Does Vision Inevitably Have to be Active?

Gösta H. Granlund
Computer Vision Laboratory
Linköping University
S-581 83 Linköping, Sweden

Abstract

It has become increasingly apparent that perception cannot be treated in isolation from the response generation in vision systems. This is first of all because it has emerged that many classical aspects of perception, such as geometry, probably do not belong to the percept domain of a vision system, but to the response domain. Secondly, it turns out that the interpretation to be generated at a given instance is as much dependent upon the state of the system, as the percepts impinging upon the system. The state of the system is in consequence the combination of the responses produced and the percepts associated with these responses. Finally, a very high degree of integration is required between different levels of percepts and corresponding response primitives. It is argued that information must be acquired actively by the system itself, through response driven association with percept transformations. After an active training, however, the system can exhibit a reactive behavior to passively observed percepts.

Keywords: Active Vision, Robotics, Learning, Object Representation

1 Introduction

In robotics we have traditionally assumed a structure according to Figure 1, where a vision system is controlling an actuator system. The systems were viewed as sophisticated perception modules, to which a few actuator wires were attached, causing the requested responses.

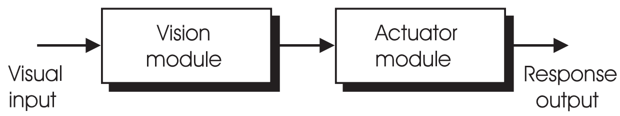

Figure 1: Simplest structure of robotics control system

A major problem in the implementation of such a system structure is that the channel between the analysis and the response generation parts is very narrow. This implies that the information available from the analysis stage is not sufficiently rich to allow the definition of a sufficiently complex response required for a complex situation.

For this reason we want to propose a different conceptual structure, which has the potential of producing more complex responses, due to a close integration between visual interpretation and response generation [4], as illustrated in Figure 2.

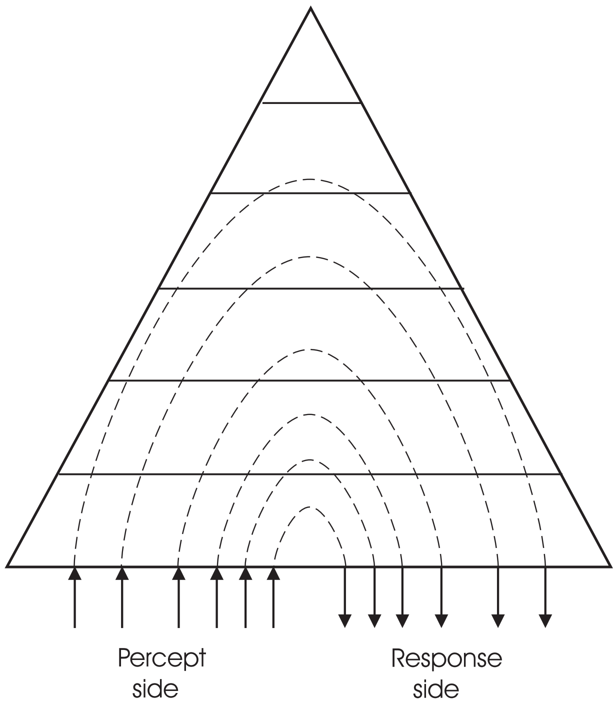

Figure 2: A stylized analysis-response structure viewed as a pyramid .

This structure is an extension of the computing structure for vision, which we have developed over the years [5]. The input information enters the system at the bottom of the processing pyramid, on the left. The interpretation of the stylized Figure 2 is that components of an input are processed through a number of levels, producing features of different levels of abstraction. These percept features of different levels, generated on the left hand side of the pyramid, are brought over onto the right hand side, where they are assembled into responses, which propagate downward, and ultimately emerge at the bottom on the right hand side. A response initiative is likely to emerge at a high level, from where it progresses downward, through stages of step-by-step definition. This is illustrated intuitively as percepts being processed and combined until they are "reflected" back and turned into emerging responses.

