QUARTERLY OF APPLIED MATHEMATICS

Vol. XXVII APRIL 1969 No. 1

FOUNDATIONS OF PATTERN ANALYSIS*

BY

ULF GRENANDER

Brown University

Contents

*Received March 15, 1967. This work was supported in part by the IBM Corporation through a contract with Brown University and by NSF Grant GP 7347.

1. Introduction. The purpose of this paper is to lay a foundation for the formal analysis of patterns. It is hoped that such an analysis would be of help in trying to recognize patterns.

What, then, is a pattern? Or, rather, what should we make the term signify? If words mean anything, it may be instructive to take a look at some of the words one uses for this notion.

In everyday language we use the word a great deal, allowing it to mean many different things. The most usual meaning of the word is perhaps "design" or "style of marking", but it can also mean "sample" or "copy". The French "dessin", German "Muster" and Swedish "mönster" can mean design but have, in addition, a flavor of something desirable, something one ought to imitate. The same is true of the Greek "παράδειγμα" and the Latin "exemplum", while the Russian "шаблон" indicates something that can be copied and reproduced in numbers. This double meaning gives us a clue: we shall have two types of pattern analysis. One of them is *descriptive*; it attempts to describe patterns mathematically and to study recognition processes observed in nature. The other is *normative*; given certain patterns, how do we construct recognition processes that are best (or good) in some sense? While being aware of this distinction, we shall try to develop a single mathematical theory that can be used for both purposes.

The word "pattern" is also used abundantly in the extensive literature on pattern recognition. There it is given a more precise meaning, but one that varies from case to case. The most usual one is perhaps that of alphanumeric symbol: letters and numerals from some specified writing system. Indeed, a considerable part of work dealing with pattern recognition has concerned itself with the problem of how to achieve automatic recognition of typed, handwritten or handprinted characters. In the same category we find the study of Chinese ideographs or other man-made symbols intended for communication, such as Morse code. A different sort of pattern is found in the study of fingerprints, electrocardiograms (ECG) and speech: these are natural phenomena that we wish to analyze into basic components.

Our goal is to develop a model flexible enough to make it possible for us to discuss patterns in general: *to give us a precise language in terms of which we shall be able to analyze and describe patterns.*

By patterns we shall understand the following. Starting from a set of objects (called *images*) and a set of rules by which we can transform an image into others, we shall say that two images are *similar* if one can be transformed into the other by applying some of our rules successively. By a pattern we could mean a *class of mutually similar images*. We shall use a more general definition below, but already this narrower concept should give a good indication of what we are looking for.

The images are formed from certain primitive building blocks or atoms that we shall call *signs*. We are allowed to combine signs according to certain rules that will be discussed in detail below. It will often be natural to group these signs into disjoint categories, called *subroutines*, and the analysis of a pattern will proceed through the images of which it consists, representing each image in terms of subroutines and signs.

A large part of current work in pattern recognition deals with the construction of learning algorithms. Starting from rudimentary assumptions about the set of objects being observed, the algorithm should modify itself successively as more experience is being collected. As time goes on the algorithm should learn to distinguish between different patterns more and more accurately. The uncertainty involved here is often described in statistical terms and much of this research effort could be classified as studies of sequential decisions functions. Note that very little structure is assumed about the patterns in such studies.

Our approach is different. The idea of sequential decision is certainly of general relevance in pattern recognition and especially so when we deal with patterns that are not well known beforehand. It is only part of the picture, though. The author believes that in order to be able to find efficient recognition schemes we must be prepared to devote more study to the origin and *generation of the objects that we observe*. Thus we shall base the construction of the recognition algorithm on *analysis of the patterns*, tailored after the particular case in which one happens to be interested. In each such case we shall restrict our attention to certain algebraically defined patterns. Similar ideas can be found scattered throughout the literature on the subject, but they are seldom pursued systematically. An unusually precise formulation of this idea can be found in Narasimhan [1], where bubble chamber tracks are analyzed into components.

In this paper the emphasis will be on the analysis part of the problem. Once this is clear it is hoped that we shall have a tool useful also for sequential procedures; the latter will not be studied here.

The way in which we try to build a mathematical model of patterns will of course lead us to algebraic considerations. But we can also look upon it from another angle. We introduce images (and operations upon them) not just through enumeration but by specifying rules for generation and transformation. A reader familiar with current trends in linguistics will see that the approach has a great deal in common with generative grammars. He may wish to think of the subroutines of signs as related to linguistic concepts such as morphemes, verbs, nouns and higher syntactic categories. If we deal with reasonably complicated classes of images, it is clear that we can hope to describe them more economically by giving the rules of generation rather than by presenting a long list of images. To cite a famous linguistic *dictum*—"make infinite use of finite means".

But let us not claim too much. The sort of patterns we have in mind may be difficult enough to handle, but their complexity does not approach that of natural language. Natural language, with its almost unlimited capability of expressing the products of human intelligence, has a logical structure that is so much richer than that of the patterns we are studying that a direct comparison would not be fair. A better analogue might be some of the formal languages. The mere fact that we are at all able to analyze our patterns in a precise way indicates that our structures are simpler than those of the linguist.

Anyway, our approach, as well as the concepts and terminology we are going to use, is influenced by linguistics. We shall call the different models to be developed *grammars of patterns*. Admittedly, these grammars bear only superficial resemblance to those of the linquist, but the attitudes they represent are similar. Such a grammar shall specify the *medium* in which we work—the signs—as well as the *rules of generating images*. But this is not enough. We also must prescribe *rules for transforming images*, the similarity transformations. The latter rules will tell us to what pattern a given image belongs and it may be tempting to look upon them as means to extract the meaning of the observed image.

The model that we have sketched roughly above is quite general. In this paper, however, we shall restrict it considerably by assuming that the rules of transformations simply consist of a *group G of transformations*. Even with this restriction the grammars obtained cover a number of widely different cases. Similarity now means invariance with respect to G, and we recognize equivalence classes, etc. Behind this there is a long history of models where invariance plays the main role. In physics we have the theories of relativity and in geometry Klein's Erlanger Program under the motto: "Geometrische Eigenschaften sind durch ihre Unveränderlichkeit gegenüber den Transformationen der Hauptgruppe characterisiert." Let us also remind the reader of Herman Weyl's beautiful book on symmetry, where he tries to describe the aesthetic quality through invariance and repetition (but not quite). Perhaps the closest model is from the theory of perception; see the celebrated paper by Pitts and McCulloch [1], and a recent attempt to analyze visual perception by Hoffman [1].

The linguistic (or algebraic) model described above will be denoted by $\mathcal{G}$. If $\mathcal{G}$ is specified in a sufficiently simple way, we can look for recognition algorithms to classify the images of $\mathcal{G}$ according to what pattern they belong to. In principle this is easy, at least if we do not go into the question of decidability. In this paper we shall not take a finitistic attitude, and decidability will not be discussed. In the opinion of the author a more serious deficiency is that we have not tried to measure the amount of computation that would be required to implement any particular recognition algorithm using given means of computation.

Unfortunately, real life patterns can seldom be described through such a grammar. Instead, we shall let $\mathcal{G}$ represent only the first stage of the analysis and we shall call it a *pure grammar*, with pure images and patterns. In the next stage we add to the grammar a *deformation mechanism* $\mathfrak{D}$ mapping the set $\mathfrak{I}$ of pure images into a set $\mathfrak{I}^{\mathfrak{D}}$ of *deformed images*. Usually $\mathfrak{I}^{\mathfrak{D}}$ is much larger and includes $\mathfrak{I}$. The procedure of recognition then looks as follows:

$$\mathcal{P}_r \ni I \underset{\mathfrak{D}}{\longrightarrow} I^{\mathfrak{D}} \underset{\phi}{\longrightarrow} \mathcal{P}_s^* \tag{1.1}$$

We start from some image I in the pure pattern $\mathcal{P}_r$, then I is deformed into $I^{\mathfrak{D}}$. The only thing that we can observe is $I^{\mathfrak{D}}$ and the recognition function ϕ should map $I^{\mathfrak{D}}$ into the set of pure patterns. Of course we want ϕ to be good in some well-defined sense, but more about that later. It is first when we take the deformation mechanism into account that the model attains its full flavor of pattern analysis. This model will be called a deformation grammar and denoted by $\mathcal{G}_{\text{def}}$. In the terminology of de Saussure, $\mathcal{G}$ would correspond to "langue" and $\mathcal{G}_{\text{def}}$ to "parole".

It is still not clear how we can compare two alternative recognition functions ϕ_1 and ϕ_2; which one should be considered better than the other? If the set $\mathfrak{I}^{\mathfrak{D}}$ of deformed images has a metric and if $\mathfrak{I} \subset \mathfrak{I}^{\mathfrak{D}}$, it may appear reasonable to ask that the recognition function should *recognize that* (or those) *pure images and pattern(s) that are at the smallest possible distance from the observed image* $I^{\mathfrak{D}}$. Behind this argument there is obviously some idea that small deformations in $\mathfrak{D}$ are more likely (in some unspecified sense) than large ones. This idea will be followed up below by introducing the notion of *effort* of deformation. A deformation with a metric as above will be called a *metric deformation grammar* and denoted by $\mathcal{G}_{\text{met}}$.

The trouble with the metric deformation grammar is that it assumes that we are given a metric naturally belonging to the deformation mechanism. If we do not have access to a metric tailored to the grammar $\mathcal{G}_{\text{def}}$ then we must look for other avenues of approach.

A most promising but also rather difficult way out is to postulate the existence of a probability distribution P on the set $\mathfrak{D}$ of all deformations. It should be borne in mind that if we aspire to some degree of realism when building such a *probabilistic deformation grammar* $\mathcal{G}_{\mathrm{prob}}$ we must be prepared to deal with some rather sophisticated statistical structures. Actually, the sort of problem met with will then be of the type "inference in stochastic processes", if "stochastic process" is interpreted in a very broad sense. Indeed, we may now look upon $\mathfrak{J}$, the set of pure images, as a parameter space, and upon $\mathfrak{J}^{\mathfrak{D}}$, the set of deformed images, as the sample space. The recognition problem then appears in the guise of an estimation or multiple decision problem with certain invariances. Both the parameter and the sample space can be extremely complicated; note the way in which they have been constructed starting from the original signs. We could not hope to be able to deal with this large family of models merely by applying decision theoretic techniques in a routine manner. On the contrary, much of our effort will be spent in formulating such models in concrete cases, to make them flexible enough to fit real patterns. We have a long way to go before we have such an arsenal of tools, and this paper should be considered only as the first systematic attempt to arrive at this goal.

The set of probabilistic deformation grammars presents us with an arsenal of attractive models. It should not be assumed, however, that a $\mathcal{G}_{\mathrm{prob}}$ is always the thing to look for. It may very well happen that it does not make sense to impress any probability measure P on $\mathfrak{D}$, or, in case it does, it may be hard to find P. Actually, this last problem (exploratory grammar) is of considerable practical interest, but will not be pursued systematically here.

Let us take a look at the content of this paper. It is divided into four parts. The first two of these deal with pure patterns and the last two with deformed patterns. The first two are quite elementary and only give a simple and natural background to parts III and IV.

We start in sections 2 and 3 by introducing, in a precise manner, what we mean by subroutines and configurations, and discuss the rules of generating the image algebra: the set of objects that we can observe. One has to be careful here to distinguish between the way of generating these objects and what properties of the objects that are available to the subject who plays the role of the observer. This distinction forces us to formulate explicit rules of identifying configurations: if two configurations cannot be distinguished by the observer, we will consider them identical even if they have been generated in different ways. This leads to the concept of an *image*, incorporating all the observable quantities.

Let us try to analyze the observable images into constituent parts. This will parametrize the images and we look for parameters that are sufficient for this but at the same time are as simple as possible. It may happen that these parameters can be chosen as the signs of the configuration. If so, then we may be satisfied, but in general the signs are not observable themselves. It is then necessary to look for other *representations of the image* and this will be studied in section 3, especially when the representation can be expressed in terms of the subroutines of the configurations. This problem requires more attention than we have given to it.

By a *pattern* we shall mean a set of images which is invariant under similarity transformations. To exemplify this, let us choose these sets as the equivalence classes modulo G. We could then write the set $\mathcal{P}$ of patterns as $\mathcal{P} = \mathfrak{J}/G$. Very often in this paper we shall assume that the patterns have been introduced in just this way; the reader should realize, however, that it is a special class. At any rate, in section 5 we study the notion of patterns and discuss how to recognize the pattern to which an observed image belongs.

Up to here the discussion has been general and a bit abstract. In Part II we shall study particular grammars of patterns, and we start in section 6 by examining some practical cases. We do not intend to study any of them in detail; we merely want to give the reader some feeling for what sort of models may be required. In sections 7–12 we present a number of such models, but, of course, this list is far from complete.

The simplest possible case, that of *permutation patterns*, is considered in section 7. This section serves mainly to make the reader familiar with the application of the general notions from Part I.

In section 8 we get to more interesting structures, models for *line patterns*. Both *Euclidean* and *differential line patterns* seem to present useful and flexible models and should be developed further. Especially, one should be aware of the importance of the singular points of the patterns for recognition purposes, both in pure and deformation patterns.

Set patterns, for which the boundary of the image is well described through a tractable grammar, are discussed in section 9. Actually, such problems are best expressed in the terminology of line pattern grammars, and that is the approach which we shall choose.

By a natural generalization we get from line patterns to the important class of *boundary patterns* (see section 10). Here we start from the specification of certain boundary conditions, which play the role of the singular points earlier. In addition, we must prescribe what subroutine has generated the sign associated with a given boundary condition.

An even more general notion is that of *contrast patterns*, which includes all the cases that we have discussed. Admittedly, it is a natural model only when its basic elements, the background and contrast, are simple enough. This is the topic of section 11.

Finally we examine some non-commutative algebras forming the *time patterns* of section 12. This concludes our treatment of the pure images and patterns.

In this way we have arrived at a natural and attractive model of pure patterns. Although it is certainly far from complete, its main outline is clear and simple. When we turn to deformed patterns, and these are the truly important ones from a practical viewpoint, we meet difficulties of a different magnitude. At the same time we meet a number of exciting problems that present a challenge to the model builder as well as to the pattern analyst.

In Part III we shall formulate some deformation models and indicate how these models can lead us to sensible recognition algorithms. But first we must get some idea of how such deformation grammars can be constructed, and this is discussed in general in section 13, as well as desirable qualities for the recognition functions such as invariance of different kinds.

When we are given a distance function in the image algebra $\mathfrak{J}^{\mathfrak{D}}$ we can study minimum distance recognition. Of special interest is the case in which the recognition function is *invariant*, since then our mapping leads directly back to pure patterns rather than *via* the detour of a pure image. Sometimes the distance can be introduced in a convincing way through an *effort function* on $\mathfrak{D}$. This describes how different deformations may require different power, energy, etc.

In section 15 we get to the core of the theory: the *probabilistic deformation grammars*.

Now it depends on what sort of *a priori* knowledge we have concerning the frequency of occurrence of the different grammatical entities. If we know both the frequencies of the pure images and of the various possible deformations things are simple, at least in principle (computationally they may be cumbersome). Unfortunately, the first of these distributions is likely not to be known; it may not even make sense to talk about it. Instead we may only know the distribution of the pure patterns, and then it takes some care in formulating the problem of optimum recognition.

Once this problem is well understood we can go ahead to study particular cases, which is done in Part IV. It is not very profitable to do this *in abstracto*, but a few general principles are advanced in section 16. These will not lead to completely specified probabilistic deformation grammars, but serve only as a guide in choosing the form of the model.

This is illustrated for permutation patterns in section 17, and it is shown how some reasonable recognition schemes for such a $\mathcal{G}_{\text{prob}}$ can also be viewed as originating from a $\mathcal{G}_{\text{met}}$ with the criterion minimum distance recognition. This reduction has also been carried out in some of the following sections.

Now the grammars become more sophisticated. To reduce the recognition to as tractable a form as possible, we shall look for simple parameters that may describe the deformed images only partially but contain all the relevant information as far as recognition is concerned. This will be done for some probabilistic deformation grammars, both for Euclidean and differential line patterns, in section 19. In that connection we also treat some set and boundary patterns, especially the case in which the signs are given as convex sets and the image algebra $\mathfrak{J}^{\mathfrak{D}}$ is of a quite different nature from that of $\mathfrak{J}$ and does not contain the latter.

In a sense the contrast patterns are the most general ones of those that we study. Therefore it seems reasonable to pay special attention to their deformation grammars. A few such grammars are studied in section 19, both when the background space and when the contrast space are subject to deformation. We shall find some remarkably simple reductions to recognition by maximum correlation, etc.

The only non-commutative image algebra that we have investigated is that of time patterns. A few probabilistic deformation grammars for time patterns are suggested and analyzed in section 20. One way of doing this is *via* the notion of *subjective time*. Another is by allowing the *machine producing the subroutines to be imperfect*: this introduces deformations in a way that is believed to have considerable interest.

The many deformation grammars suggested in Part IV should help the reader to see the possibilities of the present approach to pattern analysis. It is also clear, though, that this is only a beginning, and that we probably need grammars that are a good deal more sophisticated than these to tackle the really hard pattern recognition problems with success.

The restriction of these grammars to groups of similarity transformations is severe. The author is working on more general grammars in which this restriction has been removed; this work will be reported in a separate publication (Grenander [2]).

Several other parallel studies are in progress. One, in cooperation with W. Freiberger, is concerned with computer realization of certain probabilistic deformation grammars. After generation, the images will be displayed on the CRT of an I.B.M. 2250 (see Freiberger and Grenander [1]).

A more general treatment of patterns (not assuming similarity groups) will appear

(Grenander [3]) in which a series of particular patterns appearing in applications will be discussed.

PART I. PURE PATTERNS

2. Signs and Configurations. The elements from which we start when forming patterns will be called *signs*. The single sign, generically denoted by s, from the set S of all signs that we consider, is quite arbitrary in character. This and other notions to be introduced here will be exemplified in Part II; here we will deal with them abstractly, and we only require that they satisfy some general and innocuous conditions. Let us think, however, of the sign being produced by a machine of some kind. By manipulating the controls of the machine we get as output the various signs that we need to generate our patterns.

In a practical study of patterns the signs are usually not given to us *a priori* and we may have considerable leeway in choosing them. To help us choose we may use criteria such as:

a) the signs should be as simple as possible in some sense
b) the number of signs should be small
c) the ensuing analysis should be simple
d) the signs should have a direct interpretation in subject matter terms.

Since these and other related and reasonable criteria may be antagonistic to each other, it can very well happen that they do not lead us to a uniquely determined set of signs. Instead, the choice becomes a question of convenience and of the degree of approximation to which we aspire.

What sign will be produced by our machine will depend upon what program we feed into it. Say that we have certain *subroutines* available, labelled σ, where σ is chosen from some set Σ. Any sign $s \in S$ shall be generated by one and only one subroutine:

$$s \to \sigma(s) \in \Sigma. \tag{2.1}$$

On the other hand, we do not assume that knowledge of $\sigma(s)$ determines s. On the contrary, the subroutine of a given sign usually contains only crude information, easily discernible, about the sign. The set of signs belonging to subroutine σ will be denoted by S_σ. Sometimes it will be convenient to allow for the existence of the empty sign (or null-sign), denoted by ϕ.

What do we do with our signs? Well, first of all we can take a sign, change it in some way and display it in its new inflected form. In other words, we possess transformations that take signs into signs, and we shall assume that they are defined throughout S. Actually, this is not always the case: in a more general model we would ask only that each of our transformations be defined in some subroutine, or they may be mappings of the form

$$S_{\sigma_1} \times S_{\sigma_2} \times \cdots \times S_{\sigma_m} \to S_\sigma . \tag{2.2}$$

It is adequate for the present purpose, though, to assume that what we shall call a *similarity transformation* should be defined in the whole of S.

The similarity transformations will form a group under the usual composition of transformations. Again, this is really too drastic an assumption for the general theory of patterns, but it will simplify much of what follows and lead to basic and elementary notions, while at the same time the model is wide enough to be able to deal with a multitude of concrete cases.

Finally, we shall ask that the subroutines S_σ have a certain stability in the sense that they are invariant with respect to similarity transformations. Denoting the similarity transformation by g, the element of a group G, we should have

$$gs \in S_\sigma \text{ if } s \in S_\sigma . \tag{2.3}$$

In the empty sign ϕ we adopt the convention $g\phi = \phi$, $\forall g \in G$.

Let us sum up what we have discussed so far in

Postulate 2.1 *on signs. The set S of signs is divided into subsets S_σ, $\sigma \in \Sigma$, the subroutines, so that different subroutines have no sign in common except for the empty sign ϕ. The set G of similarity transformations forms a group and the subroutines are invariant with respect to G. Any $g \in G$ is defined on S; $g\phi = \phi$, $\forall\ g \in G$.*

Two signs s_1 and s_2 are said to be *similar* if there is a similarity transformation $g \in G$ such that $s_1 = gs_2$; this is written as $s_1 \equiv s_2 \pmod G$. The reader can now see why it was convenient to assume that G is a group: the similarity relation becomes (a) reflexive $s \equiv s \pmod G$ (b) symmetric $s_1 \equiv s_2 \pmod G \rightarrow s_2 \equiv s_1 \pmod G$ and (c) transitive $s_1 \equiv s_2 \pmod G$ together with $s_2 \equiv s_3 \pmod G \rightarrow s_1 \equiv s_3 \pmod G$. This means that it is an equivalence relation so that S is divided into disjoint equivalence classes. From each equivalence class let us select a representative, a *template.* For the moment we do not have to bother about how this should be done, but we assume that in any case we deal with some sort of standardization technique which has been developed. Once this is done it leads us to a unique analysis of any given non-empty sign $s \rightarrow (t, g)$, where t is a template and $s = gt$.

The next stage in the construction of our model starts from the set of all finite ordered sequences of signs: $c = (s_1, s_2, \cdots s_n)$, $s_1, s_2, \cdots s_n \in S$. The reason that we consider only finite sequences is simply that we wish to void topological complications. At present there is nothing finitistic in the approach. On the contrary, it will probably turn out to be convenient to extend the model to one that admits infinite sequences of signs.

