AD-759 712

COMPUTER IDENTIFICATION OF TEXTURED VISUAL SCENES

Ruzena Bajcsy

Stanford University

Prepared for:

Advanced Research Projects Agency

October 1972

DISTRIBUTED BY:

NTIS
National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

STANFORD ARTIFICIAL INTELLIGENCE LABORATORY
MEMO AIM-180

STAN-CS-72-321

COMPUTER IDENTIFICATION OF TEXTURED VISUAL SCENES

BY

RUZENA BAJCSY

SUPPORTED BY

ADVANCED RESEARCH PROJECTS AGENCY

ARPA ORDER NO. 457

OCTOBER 1972

COMPUTER SCIENCE DEPARTMENT

School of Humanities and Sciences

STANFORD UNIVERSITY

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Stanford University Computer Science Department Stanford, California 94305	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

Computer Identification of Textured Visual Scenes

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

technical, October 1972

5. AUTHOR(S) (First name, middle initial, last name)

Ruzena Bajcsy

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
October 1972	~~157~~ 167	79

8a. CONTRACT OR GRANT NO. SD-183 b. PROJECT NO. ARPA Order No. 457 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) STAN-CS-72-321 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AIM180

10. DISTRIBUTION STATEMENT

Distribution Unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
Details of illustrations in this document may be better studied on microfiche.	

13. ABSTRACT

This work deals with computer analysis of textured outdoor scenes involving grass, trees, water and clouds. Descriptions of texture are formalized from natural language descriptions; local descriptors are obtained from the directional and nondirectional components of the Fourier transform power spectrum. Analytic expressions are obtained for orientation, contrast, size, spacing, and in periodic cases, the locations of texture elements. These local descriptors are defined over windows of various sizes; the choice of sizes is made by a simple higher-level program.

The process of region growing is represented by a sheaf-theoretical model which formaliz the operation of pasting local structure (over a window) into global structure (over a region). Programs were implemented which form regions of similar color and similar texture with respect to the local descriptors.

An interpretation is made of texture gradient as distance gradient in space. A simple world model is described. An interpretation of texture regions and texture gradiant is made with a simulated correspondence with the world model. We find that a problem-solving approach, involving hypothesis-verification, more satisfactory than an earlier pattern recognition effort (Bajcsy 1970) and more crucial to work with complex scenes than in scenes of polyhedra. Geometric clues from relative sizes, texture gradients, and interposition are important in interpretation.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6801

Unclassified
Security Classification

STANFORD ARTIFICIAL INTELLIGENCE LABORATORY		OCTOBER 1972
MEMO AIM-180

COMPUTER SCIENCE DEPARTMENT
REPORT CS- 321

COMPUTER IDENTIFICATION OF TEXTURED VISUAL SCENES

by

Ruzena Bajcsy

ABSTRACT: This work deals with computer analysis of textured outdoor scenes involving grass, trees, water and clouds. Descriptions of texture are formalized from natural language descriptions; local descriptors are obtained from the directional and non-directional components of the Fourier transform power spectrum. Analytic expressions are obtained for orientation, contrast, size, spacing, and in periodic cases, the locations of texture elements. These local descriptors are defined over windows of various sizes; the choice of sizes is made by a simple higher-level program.

The process of region growing is represented by a sheaf-theoretical model which formalizes the operation of pasting local structure (over a window) into global structure (over a region). Programs were implemented which form regions of similar color and similar texture with respect to the local descriptors.

An interpretation is made of texture gradient as distance grandient in space. A simple world model is described. An interpretation of texture regions and texture gradient is made with a simulated correspondence with the world model. We find that a problem-solving approach, involving hypothesis-verification, more satisfactory than an earlier pattern recognition effort (Bajcsy 1970) and more crucial to work with complex scenes than in scenes of polyhedra. Geometric clues from relative sizes, texture gradients, and interposition are important in interpretation.

This research was supported in part by the Advanced Research Projects Agency of the Office of the Secretary of Defense under Contract No. SD-183.

ACKNOWLEDGEMENTS

This study would neither have been started nor completed without the advice and endless support and encouragement of my teacher, Professor John McCarthy.

