ESPRIT '86: RESULTS AND ACHIEVEMENTS
Directorate General XIII (Editors)
Elsevier Science Publishers B.V. (North-Holland)

TOOLS FOR INTEGRATION OF PERCEPTUAL DATA

Giulio Sandini* and David Vernon**

**University of Genoa, Dept. Communication, Computer and Systems Science*
***Trinity College, Dublin, Dept. Computer Science*

Esprit project P419 is concerned with the elucidation of the computational base inherent in perceiving general 3-D scenes and in perceptual-motor skills associated with cursive script understanding. This paper focuses on a specific result of our research, a software tool - VIS *(Virtual Image System)*- which allows the use of multiple representations of image data. This tool is conceived for a highly interactive environment in which the user - an artificial vision scientist - may conceive and carry out experiments with sets of image data, including time sequences of stereo image pairs, for investigating strategies of integration of perceptual data. The software tool, originally developed in the context of a specific task of P419 that deals with early processing of visual information, is now being expanded in the direction of three-dimensional integration, cognitive computation and also cursive script analysis. At present, the system is at an advanced specification stage and preliminary implementations are being used.

1. INTRODUCTION

1.1 ORGANISATION

Project 419 is directed toward the understanding of the computational base inherent in perceiving general 3-D scenes and in perceptual-motor skills associated with cursive script generation. An underlying consideration is that there is, indeed, a commonality of function in these two seemingly disparate processes. With particular reference to the image understanding aspect of the project, a central theme of the research is the requirement of utilising several low-level vision modules in a coherent manner by incorporating some form of cognitive modelling. The importance of cognition and appropriate organisation is echoed in the handwriting research as a central tenet of the work suggests that understanding the cognitive computational base of handwriting is an essential pre-cursor to successful recognition. This paper is addresses one aspect of low-level vision and its interaction with cognitive processes.

The prime contractor for this project is the Department of Communications, Computer, and System Science, University of Genoa, and there are two industrial partners and two academic partners. These are, respectively, Video Display Systems Ltd., Florence, Computer Applied Techniques Ltd., Dublin, Department of Experimental Psychology, Catholic University of Nijmegen, and Department of Computer Science, Trinity College, Dublin.

1.2 RATIONALE

Perception facilitates the interaction of a person with his world through the construction of a useful "mental model" of the local environment. It involves a robust ability to assimilate information and to interpret it in a manner which is valid, coherent, and consistent with both past experience and expectations of future occurrences. Perception may involve information presented in several ways (visual, cutaneous, kinesthetic, aural) and much of this information is inter-dependent in content and presentation. It also involves geometeric reasoning, in particular the ability to deal with three-dimensional representations, as well as cognitive skills, such as the formation and validation of hypotheses, reasoning, and an active seeking of stable consistency between interpretation and actuality.

We see, then, perception as being more than an instantanious information processing task, but also one in which the explicit and concurrent flow of information between processes, and the perceptual organisation which facilitates that flow, is equally important. Further, perception and motor-skills are intrinsically linked and complementary.

The pursuit of image and movement understanding (which may be viewed as a limited case of the study of anthropomorphic perceptual-motor- skills) necessitates the development of tools which may facilitate the integration and interaction of perceptual information, i.e. sensory data, and in addition, their presentation and interaction with cognitive capabilities. As far as possible, these tools should not prejudice the research philosophy or any investigation of computational theories of perception. This paper presents the fruit of the first eighteen months of work in one area of project IMU-P419 and addresses the generation of a tool for integration of perceptual data in this manner.

2. THE VIRTUAL IMAGE SYSTEM

VIS was conceived as a leading tool in task #1 of P419 which investigates low-level, or "early", visual processes. In particular, the system was to incorporate features to enable integration of distinct low-level visual cues and to allow interactive investigation of the characteristics of such integration. As the approach to low-level vision taken in this project is much in sympathy with that of David Marr and his associates at M.I.T.[1,2], many of these cues are based on object boundaries, stereopsis, and motion derived from the zero-crossings of the Laplacian of an image which has been convolved with a Gaussian mask (for example, see diagram 2). Thus the system was developed with reference to the subject matter of low-vision while allowing more global design criteria to be finalised as the project proceeded. The experience gained so far with the system meets these requirements.