The machine producing the signs may be able to produce only certain sequences c, or, from another point of view, only certain sequences may appear as meaningful to us. To answer the question of what sequences should be considered meaningful we must appeal to some set $\mathcal{R}$ of rules. The sequences accepted by $\mathcal{R}$ will be called the *grammatical configurations* and the set of all of them is denoted by C. We impose certain mild restrictions on $\mathcal{R}$ through the following

Postulate 2.2 *on configurations. The set C defined through $\mathcal{R}$ shall have the properties*

i) *if* $c = (s_1, s_2, \cdots s_n) \in C$ *then* $gc = (gs_1, gs_2, \cdots gs_n) \in C$ *for* every $g \in G$.
ii) *if* $(s_1, s_2, \cdots s_n) \in C$ *then* $(s_\alpha, s_{\alpha+1}, s_{\alpha+2}, \cdots s_\beta) \in C$ *for* $1 \leq \alpha \leq \beta \leq n$.
iii) $(\phi) \in C$.

Of course, it can happen that $\mathcal{R}$ does not restrict C at all, so that all finite sequences of signs from S are considered as meaningful. We then speak of *free configurations.* In this paper we shall focus our attention on the free configuration case, but the reader should be aware that occasionally we may have to impose further restrictions to arrive at interesting structures.

If $c = (s_1, s_2, \cdots s_m, s_{m+1}, \cdots s_n) \in C$ then we shall write $c = c_1 + c_2$ with $c_1 = (s_1, s_2, \cdots s_m) \in C$, $c_2 = (s_{m+1}, s_{m+2}, \cdots s_n) \in C$. Also $c + \phi = \phi + c = c$ for any configuration $c \in C$.

Here signs play the role of atoms; they have no inner structure. Sometimes, however, we may form our signs from more elementary signs. Then we would have a hierarchic *ordering* of signs. The reader may wish to think of the signs as corresponding to the words of natural language from which syntactic notions may be formed on different levels of logical complexity. The subroutines may then be taken as nouns, verbs, adjectives, etc. This linguistic parallel will not be explored in this paper, and is mentioned only for its possible suggestive value.

3. Images. In the third step toward the completion of our grammar, we introduce the observer. When confronted by configurations from C he may not always be able to distinguish between different configurations. Two configurations, $c = (s_1, s_2, \cdots s_n)$ and $c' = (s_1', s_2', \cdots s_m')$, may carry the same meaning to him although they are different considered as ordered sequences of signs. He *identifies* them, and we shall write this cRc'. The relation R will be assumed to satisfy

Postulate 3.1 *on images. The identification relation R shall be such that*

i) if cRc' then $gcRgc'$ for any $g \in G$.
ii) *if $c = c_1 + c_2$, $c' = c_1' + c_2'$, where all the c' s are from C, and if c_iRc_i' for $i =$ 1, 2, then cRc'.*
iii) *R should be an equivalence relation.*
iv) *if no subroutine of c is equal to any subroutine of c', then either $c\not{R}c'$ or $c = c' = \phi$.*

The intuitive motivation behind our choice of the three postulates will be clear after examining some special cases in Part II of this paper.

Theorem 3.1 *Consider a set C of free configurations and the set $\mathfrak{I}$ of all equivalence classes I generated by the identification relation R. On $\mathfrak{I}$ is defined a unary operation gI and a binary operation $I_1 + I_2$ such that*

i) *I forms a semigroup with unit under the binary operation.*
ii) *the distributive law $g(I_1 + I_2) = gI_1 + gI_2$ holds.*
iii) *the associative law $g_1(g_2I) = (g_1g_2)\, I$ holds. The classes I will be called images.*

Proof: Because of (iii) of Postulate 3.1 the classes I are well defined, disjoint and cover the whole of C. If $c = (s_1, s_2, \cdots s_n) \in I$ we define gI as the image containing the configuration $gc = (gs_1, gs_2, \cdots gs_n)$. This definition is unique because of (i) of the postulate. If $c_1 \in I_1$ and $c_2 \in I_2$ then we define $I_1 + I_2$ as the image containing the configuration $c_1 + c_2$. The definition is unique because of (ii) of the postulate. Since composition of configurations is associative, the same is true for images. Hence I forms a semigroup. The image E containing the configuration (ϕ) is a unit element of the semigroup. The distributive (ii) and associative (iii) laws follow from the fact that they are true for configurations.

After identification C turns into $\mathfrak{I}$, an *image grammar* (*image algebra*). The mathematical properties of $\mathfrak{I}$ bear some resemblance to a vector space and more to a module. Note, however, that G does not necessarily form a field or ring.

Let us turn for a moment to the important case when the image grammar $\mathfrak{I}$ is commutative so that $I_1 + I_2 = I_2 + I_1$ holds. Consider the set C_σ of all free configurations with signs from the subroutine σ. After identification it turns into a subset I_σ of I. Since C_σ is a sub-semigroup, the same is true of I_σ. For configuration $c = (s_1, s_2, \cdots s_n) \in C$ we can form the configuration $c = (s_{i_1}, s_{i_2}, \cdots s_{i_k})$ consisting of those signs $s_{i_1}, s_{i_2}, \cdots s_{i_k}$ that belong to subroutine σ. The corresponding image $I_\sigma(c)$ is in I_σ and $I = I_{\sigma_1}(c) + I_{\sigma_2}(c) + \cdots$ where the composition is taken over the possible subroutines $\sigma_1, \sigma_2, \cdots$. Of course many $I_\sigma(c)$ may be equal to E. We have then represented the original image as a composition of its *subroutine parts*, but it should be noted that this representation is not always unique, since it depends upon what c and I we have started from. We will return to such representations in section 4.

It can happen that a configuration $c = (s_1 \cdots s_n)$ is identified with some $c' = (s_1' \cdots s_m')$ with fewer signs, $m < n$: c *simplifies* into c'. The following is not unusual, as we shall see later on. Consider a free commutative image algebra I and two configurations c and c' that each can be simplified into single signs and such that the set of signs common to c *and* c' also simplifies into some single sign: assume that it then follows that the configuration consisting of all the signs appearing in c *or* c' simplifies into some single sign. If this is true, we say that the signs in S are *collapsible*. We can then repeat this process until we arrive at a configuration that cannot be simplified any further. Applying the simplification procedure in a different way, however, we may arrive at different end results.

Let us end this section by considering generation of sets of images in analogy with vector spaces. Let H be a subset of $\mathfrak{I}$ and introduce the set H^G consisting of all images of the form

$$g_1 I_1 + g_2 I_2 + \cdots + g_n I_n \tag{3.1}$$

where n is arbitrary, $g_k \in G$, $I_k \in H$, $k = 1, 2, \cdots n$. The images in H^G are those spanned by those in H. In particular we shall be interested in the sets $I_\sigma = S_\sigma^G$ spanned by all the signs from a certain subroutine σ.

Definition 3.1 *By the dimension Dim* ($\mathfrak{I}$) *of* $\mathfrak{I}$ *we shall mean the smallest integer* n *such that there exist* $I_1, I_2, \cdots I_n$ *for which* $\mathfrak{I} = \{I_1, I_2, \cdots I_n\}^G$. *In case there is no such* n *the dimension is said to be infinite. The set* $\{I_1, I_2, \cdots I_n\}$ *is called a basis of* $\mathfrak{I}$.

If the set S can be analyzed in terms of a finite number of templates, then I has finite dimension. The opposite need not hold, but, if not, then the choice of signs has not been made in an economical way.

Definition 3.2 *The images* $I_1, I_2, \cdots I_n$ *are said to be independent if for no* $m < n$ *does there exist a set* $H = \{I_1', I_2', \cdots I_m'\}$ *such that* $\{I_1, I_2, \cdots I_n\} = H^G$. *A set* $\{I_1, I_2, \cdots I_n\}$ *is said to be minimal if it has no proper subset K with* $K^G = \{I_1, I_2, \cdots I_n\}$.

A basis always consists of independent images.

Theorem 3.2 *If the non-null images* $I_\nu \in \mathfrak{I}_{\sigma_\nu}$, $\nu = 1, 2, \cdots n$, *where all the* σ_ν *are different subroutines, then the set* $\{I_1, I_2, \cdots I_n\}$ *is always minimal.*

Proof: Let $I_1', I_2', \cdots I_m'$ ($m < n$) be a subset of $\{I_1, I_2, \cdots I_n\}$ such that

$$\begin{cases} I_1 = g_{11}I_1' + g_{12}I_2' + \cdots + g_{1m}I_m' \\ \cdots \\ I_n = g_{n1}I_1' + g_{12}I_2' + \cdots + g_{nm}I_m' \end{cases} \tag{3.2}$$

Represent the I_ν and I'_μ by configurations c_ν and c'_μ. Then the above equalities hold when images are replaced by configurations and the equality sign by the relation R. It follows from (iv) of Postulate 3.1 that at least one of $I'_1, I'_2, \cdots I'_m$ must be in I_{σ_1}, at least one in I_{σ_2}, and so on. But this is impossible for $m < n$, so the $I_\mu s$ must form a minimal set.

Extending the definition of similarity from signs to images, we shall say that two images I_1 and I_2 are similar if there exists a similarity transformation $g \in G$ such that $I_1 = gI_2$.

4. Representation of Images. It is inherent in the definition of an image that the identification mapping $C \to I$ need not have (and usually does not have) a one-valued inverse. We must therefore look for ways of analyzing images into certain configurations in as simple and compact a way as possible. For this sort of *morphology of images* we need a notion of simplicity of a configuration.

Let us assume for the moment that the elements of C are partially ordered by a relation "$<$" in that for any pair of configurations c and c' from the same image we have either $c < c'$ or $c' < c$ or both. We may not be able to compare configurations from two different images. Add to this the lattice property that in any set K of configurations of one image there is at least one element $c_0 \in K$ such that $c_0 < c$ for all $c \in K$: c_0 is *at least as simple* as any other configuration in the image. As an example the reader may think of the ordering introduced as $c < c'$ if the number of signs in c does not exceed that in c'. More generally, we may have additive complexity: the complexity of c is the sum of the complexities of its signs.

From any image $I \in \mathfrak{J}$, $I \subset C$ we can then select a configuration c_I at least as simple as the other configurations in the same image. This defines a mapping μ: $C \to C$, putting $\mu(c) = c_I$ if $c \in I$, with the properties

(4.1)
- (i) μ is an image preserving mapping in the sense that $\mu(c) = \mu(c')$ implies $\mu\ (gc) = \mu(gc')$ and $\mu(c_i) = \mu(c'_i)$, $i = 1, 2$, implies $\mu(c_1 + c_2) = \mu(c'_1 + c'_2)$.
- (ii) μ is idempotent: $\mu(\mu(c)) = \mu(c)$.
- (iii) μ is monotonic: $\mu(c) < c$.
- (iv) μ preserves subroutines partially: $\mu(c)$ has at least one subroutine in common with c.

Indeed, if $\mu(c) = \mu(c')$ then c and c' belong to the same image I, so that for any $g \in G$ the configurations gc and gc' belong to gI and $\mu(gc) = \mu(gc') = c_{gI}$. Similarly, if $\mu(c_i) = \mu(c'_i)$, $i = 1, 2$, then $c_1 + c_2$ belongs to the same image $I_1 + I_2$ as does $c'_1 + c'_2$, so that $\mu(c_1 + c_2) = \mu(c'_1 + c'_2) = c_{I_1} + c_{I_2}$. This proves (i). That μ is idempotent and monotonic is obvious. Property (iv) holds since $cR\mu(c)$ always holds and implies, because of Postulate 3.1 (iv), that c and $\mu(c)$ have at least one subroutine in common. This completes the proof of (4.1).

On the other hand, let us start from C (before identification has been effected) and with access to a simplification operator μ with the properties

(4.2)
- (i) μ is an image preserving mapping $C \to C$.
- (ii) μ is idempotent: $\mu(\mu(c)) = \mu(c)$.
- (iii) μ preserves subroutines partially.

Then let us agree to define the identification relation R through

$$cRc' \text{ if and only if } \mu(c) = \mu(c'). \tag{4.3}$$

Then Postulate 3.1, (i) and (ii) follow from the image preserving property and (iii) is obvious. Condition (iv) of the Postulate follows from condition (iii) above. Hence an image will always be of the form

$$I = \{c | \mu(c) = a\} \tag{4.4}$$

but since μ is idempotent we have $\mu(a) = \mu[\mu(c)] = \mu(c) = a$, so that $a \in I$. To each image I there is then a uniquely determined $a \in I$ such that (4.4) holds. Defining a partial order relation $<$ by the condition $c < c'$ if $\mu(c') = c$, it follows that $a = c_I$ is at least as simple as any other configuration in I, so that we have arrived at essentially the same situation that we had when we started with a partial order. Therefore we can either start with images and a notion of simplicity or with configurations and a simplification operation and use it to identify images. Let us sum this up concisely as follows:

Theorem 4.1 *A partial order on the set C of configurations such that any c and c' are comparable if they belong to the same image leads to an image preserving, idempotent, monotonic and partially subroutine preserving mapping of C into C. Conversely, if a simplification operator μ of C into C is image preserving, idempotent and partially subroutine preserving, it leads to an identification into images: c and c' are similar if they are simplified into the same configuration.*

What this amounts to more generally is that the existence of certain types of standard representations is intimately related to what sort of identification rule we apply to C in order to arrive at the images. Let us consider as an example the case when the signs are collapsible (see section 3). We then have a simplification rule (although it can be many-valued) consisting of collapsing signs as far as possible and it is natural to identify two configurations c and c' if they can be simplified into one and the same configuration. What we are doing then could be called *identification via the maximal connected components* of the configurations. More about this later.

We shall occasionally identify configurations by considering their subroutines c_σ : *identification via subroutines.*

Theorem 4.2 *Consider a free commutative C and let there be given identification rules R_σ (satisfying Postulate* 3.1 *(i), (ii), (iii) in C_σ) in C_σ for each subroutine σ. Define R by: cRc' if $c_\sigma R_\sigma c'_\sigma$ for each σ. The image algebra $\mathfrak{J}$ is then a direct product $\mathfrak{J} = \mathfrak{J}_{\sigma_1} + \mathfrak{J}_{\sigma_2} + \cdots$ and the subroutine images $I_\sigma(c)$ are uniquely determined. Conversely, given R and $\mathfrak{J} = \mathfrak{J}_{\sigma_1} + \mathfrak{J}_{\sigma_2} + \cdots$ in terms of uniquely defined subroutine images implies identification via subroutine images.*

Proof: If R is given as above, cRc' means that $c_\sigma R_\sigma c'_\sigma$, $\forall\ \sigma \in \Sigma$. But for any $g \in G$ we then have $g\ c_\sigma R_\sigma g c'_\sigma$, $\forall\ \sigma$, which implies $gcRgc'$, which is (i) of Postulate 3.1. In the same way one proves (ii) and (iii) of the Postulate. Finally, (iv) is a direct consequence of the way in which R was introduced, so that R defines an image algebra $\mathfrak{J}$. If cRc', it follows that $c_\sigma R_\sigma c'_\sigma$, so that $I_\sigma(c) = I_\sigma(c')$: the subroutine images are uniquely defined so that we have a unique representation $I = I_{\sigma_1}, I_{\sigma_2}, \cdots$ and $\mathfrak{J}$ is a direct product of I_σs. To prove the converse we observe that cRc' means that $I_\sigma(c) = I_\sigma(c')$, $\forall \sigma$, which can be taken as the definition of R_σ , which proves the statement.

The following is a special case of the last theorem.

Theorem 4.3 *Let $\mathfrak{J}$ be a free commutative image algebra forming a group with the $\mathfrak{J}_\sigma$ as subgroups under composition of images. Then the subroutine images are uniquely determined.*

Proof: Consider two similar configurations c and c': cRc'. Write them in terms of identified configurations

$$\begin{cases} c = (c_{\sigma_1}, c_{\sigma_2}, \cdots) \\ c' = (c'_{\sigma_1}, c'_{\sigma_2}, \cdots) \end{cases} \tag{4.5}$$

where some of the components may be empty. Let $c_{\sigma_1} \in I_1$, $c'_{\sigma_1} \in I'_1$, and so on. Then

$$I_{\sigma_1} + I_{\sigma_2} + \cdots = I'_{\sigma_1} + I'_{\sigma_2} + \cdots,$$

and using the inverse "$-$" of the group operation "$+$",

$$I_{\sigma_1} - I'_{\sigma_1} = I'_{\sigma_2} - I_{\sigma_2} + \cdots \tag{4.6}$$

The left member belongs to $\mathfrak{I}_{\sigma_1}$ and the right member to the set of images spanned by $I_{\sigma_2}; I_{\sigma_3} \cdots$. Let $c_L \in C_{\sigma_1}$ and $c_R \in C_{\sigma_2} + C_{\sigma_3} + \cdots$ be configurations from the left- and right-hand side of (4.6). They are identified but, according to Postulate 3.1 (iv), they must be empty, so that $I_{\sigma_1} = I'_{\sigma_1}$ and so on, which proves the statement.

To keep the amount of data processing within reasonable limits, it is important to parametrize the images efficiently by choosing their representation through configurations well. As long as we consider only pure image algebras this need not be crucial, but when we proceed to deformation grammars we cannot neglect this aspect of the analysis.

5. Patterns. By now we have a rough but fairly general idea of how our model should look. The building blocks to form an image are the signs generated by some machine. A particular sequence of outputs of the machine, a configuration, may not be exactly identified by the observer. It should be emphasized that this is not due to distortion by noise or anything like this (such deformations will be discussed in Parts III and IV), but is caused by the way in which the subject receives and organizes his observations. The mathematical formalization of this fact led us to the notion of an image.

Different images are certainly always distinguishable. On the other hand, I_1 and I_2 may carry the same "meaning" although $I_1 \neq I_2$. This means that we may be able to understand the "meaning" of an image without knowing it in all its details: there are certain characteristic traits in it that tell us enough about it. These traits could be things like the different subroutines that make up the image or the different g's that were used to get the component signs from their templates. We express this in a definition.

Definition 5.1 *By patterns we mean a family $\mathcal{P}$ of disjoint and G-invariant subsets $\mathcal{P}_1, \mathcal{P}_2, \cdots \mathcal{P}_n \cdots$ of the image algebra $\mathfrak{I}$.*

It is obvious that the $\mathcal{P}_\nu$ should be assumed to be disjoint, but why do we not also assume that they together exhaust all of $\mathfrak{I}$? There is a very good reason for this. In almost any real-life situation the subject will not be confronted by these (pure) images. Instead, they have been deformed, and what he perceives will be a corrupted version of the pure image. To facilitate recognition of the patterns redundancy is introduced, and it should be intuitively plausible that recognition is easier when the subsets $\mathcal{P}_\nu$ are well separated from each other in $\mathfrak{I}$ and do not cover the whole image algebra. The reason why we want the $\mathcal{P}_\nu$ to be G-invariant is that two images I_1 and I_2 will be considered as carrying the same "meaning" if there is a $g \in G$ such that $I_1 = gI_2$, or, in terminology already introduced, if the images are similar.

Problem of recognition of pure patterns: To find a mapping (or recognition function) μ of $\mathfrak{I}$ into $\mathcal{P}$ classifying the image according to what pattern (if any) it belongs to.

The simplest case is when we can find one image I_ν in each pattern $\mathcal{P}_\nu$, a *prototype*, such that gI_ν runs through all the elements of $\mathcal{P}_\nu$ once and only once when g runs through G. Then the classification consists simply in the unique analysis $I = gI_\nu$. In this special but important case we write naturally $\mathcal{P} = \mathfrak{I}/G$ and it will be studied a great deal later on.

More generally, there may be given a set T of transformations t mapping $\mathfrak{I}$ into itself and a set of prototypes I_ν such that $TI_\nu = \mathcal{P}_\nu$. Of course, T should contain the similarity transformations of G, but it need not itself be a group. Images belonging to the same $\mathcal{P}_\nu$ will be called *synonymous*; especially when $\mathcal{P} = \mathfrak{I}/G$, this reduces to similarity.

This problem of recognition may appear simple in principle but gets its real interest in concrete situation when it becomes a matter of organizing the computations in a feasible way. This is sometimes done by considering *probes*, functions f defined in I, and such that $f(I_1) = f(I_2)$ if I_1 and I_2 are *synonymous* (i.e., they belong to the same pattern). One could say that a probe singles out one of many characteristic traits. A set F of probes $F = \{f\}$ is said to be *complete* if it separates patterns: $f(I_1) = f(I_2)$, $\forall f \in F$ should imply that I_1 and I_2 are synonymous.

Finally, the reader should be aware of the following problem, which we mention here only in passing.

Problem of segmentation. For any given image I, to analyze it into $(\nu_1, \nu_2, \cdots)$ if I can be written as $I = I_1 + I_2 + \cdots$ with $I_1 \in \mathcal{P}_{\nu_1}, I_2 \in \mathcal{P}_{\nu_2}, \cdots$.

This problem arises especially when the ensuing analysis has some reasonable degree of uniqueness. As an example we mention Morse code (see section 12) and comma-free codes.

So far we have arrived at a set

$$\mathcal{G} = (S, \mathfrak{I}, G, \mathcal{P}) \tag{5.1}$$

which will be called a (pure) *grammar of patterns*. Now we shall turn to the formulation of such grammars in special instances, keeping in mind that the model will achieve its full potential only in Part III when deformations are allowed.

Part II. Grammars of Patterns

6. Taxonomy of images and patterns. Up to now the discussion has been fairly general except for certain restrictions on the form of the grammar mentioned in the text, and we have been mainly concerned with building an all-purpose model for pattern analysis. The treatment has been quite elementary but also somewhat abstract, with little attention paid to the intuitive motivation that has led to this model. The latter will be more easily understood when examining a number of special patterns, and this will be started in this part of the paper and continued in greater detail in Grenander [3].

We shall attempt to *classify patterns* into families of related types according to what *logical structure* they have expressed in terms of our model. It would also be possible to group them according to their *subject matter interpretation*, but this will not be done here. We shall, however, take a quick look at some naturally occurring patterns in order to understand better the motivation behind this approach.

Let us think first of all of *patterns in time*, a familiar case being presented by *Morse* *code*. Here the signs consist of dot, dash and spaces of various length within letters, between letters and between words. The images are formed simply by concatenation, so that the order between successive signs is relevant: this image algebra is non-commutative. Note that we have no natural origin on the time-axis, nor is the time scale of any importance. The only thing that matters is that the ratios between the lengths of different signs take the values prescribed by the rules defining Morse code. The similarity transformations are then given by changing the time parameter through relations of the form $t \to a\,t + b$ where $a > 0$.

