Inductive Theory of Vision

Lev Goldfarb, Sanjay S. Deshpande, Virendra C. Bhavsar

Faculty of Computer Science
University of New Brunswick,
Fredericton, N.B., Canada E3B 5A3,
[Ph: (506)453-4566, FAX: (506)453-3566,]
E-mail: goldfarb, d23d, bhavsar@unb.ca

Abstract

In spite of the fact that some of the outstanding physiologists and neurophysiologists (e.g. Hermann von Helmholtz and Horace Barlow) insisted on the central role of inductive learning processes in vision as well as in other sensory processes, there are absolutely no (computational) theories of vision that are guided by these processes. It appears that this is mainly due to the lack of understanding of what inductive learning processes are.

We strongly believe in the central role of inductive learning processes, around which, we think, all other (intelligent) biological processes have evolved. In this paper we outline a (computational) theory of vision *completely* built around the inductive learning processes for all levels in vision. The development of such a theory became possible with the advent of the formal model of inductive learning—evolving transformation system (ETS). The proposed theory is based on the concept of *structured* measurement device, which is motivated by the formal model of inductive learning and is a far-reaching generalization of the concept of classical measurement device whose output measurements are not numbers but structured entities ("symbols") with an appropriate metric geometry.

We propose that the triad of object structure, image structure and the appropriate

mathematical structure (ETS), to capture the latter two structures, is precisely what computational vision should be about. And it is the inductive learning process that relates the members of this triad. We suggest that since the structure of objects in the universe has evolved in a *combinative* (agglomerative) and hierarchical manner, it is quite natural to expect that biological processes have also evolved (to learn) to capture the latter *combinative* and hierarchical structure. In connection with this, the inadequacy of the classical mathematical structures as well as the role of mathematical structures in information processing are discussed.

We propose the following postulates on which we base the theory.

Postulate 1. The objects in the universe have emergent *combinative* hierarchical structure. Moreover, the term "object structure" cannot be properly understood and defined outside the inductive learning process.

Postulate 2. The inductive learning process is an evolving process that tries to capture the emergent object (class) structure mentioned in Postulate 1. *The mathematical structure on which the inductive learning model is based should have the intrinsic capability to capture the evolving object structure.*

(It turns out that the corresponding mathematical structure is fundamentally different from the classical mathematical structures.)

Postulate 3. *All basic representations* in vision processes are constructed on the basis of the inductive image representation, which, in turn, is constructed by the inductive learning process (see Postulate 2). Thus, the inductive learning processes form the core around which all vision processes have evolved.

We present simple examples to illustrate the proposed theory for the case of "low-level" vision.

Keywords: vision, low-level vision, object structure, inductive learning, learning from examples, evolving transformation system, symbolic image representation, image structure, abstract measurement device.

"If the universe is not meaningless, what is its meaning? For me, this meaning is to be found in the structure of the universe, which happens to be such as to produce thought by way of life and mind. Thought, in turn, is a faculty whereby the universe can reflect upon itself, discover its own structure ... ".

Christian de Duve, *Vital Dust*, 1995.

"The final results of the experience and reflections just presented may, I believe, be summarized as follows:

1. In human beings we find reflex movements and instincts as effects of innate organizations. Instincts act in the interest of the pleasure of some impressions and in avoidance of the discomfort of others.

2. *Inductive inferences, executed by the unconscious activity of memory, play a commanding part in the formation of intuitions* [our italics]."

Hermann von Helmholtz, *Selected writings of Hermann von Helmholtz*, 1971.

1 Introduction

So far, most of the formal developments in computer vision have proceeded under the implicit assumption that inductive learning processes are not very relevant to the extraction of basic image information and construction of the corresponding representations, particularly for low-level vision [1], [2], [3]; some recent exceptions include the work by Pachowicz [4] and Bala [5]. However, the latter work does not at all attempt to present any *theory* of vision. All the core research work in computational vision (e.g. the "school of D. Marr") absolutely ignores the inductive learning processes. And this has been already noted by some researchers in perception and sensation: e.g. Coren [6] p. 14, notes that "David Marr [1] ... began with the general presumption made in direct perception that *all of the information needed is in the stimulus inputs* [our italics]", i.e., no dependence on the learned information.

The inductive theory of vision proposed in this paper *postulates that the vision processes get all the basic representations via inductive learning processes* (Sections 4, 9). Why do we believe in (and propose) the absolutely central role of inductive learning processes in vision? The answer can be seen from two sides. First, the hope is that the inductive processes embody the universal and efficient means for extracting and encoding the *relevant* information from the environment. In other words, we believe (together with Helmholtz [8] and Barlow [9]) that there is a universal mechanism for the extraction of relevant features from the environment (and therefore for building representations of the environment). This mechanism operates across various senses and at all levels of information processing. Second, the evolution of intelligence could then be seen, more or less, not as ad hoc but as a result of interactions of such a mechanism with the environment.

Why are most of us ignorant of the quintessence of the scientific experience of the last four centuries which is mathematical parsimony? More specifically: Why have the corresponding inductive learning models not been developed and/or perceived as central to AI and cognitive science in general, and vision in particular?

It appears that the main obstacle to the development of an adequate inductive learning model has been the absence of the satisfactory mathematical structures. Any modeling begins with the choice of an appropriate mathematical structure (see the quotation by John von Neumann at the end of Section 3). In case of inductive learning, in spite of extensive investigations over the last 40 years, it appears that none of the existing mathematical structures are appropriate to model these processes (see Section 6). Unfortunately, the latter fact has been hardly recognized. This is basically due to the improper understanding of the role and the implications of the choice of the underlying *mathematical structures* in information processing (see also companion paper [10]). It was conjectured that the problem of inductive learning requires for its solution a new mathematical model based on a fundamentally new mathematical structure which allows for a dynamic update of the class structure being constructed during the inductive learning process ([11] Section II).

We base the proposed theory of vision on the evolving transformation systems (ETS) model for inductive learning originally proposed in [12] (see also [11] [13] [14]). The model represents a very natural symbiosis of symbolic and numeric formalisms and clarifies the role of each one in the inductive learning process. Moreover, it suggests a new form of inductive class representation which also represents a symbiosis of the classical symbolic (discrete) and numeric (continuous) representations. ETS also allows for a very "natural" learning process in which the *structure of the class* is regularly updated.

We strongly believe that any theory of vision will stand or fall based on the appropri-

ateness of the chosen mathematical structure that is chosen to model the central perceptual process—inductive learning. To better clarify the relationships between the theory and the corresponding mathematical structure involved, i.e. that the theory *should be built on top* of the mathematical structure, the paper is composed of two parts. In the first half of the paper (Sections 2–6) we justify and outline the formal foundations of the theory, which includes a description of the relevant underlying mathematical structure. In Section 2 we emphasize that the structure of objects in the universe is *combinative* and in order to capture this combinative structure, a fundamentally new mathematical structure may be necessary. What is a mathematical structure? The answer to this question and the role of mathematical structures are discussed in Section 3. The central role of inductive learning processes in capturing the evolving combinative structure is outlined in Section 4. In Section 5 we outline a new mathematical structure—evolving transformation system (ETS)—and in Section 6 the inadequacy of the classical mathematical structures for modeling inductive learning processes is discussed in light of the ETS.

In the second half of the paper (Sections 7–11) we outline the basic ideas of the inductive theory of vision, including the goal of vision, fundamental limitations of the current approaches to low-level vision, the role of inductive learning in low-level vision, structured and inductive image representation and signal to symbol transformations and present few examples. The central concept of the second half of the paper is that of *structured measurement device* which appears to be a revolutionary but necessary (symbolic) generalization of the corresponding classical concept.

2 Evolution of the universe and the structure of objects

We begin the first part of the paper by examining the evolving structure of objects in the universe. The following extended quote by the astrophysicist Hubert Reeves, who is an outstanding popularizer of science, gives a *very appropriate* summary of the present understanding of the evolution of the universe.

> Let's imagine that Aristotle, one of the pioneers of the enterprise, were to return among us and ask: "What do you know about nature that we didn't know in our time? What have you learned that is new since we walked upon the earth?"
>
> ....

We could answer Aristotle's question with two sentences: 1. "Nature is structured like a language"; and 2. "A pyramid of complexity has arisen over the ages."

. . . .

A written language is made up of letters, words, sentences, paragraphs, chapters, books, collections. The basic formula is combinative. Words are combinations of letters, and a combination of words gives us sentences.

Here we encounter an important concept: the "emergent property." The combination of letters in a specific order results in the appearance of something "new," something we can't find when we look at each element separately. . . .

These emergent properties come into play at every level. . . .

We can therefore use the image of a pyramid with superimposed "alphabets" (letters, words, sentences, paragraphs, etc.) to describe a written language. Each element at a particular level is composed of elements form the level below and makes the elements of the level above. Words are the "letters" of the sentences, sentences are the "letters" of paragraphs, and so on. The basic principle of the construction is, once again, a combinative process that generates emergent properties.

. . . .

. . . We have come to realize that this formula (writing), invented by human beings some five or six thousand years ago, has been used in nature for fifteen billion years.

. . . .

In parallel with the pyramid of written language (letters, words, sentences, paragraphs, chapters, etc.) we can today raise up the pyramid of nature's superimposed alphabets.

The lowest level is home to the nuclear force, responsible for combining quarks into nucleons into nuclei. Higher up, we enter the territory of the electromagnetic force, in charge of the formation and workings of molecules, cells, and living organisms.

. . . .

We have now explained the meaning of our first key statement: "Nature is structured like a language." . . .

> The second key statement in our message to scientific pioneers comes to us from astronomical and cosmological research. It goes like this: "The pyramid of complexity was erected in the course of time."
>
> . . . .
>
> ... The corresponding structures – nucleons, atoms, molecules, cells, organisms – have all appeared one after the other. We can thus describe the history of the universe in terms of the ascent of matter toward higher levels. [15] pp. 105-110.

Thus for our purposes we may think of an "object" structure[1] as being evolved in a combinative (agglomerative) manner starting from some initial simple "objects". Moreover, the biological evolution also points to the same, "combinative" direction:

> These and other living relics of once-separate individuals, detected in a variety of species, make it increasingly certain that all visible organisms evolved through symbiosis, the coming together that leads to physical interdependence and the permanent sharing of cells and bodies. Although, as we shall see, some details of the bacterial origin of mitochondria, microtubules, and other cell parts are hard to explain, the general outline of how evolution can work by symbiosis is agreed upon by those scientist who are familiar with the lifestyles of the microcosm.
>
> . . . .
>
> This revolution in the study of the microcosm brings before us a breathtaking view. It is not preposterous to postulate that the very consciousness that enables us to probe the workings of our cells may have been born of concerted capacities of million of microbes that evolved symbiotically to become the human brain. [16] pp. 21-22.

In view of the above it is quite natural to expect that biological information processes have evolved to capture the latter combinative and hierarchical structure. In other words, one can think of the central biological processes as those processes in the universe that try to capture the evolving structure of the universe.

If we look at vision processes from this perspective, we can hypothesize that all levels in vision try to capture this combinative "object" structure (at an appropriate level).

[1] The word object refers also to an event, a process, etc.

Postulate 1. The objects in the universe have emergent *combinative* hierarchical structure. Moreover, the term "object structure" cannot be properly understood and defined outside the inductive learning process.

3 Mathematical Structures and their role

A group of outstanding French mathematicians, who took the pseudonym of Nicolas Bourbaki [17], contributed significantly to the popularization of mathematical structures whose understanding was emerging during the first half of this century.