The number of levels involved in the generation of a response will depend on the type of stimulus input as well as of the particular input. In a comparison with biological systems, a short reflex arch from input to response may correspond to a skin touch sensor, which will act over interneurons in the spinal cord. A complex visual input may involve processing in several levels of the processing pyramid, equivalent to an involvement of the visual cortex in biological systems.

A characteristic feature of this structure is that the output produced from the system leaves the pyramid at the same lowest level as the input. This arrangement has particular reasons. We believe that processing on the percept side going upward in the pyramid, usually contains differentiating operations upon data which is a mixture between input space and property space. This means that variables in the hierarchical structure may not correspond to anything which we recognise at our own conscious level as objects or events. In the generation of responses on the right hand side, information of some such abstract form is propagated downward, usually through integrating operations. Only as the emerging responses reach the interface of the system to the external world, do they have a form which is in terms of objects as we know them. In conclusion, this is the only level at which external phenomena make sense to the system; be it input or output.

This mechanism has far-reaching consequences concerning programming versus learning for intelligent systems. If we try to "push" information directly into the system at a higher level, it must have the correct representation for this particular level, or it will be incomprehensible to the system. A more serious problem, which we will deal with later, is that new information will have to be related to old information, on terms set by the system and organized by the system. It will require the establishment of all attribute links and contextual links, which in fact define the meaning of the introduced item. It is likely that information can only be input to a system through the ordinary channels at the lowest level of a feature hierarchy system. Otherwise it cannot be recognized and organized in association with responses and other contextual attributes, which makes it usable for the system.

In biological systems, there appear to be levels of abstraction in the response generation system as well, such that responses are built up in steps over a number of levels [14, 15, 1, 8]. Arguments can be made for the advantage of fragmentation of response generation models, to allow the models to be shared between different response modes.

1.1 Organization of Higher Level Processing

It seems that there has to be two tiers of organisation within an effective computation structure for spatial information, see Figure 3. Within the lower tier there is an organisation of data in relation to external geometry. This is also true for biological systems, where low level orientation description is mapped upon the cortex in accordance with position in the visual field. Similarly for motor functions and other features, which are mapped correspondingly between the body and the cortex. For technical systems, it can be assumed that computations to produce these low level features can be made in parallel, and that influences on earlier levels of computation are at least very local.

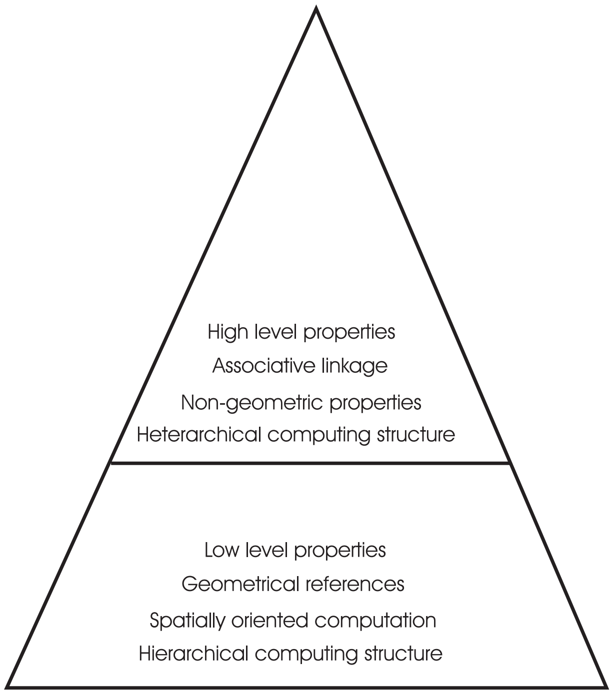

Figure 3: The two-tier pyramid

It is postulated that at some level of abstraction, local geometrical relations become less important, and other non-spatial and non-local relations become essential. The separation into distinctive paths for WHAT and WHERE information is one indication of this in biological systems. This forms the upper tier of the computation structure for spatial information, see Figure 3. In this part, the computation structure has to be formed through association based upon properties of the signals themselves as they are driven by stimuli, rather than by geometric adjacency. We will in this paper deal with issues concerning computation structures for the upper tier.

