of momentary glimpses" (7,p322)

The third goal, that of computer Implementation of the theory, had two primary motivations. One was to force an operational specification of the theory, somethInq which is often elusive for concisely stated. Intuitively understood, descriptive theories. The other was that features of the implementation could be expected to be naturally applicable in extension to the realization of an adequate, generalized computer vision system.

THE EXPERIMENTS

Two experiments were conducted to provide data from which to Infer characteristics of the visual image representation and rules of the perceptual processes. The first experiment presented subjects with a task situation which forced them to perform a perceptual actlvltv over an extended (cognitive) time frame. VIPS has been Implemented to explain this behavior. The second experiment presented task situations nearer to that of "normal" visual form perception. These results are considered In light of the theory embodied by VIPS and as bases for Its necessary extension and modifIcation.

The main task of Experiment I presented the subject (4 subjects) with a line drawing taped to a table which was covered by a large paper mask with a hole In It. The hole (of approx. 3 degrees visual angle) allowed the subject to view at most one vertex at a time. The subject's task was to move the hold about the drawing until being capable of verbally descrlbing and drawing the picture The subject was also Instructed to "think aloud"

during the hole movement sequence. This task is an extension and modification of a partial, sequential viewing task described by Hochberg ( ),

Transcriptions of the video-taped protocols served as a basis for perceptual process rule specification, for the structure of recognition long term memory, and, together with the verbal descriptions and drawings, for image representation speclfIcatlon. Figure I is the initial segment of a transcribed protocol. B' illustrates the view through the hole (circle) as then seen tv the subject. If this Is a vertex, a label number is given in the lower right. C' Indicates the direction of hole movement. If any, by arrow *and* symbol. D' Is the transcribed verbalization which occurred at that point. This example Illustrates the type of behavior this task situation produced.

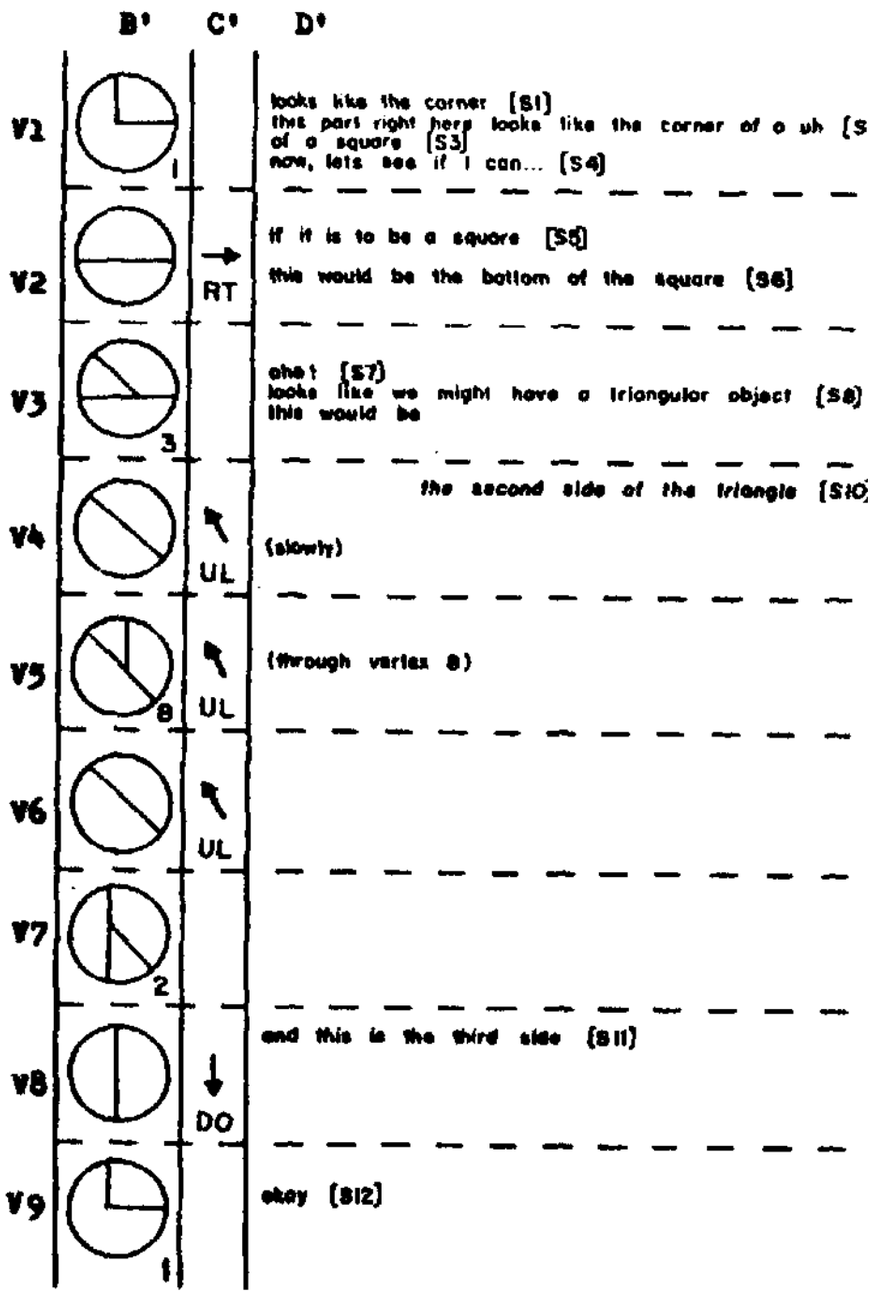


Figure 1

Experiment II was conducted upon a corneal reflecting eye movement tracking and recording (video-taping) system. Two subjects participated; both had been subjects for Experiment I. In the first part, a line drawing (of approx. 20 degrees visual angle) was presented for 250 to 350 milliseconds while the subject fixated a pre-set point of the visual field. Immediately upon removal of the picture, the subject proceeded to verebllv describe and to draw what he now knew of the line drawing. This data served as basis for inferring the processing that an Initial fixation receives and characteristics of the resultant representation.