This tool for integration of visual information has been dubbed a Virtual Image System (VIS) as it allows a user to build an "image system", i.e. a hierarchical structure of representations where the computational transformations are implied by the "transfer" from one representation to another. This hierarchical structure is generically referred to as a Virtual Image Structure.

The global Virtual Image Structure managed by VIS may contain multiple component Image Structures: each one is identified by such parameters as the time instant (for an image sequence), by the left/right specification of stereo pair, etc.. Image Structures are organised into one or more colour channels according to different color coding schemes (RGB, HIS, HSV etc). Each colour channel may be organized into different spatial frequency channels using a pyramidal structure*, with image resolution varying from 1024x1024 to 64x64. The Virtual Image Structure is simply a hierarchy of these pyramids. Moreover, each pyramid can store three levels of representations: iconic, regional, and boundary [3,4,5].

Each Virtual Image Structure hierarchy may comprise many iconic representations: framestores, intensity images, convolution images, etc. Lower in the hierarchy, the iconic representation of a channel may be transformed into a tree-structured regional representation which makes explicit the topology of "regions" of uniform convolution sign. From the regional representations it is possible to build, further down the hierarchy, several contour representations which make explicit, for the boundary of a region such features as the zero-crossing slope and direction, boundary motion, stereo disparity, etc.. Diagram 1 depicts the top-level conceptual organistion of a typical system, configured for the analysis of scenes derived from a single colour channel.

Image representations in computational domains, different from the cartesian plane, are also being considered: e.g. the log-polar plane is well suited to scale and rotation invariant filtering and recognition [6] while the spatial frequency space is basic for Fourier-based processing.

*A pyramidal structure is a set of iconic representations of the same image with variable resolution. The "pyramid" is a metaphor of such arrangement - a sort of stack of icons.

A Visual Image Structure can be configured so that it comprises as many pyramids as necessary and representations may be added "with" selective parameters and selective windows, which are specified dynamically, at run time. This feature, together with the possibility of deleting representations, provides the basic framework for a cognitively tunable (or "penetrable", according to Pylyshin, 1984) early perceptual processing - a computational formulation of "attention".

The Virtual Image System (VIS) uses menu-based user/system dialogue and a facility is provided which allows a user to teach the system a sequence of menu selections which may be saved on file and recalled later for replay. Several such files may be generated and a directory of these "command" files is available on request. All menu slections are suported by on-line texts. Further, the system is designed in such a way as to enable a processing and analysis session to be suspended and resumed at a later date, saving in a file the current virtual image structure.

A data-flow paradigm was adopted to allow the user to manipulate images merely by specifying source and destination representation, the types of these representations defining an implicit transformation. For example, if the source image is an intensity image type and the destination is a convolution image type (i.e. one that has been convolved with a Laplacian of Gaussian mask) then the appropriate transformation is effected implicitly by the system as part of the transfer, prompting the user for auxiliary information, as appropriate. In addition, the system allows extensive windowing in both source and destination images so that, in addition to being transformed, the destination image may be a scaled and translated version of the source window.

The remainder of this text deals in a little more detail with the menu system and user dialogue, the virtual image system data-structures, system processing functions, examples of preliminary applications, and the system extensions to cognitive reasoning and cursive script analysis, respectively. The presentation is mainly at the level of conceptual specification.

3. VIRTUAL IMAGE STRUCTURES

As mentioned in the introduction, VIS is designed to allow dynamic, i.e. run-time, configuration of Virtual Image Structures (VISs) used in any one processing session. Thus, a user can build a virtual image structure from scratch, beginning with an empty structure and subsequently adding, arborising, and/or deleting representations as desired and as the situation demands.

Formally, VISs can be defined, using a BNF-like notation, as follows:

<Virtual Image Structure> ::= {<Image Structure>}*
<Image Structure> ::= {<Colour Channel>}*
<Colour Channel> ::= {<Spatial Resolution Channel>}*
<Spatial Resolution Channel> ::= <iconic rep> | <iconic rep> <regional tree>

It should be noted that sequence of BNF syntactic constructs, e.g.
<...> <...> <...>, does not necessarily imply temporal or spatial ordering (as it would in the linguistic formulation) but, more importantly, that the constructs are mutually associated, i.e. logically related. A Virtual Image Structure may contain several Image Structures, each of them being identified by a specification of time parameters, spatial parameters (coordinates of the camera, identification of right/left camera etc.), and optical parameters. An Image Structure is possibly split into colour channels and each one may be divided into several spatial frequency channels. The spatial frequency channels are organized in a pyramidal structure (i.e. using a multiple resolution scheme).