Presently a mathematical structure, e.g. totally ordered set, group, vector space, topological space, is understood as a set—carrier of the structure—*together with* a set of operations, or relations, defined on it (and/or on its power set where the power set is a set of all subsets of the given set). The relations/operations are actually specified by means of axioms [17] [18] and describe (axiomatically) the interrelationships among the elements of the carrier set. In other words, mathematical structures, essentially, postulate various kinds of abstract relations among the objects in the set, i.e, one *postulates* the rules for manipulating, or working with, the objects in the set. An example of a typical algebraic structure is that of a group, which is defined as a carrier set G *plus* a binary operation "$\circ$" satisfying the following axioms:

(i) $a, b \in G \Rightarrow a \circ b \in G$,

(ii) the operation is associative,

(iii) the operation has a *neutral* element e: $\forall a \in G \qquad e \circ a = a$,

(iv) every element of G has an inverse with respect to e: $\forall a \in G \qquad \exists a^{-1} \in G \qquad a^{-1} \circ a = e$.

It is not surprising that some structures are more similar to each other than to other structures, e.g. fields and vector spaces, topological spaces and uniform spaces, because such structures were "designed" to capture similar types of relationships.

It appears that in relation to reality mathematical structures play the following role. If the axiomatic structure captures optimally the generative abstract structure of the natural phenomenon, then the mathematical results obtained within the postulated structure also capture important features of the phenomenon and thus can predict these features. It is important to note that the same phenomenon may be captured to a *more* or *less* satisfactory degree by various mathematical structures. In other words, a mathematical structure could be thought of as eyeglasses through which we can view the phenomenon. It goes without saying, that as the need arises, new eyeglasses may need to be constructed, since the eyeglasses may change *completely* our view of the phenomenon (compare the mathematical model of classical mechanics with that of quantum mechanics; see also companion paper [10]).

In view of the above, we must emphasize the absolutely critical role of (mathematical) structures in the study of nature. This is simply because all our formal (and therefore informal) understanding of the phenomenon is completely dependent on the postulated structure. In this connection, it is useful to keep in mind the view of John von Neumann, one of the leading scientists of this century:

> "To begin, we must emphasize a statement which I am sure you have heard before, but which must be repeated again and again. It is that the sciences do not try to explain, they hardly try to interpret, they mainly make models. By model is meant a mathematical construct which, with the addition of certain verbal interpretations, describes observed phenomena. The justification of such a mathematical construct is solely and precisely that it is expected to work – that is, correctly to describe phenomena from a reasonably wide area. Furthermore, it must satisfy certain aesthetic criteria – that is, in relation to how much it describes, it must be simple." [19] p. 492.

4 On the central role of inductive learning processes

We strongly believe that inductive learning processes are the only central processes around which all other biological information processes have evolved. Why? The answer should become clear if we re-think carefully a well known fact captured by Schrodinger in the following quotation.

> "A single experience that is never to repeat itself is biologically irrelevant. Biological value lies only in learning the suitable reaction to a situation that offers itself again and again, in many cases periodically, and always requires the same response ...." [20] p. 96.

The same point has been emphasized by a number of outstanding physiologists such as Helmholtz, Barlow [9] and others. Here is a quotation from the last paper of Helmholtz.

"The final results of the experience and reflections just presented may, I believe, be summarized as follows:

1. In human beings we find reflex movements and instincts as effects of innate organizations. Instincts act in the interest of the pleasure of some impressions and in avoidance of the discomfort of others.

2. *Inductive inferences, executed by the unconscious activity of memory, play a commanding part in the formation of intuitions.*

3. It may be doubted that there is any indication whatsoever of any other source or origin for the ideas possessed by a mature individual." [8] p. 512; [our italics].

We believe it is the inductive learning processes that allow a biological agent to extract the necessary information from the observed phenomenon and consequently to encode it, i.e., to commit it to memory.

What is the correct formulation of the inductive learning problem? The *basic* problem may be stated as follows:

Given a finite set C^+ of positive training objects that belong to a (possibly infinite) set C (concept) to be learned and a finite set C^- of negative training objects that do not belong to the concept C, find an analytical model that would allow one to construct the class representation and, *as a consequence*, to recognize if the new element belongs to C. In other words, on the basis of the finite training set $C^+ \cup C^-$, such that $C^+ \cap C^- = \phi$, the agent must be able to form an "idea" of the inductive generalization corresponding to the concept[2] C. The potential model should be able to construct the (inductive) structure of the class C based on its finite subset C^+. By *the structure of the class* we mean: (1) the *symbolic* features that make the objects of the class similar to each other and/or different from other objects outside the class, and (2) the *emergent* interrelationships among these features (see Section 5). The inductive learning process would then involve the discovery and encoding of the (inductive) structure of the class, which during the *consequent* recognition stage is used in the resulting distance function to compare a new object to some fixed objects from C^+.

We believe that the question of representation is directly related to that of the class structure and therefore cannot be addressed properly outside the inductive learning model. Moreover, we believe (and illustrate this in Section 10) that *the construction of the repre-*

[2] Objects in C^+ could be different "views" of the same object.

sentation should be the main part of the inductive learning process.

It appears that one of the basic reasons for a peripheral role of inductive learning in artificial intelligence in general, and vision in particular, is the lack of an adequate formal model for inductive learning. At the same time, we believe that the absence of the model could be partly explained by the lack of an adequate understanding of the role of inductive learning processes by the three founding schools of AI (viz. Massachusetts Institute Technology, Carnegie Mellon University and Stanford University).

5 A new mathematical structure: The Evolving Transformation System

We now turn to a *very brief* description of the evolving transformation system (ETS) model of inductive learning proposed in [11]-[14], [21]. The model could be thought of as the formalization of the concept of *symbolic system* whose role in AI has been so pervasive. ETS is the mathematical structure that is "built from more primitive" mathematical structure—the transformation system, whose axiomatic definitions we give next.

Definition 1 *A* **transformation system** *(TS) is a triple*

$$\mathbf{T} = (S, O, D)$$

where:

S is a set of **structs** *over a finite struct alphabet A; structs are analogously structured discrete representations of objects (e.g. strings, trees, etc.) ;*

$O = \{o_i\}_{i=1}^m$ is a finite set of operations that are multivalued functions, $o_i : S \to S$, satisfying the following two axioms:

(i) $\forall o \in O \quad \forall s \in S \quad \exists o^{-1} \in O$ *such that* $\forall s' \in o(s) \quad s \in o^{-1}(s')$

(ii) for every pair of structs there exists a sequence of operations that transforms one into the other;

the set O specifies permissible operations for transforming one struct into another (e.g. deletion-insertion, substitution operations) and can be thought of as *a postulated set of basic/primitive "object features";*

$D = \{\Delta_\omega\}_{\omega \in \Omega}$ *is a (competing) family of distance functions defined on set* S *whose parameter set* Ω *is the* $(m-1)$ *-dimensional unit simplex in* $\Re^m$

$$\Omega = \left\{ \omega = (w^1, w^2, \ldots, w^m) \ \mid \ w^i \geq 0, \ \sum_{i=1}^{m} w^i = 1 \right\}$$

and each of the distance functions Δ_ω *is defined as follows: weight* w^i *is assigned to the operation* o_i *and*

$$\Delta_\omega(s_1, s_2) = \min_{\mathbf{o}_j \in \mathbf{O}} \sum_{n=1}^{k} w^i_{(j)}$$

where the minimum is taken over the set $\mathbf{O}$ *all possible sequences* $\mathbf{o}_j = \left(o_i^{(j)} .... o_k^{(j)}\right)$ *of operations that transform struct* s_1 *into struct* s_2.

To compute the distances $\Delta_w(s_1, s_2)$ the system must use its set of operations in a *"cooperative and competitive"* manner. Thus, all the properties of the system resulting from this definition can be viewed as *emergent* properties.

If $\mathbf{T}_1 = (S_1, O_1, D_1)$ and $\mathbf{T}_2 = (S_2, O_2, D_2)$ and, moreover, $S_1 \subseteq S_2$, $O_1 \subseteq O_2$, then we will say that the transformation system $\mathbf{T}_1$ is a *transformation sub-system* of $\mathbf{T_2}$: $\mathbf{T}_1 \subseteq \mathbf{T}_2$.

Since the set of real numbers is constructed form the set of rational numbers, which is, in turn, constructed from the set of natural numbers (see, for example, [22]), it is not difficult to see that for all practical purposes the set of real numbers can be viewed as a TS: the set of natural numbers is a Peano TS where the alphabet $A = \{a\}$, S is the set of strings over A (i.e. $S = \{\theta, a, aa, aaa, \ldots\}$), and O is the set consisting of a single operation $o : a \leftrightarrow \theta$ ($\leftrightarrow$ denotes a pair of operations $a \rightarrow \theta$, $\theta \rightarrow a$).

One should also note that since from computational perspective the class of TSs is equivalent to the class of Turing machines ([23], Chapter 13, Theorem 2), in a TS we have as powerful computational "device" as is known at present.

Learning in a TS can be reduced to the following optimization problem:

$$\max_{\omega \in \Omega} f(\omega), \qquad f(\omega) = \frac{f_1(\omega)}{c + f_2(\omega)}$$

where $f_1(\omega)$ is the Δ_ω-distance *between* C^+ and C^-, $f_2(\omega)$ is the average Δ_ω-distance *within* C^+, and c is a small positive constant to prevent the overflow condition (when the values of $f_2(\omega)$ approach 0). Since $f(\omega)$ gives both the measure of compactness of C^+ as well as the measure of separation of C^+ from C^-, it is called the *quality of (learning) class perception.*

For a given concept C and *given set O of operations*, every optimal weighting scheme $\omega^* \in \Omega$ generates the 'best' metric configuration of the training set C^+. In other words, under the given set of operations, the corresponding distance function Δ_{ω^*} gives the 'best' separation of the positive training set with respect to the negative one.

Evolving transformation system (ETS) is a mathematical structure that involves a finite or infinite sequence of TSs with a common set S of structured objects

$$\mathbf{T}_i = (S, O_i, D_i, R)$$

in which each set of operations O_i, except O_0, is obtained from O_{i-1} by adding to it one or several operations that are constructed from the operations in O_{i-1} with the help of a small set R of *composition rules.* Each rule $r \in R$ specifies how to (systematically) construct a new operation from the existing operations, e.g. to the set $\{a \leftrightarrow \theta, b \leftrightarrow \theta\}$ one can add operation $ab \leftrightarrow \theta$, or $a \leftrightarrow b$.

From the above definition of ETS it follows that at stage t we have

$$O_0 \subseteq O_1 \subseteq O_2 \subseteq ... \subseteq O_t,$$

$$\Omega_0 \subseteq \Omega_1 \subseteq \Omega_2 \subseteq ... \subseteq \Omega_t.$$

Roughly speaking, the inductive learning process for the ETS proceeds by constructing a sequence of O_i's in such a way that, for each consecutive $\mathbf{T}_i$, the minimum value of f_2 decreases (while making sure that the value of f_1 is not zero), i.e., the inter-distances in C^+ gradually shrink to zero while the distance between C^+ and C^- remains non-zero. We strongly recommend to see [11] for an example of the inductive learning process with ETS. The fundamental difference between the ETS model and the other inductive learning models is due to the intrinsic capability of the former to evolve through the above sequence of TS's, each of which offers a fundamentally different and "more optimal" class of distance functions.

Definition 2 *In view of the above model of inductive learning process,* **the inductive class structure Π** *is defined by the following triple*

$$\Pi = (\bar{C}^{+}, O_{fin}, \bar{\Omega}_{fin})$$

where $\bar{C}^{+}$ is a small subset of C^{+}, O_{fin} is the final set of operations at the end of the learning process, and $\bar{\Omega}_{fin} \subseteq \Omega$ is the set of optimal weight vectors for the final transformation system (relatively often $\mid \bar{\Omega}_{fin} \mid = 1$).