Or think of the electrocardiogram with its typical sequence of marked maxima and minima labeled P, Q, R, etc. It would be natural to attempt to describe the ECG in terms of concatenation of signs, each sign being generated through a certain subroutine. Attempts have been made instead to express the observed time function as linear combinations of certain functions, orthogonal to each other on the interval representing one cycle or period of the ECG. But this is open to criticism. First: the period is not quite easy to define, and it can fluctuate considerably, which makes this sort of analysis awkward. Second, and more important: it seems reasonable to believe that by splitting the time function into consecutive signs (instead of linearly into orthogonal components) one would get a more intrinsic characterization of the physiological phenomenon under study. What this amounts to is, in the last analysis, what is the most compact and economical grammatical formalization of the phenomenon studied, a question that can only be answered with the help of empirical data. On the other hand, for the *encephalogram* the linear representation *via* Fourier analysis seems to work fairly well, and there is less reason to try an analysis in terms of consecutive signs.

Fingerprints can be idealized as *line patterns*: the signs here are curves if we neglect the width and porous structure of the ridges making up the print. The images are now formed simply by superposition. What makes the fingerprint useful as identification is the long-known fact that certain geometric invariants exist. The distance between two distinguished points of the print of a certain individual may very well change when he grows older or when the finger is pressed in different ways against the recording surface. This change need not even be uniform so that the similarity transformations should consist not only of translations, rotations and uniform dilatations but also of more general topological mappings of the plane onto itself. The same is true about printed or handwritten *alphanumeric symbols*, at least if we try to work with a family of alphabets rather than with one particular type font.

But if we also take into account the width of the ridges of the fingerprints we can no longer consider the image as made up of curve segments. Instead, the signs will consist of other subsets of the plane and we will speak of *set patterns*. This also occurs when studying block letters with solid areas rather than thin lines. We shall pay a good deal of attention both to line and set patterns.

If the information is spread out in a more continuous way over the plane, other image algebras will be more useful. We may measure the darkness of the picture locally and study how it varies over the plane. The signs will be functions defined in the plane combined into images just by addition. The values taken by these functions could be real or, for colored pictures, vector-values. A three-color print could be analyzed in terms of three-dimensional signs. Or we may have values that combine not as vectors but with a more general semi-group operation. Mathematically this will be formalized as *contrast patterns* (see section 11), a notion of considerable generality.

Many other types of natural patterns could be mentioned, e.g. in connection with questions of human perception, but the above should be enough to give the reader some idea of what has motivated the introduction of some constructions below.

Leaving aside the practical background, we should ask: what constitutes the logical structure of a particular set-up for pattern analysis? Three things do: *material, perception and interpretation.* The *material* we use to build images are the signs and the similarity transformations. The *perception* that leads us to consider certain different configurations as indistinguishable objects is given by the rules R of identification. The *interpretation* of the images is expressed *via* the family $\mathcal{P}$ of possible patterns. Now we will go ahead and specify these three notions in different ways.

Recognition algorithms will be based on a *grammatical analysis of the image.*

7. Permutation Patterns. Let us start with the simplest possible case, when the signs are vectors from R^n, and images are formed via vector addition, so that two configurations $c = (s_1, s_2, \cdots s_n)$, $c' = (s_1', s_2', \cdots s_m')$ with $s_1, s_2, \cdots s_1', \cdots s_m' \in R^n$ are identified if

$$s_1 + s_2 + \cdots s_n = s_1' + s_2' + \cdots + s_m' . \tag{7.1}$$

That means that we can regard an image I as a vector $I = (x_1, x_2, \cdots x_n)$ in R^n, and it can be viewed as the sum of the signs.

Let the similarity transformations g be the cyclic transformations

$$gI = g(x_1, x_2, \cdots x_n) = (x_{1+g}, x_{2+g}, \cdots x_{n+g}), \tag{7.2}$$

where we have used the same symbol g to denote both the group element and the amount of the cyclic translation. Of course we count modulo n in the subscripts of the x's.

Say that the patterns are defined through the condition that I and I' belong to the same pattern if there is a $g \in G$ such that $I' = gI$. Recognition could be achieved simply by comparing a given $I = (x_1, x_2, \cdots x_n)$ with a prototype $I' = (x_1', x_2', \cdots x_n')$ from an arbitrary pattern $\mathcal{P}_k$. Starting with x_1, consider the set $T = \{0, 1, 2, \cdots n - 1\}$ of g's for which $x_1 = x_{1+g}'$. For any $g \in T$ verify if $x_v = x_{v+g}$, $\forall v$, or not. If this does not occur I does not belong to the pattern $\mathcal{P}_k$. This algorithm is completely straightforward, but one could ask if it would not be possible to recognize patterns by choosing the prototype images to be of a certain standard form. Say, for example, that $\max_v x_v$ is attained for a single value $v = v(x)$. Let us then use as a prototype (for the pattern to which I belongs) the pattern $I' = (x_v, x_{v+1}, \cdots x_{v+n-1})$. We would then immediately be able to recognize patterns by rearranging the images so that they start with their largest component. But this will not work if the maximum is realized for two (or more) different values v and μ. But let us then choose x_v as the first coordinate of the prototype if $x_{v+1} > x_{\mu+1}$ and so on. If $x_{v+1} = x_{\mu+1}$ we go one step further. If all the "followers" of x_v and x_μ are the same it follows that $(x_v, x_{v+1}, \cdots x_{v+n-1}) = (x_\mu, x_{\mu+1}, \cdots)$ so that in any case this procedure leads us to a well-defined prototype for each pattern.

In the terminology of section 5, we could say that we based our scheme of recognition in the probes

$$\begin{cases} p_1 = \max\limits_v x_v \\ p_2 = \max\limits_\mu x_{\mu+1} \quad \text{if} \quad p_1 = x_\mu \\ \text{etc.} \end{cases} \tag{7.3}$$

Another set of probes, based on Fourier analysis, was discussed in Grenander [2].

So far we have not restricted the choice of the signs from R^n, but now let us assume, to be concrete, that the device that produces the signs has the form of a difference equation

$$Lx_v = a_0 x_{v+p} + a_1 x_{v+p-1} + \cdots + a_p x_v = 0; v = 0, 1, \cdots n-1. \tag{7.4}$$

In order that the solution x_v be periodic with period n we shall assume that the roots $z_1, z_2, \cdots z_p$ of the characteristic equation

$$a_0 z^p + a_1 z^{p-1} + \cdots + a_0 = 0 \tag{7.5}$$

have modules one and argument equal to some multiple of $2\pi/n$. All the solutions (signs) of $Lx = 0$ will be taken together as a subroutine S_σ. This is legitimate since it satisfies Postulate 2.1: S_σ is G-invariant. In the same way the other subroutines are introduced by other difference operators L_σ, and we shall assume that, for any $\sigma_1 \neq \sigma_2$, no characteristic root of L_{σ_1} coincides with any characteristic root of L_{σ_2}. Then Postulate 3.1 is satisfied. Given an image $I = (x_1, x_2, \cdots x_u)$ we can determine the subroutine image I_σ by Fourier analysis, calculating the expressions

$$\sum_{v=1}^{n} x_v z_{\sigma k}^{-v}, \tag{7.6}$$

where $z_{\sigma_1}, z_{\sigma_2}, \cdots$ are the characteristic roots of the subroutine σ. In this way Fourier analysis enters as a natural tool for pattern analysis in this particular image algebra.

In this special case the recognition problem is so simple that we would not get much help computationally from using standard representations of the prototypes and our discussion should be considered as merely an illustration of the general approach.

Let us now turn to the image algebra where G is the permutation group of n objects operating on the subscripts of the x's and the rest of the set-up is just as it was before. In principle recognition could again be achieved by straightforward algorithms, but now this will be more difficult computationally since to a given image we can have as many as $n! - 1$ similar images (if the x's are all different). Therefore we have reason to look for some simple probes and these are easily found in the elementary symmetric functions, or, equivalently, in the sums

$$\begin{cases} p_1(I) = \sum_{\nu=1}^{n} x_\nu \\ p_2(I) = \sum_{\nu=1}^{n} x_\nu^2 \\ \cdots \\ p_n(I) = \sum_{\nu=1}^{n} x_\nu^n \end{cases} \tag{7.7}$$

This set is known to be complete: if two images I and I' give the same values for these probes then they are similar. Another way of organizing the recognition algorithm would be by rearranging the x's in non-decreasing order. With this standard representation recognition is achieved just by inspection in terms of coordinates.

Finally, consider non-directed graphs with the n edges $1, 2, \cdots n$. The signs consist of edges, say between i and j, which we will write as $x_{ij} = x_{ji} = 1$; and $x_{ij} = x_{ji} = 0$ if these edges are not joined together. We put $x_{ii} = 0$, and signs are combined by superposition without counting any multiplicity. An image I can then be represented by an array

$$I = \begin{array}{rrrrr} x_{12}, & x_{13}, & x_{14}, & \cdots & x_{1n} \\ & x_{23}, & x_{24}, & \cdots & x_{2n} \\ & & x_{34}, & \cdots & x_{3n} \\ & & & \cdots & \\ & & & & x_{n-1n} \end{array} \tag{7.8}$$

As similarity transformations we shall take the transformation of an image due to a permutation of the indexes of the edges $\nu = 1, 2, \cdots n$. The order of G is then $n!$ and not $N = [n(n-1)/2]!$; the latter corresponds to the symmetric group operating on the x's. To decide whether two given images are similar or not we can use the following probes.

Consider the set of G-invariant polynomials in the x's of the order at most n. They separate images which are not similar. Indeed, suppose that $I \not\equiv I'(\text{mod } G)$. Then, for any $g_k \in G$, we have $I \neq g_k I'$, so that there is a pair i_k, j_k of indices so that $x_{i_k j_k} \neq (g_k x)'_{i_k j_k}$ where the x_{ij} and $(g_k x)'_{ij}$ are the entries in the arrays for I and $g_k I'$ respectively. Say that $x_{i_k j_k} = \delta_k$ and $(g_k x)'_{i_k j_k} = 1 - \delta_k$ and form the linear expression

$$L(z) = \sum_k (1 - \delta_k) z_{i_k j_k} + \delta_k (1 - z_{i_k j_k}) \tag{7.9}$$

where $z = (z_{11}, z_{12}, \cdots)$ is a triangular array. Then

$$\begin{cases} L(I) = 0 \\ L(gI') > 0, \qquad \forall\, g \in G \end{cases} \tag{7.10}$$

so that the invariant polynomial

$$Q(z) = \prod_{g \in G} L(gz) \tag{7.11}$$

satisfies the relations $Q(I) = 0$, $Q(I') > 0$, separating I and I' as stated. It will be enough to use as probes

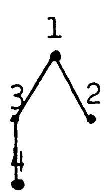
$$\prod_{i<j} x_{ij}^{\alpha_{ij}}$$

and, since the x's take only 0 and 1 as values, this reduces to $\prod_{i<j} x_{ij}$. We now start with monomials involving one x, then those with two x's, etc., all the time leaving out monomials that can be expressed in terms of the ones already introduced. Say we consider the monomial $x_{12} \cdot x_{13} \cdot x_{23} \cdot x_{34}$, and form the invariant polynomial

$$P(I) = \sum_{g \in G} (gI)_{12} \cdot (gI)_{13} \cdot (gI)_{23} \cdot (gI)_{34} \tag{7.12}$$

The value of $P(I)$ is simply the number of similar images gI that contain the subgraph

(7.13)



As an example consider $n = 4$ when the 64 images can be grouped into eleven patterns that can be expressed through the prototypes which can be recognized using four probes generated by the monomials (or the corresponding subgraphs) in Fig. 2.

8. Line Patterns. A more important sort of pattern is that where the signs consist of arcs in the plane and images are formed by superposition without counting multiplicity. The subroutines will correspond to different types of curves (e.g., line segments and circular arcs) and we shall assume that they are analytic. As far as the similarity transformations are concerned, let us start by examining the important case when they consist of the Euclidean motions of the plane. Then an arc could be written in its intrinsic form as

$$\frac{d\phi}{ds} = l = l(s), \tag{8.1}$$

where l is the curvature, s is the arc length and ϕ is the angle between the tangent and some fixed direction; in this way we get an invariant representation. In other words, we treat the signs as *geometric objects*, in coordinate-free terms. This leads us to propose the following

Model for Euclidean line patterns: We use the following image algebra:

a) *For each subroutine σ there is a real-valued function $l_\sigma(s)$ analytic on the real line. If $l_{\sigma_1}(s) \equiv l_{\sigma_2}(s + h)$ for some h and all s then $\sigma_1 = \sigma_2$ should be implied.*
b) *A sign in S_σ consists of an arc with the intrinsic equation $l(s) = l_\sigma(s)$.*
c) *For $c = (s_1, s_2, \cdots s_n)$ and $c' = (s_1', s_2', \cdots s_m')$ we have cRc' if $s_1 \cup s_2 \cup \cdots \cup s_n = s_1' \cup s_2' \cup \cdots \cup s_m'$.*

A given configuration $c = (s_1, s_2)$ can be simplified into a single sign if, at some point of intersection between the arcs s_1 and s_2, all derivatives are the same, and similarly for components with several signs. The signs are collapsible (see section 3) and we have identification *via* the maximal connected components (see section 4). But these can be recognized through a flying spot technique by scanning the image. When the spot hits

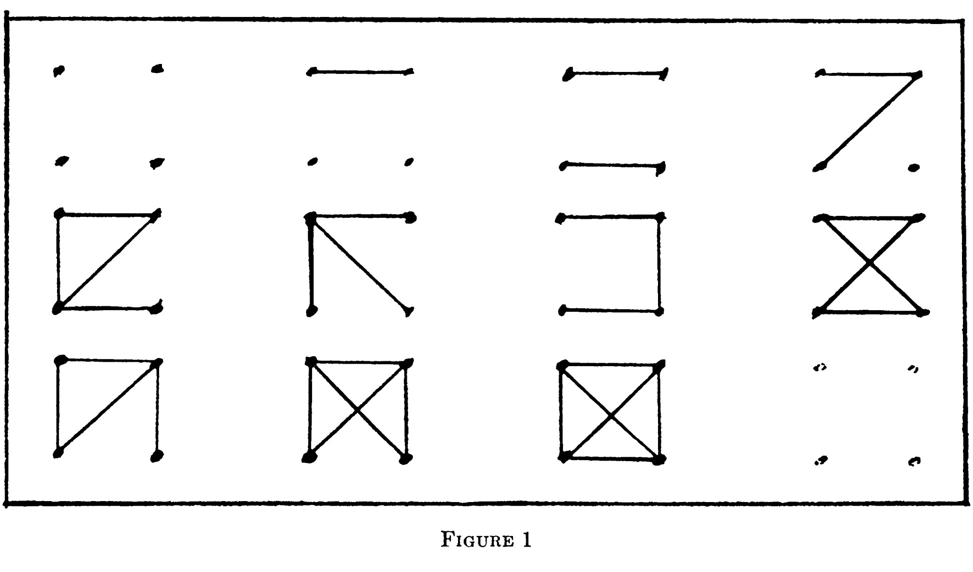


FIGURE 1

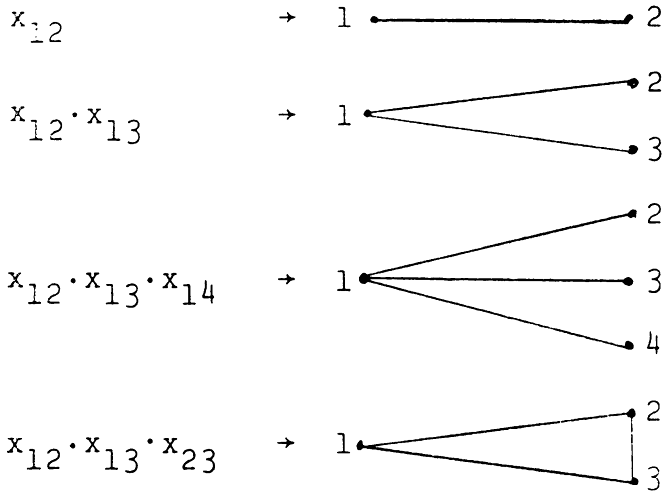


FIGURE 2

an arc it locks onto it and follows it in both directions until it hits a singular point. A point of I is said to be regular if it is an inner point of exactly one analytic arc belonging to I; otherwise it is singular. The singular points are of two types:

a) end points (which includes cusps)
b) branch points.

The arc is followed in both directions (until a singular point is met) and the arc thus obtained is classified according to what l_σ enters into its intrinsic equation. Then the scanning continues until all the maximal connected components have been identified and classified according to σ, g and end points.

This grammar was formulated for the important but special case in which the similarity transformations were the Euclidean ones. We now turn to more general line patterns and at the same time we introduce another device for generating the signs.

For each σ let there be given a differential operator $F_\sigma(x, y, y')$ of the first order (first order is assumed merely for simplicity and is not essential) such that the differentiable line element field given by $F_\sigma(x, y, y') = 0$ has exactly one integral curve passing through each point of R^2. Let G be a Lie group of transformations of R^2 onto itself and assume that $F_\sigma = 0$ is G-invariant for every σ. Now we have a natural modification of the previous model.

Model for differential line patterns. Use the following image algebra:

a) *For* $\sigma_1 \neq \sigma_2$ *the set of* R^2, *where the line element field of* F_{σ_1} *coincides with that of* F_{σ_2}, *must not contain any part of a sign belonging to* S_{σ_1} *or* S_{σ_2}.
b) *The signs in* S_σ *consist of arcs satisfying* $F_\sigma(x, y, y') = 0$.
c) *If* $c = (s_1, s_2, \cdots s_n)$, $c' = (s_1', s_2', \cdots s_m')$, *then* cRc' *if* $s_1 \cup s_2 \cup \cdots \cup s_n = s_1' \cup s_2' \cup \cdots \cup s_m'$.

Although we have dropped analyticity from the requirements, identification can still be carried out *via* the maximal connected components, and the recognition scheme looks as follows. Scan the picture, lock to an arc and classify its line elements according to subroutine by the value of the probe $F_\sigma(x, y, y')$. In this way we get arcs bounded by singular points. We describe them through their subroutine and their end points.

In passing we note that a sign s which is G-invariant, $gs = s$, $\forall\ g$, cannot be used to carry information on what similarity transformation has been applied to it. From this point of view it is useless and could be discarded. We would then modify condition (b) in the above model to exclude G-invariant signs.

A simple example may be illuminating. Let G consist of uniform dilatations with the infinitesimal operator $x(\partial/\partial x) + y(\partial/\partial y)$. A first-order differential equation invariant under G has the form $y' = F(y/x)$. The invariant signs are $y = c\cdot x$ and should be excluded. Introduce two subroutines through

$$\begin{cases} \sigma = 1 : y' = -\dfrac{y}{x} \\ \sigma = 2 : y' = \dfrac{x}{y} \end{cases} \tag{8.2}$$

then the signs of S_1 are arcs from the hyperbolas $xy = a$ and those of S_2 are arcs from the hyperbolas $y^2 - x^2 = b$. An image will consist of arcs from these mutually orthogonal families of hyperbolas.

9. Set Patterns. How about images formed in R^2 by superposition (set unions) as above but not with arcs but with more general sets? Let us distinguish three cases depending upon the degree of complexity of the boundary of the set.

If the boundary is very complicated and fuzzy, it would not be useful as a description of the set and we would look for more economical ways of defining the image. This will not be attempted here.

If, on the other hand, the boundary is sufficiently well behaved, we could use it as an economical way of characterizing the set. In the construction of the model we would use a hierarchy of two sets of signs (see section 2). The signs $^{(1)}$ would be chosen as certain directed arcs and the restriction $\mathcal{R}$ governing the formation of a configuration$^{(1)}$ of $c^{(1)}$ would guarantee that the component signs$^{(1)}$ of $c^{(1)}$ could be joined together (without violating the sign of the directions). As signs $^{(2)}$ we would take those configurations$^{(1)}$ that separate the plane into two simply connected parts. The signs$^{(2)}$ are combined to form configurations$^{(2)}$ under the restriction in $\mathcal{R}$ that the result defines a finite number of connected regions in the plane; we could use the convention that when we run through a contour we have the region to our left. In this way we would get a *model of contour sets*, the further specification of which could proceed as in the previous section. Recognition will be organized by first recognizing the signs$^{(1)}$ and then joining them into signs$^{(2)}$ which gives us the boundary.

Note that since the independent signs were chosen as the arcs from which contours can be formed we get a very flexible model. Actually, it may be too wide to give us an economical description of a phenomenon that should be described *via* a more rigid grammatical structure. In that case we may *take as signs certain given sets* (rather than arcs) and we shall discuss this model in the case of G-translations of the plane. Let us start with a finite set of templates t_ν, $\nu = 1, 2, \cdots p$, and let each t_ν be a closed set bounded by a finite number of finite analytic arcs $a_{\nu 1}$, $a_{\nu 2}$, $\cdots$. The images are formed by superposition of translates of these templates, so that no two signs of the image have any inner point in common. Recognition is most easily accomplished when no $a_{\nu\mu}$ can be analytically continued into the translate of itself (with $g \neq e$) or into any other a_{um}. Consider the boundary ∂I of a given image I. It will consist of a finite number of arcs to each of which is uniquely associated a sign. Subtract all these signs from I and denote the result by I'. Then we apply the same procedure to $\partial I'$ and so on until we have a complete grammatical analysis of I which will be used for recognizing the pattern to which I belongs.

In case some arc can be analytically continued into a translate of itself or into some other arc, the procedure should be modified as follows. Consider ∂I again and one of the analytic arcs A making up ∂I. Now we cannot rule out the possibility that A is the union of several a's. If A is not closed we start at an endpoint of it and decompose it into a's as we go along it; this may be possible in more than one way but the number of possibilities is finite. If A is periodic this number is also finite. Indeed, assume the contrary and use the compactness of A. Writing arcs in the form $(x, y) = (f_1(s), f_2(s))$ this implies that some segment of A can be written as

$$(f_1(s), f_2(s)) = (h_1^{(\nu)}, h_2^{(\nu)}) + (\phi_1(s + k_\nu), \phi_2(s + k_\nu)) \tag{9.1}$$

with $\nu = 1, 2, \cdots$ and the $k_\nu s$ are infinitely many. But this implies that the ν_1 and ν_2 are constants, which is impossible; hence only a finite number of alternatives exist. Now we proceed as before. First ∂I is represented in at most a finite number of ways as unions of a's. For each such possibility we subtract from I the corresponding signs if their union is a subset of I; consider $\partial I'$ and so on, continuing each branch of the decision tree until it ends. Although the resulting grammatical analysis is not always unique it simplifies the decision problem a good deal.