Each pyramid, in turn, contains different types of iconic representations (organised as 2D arrays), whose topological arrangement can be made explicit by means of a tree-like structure or regional representations:

<iconic rep> ::= <intensity image> |
<intensity img> <convolution img> |
<intensity img> <convolution img> <region crossing img>
<regional tree> ::= tree of <regional representations>
<regional representation> ::= <regional descriptor> |
<regional descriptor> <contour representation>

where the <contour representation> is a rich specification of the shape of region boundaries (stored by means of boundary chain codes - BCCs) and of the distribution along these boundaries of several image features. The result is a set of "profiles": slope profile, orientation profile, velocity profile, disparity profile, etc.. The slope profile stores the slope, or steepness, of the zero-crossing of the convolution (a measure of "edge strength"); the orientation profile stores the local orientation of the contour (useful, for example, to detect straight segments on a contour); the disparity profile stores the result of stereo-matching; the velocity profile gives a measure of optic flow.

<contour representation>::= <contour descriptor> <BCC> <feature profiles>
<feature profiles> ::= <slope profile> and/or
<orientation profile> and/or
<velocity profile> and/or
<disparity profile> ...

Let us now comment on some aspects of the Virtual Image Structure outlined above. Since VISs are tree-structured, it is important to store a number of descriptors at the different levels of the hierarchy, in such a way that deeper levels inherit the descriptions of the higher levels, so that each representation is coherently integrated and so that the structure can be usefully traversed.

We said already about the descriptor at the Image Structure level which stores "environmental-proprioceptive-optic" information (currently stored into the framestore descriptor). For example, the "pyramid descriptor" which details the exact nature and structure of its constituent representations. In particular, a pyramid descriptor will point to several "descriptors" each of which details the exact make-up of the representation at that level: size, level number, the number of bits, and the associated mask. In addition, each descriptor will point to a structure which is appropriate to the type of that representation, for example, a descriptor of a grey-level image would be linked to a 2-D array of bytes representing that image. Alternatively, a descriptor of a framestore will be linked to a framestore descriptor which, in turn, indicates the global structure detailing the operational characteristics of a particular framestore type. Finally a descriptor of a contour would be linked to a BCC representation of a specific feature of the contour (e.g. slope).

Let us consider each conceptual level in the VIS hierarchy, beginning with iconic representations. A framestore is more of a "device" than a representation. It is meant to interface VIS in an explicit way with different types of frame grabbers. After a given frame has been grabbed, then such iconic information is distributed to the different colour channels and pyramids, according to the "virtual" philosophy of VIS, as multiple intensity images. From each of them a convolution image can be obtained whose zero-crossings are the bases of the Marr formulation of early visual processing. The nature of the convolution process determines a nested structure of the convolution images where, recursively, regions of positive/negative convolution may contain one of more regions with the opposite sign, their boundaries being marked by the zero-crossing contours. VIS makes explicit this kind of structure by means of a particular iconic representation - the region-crossing image - and by means of the tree-structured regional representation. The latter is a tree-structure of regional descriptors and the former is an image of pointers, the pixels of the same region pointing to the same descriptor. This kind of arrangement is convenient from the computational point of view, because the natural way to extract the regions is to "fill" them with a uniform code, but also from the representation point of view because it establishes a powerful relational link between the iconic representations and the tree-structured regional representations.

Region descriptors contain gross regional statistics regarding a single region-crossing image and augment it with information pertinent to the corresponding contour representation of that same region. This contour information is exactly the information contained in the contour descriptors.

In addition, the region descripotors comprises two sets of link fields; the links of the first set point to the corresponding contour in the slope, orientation , velocity, and disparity profiles and the links of the second set may be used in a general purpose and open-ended way, relating to "structured region descriptors" which may embody perceptual concepts from developing theories of visual perception.

Pyramid descriptors comprise a header node, containing the pyramid identification number and an alpha-numeric description, a link to the next pyramid in the system, and a linked list of pointers to image descriptors.