1.2 A Few Definitions

We will use the following terminology:

By *selfgenerated action* we denote an activity of a system which is selfgenerated without any apparent external influence. The action is assumed to be internally produced in some way, often through a random noise signal creating an activity in parts of the system.

By *reaction* we mean that the system performs an activity which is initiated or modified by a set of percepts.

It should be noted that a particular action or reaction under consideration may be composed by both selfgenerated active components and reactive components. A selfgenerated high level action may well contain reactive subcomponents available from the system's earlier experience. The full implication of this will not be apparent until we have dealt with the combination of established percept-action procedures and random exploratory influences.

The structure which we will see developed, has two major modes of operation:

- **Exploratory behavior**: Selfgenerated action → Percept → Association
- **Reactive behavior**: Percept → Reaction

A selfgenerated action by the system causes a particular set of percepts to appear. The actions and percepts are linked to each other in an association process.

2 Responses versus Objects – The Curse of Generalization

There has traditionally been a belief that procedures would be simpler if an abstraction into objects is performed as an intermediary step, before a response is synthesized. A great deal of the robotics field has been devoted to the generation of such generalized descriptions [6]. The classical model for image analysis in robotics has been as follows:

1. Identify what an object *is*
2. Deduce from the object description what action to perform upon the object

It has become increasingly apparent that the task of a reactive robotics system is at the fundamental level to know what actions to perform as a particular scene or object appears, not to point out an abstract symbol representing the object. We are now convinced that the opposite approach is advantageous, and that the proper sequence of order should be:

1. Derive the action to perform upon the object from percepts
2. Generalize from the actions what an object *is*, if needed

It is again emphasized that the primary result from processing is the production of a proper, context sensitive *action*. The characteristic of objects is to a reactive system, the way in which it can handle them. This is what requires a highly integrated structure between the input and the output sides. We can never deduce from an object description what action can or should be performed with respect to the object.

A description of object class, or any other more generalized representation, can be produced for the purpose of *communication*, if required. This requires a second stage, implementing the removal of certain context specific links, but also adding other types of contextual information, which allows the information to be interpreted after communication.

The preceding is apparent also from an evolutionary point of view, in that simple organisms have built up percept-action structures, long before they have developed any structures for generalization for communication between individuals. The first level of communication between systems is also unintentional. A system can manipulate, observe and interpret the appearance and the behavior of another system, just like any other object. From a real, as well as operational point of view, it does not know the difference. The generation of intentional messages, which operationally is no different from other types of action, will follow as soon as the interpretational capability reaches a sufficient level.

The basic element is however action and this can be emphasized as:

> *What has hampered the development of vision over the years is the erroneous preconceived notion that what the system shall do is to describe what something* is*, rather than what it should* do*. This is a case where consciousness and language fool us. The earlier notion is fully correct however, if we interpret it in the fashion that what something* is*, is defined by what we can do with it, not e.g. its geometrical properties.*

Another common misunderstanding by people in robotics and control is that this is an issue of estimation and parameterization of a set of signals to obtain a description. It is instead an issue of association between a set of actions and the related set of percepts. In that process, however, estimation methods can be employed.

A particular response output node will be common to objects which should evoke the same response domain or component. This means that convergence is not to an abstraction of object, but includes a wide class of objects, characterized by the fact that all elements in the set associate to the same response. Convergence is to a response output node. For something which we conventionally denote an object, there will be several such response nodes or object response properties which are activated, the combination of which we can say are the characteristics of an object.