Thus, since, during the recognition stage, the class membership of a struct s will be determined on the basis of its Δ_{ω^*}-distance to $\bar{C}^{+}$, the class structure embodies the symbiosis of both the classical discrete (O_{fin}) and continuous ($\bar{\Omega}_{fin}$) formalisms (see also [12]).

Postulate 2. The inductive learning process is an evolving process that tries to capture the emergent object (class) structure mentioned in Postulate 1. *The mathematical structure on which the inductive learning model is based should have the intrinsic capability to capture the evolving object structure.*

6 Inadequacy the classical mathematical structures as candidates for inductive learning model[3]

Classical mathematical structures such as the ordered set, the field, the vector space, etc. have emerged as a result of investigation of *numeric* systems. On the other hand, from the very beginning AI has been dealing with symbolic systems and the feeling of the practitioners of AI has been that the symbolic systems are fundamentally different from the numeric systems. In [14] a basis of an argument that the vector space based (numeric) learning systems cannot learn in symbolic environments was presented, suggesting that, indeed, the two types of mathematical structures are quite different.

It appears that one of the basic facts contributing to the fundamental differences between the symbolic and numeric systems is as follows: while the binary relation of order plays an absolutely critical role in numeric structures, it does not play any substantial role in

[3] See also companion paper [10].

symbolic structures (see definition of a transformation system in the last section). Thus, for example, the Peano axioms for natural numbers simply specify a very restricted type of a transformation system, i.e., one over a *single letter alphabet* and with a single operation—successor function. The latter represents a mathematical structure that can completely and equivalently be specified in the language of ordered set. As a consequence, while the topology in the numeric system is intimately related to the corresponding ordering, the topology in the symbolic system is not.

It turns out that the structural limitation of vector spaces—the basic mathematical structure used to model inductive learning processes—is of the nature that prevents the learning agent from being able to modify, if necessary, *the structure* of the inductive class representation: since the space where objects are represented is linear, the only "legitimate" class structures that can (inductively) be constructed based on the finite training set are linear.

In reality, the corresponding linear subspaces induced by the training vector data representation hardly ever capture the structure of the class from which the training set was extracted. Thus, to overcome the above structural limitation of the vector space, practically all vector space based models introduce in a (necessarily) arbitrary manner (i.e., in a manner not related to the mathematical structure of the underlying vector space as well as to the structure of the data) some class of non-linear functions that is used to approximate "class" boundaries based on the vector representation of the training data. *The "class" constructed in this manner has little to do with the class from which the data was collected.*

In contrast to the above structural limitations of the classical numeric (continuous) structures, the axiomatic structure of the proposed evolving transformation system is such that it does allow the agent to modify the structure of the inductive class during the learning process (see previous section), by modifying the set of operations.

This concludes part 1 of the paper in which we have very briefly outlined a formal model for inductive learning (ETS) and mentioned the fundamental limitations of the classical mathematical structures that ETS overcomes. We have also emphasized the fundamental role that *mathematical structures* are destined to play in the construction of models for intelligent information processing systems. Finally, the absolutely central role of inductive learning processes in extracting the information from the environment and constructing the corresponding symbolic and numeric representations at all levels of biological information processing in general, and vision in particular, was postulated.

7 Vision and Symbols: A New Perspective

In this, second part of the paper, we outline a fundamentally new approach to vision (with a particular focus on "low-level" vision) *based on* the inductive learning model discussed in part I. As discussed earlier, the lack of appropriate mathematical models has resulted in little or no understanding of the basic processes involved in vision. In this connection, it is important to stress that the foundations of vision lie "outside" vision proper, since, vision processes have evolved out of earlier physical processes to capture the inductive (combinative) structure of objects. Our division of the paper in two parts is also motivated by this fact. We lay the groundwork for such a foundation by, first, identifying the goal of vision (see last paragraph of Section 7.1) and limitations of the current approaches to low-level vision, and, then, proposing an appropriate framework to accomplish it.

7.1 What is the goal of Vision?

Understanding human vision has been one of the major quests of science. However, we think, that such an endeavor is absolutely futile without an *adequate* understanding of the *objective* (goal) of vision processes. What is the "goal of vision"? Is it, to produce "a compact representation of the image data" [24], or is it to "produce from the images of the external world a description that is useful to the viewer" [25], or is it "to exploit the light emitted or reflected by the environment in order to improve the chance of survival" of the viewer [26]? These and many other descriptions of the purpose of vision can be found in the vision literature [2]. None of these goals are adequate enough to direct us to and focus on the *basic/central/driving processes* in vision. We believe that the focus on the central processes, as has been the trend in physics, is critical to the entire enterprise.

We propose that the objective of vision processes is to extract the combinative structure of objects and scenes (postulate 1), and interpret it (postulate 2) in order to "successfully" interact with the external world. What has been insufficiently understood so far is the role of mathematical structures that capture the combinative object structure in vision. We believe that it is the knowledge of such object structure which is necessary and sufficient to interact successfully with the external world. In other words, the goal of vision that we propose is to construct a representation of the external objects based on their combinative structure. Moreover, such an internal representation should also be combinative and can be captured only by an appropriate formal model. We believe that the above objective is not controversial. However, what is controversial is whether the current approaches to vision are adequate or not to capture the combinative object structure. Thus, two main objectives of this paper are: to simply stress the latter point and to outline an inductive framework for vision that can capture the combinative structure of objects, at all levels of vision.

An immediate and very natural question is: How do vision processes capture structure of objects? The answer can be intuited by focusing on what is meant by *image structure* and its role in vision. The term image structure has been used in the vision literature quite frequently and has been understood to play an important role. We believe that the structure of image has to play not just an important but the driving role. Moreover, we propose that the structure of objects should be captured through the extraction and representation of image structure. The (low-level) structure of image is conventionally understood (mainly based on the foundational work of Marr and co-workers [1]) as defined in terms of edges, blobs, etc. Recently, Koenderink proposed a new image representation scheme for describing image structure [27]. In this scheme the image is represented at different resolution levels and such a representation "defines" the image structure. However, in the schemes of Marr and Koenderink the interpretation and understanding of the term structure and hence "image structure", are quite imprecise and inadequate, since the schemes are not (explicitly) guided by postulates 1 and 2.

As far as the measurement devices used in vision are concerned, for example a CCD camera, they are also objects with the combinative structure determined by the structure of the array of sensors (see Section 9 for a proposed description of measurement devices and their structure). Since both the objects and a camera have a combinative structure, the image structure, thus, should also be interpreted as a combinative structure. It follows that the goal of vision is to determine this combinative image structure which captures the combinative object structure.

We propose to capture the concept of combinative structure by means of the axiomatically introduced new mathematical structure—transformation system (TS), since we believe that the language of mathematical structures is the *only* legitimate means for describing the structure of objects. This abstract structure, in turn, should help one to *interpret* the inductive learning processes[4].

Until recently no formal mathematical framework has been proposed that can extract the (combinative) image structure. As mentioned above, the term "structure" can get a satisfactory formal interpretation only through the corresponding mathematical structure, and such a precise interpretation of the term (combinative) structure is what has been lacking. We propose to *interpret "structure" in light of a mathematical structure only, and that no other interpretation will be adequate. Moreover, the corresponding mathematical structure should be appropriate for capturing the combinative nature of the object structure.*

The triad of image structure, the corresponding mathematical structure, and the (combinative) structure of objects is precisely what computational vision should be about. And it is the inductive learning processes, through the discovery of the class structure, that relate the members of the triad.

To summarize, we propose that the main objective of vision processes is to determine *inductively* the (combinative) image structure. This structure, in turn, determines the (combinative) object structure and, thus, allows the agent to successfully interact with the external world. The term "(combinative) structure" should be interpreted *only* through the corresponding mathematical structure as it is currently practiced in mathematics [17] [18].

7.2 Fundamental limitations of the current approaches to low-level vision

As discussed in the previous section the fundamental goal of vision is to extract and encode the combinative image structure which corresponds to the combinative object structure. These combinative structures can only be captured if the *appropriate* mathematical structures are employed. Thus, we view the limitations of any approach to vision as following from the limitations of the corresponding mathematical structures that are implicitly or explicitly used to capture the *combinative object* structure. In this section we discuss the limitations of the current approaches to low-level vision on that basis.

As we discussed in Section 3, given a mathematical structure, the types of combinative structures that it can capture are completely specified by its axioms, or postulates. Again, the widely accepted classical mathematical structure - the vector space structure - can cap-

[4] For a discussion of the role of mathematical structures see Section 3 and [10].

ture *only* certain types of structures, i.e, linear structures. Since there are *potentially* an unbounded number of classes of combinative structures (in various "environments"), we should opt, if possible, for a mathematical framework (e.g. ETS) that allows one to *dynamically modify the form* of the basic mathematical structure (e.g. TS) that is used to capture the specific class of combinative structures. In fact all the classical (numeric) mathematical structures (e.g. group, vector space) by their very definition do not allow for such a modification. In other words, in all of the existing mathematical frameworks, the class of structures they can capture is fixed. Therefore, within any classical applied framework one can capture *only certain (postulated) class of structures*, i.e., one cannot capture any other class of combinative structures. Hence, the necessity for a fundamentally new framework and the difficulties related to its development and *application*.

In order to point out the limitations of the current approaches to low-level vision, we now review some of the more known approaches to low-level vision, *all of which are based on the classical mathematical framework*.

One of the most extensively studied problems in low-level vision is that of edge detection. Several different edge operators have been proposed, e.g. Roberts [28], Kirsch [28], Sobel [28], Prewitt [28], Robinsons [28], and others, including, the well known work of Marr and Hildreth [7]. Basically, all these schemes form a part of the first stage in low-level vision, or *filtering* stage, that extract edges, blobs, lines, points, corners, etc., from the image (at different scales and orientations). A recent work by Perona [29] describes a mathematical framework for determining families of linear filters on a continuum of orientations and scales.

Edges, blobs, points, lines and corners are only some of the infinitely many possible combinative structures in the image. Thus, all the current approaches that are proposed to detect these structures, *are in fact extracting from the image an a priori postulated class of combinative structures.*

As far as *image representation* is concerned, a new scheme is proposed in [30]. The authors suggest to represent essentially *all* "the physical structure of the image in a form that is readable by point processors". The processors, in turn, determine what to extract. This scheme *does not specify what the structure is and how the processors should extract it from such a representation.* In contrast, there are some symbolic image representation schemes that have been proposed [31]. However, the underlying framework is still based on the extraction of *postulated features* viz. edges and corners. The author of [31] proposes a scheme in which he defines "gradient operators" at different scales that can extract the features at multiple scales. Thus, the latter work can be seen as a special case of [29] and an extension of [7].

More recently, a "generalized Gabor scheme of image representation" was proposed in [24]. Such a representation has, in a certain sense, optimal resolution both in frequency and spatial domains. The representation, which was motivated by the neurophysiological studies, consists of approximations of the original intensity measurement function. Again, such a representation cannot really *discover* the object structure, it can only capture a specific form of a combinative structure that is *postulated* by the set of Gabor filters. Thus, in this case, as well in all other frequency based representation schemes [32] [33], one necessarily ignores (because of the postulates) any other classes of structures that may be present in the image and that capture some other image features.

Let us briefly analyze the above approaches to low-level vision in order to identify the *common* underlying mathematical structure—the vector space. It is the vector space of measurable, square-integrable functions with an inner product defined on them. Moreover, due to the discretization, one can assume that the vector space is finite-dimensional. Thus, it is a finite-dimensional Euclidean space. Any linear-filtering operation is the projection operation, that projects the function to be filtered onto a subspace spanned by the basis vectors, i.e. functions, that represent the filters. And as such, *each of the above schemes, by choosing a particular set of filters (or basis) extracts (or approximates) a very specific (predetermined, or postulated) linear structure.* That is, when one chooses a particular basis (say a set of Gabor filters, or a Laplacian of the Gaussian filter), one has, in fact, *postulated* a very specific type of structure that is to be extracted. Therefore, the other classes of image structures, and hence object structures, cannot be *discovered.* As was mentioned above in this section, none of the existing classical mathematical frameworks allow one to extract *various* classes of the image structures, present in practically all images, without postulating them. Moreover, practically all non-trivial images contain enormous number of various image structures.