Image descriptors comprise several fields which include an internal data-type label, an alpha-numeric description, the image size, the pyramid level number, a window specification, a bit-plane mask, the number of bits in the image, a pointer to a framestore descriptor, a pointer to the actual image, and a pointer to the parent pyramid descriptor. The internal data-type is simply an integer value which identifies the type of image referred to by the descriptor. The image size is given by the number of rows (or columns) in the image; it is assumed that the number of rows is equal to the number of columns and, further, that this number is an integer dividend of 1024. The level number is, in fact, the corresponding integer divisor: thus, an image of size 512x512 pixels would be a level 2 image since 512 = 1024/2. The window specification is determined by coordinates of the top left-hand corner and the bottom right-hand corner of a rectangle; only that section of an image enclosed by this rectangular window is the subject of system processing and analysis functions. The window concept - which is a way to represent attention - may be generalized to any contour in the image plane, possibly one of the contours of the regional representation.

Contour descriptors attempt to integrate the information given by several feature profiles related to the same contour. They use two sets of fields. The first one contains explicit links to the corresponding contour in each contour-based representation while the second set contains information regarding gross contour statistics, for example, coordinates of the contour origin, contour length, enclosed area, and a measure of the variation in local orientation. Further, there are two additional fields which provide links to the previous contour descriptor and the next contour descriptor. Thus, the descriptors themselves are organised as a linear list and it becomes possible to run these lists of descriptors searching for, say, all contours whose gross statistics satisfy some criterion and then refer to them explicitly in the appropriate image.

Two more general issues can be addressed, as regards VISs: one is redundancy and the other is registration. VISs are certainly redundant: iconic, region, and contour representations express the same information. However, redundancy is the price to be paid for making structure explicit and so interfacing early processing with geometric modelling and cognitive processing. Moreover, the hierarchy of representations stored in VISs is organised in such a way to keep them registered, i.e. to preserve their spatial coherence.

4. MENU-BASED USER-INTERFACE.

Dialogue between a user and the Virtual Image System (VIS) is effected, wherever possible, by the use of self-explanatory menus. This ensures that the system is easy to use and that fast and efficient interaction is possible. When menu-based interaction is not appropriate, the system will invoke a question-and-answer session with the user, requesting some alphanumeric reply. In general the user will have the option of aborting this dialogue via a "quit" option.

Every menu comprises between one and eight options, shown as an option number and explanatory text*. The menu option is selected by depressing a numeric key corresponding to the option number and subsequently depressing the "enter" key. In general, all common menu options are assigned the same option number. This is intended to utilise the natural conceptual association

*All screen handling primitives assume that the terminal device is VT100 compatible: this affords a relatively efficient mechanism for cursor manipulation in a standard, and portable manner.

which a frequent user will form between the menu function and the option number. Some general-purpose utilities, intended to increase the efficiency of usage of the system, are provided. These include the following.

A LEARN menu option provides the facility to generate a file containing several sequential menu selections which can then be used later to re-generate (and re-invoke) this set of selections via the REPLAY menu option. Several such command files may be generated and stored on the system: the DIRECTORY menu option enables the user to display a directory of the command files which were generated using the Learn facility.

Since an integral feature of VIS is the dynamic configuration of image structures during any interactive session, it is essential that a synopsis of the current status is available opon request. This is achieved by selecting the SYSTEM STATUS option, which scans through the Virtual Image Structure, printing a summary of the status of each representation at any level of each pyramid. The entire hierarchy of pyramids (of spatial resolution channels) can be traversed in pre-order, in-order, or in post-order. Moreover, at all stages the user is afforded the opportunity to skip to the next pyramid or terminate the summary altogether.

As VIS is an interactive system for investigating the form and integration of various perceptual cues, it is mandatory that the researcher be able to suspend his/her investigations and experiments and to resume them at a later date. This facility is provided through two menu options, SAVE and RESTORE, which allow the user to save on a named file the current status of the system, i.e. a virtual image structure, and subsequently restore it (or any other saved status).

Finally, on-line assistance is available for each menu option in the form of 'HELP' texts. This information may be obtained by typing '9' followed by the menu number and <return>. '9' or '99' <return> will invoke the display of a general help text detailing this procedure for obtaining option-specific assistance.

5. IMAGE PROCESSING PRIMITIVES AND SYSTEM FUNCTIONS

As already mentioned, a data-flow paradigm is used for image processing: the user manipulates representations merely by specifying source and destination representations and an implicit transformation is invoked on the basis of the types of these source and destination representations. Since VIS facilitates extensive windowing in both source and destination images, the destination image, in addition to being transformed, may be a scaled and translated version of the source window.