In these models of set patterns we have formed images through unions of certain sets. Nothing prevents us from taking intersections instead, and this will be described in a particular case of some interest. Let the signs consist of closed halfspaces of the plane and take *the images as the intersection of a finite number of such signs.* In other words, the images are just the convex polygons. The grammatical analysis of such simple images is straightforward. To each direction ϕ, $0 \leq \phi < 2\pi$, we determine the line L_ϕ of support of I. This gives us a finite number of angles ϕ and the corresponding half planes that support I are the component signs of I.

Line, surface and set patterns in R^3 or higher-dimensional Euclidean space will not be discussed here, since this generalization would present nothing essentially different from what happens in the plane.

10. Boundary Patterns. With the experience gained through discussing the image algebras of the last sections, we are now ready to formulate a model of general interest.

For any given subroutine the signs should be of the form

$$s = [\sigma; b_1, b_2, \cdots b_\rho] \tag{10.1}$$

where the parameters b_i take values in certain given spaces $B_{\sigma_1}, B_{\sigma_2}, \cdots B_{\sigma_\rho}$ and the number ρ may depend on σ. Consider a group G of transformations g mapping the product space $B_{\sigma_1} \times B_{\sigma_2} \times \cdots B_{\sigma_\rho}$ onto itself and we shall write

$$gs = [\sigma; (gb)_1, (gb)_2, \cdots (gb)_\rho],$$

where b is the vector $b = (b_1, b_2, \cdots b_\rho)$. If s_1 and s_2 belong to the same subroutine and we write

$$(10.2) \qquad \begin{cases} s_1 = [\sigma; b_1^{(1)}, b_2^{(2)}, \cdots b_\rho^{(1)}] \\ s_2 = [\sigma; b_1^{(2)}, b_2^{(2)}, \cdots b_\rho^{(2)}], \end{cases}$$

they will connect and reduce into a single sign if the $b_i^{(1)}$ and $b_i^{(2)}$ satisfy a set of condition c_σ. We can consider c_σ as a subset of the product space $B_{\sigma_1} \times B_{\sigma_1} \times B_{\sigma_2} \times B_{\sigma_2} \times \cdots \times B_{\sigma_\rho} \times B_{\sigma_\rho}$ and it will be assumed that it is invariant: $gc_\sigma = c_\sigma$, $\forall g$. On this product space functions $\beta_1, \beta_2, \cdots \beta_\rho$ will be given taking values in $B_{\sigma_1}, B_{\sigma_2}, \cdots B_{\sigma_\rho}$ respectively such that

$$(10.3) \qquad s_1 + s_2 = [\sigma; \beta_1, \beta_2, \cdots \beta_\rho].$$

These functions should have the property that

$$(10.4) \qquad \beta_i(gb_1^{(1)}, gb_1^{(2)}, \cdots) = g\beta_i(b_1^{(1)}, b_1^{(2)}, \cdots).$$

In this way we can simplify (see section 3) configurations $c \to c'$ by repeated reductions of the component signs. If we assume that $c_1 \to c$, $c_2 \to c$ implies the existence of a configuration $c_{12} \to c_1$, $c_{12} \to c_2$, we can define the identification relation R as follows. For two configurations c_1 and c_2 we shall say that c_1Rc_2 if there exists a configuration $c \to c_1$ and $c \to c_2$. Clearly R is reflexive and symmetric. If c_1Rc_2 and c_2Rc_3 we know that configurations c_{12} and c_{23} exist such that $c_{12} \to c_1$, $c_{12} \to c_2$, $c_{23} \to c_2$, $c_{23} \to c_3$. But then there is a $c \to c_{12}$, $c \to c_{23}$ so that c_1Rc_3; the relation is transitive. The reader may now verify that the conditions of Postulate 3.1 are satisfied and we have constructed an image algebra. We shall speak of this model as one of *boundary patterns* since the parameters b_i often play the role of boundaries or boundary values. To make the model appear more intuitive we shall consider a few examples.

Let a subroutine σ consist of rectangles (in the plane) with sides parallel to the coordinate axes and represent its signs as

$$(10.5) \qquad s = [\sigma; b_1, b_2, b_3, b_4]$$

where

$$(10.6) \qquad s = \{x, y \mid b_1 \{\leq x \leq b_2 ; b_3 \leq y \leq b_4\}.$$

We choose G as translations of the plane. The sets are combined as in section 9. The conditions c_σ take the form

$$(10.7) \qquad c_\sigma = \begin{cases} b_1^{(1)} = b_1^{(2)}, b_2^{(1)} = b_2^{(2)}, b_3^{(1)} = b_4^{(2)} \\ \text{or} \\ b_1^{(1)} = b_1^{(2)}, b_2^{(1)} = b_2^{(2)}, b_4^{(1)} = b_3^{(2)} \\ \text{or} \\ b_2^{(1)} = b_1^{(2)}, b_3^{(1)} = b_3^{(2)}, b_4^{(1)} = b_4^{(2)} \\ \text{or} \\ b_1^{(1)} = b_2^{(2)}, b_3^{(1)} = b_3^{(2)}, b_4^{(1)} = b_4^{(2)} \end{cases}$$

and the βs can easily be written down; say if the first of the four conditions is satisfied.

$$(10.8) \qquad \beta_1 = b_1^{(1)}, \beta_2 = b_2^{(1)}, \beta_3 = b_3^{(2)}, \beta_4 = b_4^{(4)}.$$

As a second example we take the signs in a certain subroutine as analytic arcs:

$$s = [\sigma \mid b_1 , b_2] \tag{10.9}$$

where b_1, b_2 are endpoints of the arc. Let G = Euclidean motions of the plane. We then get

$$c_\sigma = \begin{cases} b_1^{(1)} = b_1^{(2)} \\ \text{or} \\ b_1^{(1)} = b_2^{(2)} \\ \text{or} \\ b_2^{(2)} = b_1^{(2)} \\ \text{or} \\ b_2^{(1)} = b_2^{(2)} \end{cases} \tag{10.10}$$

and the βs are determined by the fact that the resulting arc should pass through the points $b_1^{(1)}, b_2^{(1)}, b_1^{(2)}, b_2^{(2)}$.

Finally, a third example that will be examined in section 12: G = translations of the real line. Let a sign in S_σ be defined as a solution to the first-order differential equation $L_\sigma = 0$ on the interval $b_1 \leq x \leq b_2$ and write

$$s = [\sigma \mid b_1 , b_2 , b_3 , b_4] \tag{10.11}$$

where $b_3 = s(b_1)$, $b_4 = s(b_2)$. We can join two such signs together by concatenation if

$$\begin{cases} b_2^{(1)} = b_1^{(2)} , b_4^{(1)} = b_3^{(2)} \\ \text{or} \\ b_1^{(1)} = b_2^{(2)} , b_3^{(1)} = b_4^{(1)} \end{cases} \tag{10.12}$$

and if the first condition is satisfied

$$\beta_1 = b_1^{(1)} , \beta_2 = b_2^{(2)} , \beta_3 = b_3^{(1)} , \beta_4 = b_4^{(2)} \tag{10.13}$$

Signs are used in any image algebra to carry information; information about what template it belongs to and about what similarity transformation was applied to the template. Some image algebras have the property that given any image it can be analyzed uniquely into component signs. Admittedly this is a very special case, but it deserves some attention. Say first that we use only configurations formed from n different templates $t_1 , t_2 , \cdots t_n$ so that we can write $c = (s_1 , s_2 , \cdots s_n) = (g_1t_1 , g_2t_2 , \cdots g_nt_n)$. Since the s_νs are uniquely determined by I our image can be represented just as well as

$$I = (g_1 , g_2 , \cdots g_n) \in G_n \tag{10.14}$$

where G_n is the group obtained as a direct product $G \times G \times G \times \cdots \times G$ of n copies of G. Introduce the subgroup Γ of all elements in G_n of the form $(g, g, \cdots g)$. Then the patterns $\mathcal{P}_1 , \mathcal{P}_2 , \cdots$ consist of unions of cosets to Γ and recognition means simply identifying the cosets to which the image belongs. On the other hand, if our configuration is formed from n_1 signs generated by one template, n_2 signs generated from another template and so on, $n_1 + n_2 + \cdots + n_k = n$, then we can represent an image as

$$I = (g_1, g_2, \cdots g_{n_1}, g_{n_1+1}, \cdots) \in G_n^s, \tag{10.15}$$

where G_n^s consists of the elements of G_n but where $(g_1^{(1)}, g_2^{(1)}, \cdots g_{n_1}^{(1)}, g_{n_1+1}^{(1)}, \cdots)$ and $(g_1^{(2)}, g_2^{(2)}, \cdots g_{n_1}^{(2)}, g_{n_1+1}^{(2)} \cdots)$ are considered identified if any permutation of the n_1 first indices carries $g_1^{(1)}, g_2^{(1)}, \cdots g_{n_1}^{(1)}$ into $g_1^{(2)}, g_2^{(2)}, \cdots g_{n_1}^{(2)}$ and so on. In other words, G_n^s is G_n mudolo the product of k symmetric groups $S_{n_1} \times S_{n_2} \times \cdots \times S_{n_k}$. Recognition now can be achieved by classifying the image as to what coset of Γ (in G_n^s) it belongs to. We shall return to this model of *group patterns* in Part IV.

11. Contrast Patterns. Line and set patterns are well suited as models of images whose appearance show strong contrast between their different parts: we observe the drastic dichotomy between the inside and outside of a set. This made it reasonable to concentrate our attention on the boundary. But if the variation in contrast is smoother we must modify the model. We shall do that now, arriving at a useful and general model.

Model of contrast patterns. Consider a set X, the background space, and a semi-group C, the contrast space. Certain given functions t mapping X into C are the templates. A group G of transformations g from X onto X determines the similarity transformations through the definition $gt(x) = t(gx)$. For two configurations $c = (s_1, s_2, \cdots s_n)$, $c' = (s_1', s_2', \cdots s_m')$ we put cRc' if

$$s_1(x) + s_2(x) + \cdots + x_n(x) = s_1'(x) + s_2'(x) + \cdots + s_m'(x), \ \forall\ x \tag{11.1}$$

where addition means the binary operation in C.

It is easily verified that this satisfies the postulates of an image algebra if we add the usual condition on the subroutines, and it can also be shown that this model includes the concrete cases we have been studying. Indeed, the definition is embarrassingly general: given any image algebra $\mathfrak{I}$ it can be interpreted in terms of contrast patterns. We merely have to take the background X as G, the contrast C as $\mathfrak{I}$ itself with its semi-group operation of combination of images. Let an image I be represented by a function f_I mapping $G \to \mathfrak{I}$

$$f_I(\gamma) = \gamma I, \gamma \in G. \tag{11.2}$$

We let g operate on the γ's through the mapping $\gamma \to \gamma g$. In this construction the contrast space is taken as the (often) complicated space of the image algebra. Although this notion of contrast patterns is quite general, we shall not use it unless both the background X and the contrast C are sufficiently simple.

Sometimes we modify the model slightly by choosing the similarity transformations a bit differently. Let there be given a group H of transformations h of the contrast space C onto itself and define the similarity transformations of a sign $s(x)$ through

$$s(x) \to h\ s(gx), h \in H, g \in G, \tag{11.3}$$

so that the similarity transformations are given by the direct product $G \times H$. We shall call the h's the *contrast transformations*.

Little can be said about the recognition problem for general contrast patterns. It may be pointed out, though, that if X is a Euclidean space R^n the following may work if the grammar is simple enough. Look for linear, G-invariant functions L_t in $\mathfrak{J}$ with the property that

$$(11.4) \qquad L_t s = \begin{cases} 1 & \text{if} \quad t \equiv s (\text{mod } G) \\ 0 & \text{if} \quad t \not\equiv s (\text{mod } L) \end{cases}$$

where t is an arbitrary template and s any sign. If such functionals can be found we can use them as probes. Apply these probes to an image I. The result tells us what templates have been used to form I. The only thing that remains for the grammatical analysis to be complete is to find the similarity transformations that carry these templates into the component signs of I.

12. Time Patterns. Up to now most of the image algebras discussed in Part II have been commutative. The following model intended to describe patterns in time does not have that property. Let us take the signs to be real-valued functions $s = s(t)$ defined on half-open intervals [a, b). We form images by concatenation supposing ($\mathcal{R}$) that successive intervals are adjacent, and the rule R that cRc' if the signs of the two configurations are defined on the same subsets of the real line and the values of the corresponding signs agree. Starting from a group G of transformations of the time axis onto itself, we define similarity by $I \equiv I'$ if there is a $g \in G$ such that $I(t)$ and $I'(gt)$ are defined for the same t's and $I(t) = I'(gt)$ for all these t's.

As an example of this, the reader may think of Morse code with its signs: dot, dash and various spaces. In Morse code the ratio between the time duration of the signs is fixed, so that G would consist of the transformations of the time axis of the form

$$(12.1) \qquad t \to at + b, \qquad a > 0.$$

Specializing a bit further, we consider r time-independent differential operators $L_1, L_2, \cdots L_r$ all of order p and such that the initial value problem has a unique solution. Two equations $L_i s = 0$ and $L_j s = 0$, $i \neq j$, will have the same solution except for one solution which may play the role of the empty sign. If the operators are linear this solution could be taken as the one that is identically zero. The solutions of $L_i s = 0$ are the signs of the subroutine S_i. Rule $\mathcal{R}$ may be taken to say that the solutions should be joined together so that derivatives up order $p - 1$ are continuous. In case we work with discrete time the L_i's should of course be chosen as difference operators.

It is easy to recognize the patterns of this model. A template t can be written as

$$(12.2) \qquad t = (\sigma; l; t_0, t_1, \cdots t_{p-1}),$$

where the template satisfies $L_\sigma t = 0$ on the interval $[0, l]$ and with the initial values $t(0) = t_0$, $t'(0) = t_1$, $\cdots t^{(p-1)}(0) = t_{p-1}$. The signs similar to t are obtained by translating the time axis. Given an image I we apply the operators L_σ to this time function for all the values of time for which it is defined. This will classify the time points according to what subroutine σ they belong to, $L_\sigma I = 0$. In this way we get the whole time interval where the image is defined decomposed into the intervals corresponding to the component signs of the image. Finally, we just have to read off the initial values of these signs.

In Parts I and II we have only dealt with the images in their pure form. That is the easy part of the theory. We must now turn to the study of deformations of images, which brings out the typical features of pattern analysis.

Part III. Deformed Patterns

13. Deformed Images. Let us consider a grammar $\mathcal{G}$ of patterns consisting of the basic elements

$$\mathcal{G}: \begin{cases} S, & \text{the signs arranged in subroutines} \\ \mathcal{R}, & \text{the rule of forming configurations} \\ R, & \text{the identification rule} \\ \mathfrak{I}, & \text{the set of images} \\ \mathcal{P}, & \text{the family of patterns} \end{cases} \tag{13.1}$$

We cannot observe the objects of this pattern grammar directly; their form is corrupted by deformations. These deformations act upon those objects of the grammar that can be observed: the pure images $I \in \mathfrak{I}$ which are transformed into the deformed images $dI = I^{\mathfrak{D}} \in \mathfrak{I}^{\mathfrak{D}}$. *Each element d of the set $\mathfrak{D}$ of deformations maps the (pure) image algebra $\mathfrak{I}$ into the (deformed) image algebra $\mathfrak{I}^{\mathfrak{D}}$.*

Usually the grammar $\mathcal{G}$ is a rigid logical structure, but the deformation grammar $\mathcal{G}_{\text{def}}$ can be more flexible. In most of the cases we shall consider in Part IV the deformed images include the pure ones, $\mathfrak{I} \subset \mathfrak{I}^{\mathfrak{D}}$, and $\mathfrak{I}^{\mathfrak{D}}$ is more extensive than $\mathfrak{I}$. We shall then assume that the similarity transformations g of $\mathfrak{I}^{\mathfrak{D}}$ coincide (on $\mathfrak{I}$) with the similarity transformations originally introduced on $\mathfrak{I}$.

Since dI is defined for any $I \in \mathfrak{I}$ this holds especially for the case of a sign $I = s$. But we cannot go the other way; knowing the value of ds for all $d \in \mathfrak{D}$, $s \in S$ does not enable us in general to find the value of dI, all $I \in \mathfrak{I}$.

Definition 13.1. *The deformations are said to be*

a) *covariant if $gd = dg$ holds identically.*
b) *homomorphic if $d(I_1 + I_2) = dI_1 + dI_2$ holds identically.*

It is obvious that if the deformations are homomorphic it suffices to define them for signs and that if they are also covariant it is enough to introduce the deformations for the templates.

A good deal of what follows would be considerably simplified if we could assume that the deformations were covariant, but after examining some of the cases exhibited in Part IV it should be clear that this represents a rare case indeed.

How should the recognition problem of section 5 be reformulated when we start from deformed images? It is clear that we can no longer ask for the inverse mapping from $\mathfrak{I}^{\mathfrak{D}}$ to $\mathcal{P}$ since $\mathfrak{I}^{\mathfrak{D}}$ depends not only upon I but also upon d which is not observable. In the next three sections this will be discussed, but we should point out already here that we can limit the choice of recognition functions somewhat through the following invariance considerations. Say we have a mapping $\alpha : \mathfrak{I}^{\mathfrak{D}} \to \mathcal{P}$, a recognition function. It is natural to assume that *similar images* should be mapped into the same pattern: $\alpha(gI^{\mathfrak{D}}) = \alpha(I^{\mathfrak{D}})$. On the other hand, it also seems as desirable to demand that two pure images, gI and I (belonging to the *same prototype* I) after transformation by the *same deformation* d should be mapped into the same pattern: $\alpha(dgI) = \alpha(dI)$.

Definition 13.2 *A recognition function $\alpha : \mathfrak{I}^{\mathfrak{D}} \to \mathcal{P}$ is said to be*

a) *invariant (G) if $I_1^{\mathfrak{D}} \equiv I_2^{\mathfrak{D}} \pmod{G}$ implies $\alpha(I_1^{\mathfrak{D}}) = \alpha(I_2^{\mathfrak{D}})$.*
b) *invariant ($\mathcal{P}$) if $I_1 \equiv I_2 \pmod{G}$ implies $\alpha(dI_1) = \alpha(dI_2)$.*

What is the relation between these two notions of invariance? The following result is almost obvious but not very useful, since it assumes covariant deformations.

Theorem 13.1 *If the deformations in $\mathfrak{D}$ are covariant, invariance* (G) *and invariance* ($\mathcal{P}$) *mean the same.*

Proof: If α is invariant ($\mathcal{P}$) it follows that $\alpha(gdI) = \alpha(dgI) = \alpha(dI)$ for any d, g, I so that α is invariant (G) and conversely.

A bit more useful is

Theorem 13.2 *If $\mathfrak{D}$ is a group with G as a normal subgroup invariance* ($\mathcal{G}$) *implies invariance* ($\mathcal{P}$). *The pattern family $\mathcal{P} = \mathfrak{I}/G$ is mapped by the deformations* (*via the factor group $\mathfrak{D}/G = F$*) *into the deformed pattern family $\mathcal{P}^{\mathfrak{D}} = \mathfrak{I}^{\mathfrak{D}}/G$.*

Proof: For an arbitrary pure image I and d and g we can find a similarity transformation $h \in G$ such that $dh = gd$ (this merely uses the fact that G is normal) so that if α is invariant ($\mathcal{P}$), it follows that

$$\alpha(gdI) = \alpha(dhI) = \alpha(dI) \tag{13.2}$$

which shows that α is invariant (G) as stated. Consider two similar pure images I_1, I_2 from one pattern $\mathcal{P}_1 \in \mathcal{P} = \mathfrak{I}/G$ so that we can write $I_1 = g_1I$, $I_2 = g_2I$ where I is a prototype from $\mathcal{P}_1$. But the element $h = d\, g_2^{-1}\, d^{-1} \in G$, since G is normal. Hence the deformed image $I_1^{\mathfrak{D}} = dg_1I = hgd_2I = hdI_2 = hI_2^{\mathfrak{D}}$ so that $I_1^{\mathfrak{D}} \equiv I_2^{\mathfrak{D}}$ and these two deformed images belong to the same deformed pattern $\mathcal{P}_1^{\mathfrak{D}} = \mathfrak{D}\mathcal{P}_1 \in \mathcal{P}^{\mathfrak{D}}/G$. It is clear that the only thing that matters for defining the value of the mapping $d : \mathcal{P} \to \mathcal{P}^{\mathfrak{D}}$ is what coset $f \in F$ of $G\,d$ belongs to.

It should be pointed out that $\mathfrak{D}$ is not likely to form a group when the deformations really disfigure the images. A dramatic example of this is given by the mask deformations to be introduced in section 19.

A *deformation grammar* $\mathcal{G}_{\text{def}}$ of patterns can then be written as a triple

$$\mathcal{G}_{\text{def}} = (\mathcal{G}, \mathfrak{D}, \mathfrak{I}^{\mathfrak{D}}) \tag{13.3}$$

where $\mathcal{G}$ is the grammar of pure patterns, $\mathfrak{D}$ is the set of deformations and $\mathfrak{I}^{\mathfrak{D}}$ the set of deformed images.

14. Metric Pattern Grammars. Given a deformation grammar $\mathcal{G}_{\text{def}}$, how do we choose among the many possible recognition functions? One rationale presents itself immediately when the set of images possesses a metric. In this section we shall examine this approach assuming that $\mathfrak{I} \subset \mathfrak{I}^{\mathfrak{D}}$.

A real-valued function $\delta(I_1, I_2)$ defined on $\mathfrak{I}^{\mathfrak{D}} \times \mathfrak{I}^{\mathfrak{D}}$ shall be called a *completely invariant distance* if

(i) $\delta(I_1^{\mathfrak{D}}, I_2^{\mathfrak{D}}) \geq 0$ with equality if and only if $I_1^{\mathfrak{D}} \equiv I_2^{\mathfrak{D}} \pmod{G}$,
(ii) δ is symmetric and satisfies the triangle inequality,
(iii) $\delta(g_1I_1^{\mathfrak{D}}, g_2I_2^{\mathfrak{D}}) = \delta(I_1^{\mathfrak{D}}, I_2^{\mathfrak{D}})$.

Occasionally it will be convenient to allow δ to take the value $+\infty$; the reader will then have to interpret the above conditions in the appropriate way.