At the same time, from the neurophysiological studies of low-level vision, the following two descriptions of the cortical structure have emerged: a feature detector based description as proposed by Hubel and Weisel [34] and frequency filters based description as proposed by Campbell and Robson [35]. Until the work of Campbell and Robson, it was believed that the visual cortex is composed of cells that are sensitive to certain features. Later studies also "revealed" that the cortical cells behave like spatial filters [36] [37]. After Daugman's recent proposal [38] that a Gabor-filter-based signal representation provides optimal resolution in both frequency and two dimensional spatial domains, many neurophysiologists generally believe in the validity of such a description of the visual cortex [40].

The principal justification for this description of the visual cortex (Gabor-filter-based signal representation), is that this representation removes the statistical and spatial redundancies and results in a compact code [40]. However, such a compact coding scheme is also shown to be insufficient to account for the receptive field properties of the cells. Rather, it is proposed that the cells transform the higher order redundancies to the lower order ones [41] [42].

Let us briefly analyze the above results of the neurophysiological studies. All physical, including neurophysiological, experiments are guided by mathematical models (or more precisely mathematical structures), and this fact is known to neurophysiologists. Again, the above studies have been based on the frequency-filter based mathematical model, more precisely the vector space model. Any interpretation, as well as the results, will be necessarily within the context of this mathematical structure. Thus, the term redundancy used above is also defined in the context of the vector space structure, and so the interpretation of why the cells perform a spatial frequency-based-filtering *should be understood only through such a definition of redundancy.* Of course, the response properties of the cells do show a profile that is similar to the Gabor function, but does that mean that the cells *use* the responses *as modeled by the Gabor function in perception*? As previously mentioned, such a Gabor-filter-based signal representation can only be considered as a means for functional approximation. None of the neurophysiological experiments performed so far reveal any information about the basic/central processes that capture the object structure. The existing experimental settings do not allow one to perform the appropriate experiments. Is it legitimate, then, to use the response properties of the cells as a justification for the chosen Gabor-filter-based image representation (as done in [24])?

In [43] the authors have demonstrated the differences between feature detection and structure detection, and they suggest that orientation-tuned receptive field mechanisms can be appropriate for extracting features but cannot extract structure. Moreover, as we have already discussed, since these features are postulated one can extract a very specific, predetermined class of combinative structures.

According to the proposed framework, in order to determine the object structure, vision becomes a "symbolic" process right at the transduction stage in the retina. Thus, if the latter is indeed the case, then one needs an appropriate mathematical structure to guide the experiments, and, moreover, the vector space structure is absolutely inadequate for the purpose (see Section 6). Since the neurophysiological experiments have been guided by an inappropriate mathematical structure—vector space, one cannot determine the functioning structure of the visual cortex. Thus, it should be clear that the neurophysiological studies may need a fundamentally new mathematical structure to guide the experiments, and, of course, such a structure would necessarily have profound implications to our understanding of vision processes. For example, as noted in [44] p. 95 the "relatively simple feature detectors [orientation selective neurons] cannot resolve the complexities of the everyday perception", since these simple "feature detectors must somehow be combined into the complex forms that you see every day, and we do not yet know how this is accomplished". The model that we propose may shed light on the nature of these combinative processes[5].

8 The role of inductive learning in low-level vision

Inductive learning is a process of construction of class abstraction, or class concept, or generalization. It has been mainly thought of as a high level cognitive process. We believe, however, that this is a central and driving process that forms the core around which all cognitive processes have evolved at all levels. It is precisely this view that has led us to suggest the fundamental role of inductive learning processes in low-level vision. Low-level vision studies have (practically) completely ignored the learning processes. In contrast, we propose a critical role of inductive learning processes for all levels of vision, and specifically for low-level vision, which, we believe, should in turn, resolve different issues related to representation. Exactly *what* is represented is determined by the inductive learning process.

[5] See also [44] p. 94 where the experimental observations of the neurons that respond to complex stimuli such as faces are discussed.

Moreover, the term representation is clarified in light of the formal model of this process.

To properly understand the term image structure, for our purposes we propose to interpret informally the term "structure" *in this context* as follows: *structure is a collection of objects together with the (emergent) relations between them that allows one to abstract (generalize) and associate meaning with the set of objects.* This interpretation of the term structure is closely related to the interpretation of the term mathematical structure, e.g. group, field, and vector space. Thus, in the case of low-level vision, using the famaliar language, the image structure could be thought of as the collection of the intensity measurements together with the relations between them that allows one to abstract (generalize) and associate meaning with the set of measurements. Moreover, it is the inductive learning process that determine these relations. The semantics of an image region can only be determined through such a process of abstraction (inductive learning). We believe, that there is no other way to associate the meaning to the image region. And, since, "image understanding" is essentially the discovery of the image semantics, we conclude that, inductive learning should play the central role in low-level vision processes.

Again, we propose that the role of inductive learning in low-level vision is that of discovering the structure of the image, and hence that of the original object, from the vast pool of measurements. It is a process that abstracts, or generalizes, the relationships between the measurements and so discovers the exact information that "needs" to be represented.

One of the fundamental requirements of any representation scheme is its "compactness" (ability to compress the data). It is not surprising that inductive representation and compact representation are closely related. The term "compact" should be properly interpreted as "semantically compact". It is quite clear that the inductive learning process produces compressed/compact representation which is also semantically compact. In contrast, other compression schemes, e.g. Gabor-filter-based representation, do not result in a *semantically* compact representation, because the resulting functional approximation has nothing to do with the semantic information in the data, since the choice of the class of functions (e.g. Gabor functions) is arbitrary, i.e., independent of the structure of the input function. The semantic information is *the discovered structure* of the data, of which one remains completely ignorant if a classical compression scheme is applied. Inductive learning should not only provide a compact representation, but should also discover and encode the structure of the data *without imposing an ad hoc structure on it.*

To resolve the above issues, one needs an appropriate mathematical framework. None of the existing mathematical frameworks allows one to dynamically modify the *set of operations* of the basic mathematical structure. Since mathematicians never had to face such a requirement, no ready made formal solutions have existed. In other words, in all the conventional mathematical frameworks the set of operations of the chosen mathematical structure comes with the structure itself (it is specified by the set of axioms) and cannot be modified. Furthermore, any phenomenon is conventionally studied under the assumption that the chosen mathematical structure is completely adequate for its study. It appears that in order to discover the image structure and hence the object structure such an assumption proves to be too restrictive, because the corresponding (underlying) combinative mathematical structure must be dynamically updated.

Note that in general such a requirement can be realized only within the new mathematical structure, evolving transformation system (ETS). Moreover, for such a dynamic modification, one needs a process of abstraction, or generalization, that will guide the modification and, in turn, determine the corresponding final (basic) mathematical structure—transformation system (TS). It is inductive learning that plays the role of the process of abstraction that guides the dynamic modification of the basic mathematical structure. ETS, thus, is a mathematical structure which emerged from the basic mathematical structure, TS, that allows one to dynamically modify the set of operations of TS. ETS is a mathematical model for inductive learning and provides a mathematical framework for *discovering* the combinative image structure and hence the corresponding combinative object structure.

We believe, thus, that the inductive learning plays an absolutely central/driving role in vision, and TS and ETS provide an appropriate framework for modeling inductive learning processes. In the following section, we elucidate these facts by proposing a new scheme for *image representation.*

9 What are symbols: A new approach to image representation

Earlier (Sections 7.1, 8) we discussed the fundamental goal of vision which is to construct a (combinative) representation of objects. Now we will discuss how one can construct such a representation starting from the initial measurements. It is well known that measurements provide a "universal" means to study a phenomena. In case of "image understanding", it is conventionally (implicitly) assumed that the *vector representation* of the intensity measurements—i.e., the representation that is independent of *the combinative structure of the measurement device*—is adequate for understanding the structure of the object. It should now be clear, however, that the combinative structure of the measurement device has to be represented also. Moreover, it is important to note that numeric measurements (light intensity in particular) are hardly the only means to represent the environment: the counter examples are all biochemical systems. It is also important to keep in mind that in contrast to (e.g. chemical) symbols, the number itself is a human creation and does not exist in nature. As far as the relationships between various mathematical structures are concerned, the numeric mathematical structures can be viewed as a very restricted form of the transformation system (see Section 5 and [21]).

In this section we propose a new form of image representation that *inductively* constructs (from the initial measurements) a combinative object representation. The image representation is constructed in two stages: the first stage is called the structured image representation (SIR) and the second stage is called the inductive image representation (IIR).

9.1 The Structured Image Representation (SIR)

In order to construct the structured image representation (SIR) one requires an *adequate abstract specification of the measurement device*, since the device forms a fundamental integral component of any (human or machine) vision system. We define measurement device in a manner more general than is currently understood.

Definition 3 *A* **measurement unit** *is an abstraction of the elementary/atomic measurement device and will be denoted by* u. *A* **(structured) measurement device** $\mathbf{M}$ *is a triple* $(U, m, \mathbf{T})$, *where*

U is a finite set of measurement units,

$m : U \to AT$ is a mapping from U into the attribute set AT. each element of which is an n-tuple $m(u) = < a^1, a^2, ..., a^n >$, a^i characterizes one aspect of the unit (see below), and

$\mathbf{T} = (S, O, D)$ is a transformation system whose structs are built from units in U.

Each measurement unit is completely characterized by $m(u)$—an n-tuple of *attributes of* u. A chosen measurement unit can perform only one type of measurement, numeric or non-numeric (i.e., a symbolic struct measurement). Moreover, the device $\mathbf{M}$ may have units of different types, each performing, for example, different type of measurement at the *same* location. For example, attribute a^1 may be unit's type (thermal, light, acoustic, etc.), a^2 may be the unit's location in space l_u, and a^3 may be the *range* of the unit's measurements which is defined to be a set of structs, S_u. Thus, among attributes of a unit u we have the unit's range which is defined as a transformation system $\mathbf{T}_u = (S_u, O_u, D_u)$ that specifies what the unit's measurements are, while the transformation system $\mathbf{T}$ in the above definition specifies the structure of the entire device $\mathbf{M}$. Moreover, as is the case with classical measurement devices, the struct measurements are *produced* immediately by the corresponding unit u. It is not difficult to see that all present measurement devices are special cases of the above device. All present devices have units whose ranges are the numeric transformation system (i.e., they all "produce" numbers), which is a very restricted/trivial form of the transformation system (see Section 5). It is important to note that some of the attributes are *static* (fixed) and others are *dynamic* (e.g. location).

One can easily see that our definition of the measurement device *inseparably links* the *concept of the device* to that of the corresponding *mathematical structure.* This link should make it quite clear the differences between various measurement devices. In traditional computational vision, the unit's range $\mathbf{T}_u$ is a subset of reals and $\mathbf{T}$ is a vector space. What was not previously understood at all in computational vision is that the combinative structure of the objects induces the abstract geometry, defined via the transformation system, and the latter must be inseparably linked to the structure of the measurement device. Moreover, since, as stated in Section 7.1, the objective of vision is to capture the combinative structure of the objects, one can see that the measurement device (including its structure) and vision processes form an integral whole and cannot be disassociated. In particular, it is important to note that to capture the changing object structure in the environment, *the structure of the measurement device* ($\mathbf{T}$) *must be updated during the learning process* (see Section 9.2), and this point is useful to keep in mind when viewing *all definitions in the present section.*

Definition 4 *An (instantaneous) measurement* m_t *by a measurement device* $\mathbf{M} = (U, m, \mathbf{T})$ *is a mapping*

$$m_t : U \to AV,$$

where $m_t(u) \in S_u$ *and*

$$AV = \bigcup_{u \in U} S_u.$$

Thus, if all the units measure light intensity values, $AV \subseteq \Re$. Again, one should keep in mind that each unit u *produces* the corresponding structs as its *measurements*. Also note that in the proposed theory the "image" is "formed" through/by the "produced" *symbolic structs* and not (just) *numeric structs*, as is the case with the present image models. Since the produced measurements result in the images, we turn next to the process of image formation and its combinative nature.