In the second part, the subject was again given a pre-set Initial fixation point. A line drawing (approx. 20 degrees) was presented, and the subject was allowed to view (scan) the line drawing until being capable of verbally describing and drawing It from memory. The picture was removed at the subject's signal and the subject proceeded first to describe, and then to draw, the line drawing. Eye movements were videotaped and the verbal descriptions were audiotaped. This data was transcribed as sequences of fixations specified by location and duration. Figure 2 shows one such transcription. Field A' Is fixation number, B' is fixation duration. initial segment of the sequence has been connected by dotted lines. The data has been used to Judge the feasibility of accomodating the theory of VIPS to account for eye movement data.

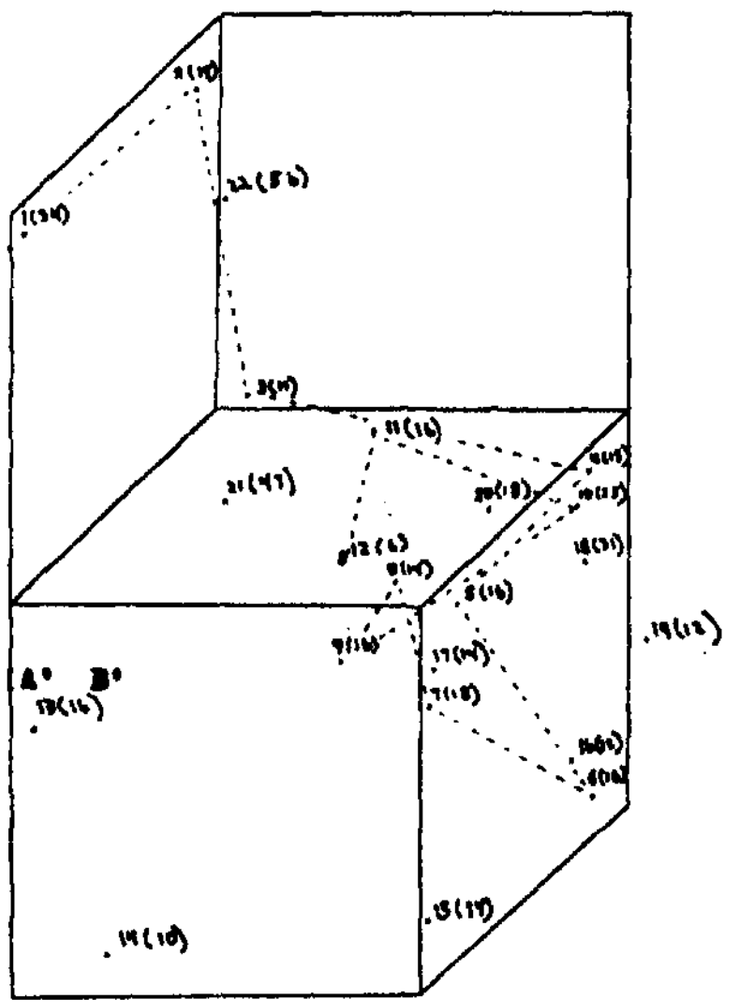


Figure 2

THE IMAGE REPRESENTATION

Imagery has recently returned as a concept under Investigation in cognitive psychology. One result of this renewed interest is the need for an adequate scientific definition of imagery. An image Is defined here to be an Internal, semantic, symbolic representation of Information which is capable of determining (guiding) behavior and

which has an Internal modality characteristic. The Internal modality characteristic distinguishes Images from other forms of Internal symbolic representation. It requires that the Image be structured so that It can be (Is accessed by processes isomorphic to those which access the external environment for the specific sensory modality (10). The units of representation, the symbols and relations, are also modality specific, for vision this being visual feature symbols and spatial relations. By further specification, a visual image is an Internal symbolic representation of visual feature and spatial relation Information which Is structured so as to be straightforwardly retrievable by Internal processes Isomorphic to those which access the visual environment during visual perception. A visual Image Is capable of guiding motor and cognitive behavior with regard to the visual environment that It represents. A primary example of such behavior is visual form perception. On the other hand, visual form perception Is a primary source of visual images. Thus, the perceptual visual Image serves as guide for its own construction.