The difference between two objects is in consequence the distinction between how we can and do handle the objects. If they should be handled equivalently, there is in our terms no difference between the objects. In addition, properties of an object upon which we can not act, are not of interest and associated features are only a load on the structure.

In order for us to be able to represent an object sufficiently well with the associated responses, as postulated, it is apparent that we need a much richer response repertoir with which we can associate an object. This richer repertoir of actions or responses will have to involve different levels of the structure and employ different modes of manipulation as well as context.

3 Response as the Organizing Mechanism for Behavior

A vision system receives a continuous barrage of input signals. It is clear that the system cannot attempt to relate every signal to every other signal. What properties make it possible to select a suitable subset for inclusion to an effective linkage structure?

It has become apparent that responses and effects thereof act as organizing mechanisms for percepts, rather than the opposite, that percepts would somehow converge onto appropriate response outputs. The reason for this is that signal structure and complexity is considerably simpler in the response domain than in the percept domain, and this fact can be used as a focusing entity on the linkage process, where the system's own responses act as organizing signals for the processing of the input.

There is a classical experiment by Held and Hein, which elegantly supports this model [7]. See Figure 4. In the experiment, two newborn kittens are placed in each of two baskets, which are hanging in a "carousel" apparatus, such that they are tied together to couple the movements of the kittens. One of the kittens can reach the floor with its legs, and move the assembly, while the other one does not reach the floor and is passively towed along. After some period of time, the kitten which can control its movements develops normal sensory-motor coordination, while the kitten which is passively following the movements fails to do so until being freed for several days. The actively moving animal experiences changing visual stimuli as a result of its own movements. The passive animal experiences the same stimulation, but this is not the result of self-generated movements.

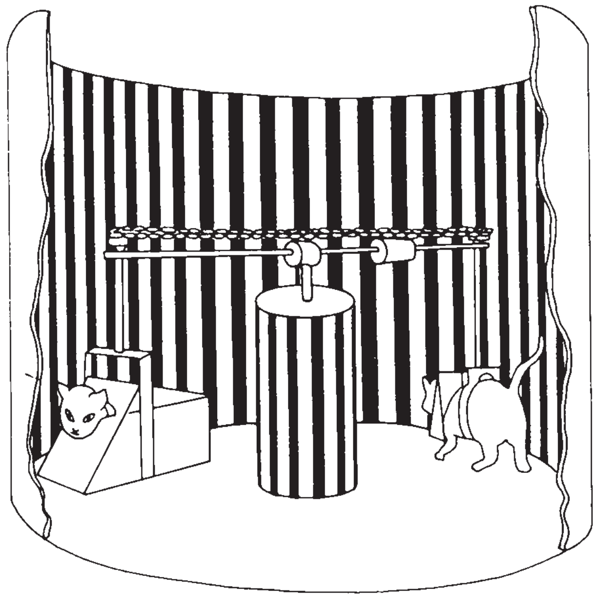

Figure 4: Experiment on active and passive perception

It is apparent that there is no basis for any estimation of importance or "meaning" of percepts locally in a network, but that "blind and functional rules" have to be at work to produce what is a synergic, effective mechanism. One of these basic rules is undoubtedly to register how percepts are associated with responses, and the consequences of these. This seems at first like a very limited repertoir, which could not possibly give the rich behavior necessary for intelligent systems. There is a traditional belief that percepts are in some way "understood" in a system, after which suitable responses are devised. This does however require simple units to have an ability of "understanding", which is not a reasonable demand upon local structures. This is a consequence of the luxury of our own capability of consciousness and verbal logical thinking; something which is not available in systems we are trying to devise and in fact a capability which may lead us astray in our search for fundamental principles. Rather, we have to look for simple and robust rules, which can be compounded into sufficient complexity to deal with complex problems in a "blind" but effective way.