We propose that images should be thought of as composed of discrete atoms (primitives) and the structure of image, as determined by a chosen combination of these atoms. Such an atomistic view was proposed and extensively studied by Grenander [45]. In accordance with Postulate 1, we suggest that the combinative structure of the image (or sub-image) can be captured by representing the image (or sub-image) by the corresponding (learned) transformation system.

Definition 5 *Given a measurement device* $\mathbf{M}$ *and its instantaneous measurement* m_t*, a (structured) image* $\mathbf{I}$ *is a triple* $(m_t(U), L, \mathbf{T})$*, where* $L = \{l_u \mid u \in U\}$ *is a set of locations and* $\mathbf{T}$ *is a transformation system for* $\mathbf{M}$ *(see Def 2); a (structured) sub-image* $\mathbf{I}_1$ *of* $\mathbf{I}$ *is a triple* $(m_t(U_1), L_1, \mathbf{T}_1)$ *, where* $U_1 \subseteq U$*,* $L_1 \subseteq L$*, and* $\mathbf{T}_1$ *is a transformation sub-system of* $\mathbf{T}_2$*, i.e.,* $\mathbf{T}_1 \subseteq \mathbf{T}$ *(see Section 5).*

Figure 1 illustrates the role of the transformation system in image representation. Conventionally, in computational vision an $m \times n$ image region is represented as a vector in a $m \times n$-dimensional vector space. Almost in all the current low-level vision approaches the vector space is used as the underlying mathematical structure for image representation, i.e., the transformation system of the corresponding measurement device is a finite-dimensional vector space. The fundamental drawback of such a vector representation (numeric struct) is related to the fact that the *spatial relations between the pixels and their intensities is almost completely lost* since the vector representation necessitates one to use the vector space operations which cannot recover this information. This is evident from Fig. 1. At the same time, the symbolic struct representation together with the operations completely (and explicitly) captures the relationship between the pixels and their intensity values.

The need for the choice of the symbolic struct representation can partly be seen from the following facts. The material properties of the objects are completely determined by their microstructural properties [46]. It is also known that the intensity values depend on the material properties and the surface geometry [47]. Moreover, the image contains objects that have combinative structure. Hence each intensity value represents the microstructural property of the object at that spatial location. Symbolic structs are tailor-made to capture this structural information.

Definition 6 *Given a (structured) image* $\mathbf{I} = (m_t(U), L, \mathbf{T})$, *a set of sub-images* $\{\mathbf{I}_k\}_{k \in K}$ *of* $\mathbf{I}$ *is called* **(structured) image partitioning** *of image* $\mathbf{I}$ *if*

$$\bigcup_{k \in K} U_k = U$$

where $U_k \subseteq U$.

One can think of the above sub-images (in neurophysiological language) as the receptive fields in the retina, or (in computational vision terms) as image windows . One should not confuse the terms structured image and image structure. The latter was *informally* discussed in Section 8.

In connection with the above (general) definitions, we will address in this paper *only one aspect* related to the low-level *feature discovery* and representation. It is useful to keep in mind that in this paper we are not addressing the issue of the relationship between the sub-image and image transformation systems as well as between various sub-image transformation systems.

9.2 The Inductive Image Representation (IIR)

Above all, it is important to stress that at a fixed "level" (see this section below) *the proposed model treats each of the sub-images and images as a class whose elements are the structs produced by the structured measurement device.*

The first main step in producing IIR is a construction of (structured) image partitioning (see Def. 4), which entails the construction of the initial transformation systems $\mathbf{T}_j$ for each of the corresponding sub-images $\mathbf{I}_j$ (see Section 10) *in order to construct their (inductive) representation.* In this paper we are not addressing any issues related to the latter step of image partitioning. The term *sub-image* is used to stress the fact that we are considering only a part of an image. An example of a sub-image (without the corresponding image) and some structs from the corresponding initial transformation system is given in Fig. 2. Each struct in the figure corresponds to a different part of the sub-image and should be considered as obtained by the measurement device. Note that the perceived global symmetry of the sub-image in Fig. 2 does not emerge at the level of "initial" (shown) measurements, or structs.

As mentioned in Section 7.2, the fundamental goal of vision is to extract and encode the image structure, which is accomplished by the processes built around the inductive learning process. The inductive learning process modifies the transformation system $\mathbf{T}$ associated with the structured measurement device (and therefore with the structured image) *at that moment* in such a way that, first, each $\mathbf{T}_j$ is modified to incorporate the inductive class structure ($\mathbf{\Pi}_j$) for each of the sub-images $\mathbf{I}_j$ of image $\mathbf{I}$, and only then $\mathbf{T}$ is constructed for the entire image $\mathbf{I}$. It should not come as a surprise that the mathematical framework we propose to model the inductive learning process is the evolving transformation system (ETS): ETS is invoked during learning to extract the inductive class structure $\mathbf{\Pi}_j$ for $\mathbf{I}_j$ (and hence the structure of the measurements) by dynamically changing the structure of $\mathbf{T}_j$ (see Section 5). That is, the set O_j of operations (and hence $\mathbf{T}_j$) evolves during the learning process through a sequence O_i^j, $1 \leq i \leq t$, resulting in the inductive class structure $\mathbf{\Pi}_j$ for the corresponding sub-image. The local inductive class structure of a (sub)image will be called the **local inductive (sub)image representation**, LIIR (see Fig. 3). Note that in the figure the set C^+ is selected from the current set of measurements, i.e. from the produced symbolic structs, and the set C^- is selected from a stored set of structs[6]. The term "local" is used above to clearly delineate the LIIR from the (**global**) IIR, which as was mentioned above, we perceive as constructed, first, through the symbiosis of LIIR's for different sub-images in the chosen image partitioning and, second, by, again, applying the inductive learning process to the resulting different sub-image representations.

Next, we give a preliminary outline of the overall image representation scheme. We propose to view the image representation as composed of many levels, that form a hierarchy. For each level i, except $i = 0$, the representation derived from those of previous level through the symbiosis of LIIR's for different sub-images[7] in the chosen image partitioning at level i might be called the *initial representation* for level i. The representation derived from the initial representation by the process of inductive learning will be the LIIR for this level, and will serve to form the initial representation for level $i + 1$. One could think of the initial representation as derived from the LIIR by re-encoding the "parts" of the structs corresponding to the operations learned at the previous level by the new symbols (see Fig. 8). It is important to not to confuse our levels with those in multiresolution image representation schemes [32]. For level 0, the initial representation is obtained directly from the structured measurement device, whereas the initial representations at the higher levels are constructed by the inductive learning process. At present, the relationships between different levels (and, in particular, to the construction of the initial representations) are not sufficiently clear to us and form an important research topic. Finally, it is worth noting that the above image representation scheme is systematic, and, moreover, there is only one central process—inductive learning—that allows one to extract the symbols and to construct the representation at every level.

In this paper we illustrate the inductive theory of vision as it applies to low-level vision; however, its extension to different levels can be viewed as a consequence of Postulate 3. The application of the theory to low-level vision should be understood as the first step towards a systematic construction of image representation. Conventionally, high level vision processes have been considered as qualitatively distinct from the low-level processes, basically, due to the absence of any learning processes in low-level vision. In light of our theory, in which

[6] For more detailed discussion see Section 10.

[7] The corresponding LIIR's come from level $i - 1$ while the "symbiosis" is constructed at level i.

learning plays a critical and similar role at all levels of vision, the distinction between low and high level vision is blurred. In view of the latter, it becomes necessary to re-interpret these two terms. The new relationship that emerges between low and high level vision is that low-level vision processes "supply" the symbols necessary to construct the representation at the next (higher) level in the hierarchy. Since in our theory low-level vision embodies (inductive) learning, the distinction between low-level and high-level vision is simply due to this hierarchical construction of image representations. It is also important to note that at each level basic mechanism for the extraction of symbols, i.e., that of inductive learning is the same. Thus, inductive learning plays a fundamental role of determining the representation at each level.

Postulate 3. *All basic representations* in vision processes are constructed on the basis of the inductive image representation, which, in turn, is constructed by the inductive learning process (see Postulate 2). Thus, the inductive learning processes form the core around which all vision processes have evolved.

10 Examples

In order to clarify the basic concepts of the theory we will now illustrate the image representation scheme on several *simple* examples (Figs 3, 4, 5, 6). These examples should be seen only as illustrations of *some* of the basic concepts, rather than a "complete" application of the theory. Each example is chosen to represent a different type of a *sub-image* with only two intensity levels, i.e., all sub-images are binary (the cardinality of the struct alphabet is 2).

Since at present the appropriate measurement devices do not exist, each sub-image $\mathbf{I}_j$, $1 \le j \le 4$, is initially encoded as a set of "linear" strings of fixed length, i.e, the chosen structs in the transformation system $\mathbf{T}_j$ (for image $\mathbf{I}_j$) are strings of length 8 over the 2-letter alphabet. Note that although it is more appropriate to encode each sub-image as a set of planar graphs (as in Fig. 2), of necessity, we chose string representation motivated by the availability of the corresponding inductive learning algorithm [11]. It is interesting to note, however, that the results obtained (even with this inadequate, string, representation) are still non-trivial and instructive. It also goes without saying that a much more satisfactory image structure can be extracted using the more adequate, planar graph, representation. All examples below address LIIR only, since, as was already mentioned, in this paper we are not dealing with issues related to the image partitioning.

In Figs 3–6 four different binary sub-images are shown. In Fig. 3 (b) some structs in $\mathbf{T}_1$ for the sub-image $\mathbf{I}_1$ are shown, and in Fig. 3 (c) we show the LIIR, or $\mathbf{\Pi}_1$, that is extracted by the inductive learning process. From Fig. 3 (a), we constructed the 8 structs in $\mathbf{T}_1$ from intensity measurements[8] by denoting the intensity value 1 by "a" and value 0 by "b". We denote the intensity values 1 and 0 by "a" and "b" respectively in order to facilitate the *symbolic* interpretation of the strings. As was mentioned in Section 9.1, presently, of necessity, we have to rely on the initial numeric intensity encoding, which, we believe, will be possible to bypass in the future, when an appropriate measurement device will produce immediately the initial symbolic encoding[9]. For example, the first row of pixels in $\mathbf{I}_1$ is encoded as the string "abababab". Let us now define the only two operations in $\mathbf{T}_1$: insertion/deletion of a, $a \leftrightarrow \theta$, and insertion/deletion of b, $b \leftrightarrow \theta$, where θ denotes the null string. Again, the choice of operations was partly dictated by the existing learning algorithm (which is applied to all 4 sub-images) and may not be the most natural for any one of the 4 sub-images. Thus, the TS of the initial (structured) measurement device (for all 4 sub-images) includes the set of strings of length 8 over $\{a, b\}$ with two basic operations $a \leftrightarrow \theta$ and $b \leftrightarrow \theta$. ETS is then invoked during learning to extract the inductive class structure $\mathbf{\Pi}_1$ for $\mathbf{I}_1$ (and hence the local structure of the measurements) by dynamically changing the structure of $\mathbf{T}_1$ (see Sections 5, 9.2). As was mentioned above, the inductive learning process constructs the inductive class structure $\mathbf{\Pi}_1$, which, in turn, determines the local sub-image structure.