I am deeply indebted to Dr. Tom Binford. He taught me how to handle computer vision problems from the A.I. point of view and helped me to clarify many confusions. His unfailing interest in the progress of my research has played a large part in bringing this work to a successful conclusion.

Consultations with Professor Robert W. Floyd are gratefully appreciated.

The discussions with Professor Jerome Feldman and the interest shown by Professor Cordell Green are also acknowledged with gratitude.

My work was dependent on the programming and system support of many people. Notably Dan Swinehart and Lynn Quam contributed greatly in several ways.

I wish to express my thanks also to Professor Ladislav Gvozdjak from the Slovak Technical University, Bratislava, whose personal support brought me to Stanford.

I am grateful to Dr. Zoltan Domotor who introduced me into sheaf theory.

Finally, the major financial support received from the Artificial Intelligence Project at Stanford University during my academic program at Stanford is also acknowledged with gratitude.

THE TABLE OF CONTENTS

1. INTRODUCTION

1.1 Statement of Problem

We intend to deal in this thesis with techniques for computer understanding of scenes with texture. We consider examples of outdoor scenes, although textured surfaces appear in almost every sort of scene, and we show some examples of isolated and artificial textures. Studies in computer vision are motivated by a wide range of applications. Those involving texture include agricultural survey and analysis of earth resources satellite pictures. Planetary exploration by remotely controlled vehicles will demand some autonomous vision because of long delay times. The social benefits of computer-controlled cars have been described by McCarthy. Industrial robots will soon acquire vision. Texture synthesis, for which we feel our techniques are applicable, is useful in computer aided design and computer aided art. Interpretation of scanning electron microscope pictures e.g. for metallurgy may be of interest. We are also interested in constructing a model of human perception. Finally, vision is one of the more interesting problem areas within artificial intelligence, and contributes to the advance in our understanding of intelligent systems.

Without undertaking a complete review of the literature, we would like to broadly contrast the work we have done with that of other work in computer vision. Several small groups have studied perception of polyhedra. Their work has been concerned with three-dimensional objects with plane, uniform faces. The limited success of these efforts has depended to a large extent upon large homogeneous areas and isolated edges. A number of prediction-verification techniques have arisen, some of which are special to the simple cases considered there. Others are more general and useful to our work. Because of the complexity of textured scenes, we feel that the prediction-verification approach to perceptual systems is even more important for our work.

Some work has been done with image processing, which is intended for improving the ease of human interpretation of images of particular value. [Quam, others]. Other work has been directed toward crop identification and other statistical summaries of the earth surface. These studies have also made limited progress and are mentioned in a survey below. However, there is much room for improvement of texture description, and those studies completely ignore scenes where the three-dimensional character is important.

It should be clear what we are really after in an interpretation of a scene. The goal is not only to get a map of colored and textured regions. We are not merely after identification of some image as a member of a class. That is, we are not out to identify the letter A. Nor do we wish to identify some region of the image as some previously seen element, although this might help us to achieve our goal. We have in mind a system with a task, to navigate, for example, and execution of the task requires understanding of the structure of the space portrayed by the image.

1.2 Outline of Thesis

In Chapter 1, after the statement of our problem, we present a review of literature that we think is relevant to an analysis of visual texture. The literature covered in this review comes from three different sources: psychology, neurophysiology and computer science. By no means is this review exhaustive. However, we hope to show the reader, through the psychological and neurophysiological review, which features in grouping are important, and thus justify the features that we use for texture description. The computer science review includes pattern recognition, linguistic and analytic approaches.

In the second chapter, instead of presenting some formal definition of a texture, which we do not believe is possible in general, we describe two concrete scenes with textured and colored regions. With these examples, we describe our representation of texture and of a real scene.

The third chapter presents the implementation of procedures which give us texture descriptors. We discuss operators in the spatial domain, that is edge and region operators. We discuss some of the techniques possible and problems to be encountered in extending these techniques to textured scenes. Then we discuss texture descriptors derived in the Fourier domain. Directionality turns out to be one of the most useful features, easily detectable in the Fourier domain. We find that the Fourier technique has many problems, and we analyze the advantages and limitations of these descriptors. We show how to compute the size of texture elements and their contrast; these analytic expressions are evaluated for several examples and appear quite useful.