Processing and analysis functions are available to the user as menu options. In particular, the set of functions which are currently available is described in the following.

(i) Framestore Initialisation.

(ii) Real-time acquisition and display.

This option provides the user with the ability to cause an incoming video signal (from a camera) to be continuously digitised and displayed, in real-time, on a video monitor. When selected, a question and answer dialogue with the user is initiated to ascertain the framestore which is to be used to accomplish the video capture. The framestore is identified by its appropriate pyramid number and level number. In general, the user may abort the operation at any point by replying to a prompt with a null entry.

(iii) Grab a frame of video.

With a semantics similar to the previous option, this option freezes a video frame in a framestore.

(iv) Transfer image.

This is the most powerful (and computationally expensive) option. It implements an interactive request to transform the contents of one level of a pyramid to a (possibly different) level of a (possibly different) pyramid. The source and destination images cannot both be at the same level of the same pyramid. If the representations associated with these levels are different, then the implicit transformation will be effected.

The spatial organisation of the destination image depends significantly on the relationship between the window associated with the source image and the window associated with the destination image and also on the hierarchical/pyramidal relationship between the two images. Thus. the destination sub-image may be an enlarged or reduced version of the source (in either the x or y directions) and is dependent on the spatial mapping dictated by the window 1 to window 2 spatial mapping and the pyramid level to level mapping. A pixel filling algorithm is used in all cases and the appropriate coordinates in the source image are determined by a two-stage process comprising spatial mapping and bi-linear interpolation. Note that, again, the user may abort the request at any point of the interactive dialogue by replying to a prompt with a null entry.

(v) Selection Mechanism

A powerful tool for guiding or aiming the transfer functions is to associate a "with" attribute to the transfer functions, i.e. a selection mechanism in which the target representation stores only those components of the source representation that comply with the constraint indicated by the "with" attribute. The specification of this mechanism is still under development, but preliminary experiences have been gained with a conjunctive specification of the following type

$$(L1 < f1 < H1) \ \& \ (L2 < f2 < H2) ... \& \ (Ln < fn < Hn)$$

where f1, f2, ... are "features" and L1, L2, ... H1, H2, ... are low and high thresholds, respectively, for the corresponding feature to pass the selection test. In this way one may select "only" the contours that have a sufficient area and sufficient energy or "all" the contours that have a given shape and color.

(vi) Windowing

Selection of this menu option allows the user to modify the window coordinates of a particular image. The modification is facilitated by question and answer dialogue between the user and the system, during which the user is prompted for the required information. The prompts are accompanied by explanatory text, and, again, the user may quit the modification at any point by replying to a prompt with a null entry. Window specification using graphic devices (e.g. mouse, digitising tablet) or self-generated by the system (as contours) are being considered.

6. VOLUMETRIC REPRESENTATIONS - VIRTUAL SOLID STRUCTURES

The representations generated by VIS (iconic, region, contour) are characterized by a common feature: they are all ego-centric, i.e. they are dependent on the position and orientation in space of the viewing camera. Perception, on the other hand, requires the capability to build eco-centric representations, i.e. representations related to the environment and therefore independent of the viewpoint. Although symbolic representations, such as semantic nets, meet such a requirement we believe that trying to infer them directly from VISs is too large a step to be accomplished in one shot, except for very special "toy worlds". More general perceptual tasks require the maturation of what, in Piagetian terms, could be called *spatial organization.*

One way to obtain a spatial organization or perception (and action) is by means of volumetric (analogic) representations of space, an extension of the concept of icon, which is ego-centric, to that of 3D-icon.

An important characteristic of icon representations (both 2D and 3D) is that they represent at the same time shapes and "free space".
From the point of view of shape representation this is obviously inefficient and indeed relational/symbolic representation have the purpose of expressing parsimoniously and explicitly only the structures of objects. However, the direct representation of spatial coherence makes 3D iconic representations extremely useful for at least two main pupsoses: (i) integration over time of the flow of VISs, (ii) formulation of reaching/navigation tasks.

As regards the former purpose, we may observe that human perceptual processes are governed by two main sampling rhythms: visual sampling (uniform and not exceeding the TV framing rate of 25 frames per second) and oculomotor sampling (nonuniform and not exceeding the rate of 3-4 saccadic eye movements per second). Between a saccadic eye movement and the next one, the brain receives a flow of smoothly variable VISs, coupled with a flow of proprioceptive information (information about ego-movements with respect to the environment). A working hypothesis that we formulate is that these two flows are used to set up and refine a stereometric representation of object shape.