In all four examples, the set C^+ consists of the shown 8 strings, while C^- consists of the two "homogeneous" strings: *aaaaaaaa* and *bbbbbbbb*. Consider $\mathbf{\Pi}_1$, which is the result of the inductive learning process: it is given by the final set of operations $O_{fin} = \{a \leftrightarrow \theta, b \leftrightarrow \theta, bababa \leftrightarrow \theta\}$, the weight vector $\omega = (0.5, 0.5, 0)$, i.e., $\bar{\Omega}_{fin} = \{\omega\}$, and $\bar{C}^+ = \{abababab\}$. Note that the operations $a \leftrightarrow \theta$ and $b \leftrightarrow \theta$ are the initial operations, while the operation $bababa \leftrightarrow \theta$ is the operation which is learned (or discovered) by the ETS.

[8] Conventionally, under binary encoding, the black pixel is denoted by the intensity value 1 and white pixel, by the intensity value 0.

[9] In other words, at present, the initial transformation system $\mathbf{T}_u$ for the range of the unit (see Def. 1) is a numeric transformation system with two numeric structs 1 and 0.

As one can see, *the class structure, or (local) image structure is not postulated but rather determined as a result of the learning process*, i.e., the relationships between the structured measurements are now determined by the operations discovered by the ETS, and thus $\mathbf{\Pi}_1$ specifies the structure of the measurements. In connection with Section 5, we note again that it is the evolving structure of *ETS that allows one to dynamically update the set of operations of the basic mathematical structure to discover the combinative structure of the measurements (data) without postulating it.* Unlike the vector space based approaches to low-level vision, where the underlying structure (vector space) is fixed and cannot be changed, the underlying structure in ETS evolves. Thus, we can claim that the structure of the measurements is discovered rather than imposed on them a priori and in an arbitrary manner.

How should one think intuitively of a local sub-image representation? It is the sub-image's inductive class structure, symbolic sub-image representation, that provides the key to understanding the LIIR. The inductive learning process extracts, or discovers, the inductive class structure, i.e., the corresponding operations, which one could think of as the corresponding *symbols*. Moreover, the operations (or symbols) play the fundamental role in capturing and describing the local image structure. Since, these operations are constructed by the inductive learning process, or by the process of abstraction/generalization, they represent the local semantic information in the sub-image. These symbols (operations), in fact, facilitate representation of "the *concept* of a given sub-image". For example, for the sub-image in Fig. 3 (a), the left-hand side of $bababa \leftrightarrow \theta$ in Fig. 3 (c) captures the combinative structure that indeed represents the corresponding local generalization of the sub-image. Thus, inductive learning process discovers the representation—in this case $\mathbf{\Pi}_1$—which is semantically compact and is, basically, the only information needed to be stored *at that level.* Moreover, ETS not only extracts the "symbols" but also discovers the metric, and hence the geometry, on the set of structs of $\mathbf{T}_j$. Such a metric forms an absolutely integral part of LIIR and can also be used for further image processing (e.g., for image segmentation).

One of the most extensively studied problems in low-level vision is that of edge detection. Moreover, as was also mentioned in Section 7.2, edges, blobs, points, corners are only some of the infinitely many possible combinative structures in the image. All current approaches proposed to detect these structures are, in fact, extracting from the image an *a priori postulated class of combinative structures.* To illustrate how one can extract the edge structure with the help of inductive learning process, we next consider Fig. 4. As was mentioned above, the ETS is invoked during learning, with the set C^+ comprised of 8 structs shown in Fig. 4 (b) and C^- as above. It is quite interesting that the edge structure that is present in the sub-image in Fig. 4 (a), *is extracted, or discovered, by the inductive learning process* and is represented in Π_2 (see Fig. 4 (c)): by the final set of operations $O_{fin} = \{a \leftrightarrow \theta, b \leftrightarrow \theta, baaa \leftrightarrow \theta, bbba \leftrightarrow \theta\}$, the weight vector $\omega = (0.5, 0.5, 0, 0)$, i.e., $\bar{\Omega}_{fin} = \{\omega\}$, and $\bar{C}^+ = \{aaaaabbb\}$. It is not difficult to see that the left-hand sides of the last two operations together capture the concept of edge. Fig. 5 and Fig. 6 illustrate the discovery of some additional and somewhat less trivial types of (combinative) structures. Note, again, *that none of these structures were prepostulated, they were rather discovered by the inductive learning process.*

We would like to reiterate that one simply loses the capability to discover the different types of combinative structures (that may be present) in the image, if one opts for a vector representation of the image, since, as was previously argued in Section 7.2, *in a vector space model one is postulating a priori, a very special (numeric) combinative structure to be extracted from the image, which, in turn, prevents one from the "discovery" of any other structure.*

We now address very briefly some issues related to the connection between two subsequent levels (see Figs. 6, 8). In Fig. 8 the level 1 representation is constructed by "re-encoding" the parts of the structs corresponding to the operations learned at level 0 by new symbols, e.g., the part $abaa$ is replaced by c, part $baab$ is replaced by d, where the "parts" are the the left-hand sides of operations $abaa \leftrightarrow \theta$ and $baab \leftrightarrow \theta$ learned at level 0 (see Fig. 6). Following the re-encoding, the inductive learning process can be invoked at level 1 to extract the operations for this level. Such a construction can be extended systematically for all levels.

The above examples illustrate the postulated fundamental role of inductive learning in vision: the inductive learning process by means of ETS discovers the image structure and, in turn, the (combinative) structure of the objects. Again, we note that the proposed image representation scheme is systematically uniform for all levels in that it relies only on one central process—inductive learning—that allows one to extract the appropriate "symbols" and to construct the corresponding representation at each level.

11 Conclusion

First, we propose to base the inductive theory of vision on the three postulates (given in the abstract). Since, we postulate that the combinative object structure lies at the foundation and forms the core of our ability to perceive the external world, we base the theory on the inductive mechanism that is capable of capturing combinative object/event structure based on a small representative set. While classical mathematical models, including ANN's, cannot support such an inductive mechanism, the evolving transformation system (ETS) model can. We propose that to embody the inductive mechanism one needs to introduce a fundamentally new concept of the structured measurement/recording device which should replace the classical measurement device. The concept of structured measurement device appears to us so revolutionary that at present we can barely imagine its full implications. It is important to note that the recordings of such a device are structured entities (resembling molecules) and not numbers as is the case with the classical measurement devices. These structured entities are much more "real" than the numbers, and so in some sense the proposed model operates with more "concrete" entities.

We have explained why the inductive learning mechanism should be considered as the central mechanism around which all other vision processes have evolved. Basically, its central role can be explained by the fact that combinative object structure at any level can be captured and encoded only with its help. Moreover, the very concept of object structure can be understood through the concept of inductive class representation. We contend that *the triad of image structure, the corresponding mathematical structure, and the (combinative) structure of objects is precisely what computational vision should be about. And it is the inductive learning processes, through the discovery of the class structure, that relates the members of the triad.* Thus, we propose that the main objective of vision processes is to determine *inductively* the (combinative) image structure. This structure, in turn, determines the (combinative) object structure, where the term "(combinative) structure" should be interpreted *only* through the corresponding mathematical structure as it is currently practiced in mathematics [17] [18].

Finally, we note that the proposed framework allows for a very uniform treatment of various levels in vision, in which each new level re-encodes the previous representation by adding new "symbols", inductively constructed at the previous level, to the previous alphabet of "symbols" in a systematic manner independant of the level involved. This leads to a systematic hierarchical framework for image representation.

Thus, we believe that this paper outlines, for the first time, a viable realization of the vision model as was envisioned by such pioneers as Hermann von Helmholtz and Horace Barlow.

References

[1] D. Marr (1982). *Vision.* W. H. freeman and Company, San Francisco.

[2] V. S. Nalwa (1993). *A guided tour of computer vision.* Addison-Wesley Publishing Company, New York.

[3] B. K. P. Horn (1986). *Robot vision.* MIT press, Cambridge, Massachusets.

[4] P. W. Pachowicz (1994). Semi-autonomous evolution of object models for adaptive object recognition, *IEEE Trans. System Man and Cybernetics,* Vol. 24, No. 8, 1191–1207.

[5] J. W. Bala (1993). Learning to recognize visual concepts: Development and implementation of a method for texture concept acquisition through inductive learning, Ph.D thesis, George Mason University, Fairfax, Viginia.

[6] S. Coren, L. M. Ward and J. T. Enns (1994). *Sensation and perception.* Harcourt Brace College Publishers, New York.

[7] D. Marr and E. Hildreth (1980). Theory of edge detection, *Proc. R. Soc. Lond.,* B, 207, 187–217.

[8] R. Kahl (1971). *Selected writings of Hermann von Helmholtz.* Wesleyan University Press, Middletown, Connecticut.

[9] H. B. Barlow (1990). Conditions for versatile learning, Helmholtz's unconscious inference, and the task of perception, *Vision Res.,* Vol. 30, No. 11, 1561–1571.

[10] L. Goldfarb (1996). What is inductive learning? To appear in the Proceedings of Workshop *What is inductive learning? A foundation for AI and Cognitive Science* held in conjunction with the *11th Canadian biennial Artificial Intelligence* conference, Toronto, May 20-21.

[11] L. Goldfarb and S. Nigam (1994). The unified learning paradigm: A foundation for AI. In *Artificial Intelligence and Neural Networks: Steps toward Principled Integration,* eds. V. Honavar and L. Uhr, Academic Press, Boston.

[12] L. Goldfarb (1990a). On the foundation of intelligent processes - I: An evolving model for pattern learning, *Pattern Recognition*, 23, 596–616.

[13] L. Goldfarb (1992). What is distance and why we need the metric model for pattern learning? *Pattern Recognition,* 25, 431–438.

[14] L. Goldfarb, J. Abela, V. C. Bhavsar and V. N. Kamat (1995). Can vector space based learning models discover inductive generalization in a symbolic environment? *Pattern Recognition Letters,* Vol. 16, No. 7, 719–726.

[15] H. Reeves (1993). *Malicorne. Earthly reflections of an astrophysicists.* Stoddart Publishing Co. limited.

[16] L. Margulis and D. Sagan (1991). *Microcosmos.* Simon and Schuster, New York.

[17] N. Bourbaki (1970). *Elements of Mathematics. Algebra, Part I.* Addison Wesley, Reading, Massachusetts.

[18] I. M. Yaglom (1986). *Mathematical Structures and Mathematical Modelling.* Gordon and Breach Science Publishers, New York.

[19] John Von Neumann (1963). *Collected works, Vol. 6.* Pergamom Press, New York.

[20] E. Schrodinger (1992). *What is life?* Cambridge University Press, Cambridge, UK.

[21] L. Goldfarb (1993). On some mathematical properties of ETS model. Technical Report TR93-079, Faculty of Computer Science, University of New Brunswick, Fredericton, Canada.

[22] E. G. H. Landau (1951). *Foundations of analysis.* Chelsea Publishing Company, New York.

[23] A. I. Mal'cev (1970). *Algorithms and recursive functions.* Wolters-Noordhoff publishing, Groningen, Holland.

[24] M. Porat and Y. Zeevi (1988). The generalized Gabor scheme of image representation in biological and machine vision, *IEEE Trans. Pattern Anal. Machine Intell.,* Vol. 10, No. 4, 452–467.

[25] D. Marr (1976). Early processing of visual information, *Phil. Trans. R. Soc. Lond. B,* 275, 483–524.

[26] H. H. Nagel. Principles of (low-level) computer vision. In *Fundamental in Computer Understanding: Speech, Vision, and Natural Language,* eds. J. P. Haton, Cambridge University Press, Cambridge, UK.