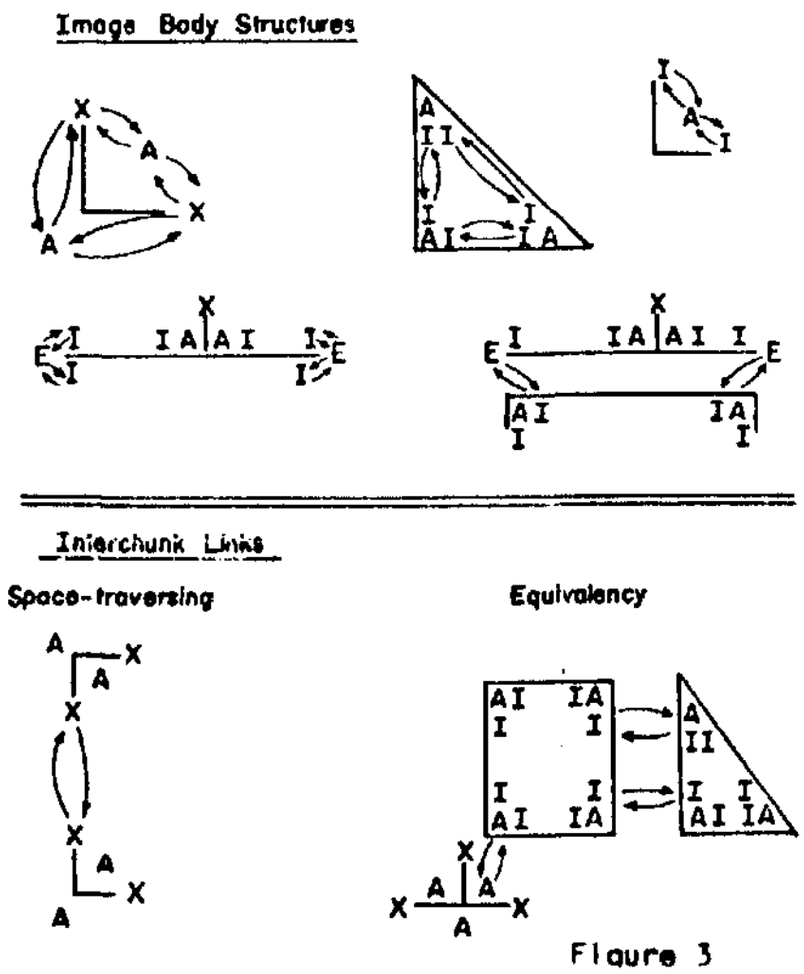
The basic unit of meaningful visual Information wlthin VIPS is the image chunk. An image chunk is a semantic structure of Interrelated symbolic elements, which represent a visual concept. image chunks are the basic structural units from which more complex Images are then constructed. VIPS uses five types of image chunks (concept types) to construct Its representations of the line-drawing environment, these being the VERTEX, OBJECT, LINE, SIDE, and FACE types.

An image chunk consists of one Chunk Header element, one or more Position elements, and several image Body elements. Properties of the Chunk Header serve to Indicate the chunk's type and general characteristics of the current instance. The Current Reference property alwavs references one imaqe Body element, therebv affording a means of access to the chunk's representational structure. A Position element serves to Indicate the perceived location of the Image Body elements which reference It. This position Is In terms of a seven-by-seven coordinate grid of locational areas. All Image Body elements reference a Position element, binding the Image Body structure, and thus the visual Image, to locations in perceptual (imaginal) space. Perceptual space is not retinally-based, but rather Is bound to the area of current Interest within the visual field. This Indirectness plays a significant role In the maintenance of a stable visual world In spite of the differing retinal states which *occur* due to fixation changes during visual form perception.

A chunk's image Body elements form a nonhlerarchlcal symbol structure which embodies the spatial configuration of visual features of the visual concept represented by the chunk. This structure is a doubly linked circular list. The relational links between Image Body elements have directional or space-traversal (direction and distance) meaning. Thus, the allowed means of accesslnq the Image, which Is by traversal of image Body elements according to existing structural links, is Isomorphic In meaning to a visual search, or scan of the external visual environment. This feature of the Image representation is the embodiment of the Internal modality characteristic for visual Imagery.

Five classes of image Body elements (XIT, ANGLE, INTERNAL, END, and QUICKSEE) are used to represent the concepts embodied by Image chunks. Each element class has an associated set of features and relations that it can embody which in turn defines possible roles for It In the Image Body *structure.* *Figure* 3 Illustrates the use of these element classes in representative examples of the different chunk type Image Body structures. A VERTEX chunk consists of alternating XIT and ANGLE elements. An OBJECT chunk consists of "corner configurations" made up of INTERNAL and ANGLE elements. An INTERNAL differs from an XIT element In that It has only one Internal vertex direction link to an ANGLE element. This Image representation feature embodies the figure-ground phenomenon of visual form perception and imagery. To traverse around the "outside" of an object corner, another related chunk must be accessed (attended). The LINE and SIDE chunks introduce the use of the END element. A SIDE chunk Is always associated with an OBJECT chunk.

A visual Image Is a structure of Interrelated Image chunks. An Image Body element of one Image chunk may be linked to one of another chunk by one of two basic relation types. One type has space traversal meaning, relating the two Image Body elements in terms of a direction and distance. The other type Is an equivalency relation. This type of link relates two Image Body elements of different chunks which represent the same aspect of the external visual field. For example, a line segment shared by two adjacent objects Is redundantly represented by both OBJECT chunks. INTERNAL elements embodying the shared line segment "side" of the objects are linked by inter-chunk equivalency relations. (See bottom Figure 3) Note that the line segment Is redundantly represented In both OBJECT chunks.