Driving the system using response signals has two important functions:

- *To simplify, learn and organize the knowledge about the external world in the form of a linked network.*
- *To provide action outputs from the network generated*

.

It is necesssary that the network structure generated has an output to allow activation of other structures outside the network. This output is implemented by the linkage to response signals, which are associated with the emergence of the invariance class. If no such association were made, the network in question would have no output and consequently no meaning to the structure outside.

In further detail, we can find some major requirements for organization:

1. A response or response equivalent signal has to be available, for three different reasons:
    - The first reason is to provide an indication of motive; to ascertain that there are responses which are associated to this percept in the process of learning.
    - The second reason is to provide a limitation to the number of dependencies and links which have to be established.
    - The third reason is to provide an output path to establish the existence of this percept structure. Without a response output path, it remains an anonymous mode unable to act into the external world.
2. Inputs must be sufficiently close in the input space where they originate, the property space where they are mapped and/or in time-space. This is both an abstract and a practical computational requirement: It is not feasible to relate events over too large a distance of the space considered. This puts a requirement upon the maps of features available, namely the requirement of *locality*.

From the preceding we postulate that:

> *The function of a response or a response aggregate within an equivalence class is to produce a set of inputs on its sensors, which similarly can be assumed to belong to a common equivalence class, and consequently can be linked.*

In consequence we propose an even more important postulate:

> *Related points in the response domain exhibit a much larger continuity, simplicity and closeness than related points in the input domain. For that reason, the organisation process has to be driven by the response domain signals.*

Driving a learning system using response signals for organization, is a well known function from biology. Many low level creatures have built in noise generators, which generate muscle twitches at an early stage of development, in order to organize the sensorial inputs of the nervous system. More generally, it is believed that noise and spontaneously generated neural activity is an important component to enable organization and coordinated behavior of organisms [9].

So far the discussion may have implied that we would have a sharp division between a percept side and a response side in the structure. This is certainly not the case. There will be a continuous mixture of percept and response components to various degrees in the structure. We will for that purpose define the notion of *percept equivalent* and *response equivalent.* A response equivalent signal may emerge from a fairly complex network structure, which itself comprises a combination of percept and response components to various degree. At low levels it may be an actual response muscle actuation signal which matches or complements the low level percept signal. At higher levels, the response complement will not be a simple muscle signal, but a very complex structure, which takes into account several response primitives in a particular sequence, as well as modifying percepts. The designation implies a complementary signal to match the percept signal at various levels. Such a complex response complement, which is in effect equivalent to the system state, is also what we refer to as *context.*

A response complement also has the property that an activation of it may *not necessarily* produce a response at the time, but rather an activation of particular substructures which will be necessary for the continued processing. It is also involved in knowledge acquisition and prediction, where it may not produce any output.

There are other important issues of learning such as representation of purpose, reinforcement learning, distribution of rewards, evolutionary components of learning, etc, which are important and relevant but have to be omitted in this discussion [10, 11, 12, 13].

3.1 Response Driven Learning, Generating Association Spaces

We have earlier stated that the learning of the external world is done through an interactive exploration of it. An apparent question is: In order for the system to be able to deal with objects, is it necessary for it to interact with every single object in every respect?

It is postulated that a response driven learning session will define a set of association spaces. The system can then be expected to deal with objects or cases which are *sufficiently similar* to what it has experienced in the response driven learning process, through interpolation and extrapolation within the defined spaces. Exactly what *sufficiently similar* implies is not clear at this point. Abstractly, it must imply that an earlier defined percept-response space is valid for the phenomenon under consideration, although it may be parametrically different from what the learning trajectory defined. In some sense, it must imply that the problem structure or topology is similar.