But (iii) implies that $\delta(I_1^{\mathfrak{D}}, I_2^{\mathfrak{D}})$ only depends on what patterns in $\mathcal{P}^{\mathfrak{D}} = \mathfrak{I}^{\mathfrak{D}}/G$ the deformed images are generated from, so that δ can be considered as some function Δ on

$\mathcal{P}^{\mathfrak{D}} \times \mathcal{P}^{\mathfrak{D}}$. Because of (ii) the function Δ has the properties of a distance; (i) guarantees that Δ separates the patterns of $\mathcal{P}^{\mathfrak{D}}$.

We shall give examples of completely invariant distance functions in Part IV, but already here we mention the following simple construction. Let δ_0 be *any* distance function in $\mathfrak{J}^{\mathfrak{D}} \times \mathfrak{J}^{\mathfrak{D}}$ with the usual properties of a metric. Put

$$\delta(I_1^{\mathfrak{D}}, I_2^{\mathfrak{D}}) = \min_{g_1, g_2 \in G} \delta_0(g_1 I_1^{\mathfrak{D}}, g_2 I_2^{\mathfrak{D}}) \tag{14.1}$$

where it will be assumed that the minimum is attained. It is then easy to show that conditions (i), (ii) and (iii) are satisfied : δ is a completely invariant distance.

Definition 14.1 *Let δ be a completely invariant distance and consider the recognition function α*

a) *for a given deformed image $I^{\mathfrak{D}}$ choose the pure pattern $\mathcal{P}_\alpha \in \mathcal{P} = \mathfrak{J}/G$ for which $\delta(I, I^{\mathfrak{D}})$, $I \in \mathcal{P}_\alpha$, is minimized (the minimum is assumed to be attained for a unique $\mathcal{P}_\alpha$)*

b) *if for the given deformed image $I^{\mathfrak{D}}$ and any pure image I we have $\delta(I, I^{\mathfrak{D}}) = +\infty$, no decision is made.*

This is called minimum distance recognition.

Since we can just as well say that α minimizes $\Delta(\mathcal{P}_\alpha, \mathcal{P}^{\mathfrak{D}})$, we get in this way a mapping $\mathcal{P}^{\mathfrak{D}} \to \mathcal{P}$. From this it follows that α is a (G)-invariant recognition function.

If (b) happens this is interpreted as an indication that the deformation has been so drastic that the original (pure) image cannot possibly be recognized; only the deformed pattern classes at a finite distance from $\mathcal{P}$ can be processed in a meaningful way.

We are led to a weaker notion of an invariant metric if we ask only that

$$\delta(gI_1^{\mathfrak{D}}, gI_2^{\mathfrak{D}}) = \delta(I_1^{\mathfrak{D}}, I_2^{\mathfrak{D}}) \tag{14.2}$$

hold for all $g \in G$. Then δ will be called an *invariant distance.* Starting from an arbitrary distance δ_0 we can construct an invariant distance as before but with the change that we now put

$$\delta(I_1^{\mathfrak{D}}, I_2^{\mathfrak{D}}) = \min_{g \in G} \delta_0(gI_1^{\mathfrak{D}}, gI_2^{\mathfrak{D}}) \tag{14.3}$$

instead of (14.1). It should be noted that an invariant distance does not generally lead to a distance on $\mathcal{P}$ as was shown to be the case for a completely invariant δ. At any rate, we can still get a recognition rule α just as in Definition 14.1. Although this function α no longer gives us a mapping $\mathcal{P}^{\mathfrak{D}} \to \mathcal{P}$ it is still (G)-invariant.

A deformation grammar $\mathcal{G}_{\text{def}}$ together with an invariant distance will be called a metric grammar of patterns and written as

$$\mathcal{G}_{\text{met}} = (\mathcal{G}_{\text{def}}, \delta). \tag{14.4}$$

Recognition by minimum distance will be helpful only if the metric introduced in $\mathfrak{J}^{\mathfrak{D}}$ is an adequate expression of our intuitive understanding of how close or distant deformed images are from each other. Very often we have some vague idea of how certain deformations require more *effort* of some kind. In the simplest case this effort would be just physical energy, but we should not restrict ourselves to this case. Generally we formalize the notion of effort of a transformation in the following way.

Definition 14.5 *Assuming that $\mathfrak{D}$ forms a semigroup, consider a real-valued function $e(d)$ on $\mathfrak{D}$ with the properties*

a) $e(d) \geq 0$, $e(d) = 0$ *iff* $d \in G$
b) $e(gdh) = e(d)$, $\forall\ g, h \in G$
c) $e(d_1 d_2) \leq e(d_1) + e(d_2)$.

Then $e(d)$ is called a completely invariant effort function.

If the deformation grammar $\mathcal{G}_{\mathrm{def}}$ possesses a naturally defined, completely invariant effort $e(d)$ we can construct a recognition function α by *minimum effort* by solving the extremum problem (for given deformed image $I^{\mathfrak{D}}$ and free pure image I)

$$e(d) = \min \text{ for } I^{\mathfrak{D}} = dI, \tag{14.5}$$

where the minimum is assumed to be attained. If (14.5) has the solution (d, I) it then follows from (b) of Definition (14.5) that (dg^{-1}, gI) is also a solution. Also any $hI^{\mathfrak{D}}$ leads to the solution (hdg^{-1}, gI) so that $\mathfrak{J}^{\mathfrak{D}}/G$ is mapped into $\mathfrak{J}/G$. This also implies that recognition by minimum effort is (G)-invariant.

This is related to recognition by minimum distance. In fact, we have

Theorem 14.1 *Let e be a completely invariant effort and put*

$$\delta(I_1^{\mathfrak{D}}, I_2^{\mathfrak{D}}) = \min_{d_1, d_2} [e(d_1) + e(d_2); \quad I_1^{\mathfrak{D}} = d_1 I_2^{\mathfrak{D}} \text{ and } I_2^{\mathfrak{D}} = d_2 I_1^{\mathfrak{D}}] \tag{14.6}$$

assumed to be attained. Then δ is a completely invariant distance.

Proof: To verify conditions (i), (ii) and (iii) of the deformation of a completely invariant distance we note that (i) holds since $\delta(I_1^{\mathfrak{D}}, I_2^{\mathfrak{D}}) = 0$ if and only if $I_1^{\mathfrak{D}} \equiv I_2^{\mathfrak{D}}$ (mod G). The function δ is symmetric in its two arguments and, to prove the triangle inequality, we consider fixed $I_1^{\mathfrak{D}}$, $I_2^{\mathfrak{D}}$, and $I_3^{\mathfrak{D}}$. If there exist d_1, d_2, d_3, d_4 such that

$$\begin{cases} I_1^{\mathfrak{D}} = d_2 I_2^{\mathfrak{D}} \\ I_2^{\mathfrak{D}} = d_1 I_1^{\mathfrak{D}} \\ I_2^{\mathfrak{D}} = d_3 I_3^{\mathfrak{D}} \\ I_3^{\mathfrak{D}} = d_4 I_2^{\mathfrak{D}} \end{cases} \tag{14.7}$$

we write $d^{(1)} = d_2 d_3$, $d^{(2)} = d_4 d_1 \in \mathfrak{D}$, so that

$$\begin{cases} I_1^{\mathfrak{D}} = d^{(1)} I_3^{\mathfrak{D}} \\ I_3^{\mathfrak{D}} = d^{(2)} I_3^{\mathfrak{D}} . \end{cases} \tag{14.8}$$

Hence, using property (c),

$$e(d^{(1)}) + e(d^{(2)}) \leq e(d_1) + e(d_2) + e(d_3) + e(d_4),$$

from which follows

$$\begin{aligned} \delta(I_1^{\mathfrak{D}}, I_3^{\mathfrak{D}}) &\leq \min [e(d^{(1)}) + e(d^{(2)}); d^{(1)} = d_2 d_3, d^{(2)} = d_4 d_1] \\ &\leq \min [e(d_1) + e(d_2); \quad I_1^{\mathfrak{D}} = d_2 I_2^{\mathfrak{D}} \text{ and } I_2^{\mathfrak{D}} = d_1 I_1^{\mathfrak{D}}] \\ &\quad + \min [e(d_3) + e(d_4); \quad I_2^{\mathfrak{D}} = d_3 I_3^{\mathfrak{D}} \text{ and } I_3^{\mathfrak{D}} = d_4 I_2^{\mathfrak{D}}] \\ &= \delta(I_1^{\mathfrak{D}}, I_2^{\mathfrak{D}}) + \delta(I_2^{\mathfrak{D}}, I_3^{\mathfrak{D}}). \end{aligned} \tag{14.9}$$

Finally, (iii) follows from condition (b) since $I_1^{\mathfrak{D}} = dI_2^{\mathfrak{D}}$ implies $g_1 I_1^{\mathfrak{D}} = (g_1 d g_2^{-1})\, g_2 I_2^{\mathfrak{D}}$ so that $e(g_1 d g_2^{-1}) = e(d)$. This proves that δ is completely invariant.

If instead of (b) we had assumed only that $e(gdg^{-1}) = e(d)$ for all $d \in \mathfrak{D}$, $g \in G$, we would only be able to prove that the resulting distance function δ is invariant.

We have tacitly assumed that the relation $I_1^{\mathfrak{D}} = dI_2^{\mathfrak{D}}$ viewed as an equation in d always has at least one solution. If this is not so, the corresponding value of $e(d)$ should be replaced by $+\infty$, in complete analogy with what was said earlier in this section.

Finally, let us mention the possibility of using a "distance" which is not necessarily symmetric in its two arguments. This may be reasonable when taking into account that in recognition we are concerned with ordered pairs $(I, I^{\mathfrak{D}})$ where I and $I^{\mathfrak{D}}$ belong (in general) to different image algebras.

15. Probabilistic Pattern Grammars. In a metric deformation grammar we use the notion of distance between images to get a reasonable recognition procedure. The rationale behind this approach is that we would believe more in an analysis that "explains" the given deformed image by means of a small deformation than in an alternative "explanation" involving a large deformation. This appeal to Occam's razor seems reasonable only when the metric fits well and naturally into the grammatical structure.

But we can make the analysis of recognition more convincing if we possess some information about how likely the different deformations are. Indeed, we could then appeal to statistical criteria and methods in order to arrive at efficient recognition techniques. The reader may be reminded of the set-up used in the statistical theory of communication theory; an added difficulty is that in the present context the parameter and sample spaces are image algebras with all their complexities.

To formalize this, let us assume that on some σ-algebra $\mathfrak{A}(\mathfrak{D})$ of subsets of $\mathfrak{D}$ we have defined a probability measure P. In $\mathfrak{J}$ and $\mathfrak{J}^{\mathfrak{D}}$ we also have σ-algebras $\mathfrak{A}(\mathfrak{J})$ and $\mathfrak{A}(\mathfrak{J}^{\mathfrak{D}})$ such that for any pair of sets $E \in \mathfrak{A}(\mathfrak{J})$, $F \in \mathfrak{A}(\mathfrak{J}^{\mathfrak{D}})$ we have

$$\{d \mid dE \in F\} \in \mathfrak{A}(\mathfrak{D}). \tag{15.1}$$

The model

$$\mathcal{G}_{\text{prob}} = (\mathcal{G}_{\text{def}}\,,\, P) \tag{15.2}$$

will then be called a *probabilistic deformation grammar*. The author believes that this set-up is adequate for most pattern recognition problems.

In section 13 we introduced the rather restrictive idea of covariant deformations. Extending this to a more useful concept, we shall say that the probabilistic deformation grammar is *covariant in probability* if for any similarity transformation $g \in G$ the stochastic mappings dg and gd have the same probability distribution. If P is covariant in probability we have for any $g \in G$, $E \in \mathfrak{A}(\mathfrak{D})$

$$P(d \in E) = P(gd \in gE) = P(dg \in gE) = P(d \in gEg^{-1}). \tag{15.3}$$

Conversely, if (15.3) holds P is covariant in probability.

Whether P is covariant in probability or not, it will sometimes be convenient to express P in terms of a frequency formation p:

$$P(E) = \int_E p(d)\, d\mu(d) \tag{15.4}$$

where μ is a σ-finite measure covariant in probability.

Before approaching the recognition problem, let us examine the following simpler question. Given a deformed image $I^{\mathfrak{D}}$ we want to find the pure image I that has generated $I^{\mathfrak{D}}$. Thus a recognition function α would define a mapping $\mathfrak{I}^{\mathfrak{D}} \rightarrow \mathfrak{I}$ and to be able to choose among such mappings we need a weight formation $w(I, J)$, real-valued and defined on $\mathfrak{I} \times \mathfrak{I}$. The value of w is the payoff when I is the true pure image and J is the one we have decided upon after observing our deformed image. We would then try to find the α that maximizes the expression

$$w_\alpha = E[w(I, \alpha(I^{\mathfrak{D}}))] \tag{15.5}$$

where E stands for expected value to be taken over $\mathfrak{I}$ and $\mathfrak{D}$. But this double integral is well defined only if we have access to a probability measure, Q, say, over $\mathfrak{I}$. We would then solve for that J that maximizes the conditional expectation

$$E[w(I, J) \mid dI = I^{\mathfrak{D}}] \tag{15.6}$$

The resulting pure images $J_0 = J(I^{\mathfrak{D}})$, considered as functions of $I^{\mathfrak{D}}$, give us the optimal recognition function; optimal in the sense of maximal expected payoff.

This brief discussion on the selection of recognition procedures should be enough to convince the reader that, at least in principle, we can appeal to the general and well-known principles of the theory of statistical decision functions. It is not necessary to mention the various possible variations of the argument above; instead, we proceed to a difficulty that is typical of the pattern recognition problem.

The trouble is that we do not always have at our disposal a probability distribution Q *on the image algebra*, so that the expression (15.5) is not always meaningful. Instead, we may have only a probability measure defined on the σ-algebra $\mathfrak{A}(\mathcal{P})$ generated by the patterns $\mathcal{P}_\alpha$. Let us use the same symbol Q for this measure and consider the case where the pure patterns are given by $\mathcal{P} = \mathfrak{I}/G$.

Theorem 15.1 *If the deformations are covariant in probability, the conditional probabilities $P(E \mid \mathcal{P}_\alpha)$ are well defined for any G-invariant set $E \in \mathfrak{A}(\mathfrak{I}^{\mathfrak{D}})$ and $\mathcal{P}_\alpha \in \mathfrak{I}/G$.*

Proof: Let I_0 be a prototype of $\mathcal{P}_\alpha$ and consider the functions of g

$$f(g) = P(dgI_0 \in E) = P(gdI_0 \in E) \tag{15.7}$$

since the deformations are covariant in probability. But

$$P(gdI_0 \in E) = P(dI_0 \in g^{-1}E) = P(dI_0 \in E) \tag{15.8}$$

since E was assumed to be G-invariant. Hence the function $f(g)$ is constant over G, which implies that

$$f(g) = P(dI \in E) = P(I^{\mathfrak{D}} \in E), \qquad I \text{ is the pure image,} \tag{15.9}$$

is constant for similar images: this conditional probability is constant over the similarity classes of $\mathcal{P} = \mathfrak{I}/G$, which proves the assertion.

This result has the following consequence. If Q is defined on $\mathcal{P} = \mathfrak{I}/G$ and if the deformations are covariant in probability, we can calculate the joint probability distribution over $\mathcal{P} \times \mathcal{P}^{\mathfrak{D}} = \mathfrak{I}/G \times \mathfrak{I}^{\mathfrak{D}}/G$. If the weight function $w(I, J)$ depends only upon what pure patterns I and J belong to and if α is a G-invariant recognition function, then the integral in (15.5) has a sense. This means that in this case *it is meaningful to ask for the best G-invariant recognition function in the sense of maximum expected payoff.*

In Part IV we shall see in a number of instances how the best recognition function can be derived. The result will depend upon what weight function w and what probability measure Q we use. Let us note already here that *the classical estimation method of maximum likelihood will lead to G-invariant recognition if the deformations are covariant in probability.* Indeed, if the frequency function p of (15.3) exists it is clear that $p(g^{-1}dg) = p(d)$ for all $g \in G$. *Maximum likelihood recognition* consists in solving the extremum problem

$$(15.10) \qquad p(d) = \max \text{ for all } i \text{ and } d \text{ satisfying the relation } dI = I^{\mathfrak{D}},$$

and recognizing that pure pattern $\mathcal{P}_\alpha$ to which I_0 belongs; I_0 is a pure image assumed to realize the maximum together with some deformation $g \in G$, we have $(gdg^{-1})gI_0 = gI_0^{\mathfrak{D}}$. Since $p(d) = p(gdg^{-1})$, it follows that the deformed image $gI^{\mathfrak{D}}$, which is similar to $I^{\mathfrak{D}}$, will be recognized as originating from the same pure pattern $\mathcal{P}_\alpha$ as $I^{\mathfrak{D}}$. The recognition function hence defines a mapping of $\mathcal{P}^{\mathfrak{D}}$ into $\mathcal{P}^\alpha$ so that it is G-invariant.

When using maximum likelihood recognition it should be remembered that the optimality properties which maximum likelihood estimates are known to have have been proved, in most cases, for the situation when the sample and parameter spaces are Euclidean spaces of finite dimension (fixed dimension in the asymptotic results). It would be worthwhile to study the sampling properties of maximum likelihood recognition when these spaces are the deformed and pure image algebras, as will be the case in pattern recognition.

IV. Deformation Grammars of Patterns

16. Constructing Deformation Mechanisms. Now we assume that the grammatical analysis of a pattern analysis problem has been carried to the stage when we have chosen the image algebras $\mathfrak{J}$, $\mathfrak{J}^{\mathfrak{D}}$ as well as the set $\mathfrak{D}$ of deformation transformations so that we have a deformation grammar $\mathcal{G}_{def}$. To make it into a probabilistic deformation grammar we must introduce a measure P on the set $\mathfrak{D}$. The choice of P will have to be made from case to case, using the available subject matter knowledge in order to arrive at the usual compromise: the model should approximate the phenomenon studied reasonably well and at the same time be analytically or numerically tractable. Nevertheless, we can state some general principles that may help us in constructing the model of deformations.

First, we should try to *separate* $\mathfrak{D}$, which may be quite a complicated space, into factors $\mathfrak{D} = \mathfrak{D}_1 \times \mathfrak{D}_2 \times \cdots$. The product may be finite, denumerably or non-denumerably infinite, as we shall see below. Sometimes we are given such a separation directly, for example when the deformations consist of a topological transformation of the background space followed by a mask deformation. We can also expect some help from the way in which the image algebras are formed from elementary parts. If we consider images whose configurations have n signs, all identifiable, we may try writing

$$(16.1) \qquad \begin{cases} I^{\mathfrak{D}} = dI = (d_1 s_1, d_2 d_2, \cdots d_n s_n) \\ \qquad\; I = (s_1, s_2, \cdots s_n) \end{cases}$$

in the hope that the properties of the factors d_ν should be tractable. But this would work only when the signs can be uniquely determined from the image. Instead we could try the corresponding separation, but for subroutine images or other components that are well determined in the image algebra under consideration.

Once we have been able to separate $\mathfrak{D}$ into reasonably simple factors we must decide upon what probability measure to introduce on $\mathfrak{D}$. To be able to do this it is important that the way in which we have separated the deformations into factors is such that the factors of d appear to behave independently of each other. The complete specification of P cannot be made without empirical measurements and in order to arrive at sufficiently accurate estimates we must have enough structure in the axiomatic model. This is a critical step in the determination of P and requires such understanding of the deformation mechanism that the ensuing grammatical analysis does not force our data into an inadequate mould. If we can actually carry out the separation so that the factors are stochastically *independent* we are left with the problem of determining the marginal distributions of the factors. To exemplify this, let us think of pure signs being generated by a mechanism of the form $L_\sigma x = 0$, where the reader may think of L_σ as a difference operator, and the deformed signs are given by $L_\sigma x = \epsilon$. The first thing we would try would then be to assume that the values of ϵ (for different arguments) are independent. If this could not be accepted as an adequate approximation we could try to remove the dependence by working not with x itself but with some transformation (possibly linear) of it. In other words, we could choose the grammar so that the deformation model has a simple probabilistic form. Or, as another example, if we deal with contrast patterns (see section 11) with a discrete background space X we may try to model P by assuming that different points of X are mapped independently of each other into the background space of $\mathfrak{J}^{\mathfrak{D}}$.

So far we have used the hierarchic construction of pattern grammars. To narrow down the choice of the marginal distributions that remain to be specified, we appeal to the role played by the similarity transformations. Again, if our $\mathcal{G}_{\mathrm{def}}$ has been well chosen we could hope that P *should have a corresponding invariance property*. Thus, if I and I' are similar pure images, $I' = gI$, we would first examine the possibility that dI and $dI' = dgI$ have the same probability distribution. Or we may try the model postulating that gdI has the same probability distribution as dgI: this leads us to covariance in probability (see section 15). In sections 17–20 we shall see in some detail how this can be done.

Applying these principles, we may be able to specify the analytic form of P and the remaining free parameters must be estimated empirically.

Here we have been discussing the specification of a probabilistic deformation grammar $\mathcal{G}_{\mathrm{prob}}$. If, instead, we are searching for a metric deformation grammar $\mathcal{G}_{\mathrm{met}}$ we would again appeal to the same principles: separation of $\mathfrak{D}$ into factors, independence of factors (here meaning additivity of sub-metrics) and invariance of the metric.

17. Deformation of Permutation Patterns. To begin with, consider the simple grammars of permutation patterns of section 7. If the coordinates x_ν are all different we can just as well let them take the values $1, 2, \cdots n$ and we could designate our signs as

$$\begin{cases} s_1 = (1, 0, 0, \cdots 0) \\ s_2 = (0, 2, 0, \cdots 0) \\ s_n = (0, 0, 0, \cdots n) \end{cases} \tag{17.1}$$

With G consisting of cyclic permutations the patterns $\mathcal{P} = \mathfrak{J}/G$ would be formed from a prototype I (a vector of zeroes and ones) and the images similar to I. Now to deformations. Let $\mathfrak{D}$ consist of transformations d that we separate into $d_1, d_2, \cdots d_n$. Each d_ν

takes values in G, and we put

(17.2) $$I_n^{\mathfrak{D}} = dI = d(x_1, x_2, \cdots x_n) = (x_{1+d_1}, x_{2+d_2}, \cdots x_{n+d_n}).$$

We make this deformation grammar $\mathcal{G}_{\text{def}} = (\mathcal{G}, \mathfrak{D}, \mathfrak{J}^{\mathfrak{D}})$ into a probabilistic $\mathcal{G}_{\text{prob}}$ by defining a probability measure over $\mathfrak{D}$ by the condition that the d_ν are independent with the same distribution $P(d_\nu = g) = p_g$. Note that we have followed closely the guidelines of section 16 in the construction of the present model.