In its simplest meaning, a 3D icon may be conceived as a 3D ternary array, one value indicating free voxels, one value occupied voxels and one value non-classified voxels. P419 has already studied in a preliminary way some aspect of this problem developing some algorithms which use the occluding contours for several views and the depth maps (from stereo matching) for the same views: many representational and computational aspects remain to be investigated [7].

In perspective, however, it is possible to conceive VSSs (Visual Solid Structures) which parallel the VISs discussed above: structures in which multiple iconic representations are linked with relational representations. The flow of VISs over time would operate on the VSS for the duration of an "ocular fixation" because VISs are changing faster than VSSs (a VSS "accumulates" a coherent stream of VISs within an ocular fixation time). We may also add that possibly "Cognitive Accumulators" may be conceived for integrating over space solid structures acquired for each fixation interval.

The coherence between the two types of structures (or better, between the current VSS and the last VIS) may be obtained by means of "relational icons", i.e. icons of pointers similar, in concept to the region-crossing images defined above. In particular, it is possible to conceive two types of such relational icons:

2D icons of pointers to the 3D icon voxels, or
3D icons of pointers to the 2D icon pixels.

Which one of such computational alternatives is better to choose is too early to say, however, we may already outline the interface between data driven processes (that produce VISs and VSSs) and cognitive perceptual processes that must interpret and use such representations according to a goal.

What is important is that the interface must be bidirectional: the cognitive processes not only "read" the perceptual data but must also "actively guide" the data gathering process. As regards the former aspect, a VSS and the current VIS may be considered as a data base which may be interrogated in order to obtain qualitative or quantitative answers Possibly more important (and certainly in a more primitive stage of conceptualization) is the dual aspect of interaction the other way around. Two main topics may be outlined: One is tuning and guiding the data gathering process in order to limit the complexity explosion by "focusing the attention" on a limited part of the world (this can be obtained using for example the Selection Mechanisms and Windowing Mechanisms already included in the VIS); the other is using conceptual hypotheses in order to "modify" the visual data base, either as regards 2D or 3D representations. In this way it is possible, for example, to complete contours or merge regions (in the 2D representations) or to "carve" volumes in the 3D representation on the basis of expectation or gestaltian concepts of continuity.

7. PRELIMINARY RESULTS

An integral part of project P419 is the application of the research to the so-called bin-of-parts problem in robotics, in which a robot must identify a single part in a bin of arbitrarily oriented and occluding parts, and remove it. This type of case study is useful in that it provides interesting feedback about the utility of the techniques being developed in what is widely recognised as being one of the most difficult industrial vision problems. In particular, one of the early objectives of the projects is to investigate the suitability of the Laplacian of Gaussian filter and the subsequently extracted zero-crossings for segmenting such jumbled parts.

As an example, an image of a basket of apples was filtered with Laplacian of Gaussian operators (having standard deviation of 3.0 and 6.0 respectively) and the zero-crossings extracted (see diagram 3, 4, and 5). Contour, slope, and orientation representations were built for each of these zero-crossing images and contour descriptors generated. Based on a manual inspection of these contour properties, individual contours of significant length and slope were conditionally selected. These contours, which represent the segmented objects, were then transferred to a zero-crossing image and subsequently displayed on the monitor (see diagram 6 and 7). Each "blob" corresponds very closely to individual apples and it is anticipated that robust segmentation would result from an analysis of the superposition of the contours in these two zero-crossing images.

ACKNOWLEDGEMENTS

The authors acknowledge significant contributions by G. Adorni, C. Braccini, M. di Manzo, P. Morasso, L. Schomaker, H.L.M. Teulings, A.J.W. Thomassen, in the framework of Esprit project P419 "Image and Movement Understanding".

REFERENCES

1. D. Marr, "Early processing of Visual Information", *Transaction of the Royal Society of London* B275, pp.483-524 (1976).

2. T. Poggio, "Marr's Computational Approach to Vision", *Trends in Neuroscience* 4 No. 10, pp. 258-262(1981).

3. G. Sandini and V. Torre, "Thresholding Techniques for Zero-Crossings", *Proc. "Winter 85 Topical Meeting on Machine Vision"*, Incline Village, Nevada (1985).