Chapter four describes a region growing algorithm applied to forming textured regions and colored regions. We present a sheaf-theoretic point of view which provides precise specification of conditions for continuity of textured or colored structures.

Chapter five is devoted to the problem of interpretation of outdoor scenes. We describe our earlier work using a pattern recognition approach to the classification of texture samples. Then we present an analysis of the use of texture gradient in determining the orientation of surfaces and relative depth. A simple world model is presented for outdoor scenes. A discussion is made of higher level procedures which make interpretation of two examples of outdoor scenes. This higher level program has not been implemented, but gives a good persepective to evaluate the modules developed, and is the target for which we aimed.

1.3 Previous Results

During the past five years or so, a great deal of work has been done in the area of computer-based visual perception, computer recognition, and computer identification of visual scenes and patterns. Several results have been obtained in computer analysis of two-dimensional images, interpreted as projections of geometrically simple real objects. See particularly, Guzman (1968), Brice and Fennema (1970), Pingle (1969), and Falk (1970).

Polyhedra and collections of polyhedra are recognized from single view projections. First the meaningful edges are recognized, then the main regions and from these, finally, in combination with the world-model, the identification of the objects is inferred.

Much less attention has been paid to computer analysis of two-dimensional pictures which depict real-world scenes. What we mean here are scenes such as forest, grass, water, and their combinations. The separate regions, formed by, say, grass and bushes, do not differ in contrast of light intensity, nor in color (as both are usually green), but rather in their texture.

A primary problem in texture is how we perceive a textured surface as uniform in a nontrivial way. Intuitively speaking, there are many levels on which one can perceive texture. In one situation we may look at the pattern showing how bricks are distributed on the wall and call that a texture. In another situation we may have a closer look at the same wall from the same distance and see the texture of the individual bricks and ignore the texture given by the architectural structure of bricks.

Flock (1965) and Freeman (1970), reporting about various experiments in connection with testing aspects of visual perception pattern into components each of which is in some way internally uniform (Werthelmer (1912)).

1.3.1 Neurophysiological Studies Relevant to Feature Extraction in Visual Perception

In what follows we shall present a review of certain experimental results dealing with visual feature detection in animals. The general scheme of experimentation is as follows. The set of stimuli consists of geometric entities such as slits, edges, bars, and corners. Recordings are made from single cells or a small number of cells in a sequence along the direction of insertion of an electrode in the visual system of animals (mostly cats and monkeys). The conclusion is that there are special neurophysiological units, identifiable in well-defined parts of the brain, capable of detecting motion, orientations, and other features of the visual stimuli.

For instance, Kuffler (1953), placing microelectrodes near retinal ganglion cells of cats, found that certain areas of the retina, when stimulated by spots of light, caused the ganglion cells to fire, while other areas inhibited firing. The shape of the excitatory areas for retinal ganglion cells was a small disk, surrounded by an inhibitory annulus or vice versa. The retinal areas, exhibiting the firing, are known as receptive fields.

Concentric receptive fields have been found also in the optic nerve and in the LGN (Lateral Geniculate Nucleus) of cats and monkeys (Hubel (1960) and Wiesel and Hubel (1960)). The only difference between retinal ganglion cells and LGN cells is that the receptive fields in the LGN are smaller. Also, the receptive fields in the LGN of monkeys are smaller than those in cats. The concentric receptive fields have a characteristic temporal behavior: If the center of the field fires for "on" responses, then the annulus fires for "off" responses or vice versa.

Only spatial changes evoke responses, while homogeneous illumination, however strong, influences very little the firing of these units. In functional terms these are "discontinuity detectors". Of course, there are "Ganzfeld detectors" in the retina, responsive to average brightness of a large region, which regulate the pupillary mechanism through the superior colliculus, but we are only interested in neural units that participate in processing patterned stimuli.