Flaure 3

The symbolic image representations of VIPS differ from the more semantic representations of visual information proposed by Winston (11). His representations do not satisfy the image definition given above. Such relations as "SUPPORTED-BY" and those accompanied by satellites, such as "MUST-NOT-ABUT", *are* not spatial or equivalency relations. Such semantic relations do appear to play a role In perception, possibly in the Incorporation process which forms more complex, "deeper" structures from interrelated image chunks. The process could provide perceptual hypotheses based upon the long term semantic knowledge to help realize the goal of efficient representation. Also, VIPS does not currently emplov a "SCENE" element, but It would exist In the guise of a contextual element (1).

THE PERCEPTUAL SYSTEM

The perceptual system of VIPS consists of six memories and four processes, as shown by Figure 4. In the figure, an arrow from a process to a memory Indicates that the process can alter the memory's contents, while an arrow from a memory to a process indicates that the process can access the memory's contents. Characteristics of the system at the architectural level reflect relevant data and theoretical proposals of cognitive psychology.

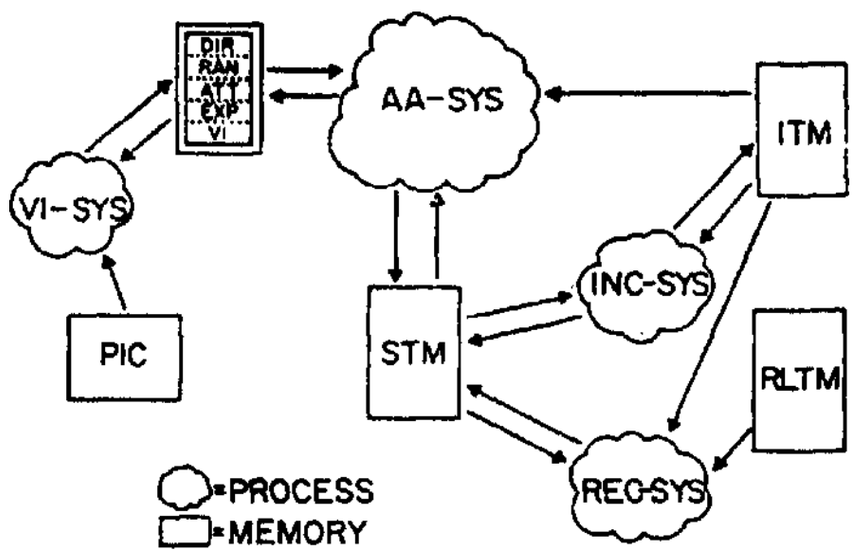


Figure 4

The Memories

Since perceptual activity involves the Interaction of organism and environment, VI PS must represent both. PIC (Picture) is the environment, defined In terms of the visual modalltv. As such, it consists of a list of vertex feature lists. Each vertex feature Iist represents the Inferred detectable visual features at one of the line drawing vertices (the observed points of "hole fixation"). These vertex feature Hsts are interrelated by links of directional and distance meaning. The Current Picture Pointer (CPP) Is associated with PIC and references one vertex feature list, representing the current hole position.

The five cells of the Visual Register (VR) serve as communication registers between the VI-SYS and AA-SYS processes. The cells of VR are set prior to VI-SYS activation by AA-SYS. VI-SYS accesses PIC according to the VR cell specifications and accordingly alters the ceils of VR prior to returning control to AA-SYS by deactivation.

The VI (Visual Information) cell Is the Iconic visual Image. It Is constructed automatically as a result of the new hole positioning. The representation of visual information In VI Is that of an image chunk, though a simple feature list appears favorable. It Is as yet unrelated to any existing Image contents produced by prior perceptual activity. Its contents and structure are not affected by any active perceptuaI goaI.

When specified, EXP (EXPectatlon) and ATT (ATTention) make possible the application of preattentive functions by VI-SYS to the newly accessed PIC Information. EXP can be specified either as an angle code (when ATT Is also specified) or as a vertex type and spec!fIcation. Upon VI-SYS deactivation, EXP will be YES, NO, or CON (CONtalned), Indicating the relationship of expectation to realization. ATT can be specified as a direction symbol. If specified It enables VI-SYS to move through any encountered vertex which is straight (has 180 degree angle) on the ATT side of the line being traversed.

DIR (DIRection) specifies the direction of the hole movement to be effected by VI-SYS, RAN (RANge) specifies the range of that move. RAN Is a value returned by VI-SYS in VIPS (hole movement); during saccadic (ballistic) eye movement behavior, it may be pre-speclfled by AA-SYS. The values of these cells are Incorporated Into the Image by AA-SYS. This reflects the Intrinsic role of motor (efferent) activity and symbols In visual perception and Imagery.

Short Term Memory (STM) consists of an ordered list of nine chunks, being the limited amount of active memory available to the perceptual (cognitive) system for image construction (12) An STM chunk is an image chunk or an Image chunk with a Special type element (GOL, LAST, COM, or 0G0L) appended. STM chunks are accessed according to image Body element containment of Special element appendment. STM is the memory in which the visual image (perception) is constructed. STM and VR *are* the active memory components or "mind's eye" during visual form perception. During recall and drawing Involving imagery, only STM is active as the "mind's eye".