[27] J. J. Koenderink (1984). The structure of images, *Biol. Cybernet.,* Vol. 50, 363–370.

[28] L. S. Davis (1975). A survey of edge detecting techniques, *Comput. Graphics and Image Process.,* 4, 248–270.

[29] P. Perona (1995). Deformable kernels for early vision, *IEEE Trans. Pattern Anal. Machine Intell.,* Vol. 17, No. 5, 488–499.

[30] J. J. Koenderink and A. J. van Doorn (1987). Representation of local geometry in visual system, *Biol. Cybernet.,* 55, 367–375.

[31] A. F. Korn (1988). Toward a symbolic representation of intensity in images, *IEEE Trans. Pattern Anal. Machine Intell.,* Vol. 10, No. 5, 610–625.

[32] S. G. Mallat (1989). A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Trans. Pattern Anal. Machine Intell.,* Vol. 17, No. 5, 488–499.

[33] P.J. Burt and E. H. Edelson (1983). The laplacian pyramid as a compact image code, *IEEE Trans. Commun.,* Vol. COM-31, 532–540.

[34] D. H. Hubel and T. N. Weisel (1968). Receptive fields and functional architecture of monkey striate cortex, *J. Physiology., Lond.,* 195, 215–243.

[35] F. W. Campbell and J. G. Robson (1968). Application of fourier analysis to visibility of gratings, *J. Physiology., Lond.,* 197, 551–556.

[36] M. B. Sachs, J. Nachmias and J. G. Robson (1971). Spatial-frequency channels in human vision, *J. Opt. Soc. Am.,* Vol. 61, No. 9, 1176–1186.

[37] L. Maffei and A. Fiorentini (1973). The visual cortex as a spatial frequency analyser, *Vision Res.,* Vol. 13, 1255–1267.

[38] J. G. Daugman (1985). Uncertainity relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, *J. Opt. Soc. Am.,* Vol.2, No. 7, 1160–1169.

[39] S. Marcelja (1980). Mathematical description of the responses of simple cortical cells, *J. Opt. Soc. Am.,* Vol. 70, No. 11, 1297–1300.

[40] T. Bossomaier and A. W. Snyder (1986). Why spatial frequency processing in the visual cortex? *Vision Res.,* Vol. 26, No. 8, 1307–1309.

[41] D. J. Field (1987). Relations between the statistics of natural images and the response properties of cortical cells, *J. Opt. Soc. Am.,* Vol. 4, No. 12, 2379–2394.

[42] D. J. Field (1994). What is the goal of sensory coding? *Neural Computation,* 6, 559–601.

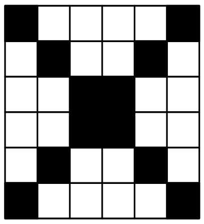
[43] K. A. Stevens and A. Brookes (1987). Detecting structure by symbolic constructions on tokens, *Computer Vision, Graphics, and Image Processing.,* 37, 238–260.

[44] E. B. Goldstien (1989). *Sensation and perception.* Wadsworth Publishing Company, Belmont, California.

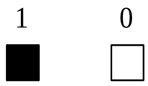
[45] U. Grenander (1981). *Regular Structures. Lectures in pattern theory, Volume 3.* Springer-Verlag, New York.

[46] W. D. Callister, Jr (1991). *Material Science and Engineering.* John Wiley and Sons, Inc., New York.

[47] B. K. P. Horn and R. W. Sjoberg (1979). Calculating the reflectance map, *Applied Optics,* Vol. 18, No. 11, 1770–1779.

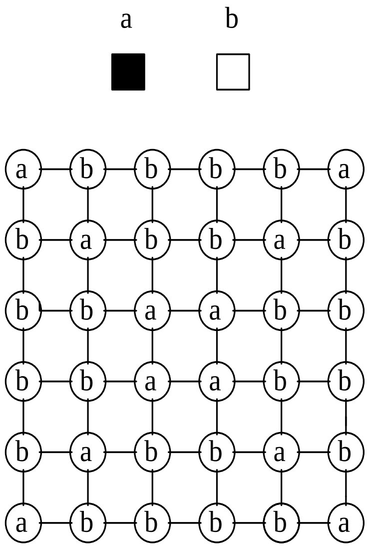


a) A sub-image



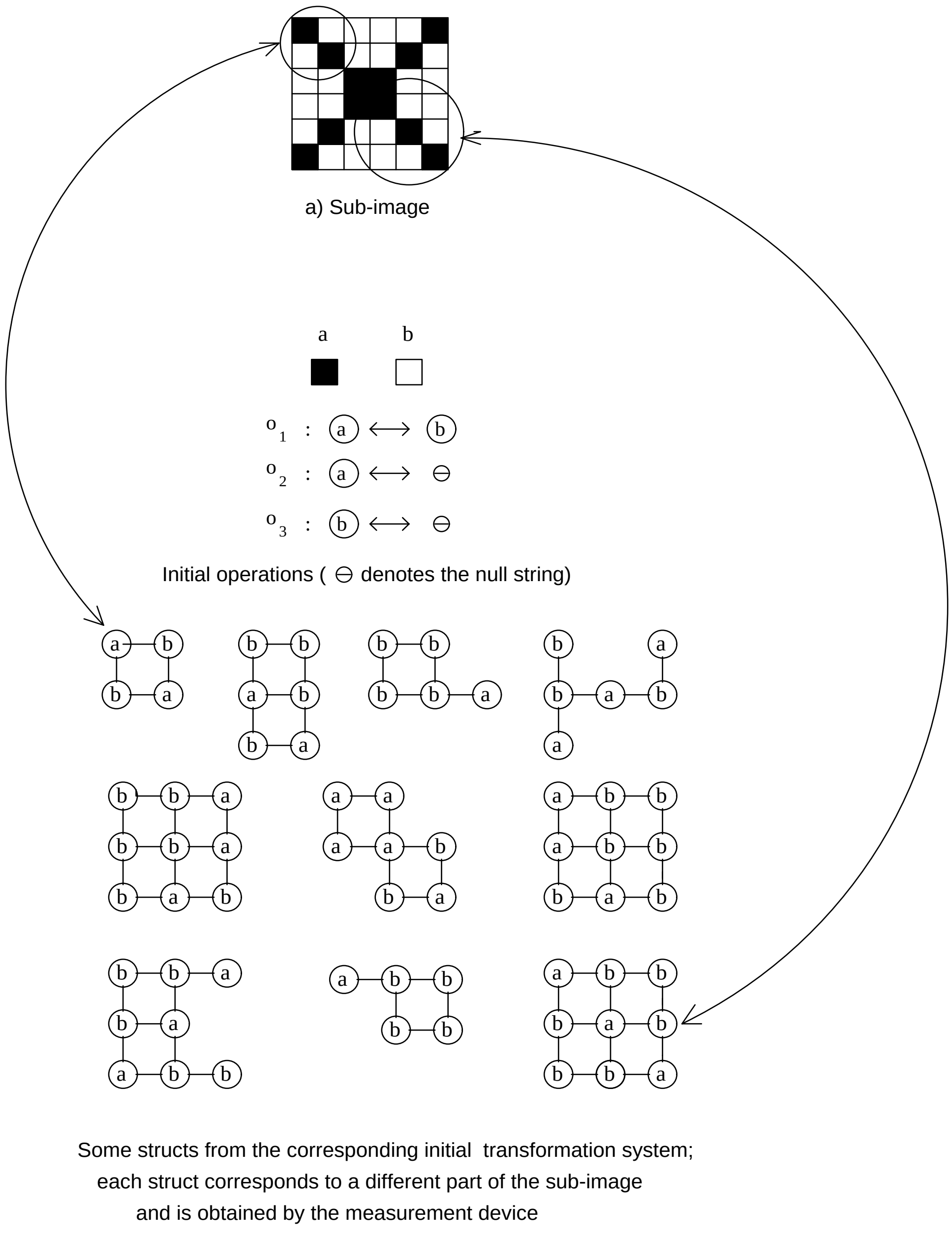
< 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1 >

b) Vector representation of the sub-image in (a) (without vector operations)



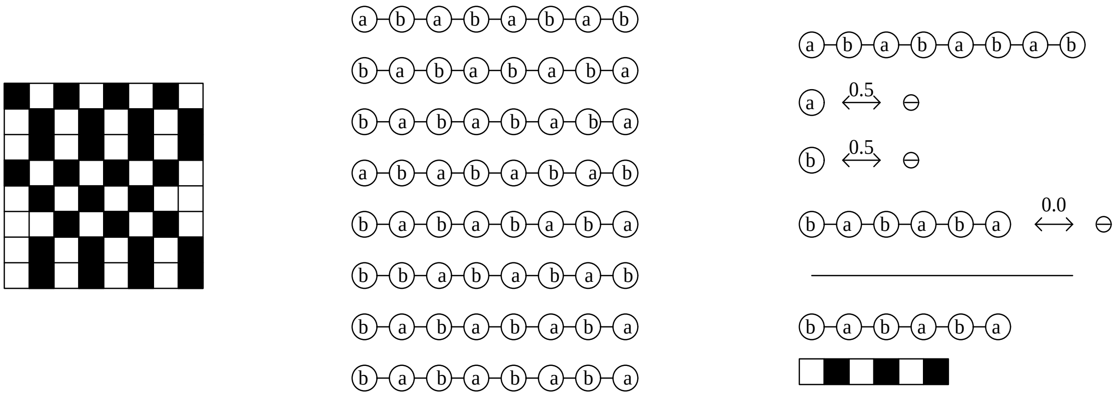
c) Struct representation of the sub-image in (a) (without the operations)

Figure 1: Sub-image and its two representations



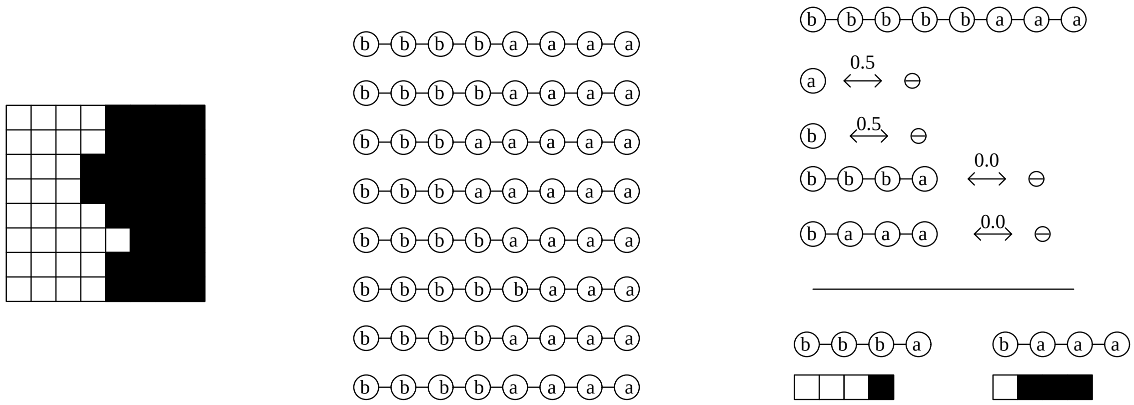
b) A initial transformation system for the sub-image in (a).