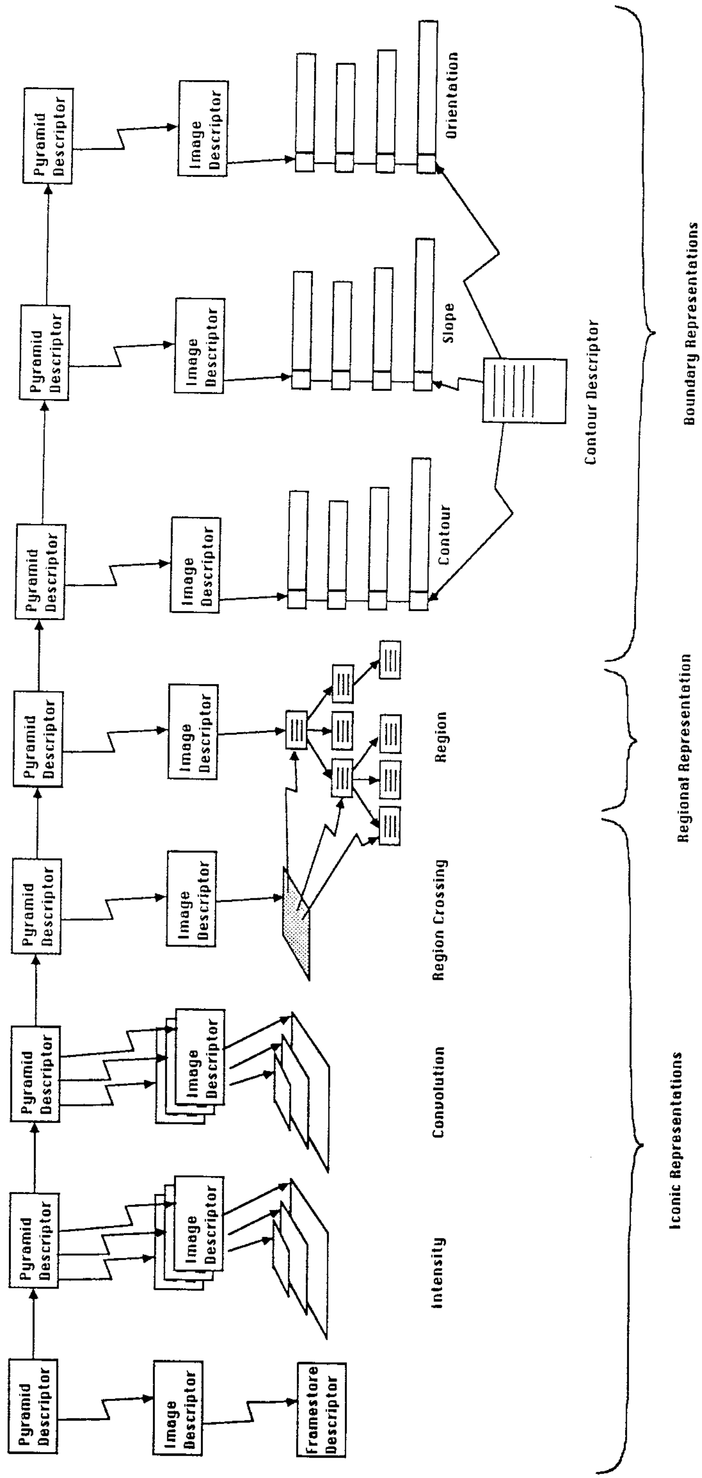
4. G. Sandini and M. Tistarelli, "Analysis of Camera Motion through Image Sequences", *Proc. Intl. Conference on "Advances in Image Processing and Pattern Recognition"* Pisa IBM, (December 10-12 1985).

5. G.Sandini, V. Tagliasco, and M. Tistarelli, "Analysis of Object Motion and Camera Motion in Real Scenes", *Proc. IEEE Intl. Conference on "Robotics & Automation"*, San Francisco IEEE-CS, (April 7-10, 1986).

6. C. Braccini, G. Gambardella, and A. Grattarola, "The use of computational spaces for 3-D object recognition", *Digital Signal Processing*, pp. 759-763 (1984).

7. P. Morasso and G. Sandini, "3D Reconstruction from Multiple Stereo Views", *Proc. III Intl. Conf. on "Image Analysis and Processing"*, (Rapallo October 1-2, 1985).