There are three components of long-term memory in VIPS. LTM Is the universe of symbols used in VIPS. The symbols are interrelated, forming a semantic network of symbols and relations. The use of any symbol in an active memory *or* by a process rule is an activation of that LTM element. (In the Implementation In LISP 1.6, It Is a pointer to that element.) As such, any active memory entity (VR cell or STM chunk) is only pseudo self-contained. A symbol Instance Implies the possible use by the active process of all the symbol's LTM relations and related symbols. A symbol remains effectively Imbedded In the LTM structure when activated.

RLTM (Recognition Long Term Memory) is a modified n-any discrimination net. The discrimination structure Is determined by angle symbols and the objects which angle configuration determine. The discrimination net is not used to directly guide perceptual activity. Rather, It Is used to name completed object Images (If possible) or to activate an available image generating process to complete a hypothesled object Image when adequate partial Information is available. RLTM is accessible only by REC-SYS and Is therefore only traversed during REC-SYS activation.

ITM (Intermediate Term Memory) Is the memory Into which selected STM chunks are Incorporated. Meaningful confirmed results of the perceptual activity are transferred to this memory, as the current meaningful processing context. Its contents are recallable into STM for use by the perceptual process. This recall requires a full system cycle thus the contents are not Immediately accessible. As a chunk is incorporated into ITM new relational properties are added to Its Chunk Header element. These Indicate the chunk's immediate imagineI context, being the image chunks to which Its image Body elements are linked. Temporal relations between OBJECT chunks are added also. Thus, the final perceptual iamge, which Is found in ITM, is a heterarchical symbol structure.

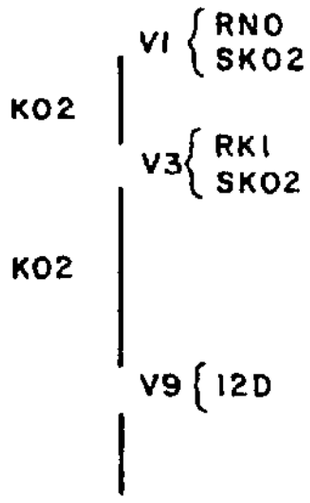
The Processes

All four processes are Implemented In the form of production systems(13). A production system consists of an ordered list of condition-action pairs. The system is cyclic in operation. Wlth each cycle, the first rule of the active process which has Its condition-half satisfied is said to "fire", resulting In Its action-half being executed. Rule firings depend upon and can alter only active memory contents. This reflects the contextual nature of human cognition.

The Assimllatlon-Accomodatlon process (AA-SYS) is named for the TWO basic classes of behavior associated with It (14). Upon Vl-SYS deactivation, this process either can assimilate the new contents of VR Into the Image In STM or must accomodate that current image in light of the conflicting contents of VR. AA-SYS Is the "main" process. It Is the source of most goals. It sets VR and activates Vl-SYS so to access new PIC contents. It activates REC-SYS to aid In Its image construction activity. It activates INC-SYS to Incorporate satisfactorily confirmed, meaningful segments (chunks) of the Image Into ITM.

As the primary source of current goals and their transitions, AA-SYS embodies perceptual strategy In the VIPS Implementetion. Two overall strategies have been inferred from the protocol data, resulting In two corresponding AA-SYS implementations. One strategy Is that of successively recognizing objects until having represented the whole line drawing. The other Is that of first attemptInq to scan and represent the entire line drawing outline and then linking up unknown inward directed exits. The second strategy diverts to the first In a number of specified circumstances.

Each rule of AA-SYS, REC-SYS, and INC-SYS has the name of the current goat as the primary consideration of its condition-half. Visual form perception Is a goal directed activity which consists of a sequence of goal-related episodes. Verbalizations and repeated patterns of hole movements were Indications of this that can be seen In the protocols. The current goal Is a primary determinant of what Image chunks are attended, what external Information Is to be accessed (fixated) next, and Into what class of imaqe structures the new Information will be Integrated and by which It will then be represented. Figure 5 Is a goal episode chart of the goal sequence as Inferred for the protocol segment of Figure I. Goals to the right (In brackets) of a protocol frame number are active at (during) that "hole fixation". A goal to the left of a vertical tine Is active durinq the hole movement sequence occurring between the protocol frames indicated. The resultant perception Is the product of the Interaction (or Interference) between cognitive goals, partial perceptions, and visual environment.



RNO-Recognize New Object
SKO2-Start Known Object 2-Dimensions
KO2-Known Object 2-Dimensions
RKI-Rerecognize, Known Interrupted
I2D-Incorporate 2-Dimensional

Figure 5

The output produced by the VIPS production systems Is a trace of rule firings and resultant active memory contents. Figure 6 presents the VIPS trace corresponding to the human behavior transcribed as frames VI and V2 of Figure I. VI has been set to the initial vertex and STM Is empty. Rule RAAO of AA-SYS fires, assimilating the Icon Into STM as a nee image chunk. The goal RNO is generated by the rule, and a new chunk Is created, marked by the special element GOL. This new chunk is to be the sight of object Image construction durlnq the ensuing REC-SYS actlvation. Rule RECO of REC-SYS next fires and traverses the vertex chunk In STM (marked by LAST), beginning construction of the new object accordingly. RECI next fires as there Is no further relevant image Information In STM. The UQX property of the node currently referenced In RlTM activates a process to complete a sauare-or-rectangte image, as is suggested by the right angle in STM. REC-SYS then deactivates, control returning to AA-SYS. Rule SK02I fires, using the new object image to specify the contents of the visual register before activatlng VI-SYS. Rule VI9 fires, moving the CCP "right" to the feature list for vertex 3.