Figure 2: An example of a sub-image and its initial transformation system



a) Sub-image $\mathbf{I}_1$ b) 8 training structs in $\mathbf{T}_1$ c) Inductive class structure of $\mathbf{I}_1$ with weights shown above each operation ($|\bar{C}^{+}| = 1$)

Figure 3: Inductive learning of local structure for sub-image $\mathbf{I}_1$



a) Sub-image $\mathbf{I}_2$ b) 8 training structs in $\mathbf{T}_2$ c) Inductive class structure of $\mathbf{I}_2$ with weights shown above each operation ($|\bar{C}^{+}| = 1$)

Figure 4: Inductive learning of local structure for sub-image $\mathbf{I}_2$

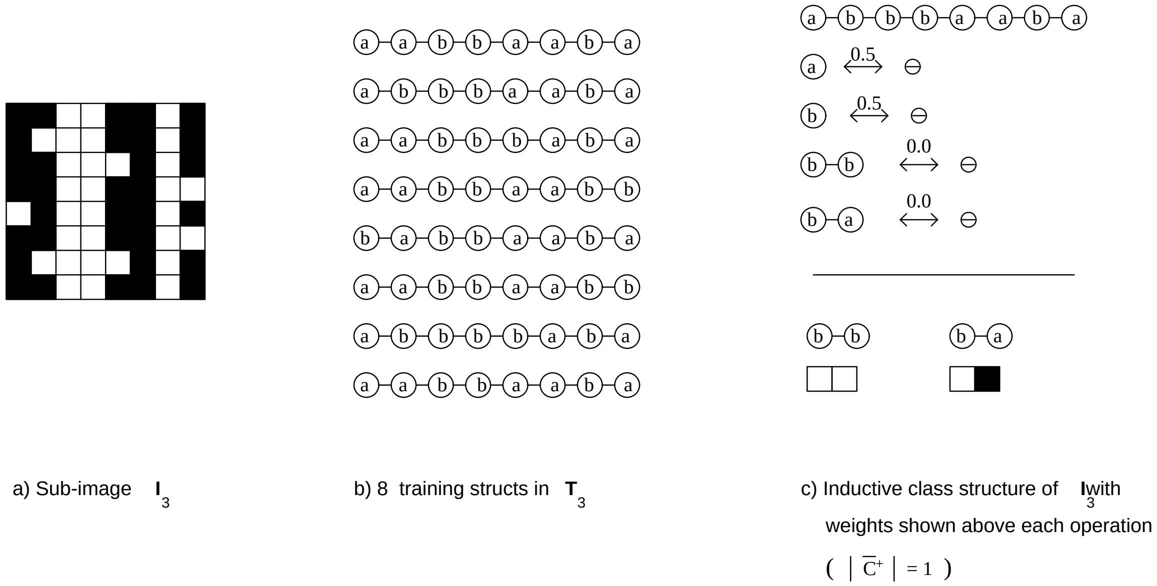


Figure 5: Inductive learning of local structure for sub-image $\mathbf{I_3}$

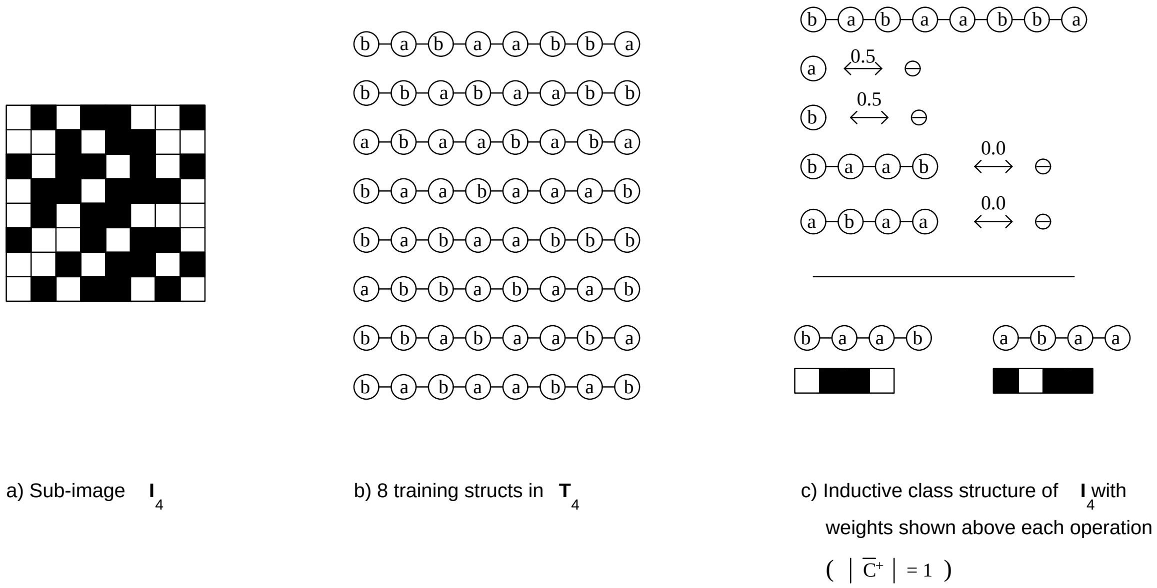


Figure 6: Inductive learning of local structure for sub-image $\mathbf{I_4}$

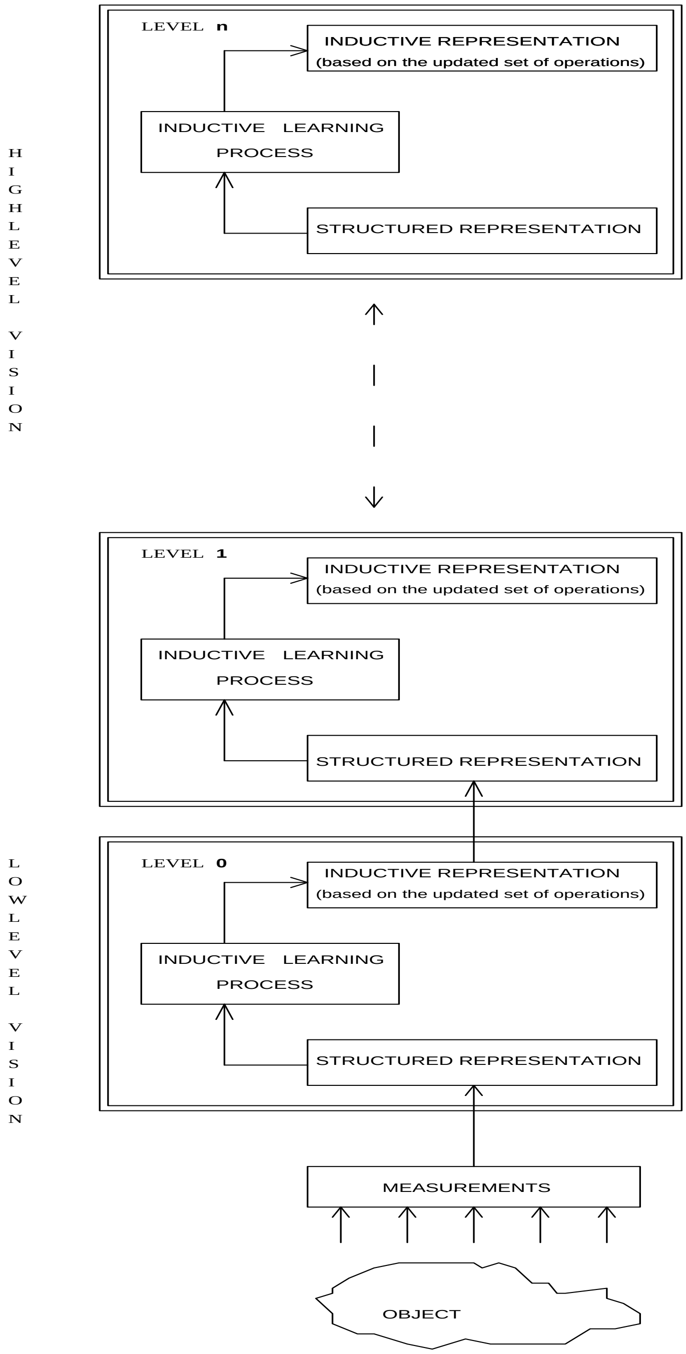


Figure 7: Hierarchical combinative image structure based on the inductive theory of vision

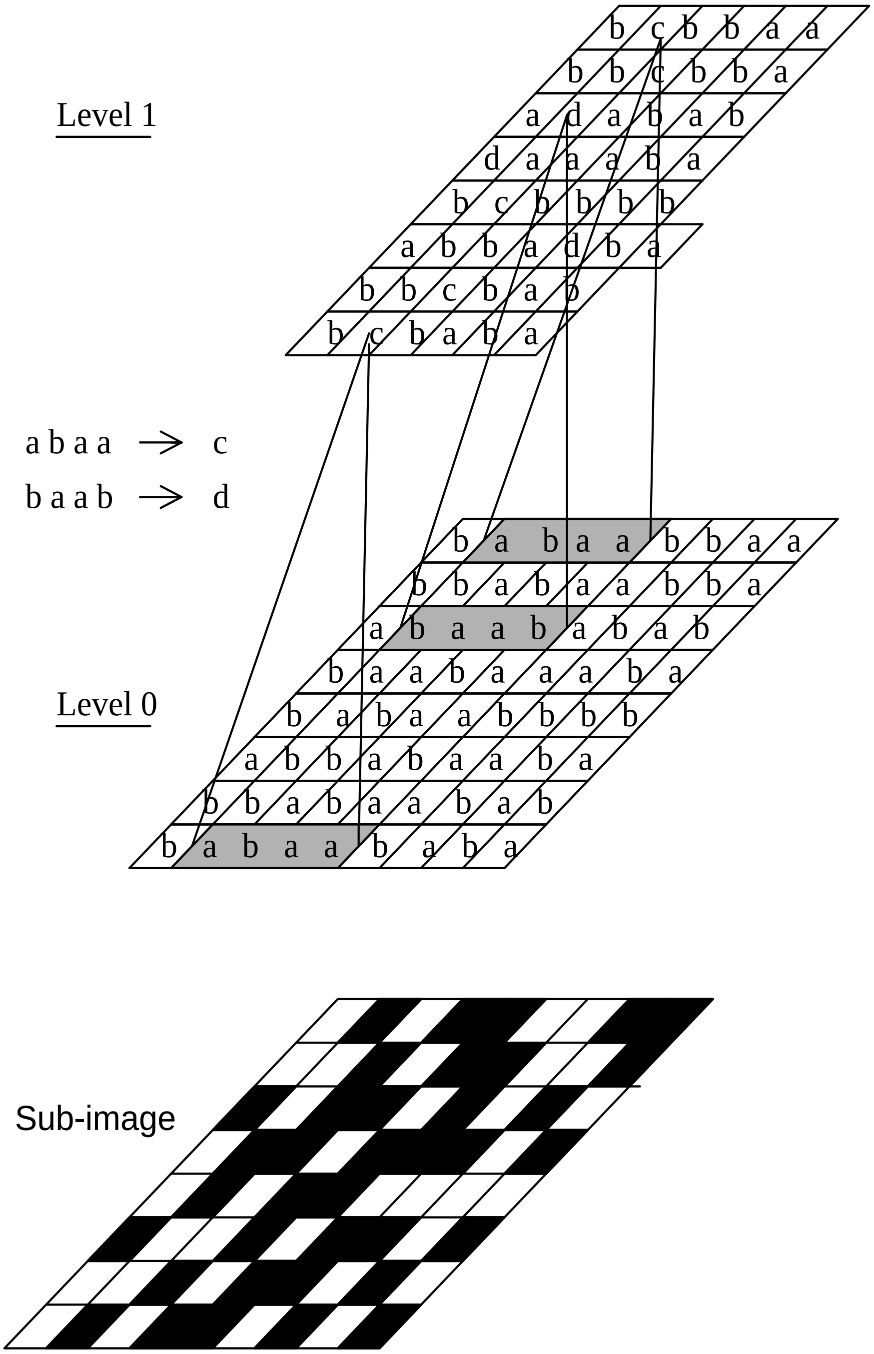


Figure 8: An example of hierarchical sub-image representation (see Fig. 6)