Things are especially simple here since $\mathfrak{D} = G \times G \times \cdots \times G$ (n times) and G is commutative so that $\mathcal{G}_{\text{def}}$ is covariant and, *a fortiori*, $\mathcal{G}_{\text{prob}}$ is covariant in probability.

If deformations are less likely when $|d_\nu|$ is large, d_ν, ± 1, ± 2, $\cdots$, the following distribution (and variants of it) may be of interest:

(17.3) $$P_g = \text{const.}\ c^{|g|},\ g = 0, \pm 1, \pm 2, \cdots$$

where the positive constant c is less than one. We shall return to it below. Of more general interest is the following construction that will be used repeatedly. Let Π be a distribution over G with the corresponding probabilities Π_g. Let $P_t(g) > 0$ be the measures of a time-homogeneous stochastic process with independent increments over G starting at $g = 0$ for $t = 0$ and with Π as its jump distribution. This means that in a time interval $(t, t + \epsilon)$ with probability $1 - \epsilon$ the process stays in the same state g, but with probability $\epsilon \Pi_h + o(\epsilon)$ it jumps to state $g + h$; $h \in G$. Then we have, using a familiar argument,

(17.4) $$\frac{dP_t(g)}{dt} = -P_t(g) + \sum_{h \in g} P_t(h) \cdot \Pi_{g-h}.$$

For given Π this system of n first-order differential equations with constant coefficients can be solved by the usual routines. The solution takes an especially simple form expressed *via* the Fourier transforms

(17.5) $$\begin{cases} \hat{P}_t(\gamma) = \sum_{g \in G} e^{(2\pi i g \gamma/n)} P_t(g) \\ \hat{\Pi}(\gamma) = \sum_{g \in G} e^{(2\pi i g \gamma/n)} \Pi(g) \end{cases}$$

where the system reduces to

(17.6) $$\frac{d\hat{P}_t(\gamma)}{dt} = -\hat{P}_t(\gamma) + \hat{P}_t(\gamma) \cdot \hat{\Pi}(\gamma), \qquad \gamma \in G.$$

This gives us

(17.7) $$\hat{P}_t(\gamma) = e^{-t(1-\pi(\gamma))}, \qquad \gamma \in G.$$

and by Fourier inversion

(17.8) $$\hat{P}_t(g) = \frac{1}{n} \sum_{\gamma \in G} e^{(-2\pi i g \gamma/n)} - t[1 - \hat{\Pi}(\gamma)]$$

Consider in particular the case when Π is uniform over G, $\Pi_g \equiv 1/n$, so that $\hat{\Pi}(\gamma) = 1$ for $\gamma = 0$ and $\hat{\Pi}(\gamma) = 0$ otherwise. Then

(17.9) $$\begin{cases} P_t(0) = \dfrac{1}{n} + \dfrac{n-1}{1} e^{-t} \\ P_t(g) = \dfrac{1}{n}(1 - e^{-t}), \qquad g \neq 0 \end{cases}$$

In the set-up (17.4) the parameter t measures the *amount of deformation* in $\mathcal{G}_{\text{prob}}$.

Say that we want to recognize a deformed image $I^{\mathfrak{D}} = dI$ as originating from one of the patterns $\mathcal{P}_1, \mathcal{P}_2, \cdots \mathcal{P}_\rho$ from $\mathcal{P} = \mathfrak{J}/G$. Assume that the weight function w (see section 15) is given by

$$w = \begin{cases} 1 & \text{correct recognition} \\ 0 & \text{incorrect recognition} \end{cases} \tag{17.10}$$

What G-invariant recognition function $\phi(I)$, taking values in $\mathcal{P}$, maximizes the expected payoff

$$E\, w(\mathcal{P}^\alpha, \phi(I)) = E\, P(I \in \phi(I^{\mathfrak{D}}) \mid I^{\mathfrak{D}})? \tag{17.11}$$

We merely have to choose $\phi\ (I^{\mathfrak{D}})$ equal to that $\mathcal{P}^\alpha$; $\alpha = 1, 2, \cdots \rho$, for which

$$P(\mathfrak{D}_\alpha \mid I^{\mathfrak{D}}) = \max_\alpha . \tag{17.12}$$

Note that the conditional probability in (17.12) depends only upon what similarity class I belongs to, so that the resulting recognition function will be G-invariant, as desired. If the patterns $\mathcal{P}^\alpha$ all have the same probability $1/\rho$ we can replace (17.12) by

$$P(I^{\mathfrak{D}} \text{ in observed similarity class} \mid I \in \mathcal{P}^\alpha) = \max_\alpha . \tag{17.13}$$

The reader will recognize this as the standard Bayes solution of decision theory and (17.13) as the maximum likelihood solution. But what becomes of this solution in the special case of cyclic permutation patterns?

Having observed $I^{\mathfrak{D}} = (y_1, y_2, \cdots y_n)$, where the y's take as values some of the numbers $1, 2, \cdots n$, we first have to solve for the d_ν and g for which there is an α such that

$$x^{(\alpha)}_{\nu+d_\nu+g} = y_\nu, \qquad \nu = 1, 2, \cdots n, \tag{17.14}$$

where $I_\alpha = (x_1^{(\alpha)}, x_2^{(\alpha)}, \cdots x_n^{(\alpha)})$ is a prototype of $\mathcal{P}_\alpha$. Compute the value of (17.13):

$$\lambda(\alpha^*) = \max_{d^*, g^*} L(d^*, g^*, \alpha^*) \tag{17.15}$$

where d^* and g^* satisfy (17.14) for $\alpha = \alpha^*$. The optimum recognition is obtained by choosing α^* so that $\lambda(\alpha^*)$ is maximized, and we need only compute

$$\lambda(\alpha^*) = \max_{d^*, g^*} \prod_{\nu=1}^{n} P_{d_\nu^* + g^*} \tag{17.16}$$

If (17.3) is our specification of the deformation mechanism, we should look for the α^*, d^*, g^* that solve

$$\sum_{\nu=1}^{n} d_\nu^* + g^* = \min. \tag{17.17}$$

On the other hand, if (17.9) is our specification of P, our problem can be reduced to solving for

$$\frac{P_\iota(0)^{N_0(d^*, g^*, \alpha^*)}}{P_\iota(0)} = \max. \tag{17.18}$$

where

$$N_0(d^*, g^*, \alpha^*) = \text{number of } \nu\text{'s for which } d_\nu^* + g^* = 0 \tag{17.19}$$

But since $P_t(0) > P_t(g)$, this is equivalent to

$$N_0(d^*, g^*, \alpha^*) = \max. \tag{17.20}$$

Note that in these two cases we have actually arrived at *minimum distance recognition* (see section 14). Indeed, the expression in (17.17) is an 1_1-norm, while (17.20) corresponds to the non-definite (but definite in $\mathfrak{J}/G$) metric in R^n

$$\delta(x, y) = \min_{g \in G} [\text{number of } x_{\nu+g} \neq y_\nu]. \tag{17.21}$$

So far this is quite a special case of pattern analysis, but we can extend it further. Say that we have, abstractly, only access to n non-similar signs $s_1, s_2, \cdots s_n$, that we only consider configurations of the form $(g_1 s_1, g_2 s_2, \cdots g_n s_n)$ and that images are identified by their component signs. Then we can just as well write an arbitrary image as $I = (g_1, g_2, \cdots g_n)$ and we have the group patterns of section 10. We make this into a probabilistic deformation grammar by postulating

$$\begin{cases} d = (d_1, d_2, \cdots d_n),\ d_\nu \in G \\ dI = (d_1 g_1, d_2 g_2, \cdots d_n g_n) \\ \text{all } d_\nu \text{ independent with the same marginal probability } p_g . \end{cases} \tag{17.22}$$

For simplicity we have assumed that G is finite or denumerable; the modification in the continuous case is straightforward. Reasoning essentially as above we are led to the following recognition algorithm: choose α so that

$$\max_{\alpha, g} \prod_{\nu=1}^{n} P_{h_\nu g_{\alpha}^{-1}, g^{-1}} \tag{17.23}$$

is realized, where $I^{(\alpha)} = (g_{\alpha_1}, g_{\alpha_2}, \cdots g_{\alpha_n})$ is a prototype of the pattern $\mathcal{P}_\alpha$ and the deformed image is $I^{\mathfrak{D}} = (h_1, h_2, \cdots h_n)$. Assume that there exists an invariant distance δ on G such that

$$\begin{cases} P_g = \phi[\delta(g, e)] \\ \phi(\delta) \text{ is positive, decreasing and subexponential:} \\ \phi(0) \cdot \phi(\delta_1 + \delta_2) \leq \phi(\delta_1) \cdot \phi(\delta_2) \end{cases} \tag{17.24}$$

Introduce the functions ψ on $G \times G$

$$\psi(g, h) = -\log \frac{\phi(\delta(g, h))}{\phi(0)} \tag{17.25}$$

and the function ϕ on $\mathfrak{J} \times \mathfrak{J} = \mathfrak{J}^{\mathfrak{D}} \times \mathfrak{J}^{\mathfrak{D}}$

$$\phi(I_1, I_2) = \sum_{\nu=1}^{n} \psi(g_\nu, h_\nu), \qquad \begin{aligned} I_1 &= (g_1, g_2, \cdots g_n) \\ I_2 &= (h_1, h_2, \cdots h_n) \end{aligned} \tag{17.26}$$

It is clear that ψ and ϕ are non-negative and symmetric in their arguments and that $\psi(g, g) = \phi(I, I) = 0$. Further, if $\phi(I, I) = 0$ it follows from the decreasing character of ϕ that $\psi(g_\nu, h_\nu) = 0$, so that $\delta(g_\nu, h_\nu) = 0$, $g_\nu = h_\nu$, $\forall\ \nu$. Finally, for any $g, h, k \in G$ we have

$$\delta(g, k) \leq \delta(g, h) + \delta(h, g) \tag{17.27}$$

so that

$$(17.28) \qquad -\log \frac{\phi(\delta(g, k))}{\phi(0)} \leq -\log \frac{\phi(\delta(g, h) + \delta(h, g))}{\phi(0)}$$

and using the subexponential property

$$(17.29) \qquad -\log \frac{\phi(\delta(g, h))}{\phi(0)} \leq -\log \frac{\phi(\delta(g, h))}{\phi(0)} - \log \frac{\phi(\delta(g, h))}{\phi(0)}$$

so that ϕ and ψ satisfy the triangle inequality. Hence ϕ is an (invariant) distance. But choosing α and g so that (17.23) is maximized is the same as choosing them so that

$$(17.30) \qquad \phi(I^{\mathfrak{D}}, gI^{\alpha}) = \min.$$

In other words, *we have again arrived at minimum distance recognition* (see section 14).

Turning to the permutation grammar described at the end of section 7, we shall introduce the deformations in an essentially different way from the one we have just studied. Let $I = \{x_{ij}; i = 1, 2, \cdots n, j = i + 1, i + 2, \cdots n\}$ we shall let the deformed image $I^{\mathfrak{D}}$ be of the form $I^{\mathfrak{D}} = dI = \{y_{ij}\}$ with

$$(17.31) \qquad y_{ij} = \begin{cases} x_{ij} \text{ with probability } p \\ 1 - x_{ij} \text{ with probability } 1 - p \end{cases}$$

for $i < j$. We need not consider $i \geq j$ since the y-matrix will be assumed symmetric just as the x-matrix. All the $n(n-1)/2$ transitions $x_{ij} \to y_{ij}$, $i < j$, should be stochastically independent. Actually we have merely followed the guidelines of section 16 in this construction. To see this, let us write $d = \{d_{ij}; i = 1, 2, \cdots n, j = i + 1, i + 2, \cdots n\}$ with $d_{ij} = 0$ or 1 and with $y_{ij} = x_{ij} + d_{ij}$, $i < j$, where addition is modulo 2. Hence the deformation d has been separated into independent elementary parts d_{ij} with the same distribution; actually the deformations are additive. This probabilistic deformation grammar is not covariant but covariant in probability. Having observed the deformed image $I^{\mathfrak{D}} = dI = \{y_{ij}\}$ let us find the maximum likelihood recognition of the pure image I. This is easy, since the expression to maximize is

$$(17.32) \qquad P(I^{\mathfrak{D}} \mid I) = P_d = p^{\delta(d)}(1 - p)^{[n(n-1)/2] - \delta(d)} = \text{constant} \cdot a^{\delta(d)};$$
$$a = \left(\frac{p}{1-p}\right);$$

where $\delta(d)$ = number of $d_{ij} \neq 0$, $i < j$. Hence $(np < \frac{1}{2})$ we should recognize $I^{\mathfrak{D}}$ as generated by that $I \in \mathfrak{I}$ for which the distance $\delta(I^{\mathfrak{D}}, I)$ is as small as possible: *in this special case we also arrive at minimum distance recognition.*

On the other hand, if we start with certain pure patterns $\mathcal{P}_1, \mathcal{P}_2, \cdots \rho \in \mathfrak{I}/G$ with the same probability $Q(\mathcal{P}_r) = 1/\rho$ and ask for the best in the sense of (17.13), we should calculate the function f defined on $\mathfrak{I} \times \mathfrak{I}$

$$(17.33) \qquad f(I_1, I_2) = \sum_{dI_1 \in GI_2} a^{\delta(d)}$$

Since f is completely invariant it is actually well defined on $\mathcal{P}^{\mathfrak{D}} \times \mathcal{P}^{\mathfrak{D}}$. The best recognition is achieved by choosing that $\mathcal{P}_r$ for which

$$(17.34) \qquad f(\mathcal{P}_r, \mathcal{P}) = \max, I^{\mathfrak{D}} \in \mathcal{P}$$

is realized.

18. Deformation of Plane Patterns. When we turn from the simple patterns of the last section to patterns in the plane we can expect to meet a larger variety of deformation mechanisms. For reasons discussed in sections 8 and 9 we shall concentrate our attention on line patterns. With the help of ideas put forward in section 16 we shall specify certain models that promise to be of use for pattern analysis.

Let us start from a pure image algebra $\mathfrak{J}$ consisting of curve segments combined according to given grammatical rules. Write the pure images as $I = I(x, y) =$ indicator function of the curve segments making up I. The simplest deformation mechanism that suggests itself immediately is the following. Let $\mathfrak{D}$ be a set of mappings d of the plane into itself and define the deformed image $I^{\mathfrak{D}}$ by the relation

$$I^{\mathfrak{D}}_{(x,y)} = dI(xy) = I(d(x, y)); \tag{18.1}$$

on $\mathfrak{D}$ there should be a probability measure P. Note that in such a probabilistic deformation grammar the deformations are introduced through *distortions of the whole plane*, the background or carrier of the images, rather than through operations on the signs and configurations as such. Since this is a special case of the important class of contrast deformations it will not be discussed here but postponed to section 19.

Instead we shall study *deformation via the separate sign*, starting from the model of pure Euclidean line patterns (see section 8), and where the image consists of a single sign. Say that the pure patterns are defined through

$$\frac{d\phi}{ds} = f_r(s), \qquad r = 1, 2, \cdots \rho; \quad s \in [0, 1] \tag{18.2}$$

where the functions $f_r(s)$ should be analytic. Since G now consists of translations and rotations, a complete specification requires that we fix the triple $b = (x_0, y_0, \phi_0)$ where (x_0, y_0) are the coordinates of the starting point, $s = 0$, of the curve and ϕ_0 is its inclination at this point. (To be precise we should separate two cases—one where the curve segment is directed, as assumed here, and another where it has no given direction.)

In Euclidean line patterns we treat the image as a geometric object and we shall do the same for the deformed images, writing

$$\frac{d\phi^{\mathfrak{D}}}{ds} = f_r(s) + \epsilon(s), \qquad r = 1, 2, \cdots \rho; \quad s \in [0, 1] \tag{18.3}$$

where the ϵ-term is a stochastic process continuous in probability. The only essential restriction is that we have chosen this additive form for the right-hand side; the fact that we shall let the probability distribution of the ϵ-process be independent of r, the pure pattern $\mathcal{P}_r$, can easily be modified. To be concrete, let $\epsilon(s)$ be a stationary Gaussian process with mean zero and covariance function $\alpha e^{-\beta' s}$, $s > 0$. The triple b is not affected by d. It is clear that this model is covariant. How do we achieve maximum likelihood recognition?

In the function space of $\Psi = d\phi^{\mathfrak{D}}/ds$, $s \in [0, 1]$, the ρ different patterns give rise to ρ probability measures, say $m_1, m_2, \cdots m_\rho$. Let us express them all in terms of Radon-Nikodym derivatives $L_1 \equiv 1, L_2, L_3, \cdots L_\rho$ taken with respect to m_1

$$m_r(E) = \int_E L_r(\Psi) P_1(d\Psi) \tag{18.4}$$

But we can use a well-established technique for solving such a problem (see Grenander

[1]). Consider the likelihood ratio 1_r for the values of

$$\Psi_\nu = \left(\frac{d\phi^{\mathfrak{D}}}{ds}\right)_{s=(\nu/n)}, \qquad \nu = 1, 2, \cdots n. \tag{18.5}$$

We get, using the Markovian property of the ϵ-process

$$1_r = \exp -\tfrac{1}{2}(Q_r - Q_1) \tag{18.6}$$

with

$$\begin{aligned} Q_1 - Q_r = {} & \tfrac{1}{2}[2f_r(0)\Psi(0) - 2f_1(0)\Psi(0) + f_r^2(0) - f_1^2(0)] \\ & + \frac{2}{\alpha(1 - e^{-2\beta/n}} \sum_{\nu=0}^{n-1} \left[\Psi\left(\frac{v+1}{n}\right) - e^{-\beta/n}\Psi\left(\frac{v}{n}\right)\right]\left[f_r\left(\frac{v+1}{n}\right) - e^{-\beta/n}f_r\left(\frac{v}{n}\right) \right. \\ & \left. - f_1\left(\frac{v+1}{n}\right) + e^{-\beta/n}f_1\left(\frac{v}{n}\right)\right] \\ & + \frac{1}{\alpha(1 - e^{-2\beta/n})} \sum_{\nu=0}^{n-1} \left\{\left[f_1\left(\frac{v+1}{n}\right) - e^{-\beta/n}f_1\left(\frac{v}{n}\right)\right]^2 \right. \\ & \left. - \left[f_r\left(\frac{v+1}{n}\right) - e^{-\beta/n}f_r\left(\frac{v}{n}\right)\right]^2\right\} \end{aligned} \tag{18.7}$$

It is known that when $n \to \infty$ the limit of 1_r exists and is L_r. After some elementary manipulations this limit can be expressed in the form $A_r - A_1$ with

$$\begin{aligned} A_r = {} & \frac{2}{\alpha} f_r(0)\Psi(0) + \tfrac{1}{2}f_r^2(0) \\ & - \frac{1}{\alpha\beta}[f_r^1(0) + \beta_r^f(0)]\Psi(0) + \frac{1}{\alpha\beta}[f_r^1[1] + \beta_r^1(1)]\Psi(1) \\ & \div \frac{1}{\alpha\beta}\int_0^1 [f_r''(s) + \beta f_r'(s)2\Psi(s)\, ds \\ & - \frac{1}{\alpha\beta}\int_0^1 [\beta f_r'(s) + \beta^2 f_r(s)]\Psi(s)\, ds \\ = {} & \frac{\Psi(0)}{\alpha}\left[f_r(0) - \frac{1}{\beta}f_r^1(0)\right] + \frac{\Psi(1)}{\alpha}\left[f_r^1(1) + \frac{1}{\beta}f_r^1(1)\right] \\ & + \frac{1}{\alpha\beta}\int_0^1 [\beta^2 f_r(s) - f_r''(s)]\Psi(s)\, ds \\ & - \frac{1}{2\alpha\beta}\int_0^1 [f_r^1(s) + \beta f_r(s)]^2\, ds + \tfrac{1}{2}f_r^2(0) \end{aligned} \tag{18.8}$$

This expression is not as complicated as it may appear at first glance. It is a linear function of the three quantities

$$\begin{cases} a_r^{(0)} = \Psi(0) \\ a_r^{(1)} = \Psi(1) \\ a_r = \displaystyle\int_0^1 [\beta^2 f_r(s) - f_r''(s)]\Psi(s)\, ds \end{cases} \tag{18.9}$$

These three probes are sufficient for the recognition procedure. The first two of them are local measurements at the ends of the curve segment constituting the sign, while the third one is a global quantity. If the deformations are irregular along the sign, large value of β, then the form of the expression (18.9) indicates that the third quantity will be the most informative for recognition purposes. At any rate, to maximize L_r we merely have to compute A_r for the observed values of Ψ and choose that pattern $\mathcal{P}_r$ for which A_r takes the largest value.

This was when the image had just one sign (in the present context we deal with identification *via* the maximal connected component (see sections 4 and 8)), but the model generalizes easily. Let the pure image I consist of n signs (curved segments) $s_1, s_2, \cdots s_n$, each of which is a maximal connected component. Write the νth sign in its intrinsic form

$$\frac{d\phi_\nu}{ds} = f^{(\nu)}(s), \qquad \nu = 1, 2, \cdots n, \quad s \in [0, 1]. \tag{18.10}$$

Here $f^{(\nu)}$ is equal to some f_r. We must also specify the initial condition $b_\nu = (x_0^{(\nu)}, y_0^{(\nu)}, \phi_0^{(\nu)})$ for each ν. So much for the pure image. The deformed image is obtained by

$$\begin{cases} \dfrac{d\phi_\nu^{\mathfrak{D}}}{ds} = f_\nu(s) + \epsilon_\nu(s) \\ b_\nu^{\mathfrak{D}} = b_\nu + \eta_\nu \end{cases} \tag{18.11}$$

where $\epsilon_\nu(s)$ is a stochastic process as before and η_ν a stochastic 3-vector. It may be instructive to examine two extreme cases.

In the first one we assume that *all* ϵ_ν and η_ν *are independent*. In the same way as before we calculate the Radon-Nikodym derivative L_r which is now the product of n factors, each factor being associated with one sign. Each such factor in its turn splits into two factors, one of which belongs to the equation for the curvature Ψ_ν and the other to the initial condition $b^{\mathfrak{D}}$. Finally we will maximize L_r to recognize a pure pattern $\mathcal{P}_r$.