Diagram 1 - Organisation of Virtual Image Structures



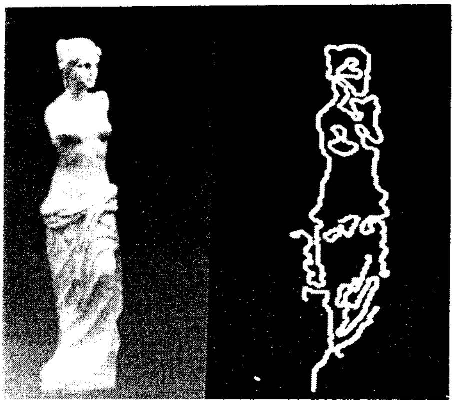


Diagram 2: Aphrodite and Zero-Crossings

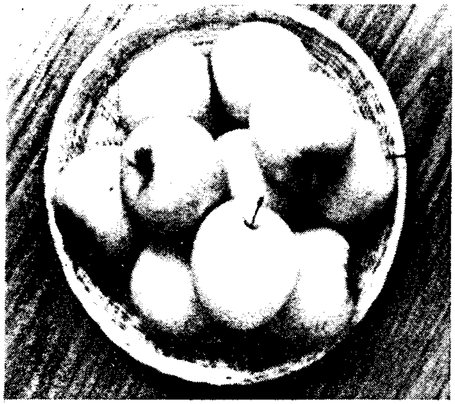


Diagram 3: Basket of Apples

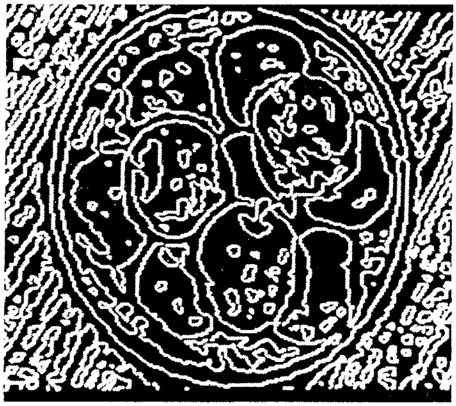


Diagram 4: Zero-Crossings (sigma = 3)

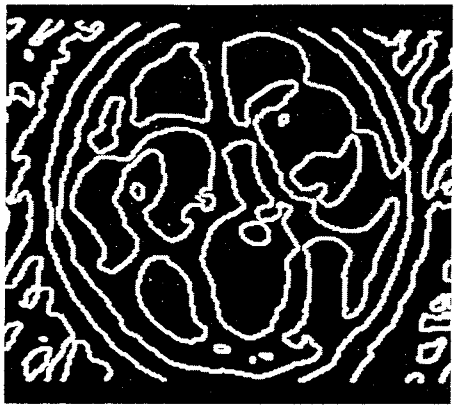


Diagram 5: Zero-Crossings (sigma = 6)

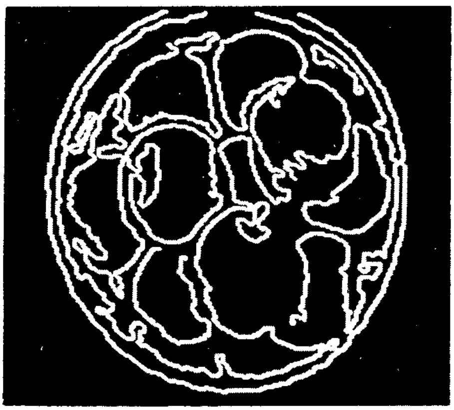


Diagram 6: Selected Zero-Crossings (sigma = 3)

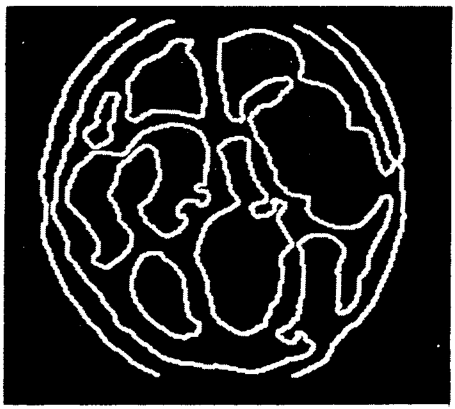


Diagram 7: Selected Zero-Crossings (sigma = 6)