It is as well postulated that humans are subject to the same limitations and possibilities. We can comprehend a phenomenon which we have not experienced before, as long as it is sufficiently similar to something of which we have an interactive experience; i.e we can deal with it as an interpolation or extrapolation within an already available percept-response space. It seems likely that at an adult age, most of the percept-response spaces used are already available, and that most of our knowledge acquisition after this deals with the implications of particular instances as interpolations or extrapolations within these avilable spaces. This can conceivably imply that cases with the same problem structure or percept-response space, may well appear very different, but we can handle them without considerable difficulty.

Similarly, this gives us the possibility to comprehend a phenomenon described to us in language or as a passively observed imagery, although we do not in any way interact with the phenomenon.

The visual characteristics of an object are consequently the way its appearance behaves under different modes of responses. These are the different percept-response invariants referred to in [3]. The crucial issue is that these invariants are determined by the response modes of the observing system. The associative learning system will create these invariants to "match" the response effects. As a consequence, the ability of the system to observe the world is given by its ability to manipulate it. The preceding leads us again to the basic principle:

> *A system's ability to interpret objects and the external world is dependent upon its ability to flexibly interact with it*

3.2 Characterization of Single Objects

From the preceding it is apparent that we require a relatively rich response structure associated with an object. Otherwise, we have no basis for a discrimination between objects, as well as no experience basis for sufficiently flexible actions with respect to the object in question.

The pattern of associated response actions of a particular object will be descriptive for the actual object, much like a set of features. The important difference is that while features generally describe properties in the percept domain, the current description is in terms of possible action states, or in consequence of earlier discourse, a *semantic* description [3]. A sufficient descriptiveness requires that actions are associated with essentially what we would conventionally term descriptive features. In the current system, rather than a description of features, we get a description of the actions related to the features.

4 Definition of Objects Versus Parts

We have in the preceding parts talked about the notion of object without strictly defining it. We can as a first approximation see an object as any discernible entity, which can be separated from the background or other objects. An object can be very simple or very complex, ranging from a line to complex composite objects of various types.

How are entire objects represented in relation to their parts? We can reasonably well see that simple primitives such as lines and edges of an object are represented as such. However, as these are combined into an object, do these combinations form an object identity of its own? Although this seems obvious there is a fundamental dilemma:

- On one hand it seems necessary that parts are combined into a closed representation
- On the other hand we know that a closed representation forced upon the system does not in itself solve our problems, as we would need a "more intelligent" system shell which looks at the structure to form an action. We know that rather we need a distributed representation in such a way that parts can themselves act into the environment

For our purpose we will give an operational definition of an object, which is consistent with a system's fragmented local comprehension of the world:

> *An entity is viewed as an object which is separate from other objects or the background, if there is an action which has a separate influence upon the particular object in relation to other objects or the background*
>
> *In earlier terms, this is a local percept-action invariance structure*

This means that if the system can perform some action which makes certain parts of the visual field behave differently to other parts, then the system will consider this a separate object. This implies that another system may well get a different division into objects. One system may pass a shelf of books, viewing this as one object, while a more educated system may know that it can pull out one of the items in the shelf, which is the object book.

We must remember that, a priori, the system has not a clue for the division into separate objects, unless it has already experienced them; the only possibility is to interact and explore.

This leads to the situation that there will be several levels of object parts, which are individually distinguishable, but in turn are parts of more complex objects, etc, forming a hierarchy of several levels. An action upon an object, which in turn acts upon another object will appear to the system as a single object, although maybe a complex, composite object with parts linked in ways which may not be fully established. Again, the system has no alternative for interpretation from its myopic view, and it may well make mistakes. From human psychology it is well known how accidental coincidences between events can lead to erroneous associations and grave misunderstandings about how the world is related, and to what is termed *superstitious behavior*.

It appears that while responses *probably* do not structure in levels, the invariances certainly do. Responses of the same type can penetrate into invariances of different levels, and thereby have different effects or meaning. On the other hand, it appears that responses have different time constants and scales. It is possible, however, that response complexes may be associated with an invariance and consequently assume a level.