In the extreme opposite case we assume that we are given *one* process $\epsilon(s)$ and *one* stochastic 3-vector η such that $\epsilon_\nu(s) = \epsilon(s)$, $\eta_\nu = \eta$, $\forall\ \nu$. Again a similar recognition procedure can be carried out, but the result will, of course, differ from the previous one. A variety of cases intermediate between these two can be written down.

As the reader may have noticed, these constructions have followed closely the general prescription outlined in section 16.

We can now go ahead to study *deformations of differential line patterns* (see section 8). Say that the pure signs are generated by first-order differential equations of the form

$$y' = f_\sigma(x, y), \tag{18.12}$$

where σ is as usual the subroutine parameter. This equation should of course be invariant with respect to the similarity transformations. Together with the values of the end-points of the arc (18.12) determines the pure sign. Let the deformed sign be given through the stochastic differential equation

$$y' = f_\sigma(x, y) + \epsilon(x, y) \tag{18.13}$$

where ϵ is a stochastic process such that the probability measure of the differential line element field of (18.13) is G-invariant. Then the deformations in $\mathfrak{D}$ are covariant in probability and we can go ahead in the same spirit as before looking for the best recognition function.

As an example we consider the case mentioned at the end of section 8. Say that $f_\sigma(x, y) = \sigma(y/x)$: any given subroutine σ has associated with it a smooth function $\sigma(u)$ of the real argument $u = y/x$. Let $\epsilon(x, y) = W(y/x)$ where W is the Wiener process. After some elementary calculations we get the Radon-Nikodym derivatives expressed through the five probes

$$\text{(18.14)} \qquad \begin{cases} a_1 = u_0 \\ a_2 = u_1 \\ a_3 = \left(\dfrac{dy}{dx}\right)_{u_0} \\ a_4 = \left(\dfrac{dy}{dx}\right)_{u_1} \\ a_5 = \displaystyle\int_{u_0}^{u_1} \sigma''(u)y'(u)\, du \end{cases}$$

where u_0 and u_1 are the values of y/x at the end points of the observed arc. In passing, let us point out that we can sometimes, just as here, find the optimal recognition function without solving the differential equation explicitly.

Actually the same idea applies to the more general notion of *boundary patterns* (see section 10). Starting with a certain pure sign $s = [\sigma; b_1, b_2, \cdots b_\rho]$ we would separate the deformation into factors. One of these factors would transform the subroutine σ into another subroutine σ^d from some space, usually much wider than Σ, and assign probabilities to the possible values of σ^d. The other factors would transform the boundary conditions $b_1, b_2, \cdots b_\rho$ according to some probability distributions. One would try to express the pure image algebra in such a way that the random quantities of these various functions would be independent, as suggested in section 17. If the reader looks back at the models described in the present section it will be clear that this is just what we have tried to do here.

As far as *set patterns* are concerned, it need scarcely be pointed out that if the information is carried by the boundary we can go ahead just as above. Instead we shall now consider a different sort of set pattern. Let the pure image consist of *convex polygons* in the plane. An image I can then be expressed through n signs $s_1, s_2, \cdots s_n$ consisting of half planes: $I = s_1 \cap s_2 \cap \cdots \cap s_n$ (see the end of section 9). The similarity transformations consist of translations and rotations. The deformation of I shall be a realization of a Poisson process with intensity λ_1 within I and intensity λ_2 outside I, say with $\lambda_1 < \lambda_2$. Then $I^{\mathfrak{D}} = dI$ cannot be described through a boundary; it consists just of isolated points. This model is one of those fairly rare cases in which $\mathfrak{J} \subset \mathfrak{J}^{\mathfrak{D}}$ does not hold.

Say that we can observe $I^{\mathfrak{D}}$ in some subset of the plane, e.g. a unit square, and that we know the inclinations of the straight lines bordering the half spaces. How do we find a good recognition function? This will depend on how restricted the signs are.

First, assume that for any particular pattern $\mathcal{P}_r$ the shape of the pure image is completely known (but, of course, not its location and orientation). The Radon-Nikodym derivative will be found by a technique similar to that employed earlier in this section

$$L_r = e^{-\lambda_1 m(gI_r) - \lambda_2 [1 - m(gI_r)]} \cdot \lambda_1^{\nu(gI_r)} \cdot \lambda_2^{n - \nu(gI_r)} \tag{18.15}$$

where I_r is a prototype of the rth pattern; g is the similarity transformation carrying this prototype to the actual pure image. By $m(I)$ we mean the area of I and $\nu(I)$ denotes the number of Poisson events observed inside I. Maximum likelihood recognition can then be based on the probe

$$a_r = \min_g \left\{ (\lambda_1 - \lambda_2) m(gI_r) + \nu(gI_r) \cdot \log \frac{\lambda_2}{\lambda_1} \right\}; \qquad r = 1, 2, \cdots \rho. \tag{18.16}$$

After computing these ρ probes, that pattern $\mathcal{P}_r$ is recognized for which a_r takes the smallest value.

Second, if the shape of the pure image is not specified in advance, one might try the same approach. Since this will lead to an unreasonably large amount of computation, it may be better to try to determine each sign separately. Consider the halfplanes associated with a certain direction and project the deformed image onto a coordinate axis orthogonal to this direction. On this axis we denote the projections of the Poisson events by $x_1, x_2, \cdots x_n$. Say that the x-variable of the unit square extends from $x = a$ to $x = b$ and that the projection of the pure image I extends from $x = \alpha$ to $x = \beta$, $a \leq \alpha < \beta \leq b$, where α and β are of course unknown (see Figure 3). If $\mu(x)\,dx$ denotes the length element of the unit square projected onto the x-axis we have the likelihood function (except for a constant)

$$\underset{\alpha\beta}{L}(\lambda) = \lambda_1^{n - \nu(\alpha,\beta)} \exp\left\{ -\lambda_1 \left[\int_a^\alpha + \int_\beta^b \mu(x)\,dx \right] \right\} \cdot \lambda^{\nu(\alpha,\beta)} e^{-\lambda} \tag{18.17}$$

where $\nu(\alpha, \beta)$ is the number of Poisson events with $\alpha < x_\nu < \beta$. We should first maximize (18.17) as a function of λ under the condition that

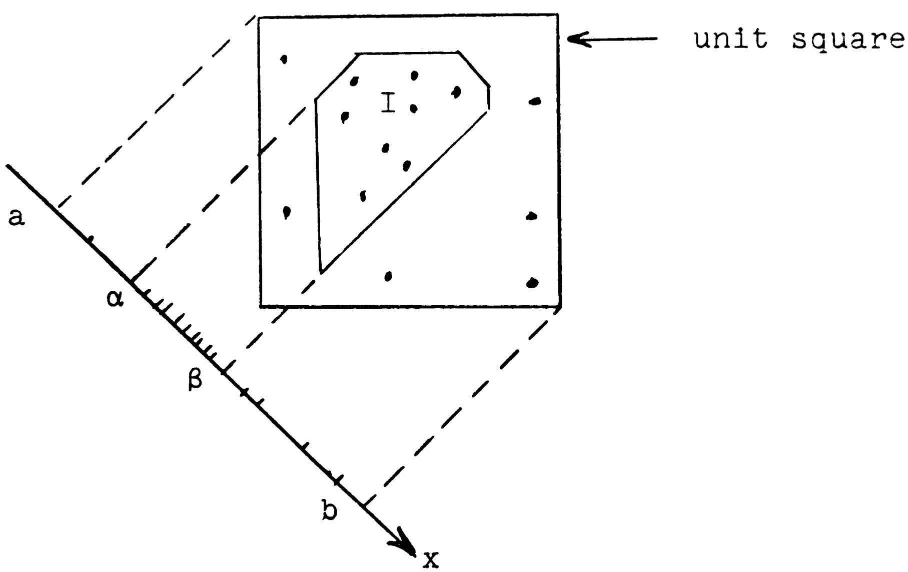


FIGURE 3

$$\lambda > \lambda_1 \int_\alpha^\beta \mu(x)\, dx \tag{18.18}$$

leading to the value of

$$\lambda^*(\alpha, \beta) = \max \left[\nu, \lambda_1 \cdot \int_\alpha^\beta \mu(x)\, dx \right] \tag{18.19}$$

We should choose the (α, β) with $a \leq \alpha < \beta \leq b$ for which

$$\lambda_1 \int_\alpha^\beta \mu(x)\, dx + \nu(\alpha, \beta) \log \frac{\lambda^*(\alpha, \beta)}{\lambda_1} - \lambda^*(\alpha, \beta) \tag{18.20}$$

takes its largest possible value. The distribution properties of this recognition function are not known and should be studied.

19. Deformation of Contrast Patterns. It was pointed out in section 11 that the idea of contrast patterns—a quite general notion—is a natural one only if the background and contrast spaces are given to us as familiar mathematical objects. We can then construct models for the deformed patterns, and we shall separate two cases. The main case is that in which *only the background space* (but not the contrast) is affected by the deformation; the other one, that in which *the contrast space is deformed.* Both these cases are subdivided according to whether the *whole image* is deformed at once or if its *signs* are deformed independently.

Model for background deformation. Consider a contrast grammar $\mathcal{G} = (X, C, S, G)$ *and let* Ξ *be a space of mappings* ξ *of* X *into itself. Introduce a probability measure* P *on* Ξ; *this carries* $\mathcal{G}$ *into a probabilistic deformation grammar by introducing the deformations as*

$$d : I = I(x) \to I^{\mathfrak{D}} = I[\xi(x)]. \tag{19.1}$$

In general we do not assume that the mappings ξ of the background map X onto the entire X. If not, then a part of a pure image $I = I(x)$ will be lost, *viz.* that part of $I(x)$ corresponding to the values of X that are not assumed by $\xi = \xi(x)$. Denote by $W(I)$ the range of the function $I = I(x)$, the "Wertevorrat", and similarly $W(I^{\mathfrak{D}})$ as the range of the deformed image $I^{\mathfrak{D}}(x)$, $dI = I[\xi(x)]$. It is clear that $W(I^{\mathfrak{D}}) \subset W(I) \subset C$.

Say, to fix ideas, that $W(I)$ is denumerable and consists of the values (in the contrast space C) $W_1, W_2, W_3, \cdots$, same for all $I \in \mathfrak{J}$ (more generally, we would decompose C into some σ-algebra of subsets). Let $X_\nu(I^{\mathfrak{D}}) = \{x \mid \xi(x) \in I^{-1}(W_\nu)$ where the inverse function I^{-1} may be many-valued. The likelihood function can then be written as

$$L(I) = P\{\xi(X_\nu(I^{\mathfrak{D}}) \subset I^{-1}(W_\nu),\ \forall\ \nu\} \tag{19.3}$$

if this is different from zero, or, more generally, as the corresponding Radon-Nikodym derivative.

Likelihood recognition would be attained by choosing that pure image for which $L(I)$ is as large as possible. By $\xi(A)$ is understood, as usual, the set of points obtained by letting ξ map each point of A separately into X. If ξ^{-1} is one-valued (19.3) can be simplified and the inclusion symbol be replaced by equality.

As outlined in section 16, it may be possible to describe these deformation grammars through dynamic models. To illustrate this, consider the following simple deformation mechanism. Start from a pure contrast grammar with $X = [0, 1]$ and $C = R^1$ with similarity transformations of the form $gx = x^g$, $0 \leq x \leq 1$, $g > 0$, and we use the symbol

g to denote both the group element and the positive number to which it corresponds. In other words, we have

$$\frac{1}{gx}\frac{d}{dx}gx = \frac{g}{x}, \qquad 0 \le x \le 1, \quad (gx)_{x=0} = 1. \tag{19.4}$$

We shall consider pure images of the form $I(x) = x^\alpha$, $0 \le x \le 1$, $\alpha > 0$. Introduce the deformation *via* the mappings $\xi = \xi(x)$

$$\frac{1}{\xi(x)}\frac{d}{dx}\xi(x) = \frac{\epsilon(x)}{x}, \qquad \xi(0) = 0 \tag{19.5}$$

where $\epsilon(x)$ is defined through the following probability measure. The interval $[0, 1]$ is divided by the events of a Poisson process with intensity μ into the intervals $i_0 = [0, x_1]$, $i = [x_1, x_2]$, $\cdots$ $i_{n-1} = [x_{n-1}, x_n]$, $i_n = [x_n, 1]$. In i_ν the ϵ-process takes the value y_ν, where $y_0, y_1, y_2 \cdots$ are independent positive stochastic variables from the exponential frequency function e^{-y}, $y > 0$. Say that the deformed image is $I^{\mathfrak{D}}(x) = f(x) = I[\xi(x)]$. If we denote the background coordinates for which the first derivative $f'(x)$ has a jump by $y_1 < y_2 < \cdots < y_{n-1}\, y_n$ and by z_ν the (constant) value

$$z_\nu = \left(\frac{d \log f(x)}{d \log x}\right)_{y_\nu < x < y_{\nu+1}}, \qquad \nu = 0, 1, 2, \cdots n \tag{19.6}$$

we can solve (19.2) and get, assuming that $I(x) = x_\alpha$, that the value of $\epsilon(x)$ between the νth and $(\nu + 1)$th discontinuity is $(1/\alpha)\, z_\nu$. Hence (19.3) gives us, after some reductions and leaving out a factor that does not depend on I

$$L(I) = \frac{1}{n+1} \exp -\left(\frac{1}{\alpha}\sum_{\nu=0}^{n} z_\nu\right) \tag{19.7}$$

Maximum likelihood recognition leads us to the pure image $I(x) = x^{\alpha*}$ for which

$$\alpha^* = \frac{1}{\alpha^{n+1}}\sum_{\nu=0}^{n} z_\nu = \frac{1}{\alpha^{n+1}}\sum_{\nu=0}^{n+1}\left(\frac{d \log f(x)}{d \log x}\right)_{y_\nu < x < y_{\nu+1}} \tag{19.8}$$

This amusing form of the recognition function leads to an image within the pattern we started from. Different patterns can be separated with complete certainty (but not images within a pattern) within this deformation grammar. This is an expression of the fact that these deformations are really too mild to be of practical interest; however, the discussion should illustrate the approach.

Instead let now the deformation $I(x) \to I[\xi(x)]$ be followed by a *mask transformation*: the deformed image will be given as $I^{\mathfrak{D}}(x) = I[\xi(x)]$, $x \in X_1$, where X_1 is a subset of X: we have no knowledge about the values of $I^{\mathfrak{D}}(x)$ outside the set X_1. The subset X_1 is selected according to some probability law P_1 and we assume that ξ and X_1 are stochastically independent and that P_1 does not depend on I. It is then intuitively clear, and it could be proved formally, that the best recognition procedure can be found by considering X_1 as fixed and equal to the observed set. On the other hand, if we relax the independence assumptions above, it can very well happen that the form and location of X_1 can help us in finding a better recognition scheme than if we had considered it as fixed.

For an image algebra with uniquely identifiable signs the model for background transformations can be modified. Say that the pure image I consists of the n signs $s_1 = s_1(x)$, $s_2 = s_2(x)$, $\cdots$ $s_n = s_n(x)$. Take *n independent versions $\xi_1, \xi_2, \cdots \xi_n$ of the mapping*

ξ corresponding to the product measure on Ξn, and define the deformed image through the configuration $(s_1[\xi_1, (x)], s_2[\xi_2(x)], \cdots s_n[\xi_n(x)]$.

The illustrative example from above would then be changed as follows. Let $C = R^n$ and define the pure pattern $\mathcal{P}_r$ as consisting of images of the form

$$I_r(x) = (I_{r_1}(x), I_{r_2}(x), \cdots I_{r_n}(x)) = (x^{g\alpha r_1}, x^{g\alpha r_2}, \cdots x^{g\alpha r_n}) \tag{19.9}$$

where the αs are given positive numbers and g is an arbitrary positive number. Deforming each component independently of the rest, we get a set of numbers derived from the deformed image

$$z_{k_\nu} = \left(\frac{d \log f_k(x)}{d \log x}\right) y_{k_\nu} < x < y_{k_{\nu+1}}, \quad k = 1, 2, \cdots n; \quad \nu = 0, 1, \cdots n_k, \tag{19.10}$$

where $I^{\mathfrak{D}}(x) = (f_1(x), f_2(x), \cdots f_n(x))$ and the y_{k_ν}'s are the points of discontinuity of $f_k'(x)$. The likelihood function is then essentially

$$L(I_r) = \frac{1}{g^n} \prod_{k=1}^{n} \left\{\frac{1}{\alpha_{r_k}} \exp\left(-\frac{1}{g\alpha_{r_k}} \sum_{\nu=1}^{n_k} z_{k_\nu}\right)\right\} \tag{19.11}$$

To maximize this in g for given r we find

$$g_r^* = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{\alpha_{r_k}} \sum_{\nu=1}^{n_k} z_k \tag{19.12}$$

and that pure pattern $\mathcal{P}_r$ should be recognized for which (19.11) takes the largest value when g_r^* is substituted for g.

Model for contrast deformation. Consider a constrast grammar $\mathcal{G} = (X, C, S, G)$ *and let* Z *be a set of mappings* z *of* C *into* C. *Introduce a probability measure* P *on* Z; *this carries* $\mathcal{G}$ *into a probabilistic deformation grammar by introducing the deformations as*

$$d : I = I(x) \rightarrow I^{\mathfrak{D}} = z[I(x)]. \tag{19.13}$$

But this is usually too restricted a model. Instead we can let the mapping of C into C depend upon x, z_x, and the measure P should then be defined on the product set C^X.

A simple but typical case of contrast deformations is obtained by choosing $X = \{1, 2, \cdots n\}$, $C = R^1$, $G =$ group of cyclic permutations of X. The signs could be generated as in section 7. Let z_x consist just of adding a real number, which we also denote by z_x, to the values of C and say that P is defined through the condition that z_x forms a Gaussian process with mean zero and positive definite covariance matrix R. This is a model of additive noise, well known from communication theory, and it is clear that maximum likelihood recognition reduces to minimum distance recognition with the metric in $\mathfrak{I}$ given as a quadratic form for the inner product

$$(I_1, I_2) = \sqrt{I_1' R^{-1} I_2} \tag{19.14}$$

If z_x is a circular process it follows that R is a cylic matrix and that the metric of (19.14) is G-invariant (see section 14). The distance (squared) between a pure image $gI_r \in \mathcal{P}_r$ and the deformed image $I^{\mathfrak{D}}$ can then be written in terms of norms and inner products

$$\begin{aligned} \delta^2(gI_r, I^{\mathfrak{D}}) &= \|gI_r\|^2 - 2(gI_r, I^{\mathfrak{D}}) + \|I^{\mathfrak{D}}\|^2 \\ &= \|I_r\|^2 - 2(gI_r, I^{\mathfrak{D}}) + \|I^{\mathfrak{D}}\|^2 \end{aligned} \tag{19.15}$$

so that if all the prototypes have the same norms recognition consists in choosing that pattern $\mathcal{P}_r$ for which the criterion

$$\max_{g \in G} (gI_r , I^{\mathfrak{D}}) \tag{19.16}$$

takes its largest value: *recognition through maximum inner product.* But if the contrast deformations z are given a bit more generally as

$$z\{I(x)\} = a\, I(x) + z_x \tag{19.17}$$

where z_x has the same properties as before and a is a real constant different from zero, then we would arrive at the following criterion:

$$\max_{g \in G} |\text{corr. coeff. between } I^{\mathfrak{D}}(x) \text{ and } gI_r(x)| \;; \tag{19.18}$$

that pattern $\mathcal{P}_r$ should be chosen for which (19.18) is as large as possible. The correlation coefficient should be interpreted in terms of the inner product associated with the matrix R^{-1}. If the norm of the pure prototypes does not vary with r we arrive again at recognition by maximum inner product; otherwise we could speak of *recognition by maximum correlation.*

Since this is quite simple we need not discuss it any further, just pointing out that the model can be extended directly to handle deformations of separate signs.

Here is one of the simplest possible deformation grammars with contrast deformations. Let $\mathfrak{I}$ consist of n-vectors whose components are zero or one. Composition of images means adding vectors componentwise modulo 2. The similarity transformations permute the components cyclically and we shall start from ρ patterns $\mathcal{P}_1 , \mathcal{P}_2 , \cdots \mathcal{P}_\rho$ in $\mathfrak{I}/G$. A deformation d is split into elementary parts (see section 16) $d_1 , d_2 , \cdots d_n$ each of which has the probability distribution

$$\begin{cases} P(d_\nu = 1) = p \\ P(d_\nu = 0) = 1 - p \\ 0 < p < \frac{1}{2} \end{cases} \tag{19.19}$$

and all the d_ν's are independent. We put

$$I^{\mathfrak{D}} = I + d \tag{19.20}$$

so that we should look for that $\mathcal{P}_r$ for which the binomial probability

$$\binom{n}{k_r} p^{k_r}(1 - p)^{n-k_r} \tag{19.21}$$

is as large as possible, where

$$k_r = \min_{g} \sum_{=1}^{n} |I^{\mathfrak{D}} - I^r_{\nu+g}| \,. \tag{19.22}$$

The reader may recognize the sum in (19.22) as the *Hamming distance* and the set-up is that familiar from algebraic coding theory. The decision rule is: minimize Hamming distance.

Another deformation grammar that bears some resemblance to the contrast deformation model starts from a linear operator L. Although what follows can be extended a great deal, the reader may prefer to think of L as one of the classical self-adjoint, second-order differential operators on some set X, say the interval $(0, 1)$. Let L have the (simple) eigenvalues $\lambda_1, \lambda_2, \cdots$ associated with the complete orthonormal system $\phi_1(x), \phi_2(x), \cdots$ of eigenfunctions. As the device generating the signs of S we shall use the equation

$$Lf = f\lambda. \tag{19.23}$$

As the group G of similarity transformations we take the operations $f \to c \cdot f$: multiplication with an arbitrary real constant c. We could also admit certain transformations of X onto itself, but this will not be done here. Images are then (twice differentiable) real-valued functions on X, the signs are multiples of the ϕ_ν's, while the ϕ_ν's themselves can be taken as templates. Starting from an image $I(x)$ we introduce the function $LI(x) = \phi(x)$ and the deformed image through the stochastic differential equation

$$LI^{\mathfrak{D}}(x) = \phi(x) + \dot{w}(x), \tag{19.24}$$

where $w(x)$ is an orthogonal process with mean zero and homogeneous on X very much along the lines of section 16. But it is possible to represent $w(x)$ as

$$w(x) = \sum_{\nu=1}^{\infty} w_\nu \int_0^x \phi_\nu(x)\, dx \tag{19.25}$$

where the w_ν's are uncorrelated stochastic variables with mean zero and constant variance σ^2. Writing $I^{\mathfrak{D}}(x)$ in terms of its coefficient $I_\nu^{\mathfrak{D}}$ and $I(x)$ in terms of coefficients I_ν we get

$$I_\nu^{\mathfrak{D}} = I_\nu + z_\nu \tag{19.26}$$

where $z_\nu = w_\nu/\lambda_\nu$ are then uncorrelated and with variance σ^2/λ_ν^2. Note that the property of being uncorrelated has carried over from the time domain of x to the "frequency domain" of ν. To specialize even further, assume that the w_ν are independent with the double exponential frequency function $e^{-|w|}$. Then maximum likelihood consists in choosing that pattern $\mathcal{P}_r \in \mathfrak{I}/G$ for which

$$\min_c \sum_\nu |I_\nu - cI_\nu^r|\, \lambda_\nu \tag{19.27}$$

is made as small as possible. In our image algebra $\mathfrak{I}^{\mathfrak{D}}$ (remember that the images have a finite number of signs) the sum in (19.27) defines a distance, so that this G-invariant decision rule is a minimum distance recognition procedure.