The rule constructs the Icon, provides range information, and returns NO In EXP.

```
VI IS (V3 P4 X5 A8 X7 A6)

 GOAL-IS: NIL
   2,0      STM: () () () () () () () () ()

     ACT #1: AA-SYS.

     RULE RAA0: SEE FIRST VISUAL INFO.

 GOAL-IS: RNO
   3,0      STM: (GOL     )
          (LAST    V3(CR X7 PNUM P4 VSPEC (RTA DL) SPEC V2 TYPE VERTEX)~
(P4 X5 A8 X7 A6)    )
          () () () () () () ()

     ACT #9: REC-SYS.

     RULE RECO: BEGIN RECOGNITION OF NEW OBJECT.

 GOAL-IS: RNO
   4,0      STM: (GOL     O1O(DIM TWO NUMS ONE TYPE OBJECTXP4 I1 I A12 I~
13)    )
          (COM     V3(CR X5 PNUM P4 VSPEC (RTA DL) SPEC V2 TYPE VERTEXX~
P4 X5 A8 X7 A6)    )
          () () () () () () ()

     RULE REC1: NO MORE MEMORY INFORMATION

 GOAL-IS: SKO2
   5,0      STM: (GOL     O1O(NAME SQUARE-OR-RECTANGLE DIM TWO TYPE OBJE~
CTXP4 I1 I A12 I13 I14 A15 I16 I17 A18 I19 I20 A21 I22)    )
          (V3(CR X5 PNUM P4 VSPEC (RTA DL) SPEC V2 TYPE VERTEXXP4 X5 A8~
X7 A6)    )
          () () () () () () ()

     ACT #1: AA-SYS

     RULE SK021: START LOOKING FOR KNOWN OBJECT INFO

     ACT #23: VI-SYS.

     RULE VI9: EXP & ATT & NO STRAIGHT SIDES FOUND

DIR IS RT

RAN IS NIL

EXP IS RTA

ATT IS UP

VI IS NIL

AT VERTEX VER3

DIR IS RT

RAN IS LO

EXP IS NO

ATT IS UP

VI IS (V24 P25 X26 A31 X30 A29 X28 A27)
```

Figure 6

Figure 7 presents plctoral representations of the image chunks created by VIPS in Its account of the remainder of the protocol section of Figure I. The NO In EXP causes object image accomodation. The new vertex is assimilated as a chunk, the square image Is altered to Include only the right and acute angles, REC-SYS Is reactivated with goal RKI and a triangle image is completed, as suggested by the partial object image in STM. As ATT will be specified as the direction to the Inside of the trianle, VI-SYS "moves" through vertex 8 during its ensuing activation. Since the new vertex is not simple, It Is assimilated Into STM and linked by equivalency to the triangle, A SIDE chunk is created with a OUICKSEE element Included, which notes only the number of exits in the by-passed vertex. Vertex Is next re-encountered In the system's attempt to confirm the triangle image. The vertex agrees with that axpected^so It is not assimilated Into STM, but is "seen" as part of the now completely confirmed triangle image. Belng completely confirmed, the triangle image and Its Immediate image environment are Incorporated into ITM. In general, VIPS accounts for over 80 percent of the hole movements. The system consistently has sufficient Image chunks In STM to be a source of the observed verbalIzatlons.