5 Sequential Versus Simultaneous Interpretation of Features

The purpose of a vision system is to produce appropriate responses. Given classical computer vision methodology, we might assume that the standard way to do this would be to consider all percepts within a particular frame, more or less in parallel, and act accordingly. A complication of this conceptual structure is that all required information will never be simultaneously available in a single view of the system. This would require that our sensor area is sufficiently large and with a large enough resolution to provide all features of an object. Even for a single view or frame of the image information input to the system, the interpretation will require various focus-of-attention and eye movement mechanisms. Even worse, not all of the information required for a response is available in the image input at some given time, but it resides in the system state and its history; something which is commonly denoted context. This means that something which we view as an object, having a complex structure of linked invariances, will as well have a complex structure of links outwards to the environment representing its context. This part of the linkage structure is as important to incorporate for an effective vision system.

This seems to be a general problem which any flexible learning system (including ourselves) will have to resolve. Actions are not solely dependent upon what is available at the percept side of the structure, but contextual information as well. It also includes the problem of convergence of several sequential response paths to a coordinated, combined result. We will see how this can be resolved using an incremental, sequential learning of invariances at different levels, as the system moves through its environment. The sequential acquisition is what makes a linkage structure up and down feasible. It could never be done using a truly parallel acquisition. It should be emphasized though that while the acquisition of information on one hand tends to be a sequential process due to the system's physically limited trajectory through the world, the processing may well be implemented in structures with a high degree of parallellism.

6 Characterization of Multiple Objects

We have seen earlier how percept-response associations can be used to keep track of the appearance of a single object under different transformations. How can we now handle multiple objects which are different, and keep track of the distinctions? According to the dogma, we have to have some continuous phenomenon which can lead us between different instances of objects.

If we were going along with conventional computer vision wisdom, we would try to find ways to go continuously from one object to another in the percept space. This is in fact what is underlying the work by Biederman[2] and others. Different objects are seen as parametric modifications of generic shapes.

It now turns out that it is not in the object percept space we can obtain a continuous transition, because there are really no physical or mass restrictions which demand this. The restrictions appear only as we try to manipulate an object:

> *We can never find continuity in percept space,*
> *but we can find continuity in response space*

There is probably no reason to expect or require continuity among object properties or any aspect of the perceptual world, except for some constancies. The provision for constancy does not reside in the perceptual domain, but in the response domain. It may, however, in certain cases map in a more simpler way to the percept domain, which makes us interpret it as a perceptual constancy. A continuous transition between different object classes in the percept domain is not any "natural process", unless we include a physical deformation which is caused by the observer. We can however hardly imagine such a process in practical situations. This is equivalent to finding a *metric* in the object representation space.

How can we find a continuity in response space, which links us to different objects? We ask ourselves what continuity is in the response space? It implies that a transition is related to a physical power/mass restriction, even if it is very fast such as in the case of switching the attention using an eye movement.

It is postulated that objects are associated with switches of attention in the view space. Eye movements and other mechanisms to switch the attention are appropriate mechanisms to go from one object to another. This implies that a system will "walk along", observe different objects along the route; manipulate them and learn their properties. This means that an object at the outset is related to a particular position in the system's observed space, and there are no free floating generalized objects.

A consequence of this is that we do not primarily deal with an object in its isolated generality, but an object at a particular place and a particular context. This is totally consistent with the view-based representation, where we have seen that we never represent an object as a generalization, but object-at-a-particular-view. Contextual information has to be maintained, and a generalization of this rule is:

> *Objects, phenomena, etc. are never represented in an isolated generalization, but are attached to a background or context*

The preceding emphasizes continuity in the response domain, but this is necessary *only* during response activated learning, as we have seen earlier for similar mechanisms. This implies that after the process of learning, a known object can be recognized as a static phenomenon, appearing out of context or in a different context. This is analogous to the earlier discussion, where we are able to recognize an object statically in a particular orientation, if its continuous transformations have been learned in a dynamic process.