It was shown in section 15 that if the deformations are covariant in probability we avoid certain difficulties and we can find G-invariant recognition procedures. There is nothing to guarantee that this is always so. To show this in a non-pathological instance it is enough to consider the case $X = N^1$, $C = R^1$, $G =$ translation group of N^1 and ρ prototypes $I_r(x)$, $r = 1, 2, \cdots \rho$, vanishing except for a finite number of x's. Let the deformations be given by

$$d : I(x) \to I^{\mathfrak{D}}(x) = I(x) + \epsilon(x) \tag{19.28}$$

where $\epsilon(x)$ is a Gaussian, stationary, ergodic process, say independent for different x's. Suppose there exists a G-invariant procedure for recognizing the ρ patterns. Then the sample space of $I^{\mathfrak{D}}(x)$, $x \in N^1$, will be split into ρ disjoint and G-invariant subsets $E_1, E_2, \cdots E_\rho$ corresponding to the ρ possible outcomes of the recognition algorithm. Because ϵ is ergodic the probability mass it attributes to a set E_ν must be zero or one.

But it can be shown that the probability measure of the $I^{\mathfrak{D}}$-process is absolutely continuous with respect to the ϵ-process, and *vice versa*. Hence all the E_ν have probability zero or one independent of what pure image $I(x)$ is true, and the recognition method is useless.

Here is something different. Let $\mathcal{G} = (X, C, S, G)$ still be a contrast grammar but do not introduce $\mathfrak{J}^{\mathfrak{D}}$ through transformations of background or contrast space. Instead, let $\zeta_1, \zeta_2, \cdots \zeta_n$ be a set of points in X and P their joint probability measure over X^n. The deformed image is defined through

$$I^{\mathfrak{D}} = (I(\zeta_1), I(\zeta_2), \cdots I(\zeta_n) \in C^n. \tag{19.29}$$

Note that this is another instance when $\mathfrak{J} \subset \mathfrak{J}^{\mathfrak{D}}$ is not true; another one was studied in section 18.

The idea behind this probabilistic deformation grammar is that we are not able to observe the image $I = I(x)$ for all values of x in X. Instead *we only observe* I *at* n *points* $\zeta_1, \zeta_2, \cdots \zeta_n$ the exact location of which is not given to us; only the probability distribution P is known. Perhaps we prefer to represent the deformed image instead by a C-valued function over the entire X obtained by some definite interpolation method form the values of the $I(\zeta_\nu)$, $\nu = 1, 2, \cdots n$. But this really changes nothing, since no information is created (nor, usually, destroyed) by interpolation.

Let us organize the recognition of the pure patterns $\mathcal{P}_1, \mathcal{P}_2, \cdots \mathcal{P}_p$ of $\mathfrak{J}/G$ corresponding to the prototypes $I_1(x), I_2(x), \cdots I_p(x)$. Consider the mapping $C^n \to X^n$

$$C^n \to (f_1, f_2, \cdots f_n) \to (\zeta_1, \zeta_2, \cdots \zeta_n) \in X^n \tag{19.30}$$

where $I_r(\zeta_1) = f_1, I_r(\zeta_2) = f_2, \cdots I_r(\zeta_n) = f_n$. In general this mapping is one-many. We have written $f_1, f_2, \cdots f_n$ for the coordinates of $I^{\mathfrak{D}}$. The probability measure P on X^n then induces a probability measure P_r (in general dependent upon I_r) on C^n which is in the range of $I^{\mathfrak{D}} : \mathfrak{J}^{\mathfrak{D}} = C^n$. Express the P_r's in terms of Radon-Nikodym derivatives $L_r(f_1, f_2, \cdots f_n)$ with respect to some fixed measure (in case absolute continuity is lacking we have a degenerated recognition problem in front of us). Maximum likelihood recognition then consists simply in finding that $\mathcal{P}_r$ for which

$$L_r(I^{\mathfrak{D}}) = \max_r. \tag{19.31}$$

To illustrate this, let X consist of the points (x, y) of the first quadrant of R^2, C be the set of real numbers, and G be the group of uniform dilatations $(x, y) \to (gx, gy)$ of X where g is also meant as a positive number. The signs are linear functions of the form $s = ax + by$; a subroutine consists of those functions with $a/b =$ constant. Let $\zeta_1, \zeta_2, \cdots \zeta_n$ be

$$\zeta_\nu = (\zeta_{\nu_1}, \zeta_{\nu_2}) \qquad \nu = 1, 2, \cdots n \tag{19.31}$$

where all the $\zeta_{\nu i}$ are independent and $\zeta_{\nu i}$ has an exponential distribution with mean value $m_{\nu i}$. If $I = ax + by$ it follows that the coordinates of $I^{\mathfrak{D}}$

$$f_\nu = a\zeta_{\nu_1} + b\zeta_{\nu_2}, \qquad \nu = 1, 2, \cdots n, \tag{19.33}$$

are all independent and that f_ν has the frequency function

$$p_\nu(f_\nu) = \frac{\exp\left\{-\dfrac{f_\nu}{am_{\nu_1}}\right\} - \exp\left\{-\dfrac{f_\nu}{am_{\nu_2}}\right\}}{am_{\nu_1} - bm_{\nu_2}} \tag{19.34}$$

unless the numerator vanishes; then the above expression should be replaced in the obvious way. This implies that we should recognize that pattern $\mathcal{P}_r$ for which

$$\prod_{\nu=1}^{n} \frac{\exp\left\{-\dfrac{f_\nu}{am_{\nu_1}}\right\} - \exp\left\{-\dfrac{f_\nu}{am_{\nu_2}}\right\}}{am_{\nu_1} - bm_{\nu_2}} = \max_{a,b} \tag{19.35}$$

The notion of covariance does not make sense here since G is not defined on $\mathfrak{J}^{\mathfrak{D}}$. However, if we apply G to X this is equivalent to changing m_{ν_i} into $1/gm_{\nu_i}$ so that the recognition rule (19.25) will lead us to (ag, bg) instead of to (a, b): the rule is G-invariant.

20. Deformation of Time Patterns. The different grammars of patterns and their deformation mechanisms are not separated from each other by strict delineations. To some extent it may be a matter of choice which one of two alternative models we prefer, and part of what will be discussed in this section could have been studied under the heading of contrast patterns. Since they have a flavor of their own, however, they deserve a separate treatment.

Let us consider a pure grammar $\mathcal{G}$ for time patterns (see section 12), with signs $s = (t)$ given in real time $t \in T$, $T = R^1$ (continuous) or $T = N^1$ (discrete). Consider a set $\mathfrak{J}$ of mappings $\tau = \tau(t)$ of T into T and with a probability measure P on τ. Putting

$$dI(t) = I^{\mathfrak{D}}(t) = I[\tau(t)] \tag{20.1}$$

we have a probabilistic deformation grammar $\mathcal{G}_{\text{prob}}$ of the background deformation type (section 19). We shall call τ the *subjective* time. Note, however, that the special way of forming images (through concatenation) is non-commutative, a case that we have avoided until now.

The transformations of time $t \to \tau$ will usually be assumed to be given through non-decreasing functions $\tau(t)$. Possibly we could tolerate a model for which the event that $\tau(t)$ decreases has a positive but small probability. If the values of $\tau(t)$ do not include a certain value of T (when t runs through T) we will miss a part of the information inherent in the pure image $I = I(t)$. On the other hand, if $\tau(t)$ is constant for some t's, then this means that a part of the pure image will be more exposed to our observation when we read the deformed image.

Let us consider two deformation mechanisms. First, can we make this $\mathcal{G}_{\text{prob}}$ covariant in probability? Consider a finite number of arbitrary time points $t_1, t_2, \cdots t_n$ and compare the two sets of transformed time

$$\begin{cases} a\colon \tau(gt_1),\ \tau(gt_2),\ \cdots\ \tau(gt_n) \\ b\colon g\tau(t_1),\ g\tau(t_2),\ \cdots\ g\tau(t_n). \end{cases} \tag{20.2}$$

The set a corresponds to the image dgI and the set b to the image gdI. We should then require that

$$\begin{cases} \text{for any } n,\ t_1,\ t_2,\ \cdots\ t_n \text{ and } g \\ g^{-1}\tau(gt_\nu) = \tau(t_\nu),\ \forall\ \nu,\ \text{distribution wise} \end{cases} \tag{20.3}$$

This is a sort of stationarity assumption. To see this take the special case g = translations of the real line. Then (20.3) says that the subjective time should obey $\tau(t + h) - h = \tau(t)$ distribution-wise, which is satisfied if and only if the subjective time can be written as $\tau(t) = S(t) + t$, with $S(t)$ = stationary (strict sense) stochastic process.

Let us specify the $\mathcal{G}_{\text{prob}}$ further. Reasoning as in section 16, we could demand that the increment of subjective time τ over a short interval $(t, t + h)$ of (objective) time t should be independent of deformations elsewhere. We would also be led to require that the distribution of the increment should be the same as for any interval $(gt, g(t + h))$. If G again consists of translations, the process $\tau(t)$ should be *time-homogeneous and of independent increments.*

To illustrate this quite flexible model, which we will call the *subjective time pattern grammar,* we will examine two special cases. In the first one we let the image simply consist of a single sign. If the signs have different range then we can clearly recognize I given $I^{\mathfrak{D}}$: perfect recognition without error. Similarly, if different signs are characterized by properties left unchanged by $\mathfrak{D}$ we can do the same. If s_1 is increasing and s_2 decreasing then there can be no doubt about which of these two signs has given rise to $I^{\mathfrak{D}}$. Of course, this should be considered a singular case, but we must be aware that such possibilities exist. Say that we have only two signs, $S = (s_1, s_2)$ with

$$s_\nu(t) = \begin{cases} 1 & 0 < t < b_\nu, \\ 0 & b_\nu < t < 1 \end{cases} \qquad 0 < b_1 < b_2 < 1 \tag{20.4}$$

We have chosen $T = (0, 1)$. Let subjective time be given as a continuous (in probability) process with independent increments such that the frequency function of $\tau(t + h) - \tau(h)$ is of Γ-type.

$$f_h(x) = \frac{1}{\Gamma(h)} x^{h-1}e^{-1}, \qquad 0 \leq x < \infty \tag{20.5}$$

The possible deformed image can then resemble those shown in Figure 4.

Let us denote (for I_ν) by β the quantity $\sup t : \tau(t) \leq b_\nu$. It is clear that in this very simple deformation grammar β contains all the relevant information. In case 3, the deformed images are identical: $I_1^{\mathfrak{D}} = I_2^{\mathfrak{D}}$. In the extreme opposite, case 1, both deformed images have $\beta < 1$. It may happen, if subjective time is severly contracted, that $I^{\mathfrak{D}}(t)$ is not defined in a neighborhood of $t = 1$, but this causes no real trouble.

To be able to recognize the underlying pure image we need the probability distribution of β. But the event

$$\{\beta \geq u\} = \{\tau(u - 0) \leq b\} \tag{20.6}$$

where b is the parameter of the true image, so that

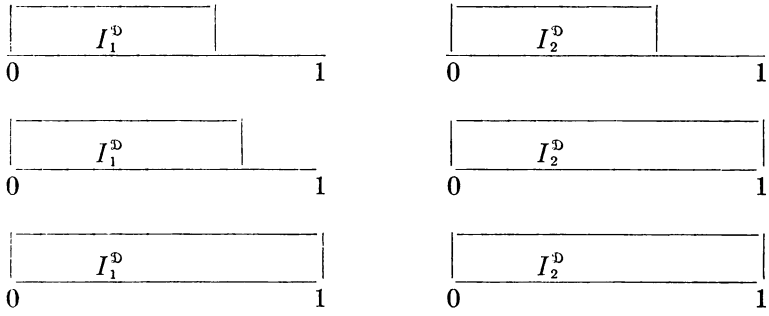


FIGURE 4

$$P(\beta \geq u) = P\{\tau(u - 0) \leq b_\nu\} = P\{\tau(u) \leq b_\nu\} \tag{20.7}$$

where the last equality follows from $\tau(u)$ being continuous in probability. But β takes values only in the interval [0, 1]. We get

$$P(\beta \leq 1) = P(\tau(1) \leq b) = \frac{1}{\Gamma(1)} \int_0^b e^{-x}\, dx = 1 - e^{-b} \tag{20.8}$$

so that the conditional distribution function, given that $\beta < 1$, is

$$F_{\text{cond}}(x, b) = \frac{P(\beta \leq x)}{P(\beta < 1)} = e^b \left[1 - \frac{1}{\Gamma(x)} \int_0^b v^{x-1} e^{-v}\, dv\right] \tag{20.9}$$

with the conditional frequency function

$$f_{\text{cond}}(x, b) = e^{-b} \frac{d}{dx} \frac{\Gamma(x; b)}{\Gamma(x)} \tag{20.10}$$

written in terms of the incomplete Γ-function $(x\,; b)$. The recognition algorithm then takes the form

$$\begin{cases} \text{if } \beta = 1 \text{ choose the largest } b_\nu \\ \text{if } \beta < 1 \text{ choose the } b_\nu \text{ making } f_{\text{cond}}\,(\beta, b_\nu) \text{ as large as possible} \end{cases} \tag{20.11}$$

The reader will be able to see how this could have been formulated for the case with several signs in the configurations, keeping the set-up the same in every other respect.

In the second example we shall allow the image to consist of several signs. Let a sign $s = s(t)$ be 0 outside an interval i_s of length n_s, a natural number. In i_s the function $s(t)$ assumes the constant value $v_s > 0$. As in a more general context, we would like (although this is not always possible) to have the pure grammar $\mathcal{G}$ to be such that images have a unique analysis in terms of their signs (see sections 3–4). Therefore, let the rule $\mathcal{R}$ say for the moment that any grammatical configuration contains only signs with different values v_s. Let $T = (0, \infty)$ and consider an arbitrary pure image $I = s^{(1)} + s^{(2)} + \cdots + s^{(l)}$ (non-commutative addition here) where $s^{(\nu)}$ has length n_ν and with the constant $\nu = a_\nu$ (see Figure 5, where we have also displayed a deformed image as it might look). We use the convention that $I(t) = 0$ for any t after the duration of the last sign. Introduce $\mathcal{G}_{\text{prob}}$ through

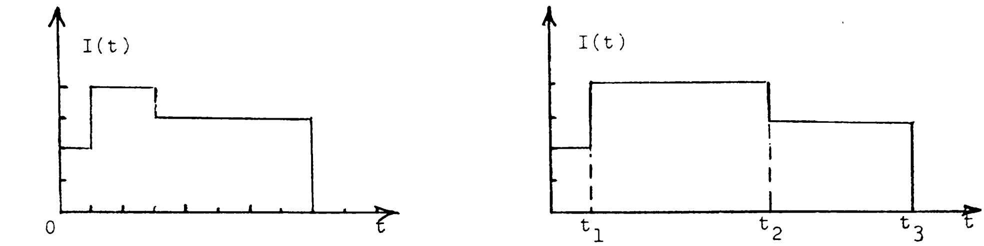
$$\tau(t) = \text{number of "points" in } (0, t) \tag{20.12}$$

where by "points" we understand the events of a Poisson process with intensity 1. Introduce the time points

$$\tau(t) = \sup t \text{ such that } \tau(t) \leq n_1 + n_2 + \cdots + n_\nu\,;\ \nu = 1, 2, \cdots l. \tag{20.13}$$

Since these parameters contain all the relevant information contained in $I^{\mathcal{D}}$ we merely have to derive the probability distribution of $(t_1, t_2, \cdots t_l) = z$ given $I = s^{(s)} + s^{(2)} + \cdots + s^{(l)} = (n_1, a_1\,; n_2, a_2\,; \cdots\,; n_l, a_l)$. Of course we need only consider those pure images I for which the a's coincide with the different ordinates of $I^{\mathcal{D}}$. The joint frequency function of z is

$$\begin{aligned} f(z) &= \prod_{\nu=1}^{l} \frac{1}{(n_\nu - 1)!} (t_\nu - t_{\nu-1})^{n_\nu - 1} e^{-(t_\nu - t_1)} \\ &= e^{-t_l} \prod_{\nu=1}^{l} \frac{(t_\nu - t_{\nu-1})^{n_\nu - 1}}{(n_\nu - 1)!} \end{aligned} \tag{20.14}$$



$\ell = 3$
$n_1 = 1,\ a_1 = 2$
$n_2 = 2,\ a_2 = 4$
$n_3 = 5,\ a_3 = 3$

FIGURE 5

We have put $t_0 = 0$. To recognize the different possible I's we start from the values n_ν^I and maximum likelihood gives us the recognition algorithm as a maximum problem over the image algebra $\mathfrak{I}$:

$$\sum_{\nu=1}^{l} (n_\nu^I - 1)D_\nu - \log (n_\nu^I - 1)! = \max_{I \in \mathfrak{I}} \tag{20.15}$$

with $D_\nu = \log (t_\nu - t_{\nu-1})$.

The time patterns studied so far in this section are characterized by (a) *signs are combined through concatenation* and (b) *deformations consist of transforming time to a subjective time.* Let us now keep the first property but introduce the deformation grammar by assuming *imperfections in the machine producing the subroutines* of signs (see section 2). To fix ideas, let the pure subroutines σ be given by the difference equations with constant coefficients (see section 12):

$$L_\sigma x_t = a_0^{(\sigma)} x_t + a_1^{(\sigma)} x_{t-1} \cdots + a_{p_\sigma}^{(\sigma)} x_{t-p_\sigma} = 0 \tag{20.16}$$

Here t runs through n_σ consecutive integers, G consists of translations of the discrete time axis. To make this definition unique we should also include initial conditions for x_t in the description of σ. Let us do that by defining these initial values as those assumed by the x's of the earlier signs; then we merely have to specify initial conditions for the first sign of the configuration. Then $\mathcal{G}$ is well defined.

We shall introduce a probabilistic deformation grammar by replacing (20.16) by

$$L_\sigma x_t = \epsilon_t \tag{20.17}$$

where the $\epsilon_t s$ are independent and identically distributed stochastic variables.

This $\mathcal{G}_{\text{prob}}$ will be called the model with *imperfect subroutines.* Note the similarity to the notion of regimes in stochastic processes. It could be combined with the earlier one of this section to embody subjective time.

How do we recognize the pure image behind an observed I? Say that t runs through the values $t = 1, 2, \cdots N$ and we are examining the possibility of representing the pure image as $I = s_1 + s_2 + \cdots + s_l$ with $s_\nu \in S_{\sigma_\nu}$ where the σ_ν-subroutines consist of sequences of length n_ν $n_1 + n_2 + \cdots + n_l = n$. Then the likelihood of observing $I^{\mathfrak{D}} = (I_1^{\mathfrak{D}}, I_2^{\mathfrak{D}}, \cdots I_n^{\mathfrak{D}})$ is given by

$$L(I) = \prod_{\nu=1}^{l} \prod_{t=n_1+n_2+\cdots+n_{\nu-1}}^{n_1+n_2+\cdots+n_{\nu-1}} p(L_{\sigma_\nu} x_t). \tag{20.18}$$

where $p(x)$ is the frequency function of the ϵ's. To fix ideas let $\epsilon_t = N(0, 1)$; then maximum recognition consists in solving

$$\sum_{\nu=1}^{l} \sum_{t=n_1+n_2+\cdots+n_{\nu-1}}^{n_1+n_2+\cdots+n_{\nu-1}} (L_{\sigma_\nu} x_t)^2 = \min_{I \in \mathfrak{I}} \tag{20.19}$$

In this expression a few x_t's will appear with negative values of the time argument and these values should be replaced by the initial values specified in the pure grammar $\mathcal{G}$ above.

Similarly, in continuous time, we would arrive at a recognition algorithm of the form

$$\int_0^T [L_{\sigma(t)} x_t]^2 \, dr = \min_{I \in \mathfrak{I}} \tag{20.20}$$

where $\sigma(t)$ is piecewise constant as specified in the description of the signs making up the pure image. This sort of recognition scheme is certainly suitable for time patterns governed by regimes.

21. Acknowledgments. This work was begun in 1965 when the author was a member of the Thomas J. Watson Research Laboratory of the International Business Machines Corporation, in Yorktown Heights, N. Y. It was initiated by a suggestion of Dr. Herman Goldstine, Director of Scientific Development, that the general problem of pattern recognition be studied. During the later stages of the work I have profited from discussions with members of the Watson Laboratory and with Dr. Walter Freiberger of the Brown University Center for Computer and Information Sciences.

References

There is a vast literature on pattern recognition. A few articles that are directly related to the present paper are mentioned below.

Freiberger, W., and Grenander, U. [1]: *Computer-generated image algebras, Proc. 4th Int. Cong. Information Proc.*, North-Holland Press, Amsterdam (1969)

Grenander, U. [1]: *Statistical inference and stochastic processes*, Ark. Mat. (1950)

Grenander, U. [2]: *Toward a theory of patterns, Symposium on probability methods in analysis*, Springer Verlag, Berlin (1967)

Grenander, U. [3]: *A unified theory of patterns*, to appear in *Advances in Computers*, Academic press (1969)

Hoffman, W. C. [1]: *The Lie algebra of visual perception*, J. Math. Psy., **3** (1966)

Narasimhan, R. [1]: *Labelling schemata and syntactic descriptions of pictures*, Inf. Cont. **7** (1964)

Pitts, W., and McCulloch, W. S. [1]: *How we know universals. The perception of auditory and visual forms*, Bull. Math. Biophysics **9** (1947)