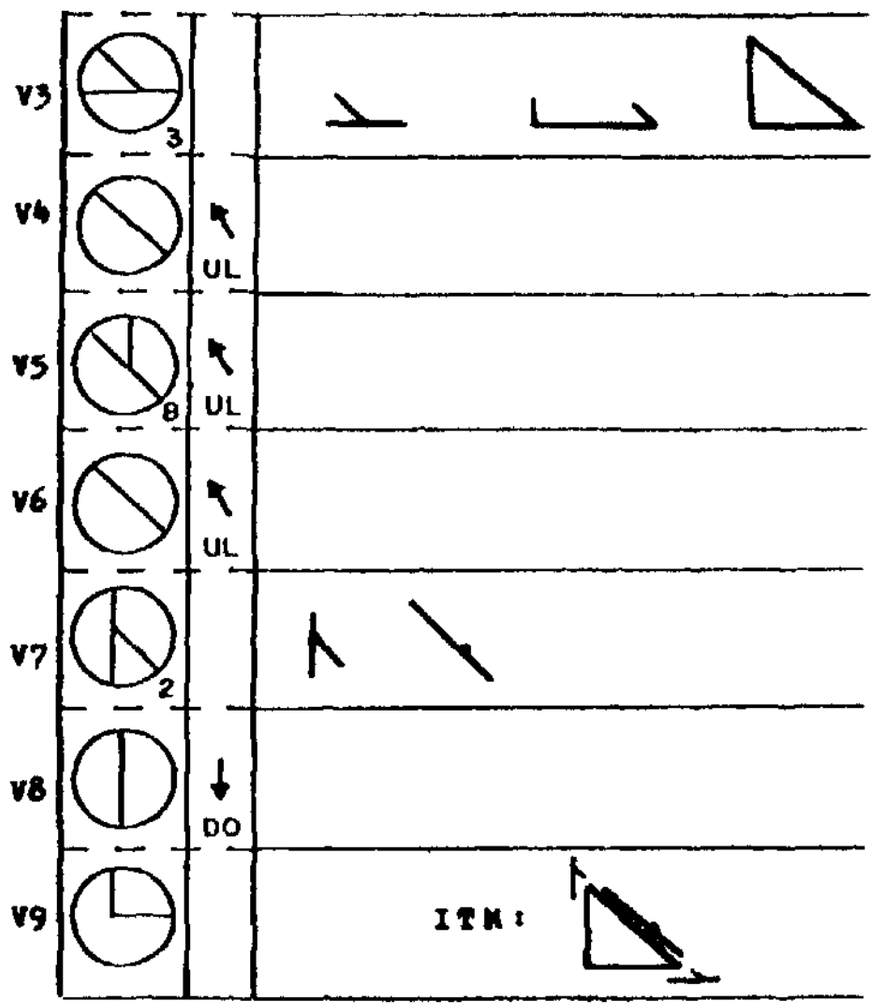


Figure 7

THE INITIAL FIXATION AND EYE MOVEMENTS

The behavior which was observed and recorded as data In Experiment II has been considered only at a general, descriptive level. No perceptual system has yet been Implemented which produces corresponding behavior and which could then serve as an operative theoretical explanation of the behavior. This general consideration of the data has been favorable to the proposition that VIPS can be transformed Into a sufficient theory of that behavior.

The verbal descriptions and partial drawings obtained as data In Part I of Experiment II provide the basis for inferrlna the state of the "mind's eye" followlnq processing of the Initial fixation by the perceptual system. The data indl cate that the Initial fixation serves the perceptual system as the source (basis) of inferences and hypotheses. General characteristics of the total extent of the line drawing are consistently inferred, thus determining (delimiting) the current area of perceptual Interest. Object hypotheses are generated for reqlons (areas bounded by spatially disjoint features) within that extent of the visual field. Such processing of the Initial fixation is consistent with an active, assimtlatlon-accomodatlon theorv of visual form perception as is embodied by VIPS.

Much of the Initial processing is based upon visual information lying In the periphery of the visual field. This Input must be realized by processes of peripheral vision. The data indicate that peripheral Information at eccentricities of up to twenty degrees of visual angle contributed to the Initial Inferences and hypotheses. Some object hypotheses were based upon visual Information which must have been entirely peripheral. Though a sufficient basis for object hypotheses, peripheral vision consistently failed to allow subjects to correctly Infer the particulars of object interactions.

The hole movement task masked peripheral vision InformatIon. Therefore, the representatIon and utilization of peripheral vision's input are not part of VIPS and are most basic and necessary extensions to the system. A class of elements which embody the position of Irregularities or discontinuities In the periphery shows promise as a possible symbolic representation. These elements would occur In VI In addition to the foveally produced feature symbols. The position Information would necessarily be retinal Iy-based (relative in direction and distance to the current fixation point). Local differencing operators have been shown to be effective as a means of detecting and locating discontinuities Q5). The operators are psychologically and physiologically feasible and readily Implementable. These elements appear to be an adequate supplement to fovea! (feature) elements as bases for the observed inferences, hypotheses, and eye movement behavior.

The recorded eye movement sequences obtained In Experiment II have been considered In terms of fixation durations and locations. Fixation durations varied from approximately one-eighth of a second to over a second. The variation In fixation length favors an active, asslmiiation-accomodatlon theory of visual form perception which predicts differing amounts of information processing per fixation. The Initial fixation was significantly longer In duration than those Immediately following it, thus upholding Its role as a highly processed source of Inferences and hypotheses. In general, the eye movement sequences consisted of an Initial, object-related, overall scan of the line drawing, perceptual construction relying heavily upon peripheral Input, followed by a rescan which concentrated primarily upon object interaction areas.

For a specific example, consider the eye movement protocol of Figure 2 and Figure 8, which is the resultant verbal description provided by the subject. Part "a" of the verbal description appears to be a conclusion resulting from the Initial segment of the protocol (that which is connected by dots). The subject's attention is primarily focused upon the ambiguous (In depth) areas of the line drawing. The remaining sequence of eye fixations indicate that the subject scans the bottom cube and finishes with a long fixation between the remaining faces. One can infer from this sequence the construction of an Image sufficient to serve as basis for part "b" of the given description.

```
        that was one of those screwy objects again
(a)     it was a
        it was attempting to be two cubes on top
                                of one another
        In order to be able to draw it I said
        there was a normal cube on the bottom
        where you can see the front
(b)     front right hand side and top
        there were two like
        sides
        two like walls
        then there was a wall sitting on top of
                                the cubes
```

Figure 8

TOWARD COMPUTER VISION

VI PS Is meant to be an operational theory of human visual form perception, but it has a natural relevance to machine vision research. Three general characteristics of VIPS are applicable, and Indlspenslble I believe, to the eventual realization of an adequate computer vision system. One Is the system architecture of VIPS as illustrated In Flqure 3. It provides a basis for the appropriate modularization of memory and process involved In perceptual activity. Memories are required (I) to hold the results of the direct analysis of environmental input, (2) In which to construct the perceptual Image, (3) to hold confirmed partial and the eventual comolete perceptions, (4) to hold applicable recognition (and general) knowledge. Processes are needed (1) to analyze environmental Input (I.e., perceptron and gradient operators), (2) to construct perceptual images and select environmental areas for analytic attention, (3) to Incorporate confirmed partial perceptions Into an overall representation.