For generalization, we can observe different levels in this dynamic-learning versus static-perception hierarchy:

- **Level 2:** Appearance of object in a particular context
- **Level 1:** Appearance of object at a particular angle

or

- **Level 2:** Object-at-a-particular-position
- **Level 1:** Object-at-a-particular-view

The identification of a particular object will in return evoke the context in which the object was learned, or the state of the system at that time. We can see that an important characterizing link of an object is the context in which the object has once been observed.

The preceding implies that we cannot simply present the system with a sequence of different objects to be learned. Objects must in the learning process be differentiated by some continuous state variable which describes in some, at least superficial, way a position, state or context to be associated with the object in the learning process. The reference will in such a case be something like: "The object we had in that position". This observation state is conceivably an important part of the object description, although a recognition can be done in other contexts.

Acknowledgements

The author wants to acknowledge the financial support of WITAS: The Wallenberg Laboratory For Information Technology and Autonomous Systems, as well as the Swedish National Board of Technical Development. Considerable credit should be given to the staff of the Computer Vision Laboratory of Linkoeping University, for discussion of the contents in this paper.

References

[1] M. F. Bear, B. W. Connors, and M. A. Paradiso. *Neuroscience. Exploring the Brain.* Williams & Wilkins, Baltimore, USA, 1996. ISBN 0–683–00488–3.

[2] I. Biederman. Human image understanding: Recent research and a theory. *Computer Vision, Graphics, and Image Processing*, 32:29–73, 1985.

[3] G. Granlund. The complexity of vision. *Signal Processing*, 74(1):101–126, April 1999. Also as report LiTH-ISY-R-2034.

[4] G. H. Granlund. Integrated analysis-response structures for robotics systems. Report LiTH–ISY–I–0932, Computer Vision Laboratory, Linköping University, Sweden, 1988.

[5] G. H. Granlund and H. Knutsson. *Signal Processing for Computer Vision.* Kluwer Academic Publishers, 1995. ISBN 0-7923-9530-1.

[6] R. M. Haralick and L. G. Shapiro. *Computer and Robot Vision, Volume II.* Addison-Wesley Publishing Company, Reading, Massachusetts, 1993.

[7] R. Held and A. Hein. Movement–produced stimulation in the development of visually guided behavior. *Journal of Comparative and Physiological Psychology*, 56(5):872–876, October 1963.

[8] I. P. Howard and B. J. Rogers. *Binocular Vision and Stereopsis*. Number 29 in Oxford Psychology Series. Oxford University Press, New York, 1995. ISBN 0–19–508476–4.

[9] L. C. Katz and C. J. Shatz. Synaptic Activity and the Construction of Cortical Circuits. *Science*, 274:1133–1138, November 15 1996.

[10] T. Landelius. Behavior Representation by Growing a Learning Tree, September 1993. Thesis No. 397, ISBN 91-7871-166-5.

[11] T. Landelius and H. Knutsson. A Dynamic Tree Structure for Incremental Reinforcement Learning of Good Behavior. Report LiTH-ISY-R-1628, Computer Vision Laboratory, SE-581 83 Linköping, Sweden, 1994.

[12] T. Landelius and H. Knutsson. Behaviorism and Reinforcement Learning. In *Proceedings, 2nd Swedish Conference on Connectionism*, pages 259–270, Skövde, March 1995.

[13] T. Landelius and H. Knutsson. Reinforcement Learning Adaptive Control and Explicit Criterion Maximization. Report LiTH-ISY-R-1829, Computer Vision Laboratory, SE-581 83 Linköping, Sweden, April 1996.

[14] R. A. Lewitt. *Physiological Psychology*. Holt, Rinehart and Winston, 1981.

[15] G. M. Shepherd. *The Synaptic Organization of the Brain*. Oxford University Press, 2nd edition, 1979.