The second characteristic is the sequential, localized, In-depth processing of environmental Information and the cyclic assimllatlon-accomodation of the symbolic representation being constructed In light of that processing's results. Early attempts at computer vision found that the information load associated with the storing and processing of a whole scene at one time was astronomical. Since than a main theme of vision research has necessarily been to reduce Internal information flow. What VIPS and other eye movement research Indicates Is that the human vision system faces the same problem and that It approaches (not always solves) the problem by the selective. In-depth processing (fIxation) of local areas within the visual environment. The non-fixated region of the visual environment Is temporarily given only minor attention and sparse representation. Percention is then accomplished by the sequential application of the In-depth process. With each new fixation, new Information is incorporated into the perception under construction. The location of the next fixation is dependent upon the perceptual state following this Incorporation. The perceptual state consists of the partial perception In active memory and the current goal.

The third applicable characteristic of VIPS Is the goal determined nature of Its activity. The current goal plays a role In limiting information flow by influencing what is fixated, how It is represented, and the determination as to when the perceptual construction Is sufficient.

thus concluslng perceptual activity. Data from visual search tasks Illustrate the effects of goals upon the location of fixations during stimulus scanning (16,17). The now classic data nresented by Yarbus (18) Indicate clearly the effects that question-answering (Informational) goals have upon the fixation of a presented visual scene. Other research in cognitive psychology has consistently indicated the selectivity of Information representation based upon task-related goals (19)•

What this should tell researchers In computer vision Is that determining one perceptual strategy and one representational scheme Is not the proper target. An adequate vision system will require the availability of several strategy/representation sub-system. Each can then be properly applied In the different goal-related situations which the associated "seeing organism" will find Itself. There Is *never* "correct perception", only "sufficient perception" for the task at hand.

CONCLUSION

VIPS has been implemented to directly account for four protocols. An evaluation of protocol-program trace correspondence Is favorable to the perceptual theory embodied In the system. The element classes of the Image representation are shown to form a basis for Image generalization. Extensions are discussed which transform the system Into a more inclusive perceptual theory. A comparison of studies yields the proposal that human Information representation In active memory If flexible, being as suited to the task as possible In light of the existing symbols, relations, and structure of an evolving long term memory upon which all cognition Is based.

The research which has been reported here has resulted in a speculative, theoretical specification of the constructive theory of visual form perception. It is not meant to be a conclusion and so this report likewise does not really have one. As an initial Implementation, VIPS can now have modifications and extensions applied (some are discussed above) so to better embody the perceptual activity It Is meant to explain. Being a theoretical statement, It can serve as a source of experimental questions, the resulting experimental findings being the basis for subseguent specification improvement. Finally, the model also embodies characteristics relevant to the development of a generalized machine vision system.

ABSTRACT

A computer implementation of the constructive theory of visual form perception and Imagery is described. An Illustrative example of the system In action and a discussion of Its correspondence to human behavior are presented. Extensions to the Implementation are discussed. Aspects of the model are presented as being relevant to machine vision research.

REFERENCES

(1) Farley, A. M. VIPS: A Visual Imagery and Perception System; the result of a protocol analysis, Ph.D. Thesis, Computer Science Dept., Carnegie-Mellon University, 1974.

(2) Helmholtz, H., Handbook of Physiological Optics, Dover reprint, 1963.

(3) Koffka, K. Principles of Gestalt psychology, New York: Harcourt-Brace, 1935.

(4) Simon, H. A. An Information-processing explanation of some perceptual phenomena. British Journal of Psychology, 1967, 58, 1-12.

(5) Gregory, R. L. The Intelligent eye. New York: McGraw-HIII, 1970.

(6) Hebb, D. O. The organization of behavior. New York: Wiley, 1949.

(7) Hochberg, J. E. In the mind's eye. In Haber, R. N. (ed.), Contemporary theory and research in visual perception, New York: Holt, Rinehart & Winston, 1968, p 309-331.

(8) Hochberg, J. E. Attention, organization, and consciousness. In Mostofsky (ed.), Attention: contemporary theory and analysis, New York: Appleton, 1970.

(9) Haber, R. N. & Hershenson, M. The psychology of visual perception. New York: Holt, Rinehart & Winston, 1973.

(10) Simon, H. A. What Is visual imagery? an information processing interpretation. In Gregg, L. W. (ed.), Cognition in learning and memory, New York: McGraw-HIII, 1972, p 183-204.

(11) Winston, P. H., Learning structural descriptions from examples, MAC-TR-76, Cambridge, Mass: MIT.

(12) Miller, G. A. The psychology of communication. New York: Basic Books, 1967.

(13) Newell, A. & Simon, H. A., Human Problem Solving, Englewood Cliffs, N. J.: Prentice Hall, 1972.

(14) Piaget, J. Structuralism. New York: Basic Books, 1970.

(15) Uhr, L. Pattern Recognition, learning, and thought. Englewood Cliffs, N. J.: Prentice Hall, 1973.

(16) Williams, L. G. The effects of target specification on objects fixated during visual search. Acta Psychologica, 1967, 27, 355-360.

(17) Gould and Dill Eyemovement parameters and pattern discrimination. Perception and Psychophysics, 1969, 6, 311-320.

(18) Yarbus, A. L. Eye-movements and vision, New York: Plenum Press, 1967.

(19) Trabasso T. & Bower, G., Attention in learning: theory and research, New York: Wiley, 1968.