A revolutionary discovery was the description of the relation between receptive field geometry and the cytoarchitecture of the cortex. Mountcastle (1957) discovered the columnar organization of the cat's somatosensory cortex. This vertical modular arrangement in the somesthetic cortex means that units along a column perpendicular to the cortical surface all give rise to the same sensory discharge. In the monkey, cells along one column respond to skin touch-pressure, and the cells along another column to joint rotation. (Powell and Mountcastle, (1959)). The interesting feature of this correlation between cortical organization and functional organization became fully apparent in the findings of Hubel and Wiesel (1960, 1962) in the visual cortex of the cat. They found feature extractors of hierarchically increasing complexity. However, as one goes from the so-called simple units, having elongated receptive fields with antagonistic surroundings - also called slit or edge detectors to complex and hypercomplex units that respond to highly special features (like movement in a certain direction, or the end of a line), one notices that despite their diversity, all of these feature extractors have a common characteristic: they all respond optimally to a certain orientation. In a vertical module (column) perpendicular to the cortical surface of the cat and the monkey (Hubel and Wiesel 1960, 1962, 1968) there are several types of units from the simple and complex or even hypercomplex kind. But in a given column, all detectors have the same preferred direction. In addition to this mapping of the orientation information, the retinal position is also maintained and units with receptive fields in neighboring retinal positions tend to lie in close proximity.

Another remarkable finding by Hubel and Wiesel is the hierarchy of feature extraction. Each unit in the hierarchy results from the outputs of units of lower complexity using both excitatory and inhibitory connections. The simple units of slit or edge detector type are built from the so-called Kuffler-units in the LGN by "summing" several adjacent Kuffler units that fall on a line of a given orientation. This summation results in a narrow elongated receptive field having elongated elliptical excitatory (inhibitory) area surrounded by an antagonistic neighborhood. Such cortical units fire optimally for those line segments (slits or edge) that fall on the proper location on the retina and have the preferred orientation. These simple units are the only ones (in addition to Kuffler - units) whose receptive fields can be plotted by luminous dots and segregated into inhibitory and excitatory areas. The complex and hypercomplex units, on the other hand, respond to such complex features as movement of an edge in a certain orientation and direction or the perpendicularity of two intersecting line segments. Here it seems that the notions of straightness, orientation, velocity, position, parallelism, perpendicularity, abrupt ending of a line, corners, and so on, appear as features.

The primary visual problem, how this information is used at later stages, is still untouched.

1.3.2 Psychophysical Experiments Suggesting the Existence of Visual Frequency Analyzers

In this section we give a brief survey of experiments concerning the alleged existence of spatial frequency analyzers, located in the neural system of human subjects. The set of stimuli consists of simple line patterns with differing orientations, contrast, and spatial frequency.

The human subjects involved in experiments are asked to respond to the threshold contrast of the stimuli. Certain aspects of response are used as arguments for and against the existence of a frequency analyzer in the subject. Enroth-Cugel and Robson (1966) and Campbell and Robson (1968) claim to have found neurophysiological evidence for a spatial frequency analyzer. Several experiments have been done to determine the properties (such as the transfer function) of the hypothetical analyzer using masking methods. For example, Pollehn and Roehing (1970) used filtered two-dimensional visual noise and spatial sinusoidal gratings. Julesz and Stromeyer (1970) used one-dimensional filtered noise for masking. The noise consisted of vertical strips whose amplitude along the horizontal axis of a CRT monitor was determined by a Gaussian process. The visibility of a sinusoidal grating is strongly dependent on the frequency of the grating. If the grating frequency overlapped the noise band, it was masked. However, the rejection band had to be at least an octave wide on either side (Blakemore and Campbell (1969) and Julesz (1971)). The frequency analyzers have such a shallow characteristic that the analogy to Fourier analysis is rather remote.

Historically, the first hint of spatial frequency analysis was made by Campbell and Kulikowski (1966) who investigated the visibility threshold of a test grating as a function of the orientation of a masking grating. They briefly mentioned that maximum threshold increase occurs when the masking and test gratings have similar geometry. That spectral analysis actually occurs in the visual system was suggested by Pantle and Sekuler (1968) using adaptation and test gratings of different frequencies and by Campbell and Robson (1968) who noted that a square grating appears as a sinusoidal grating until the higher harmonics reach their visibility thresholds.

A recent study by Nachmias et al. (1969) showed that at the threshold of visibility the various spatial frequency analyzers are statistically independent of each other (as long as the various spectral components have frequency ratios in excess of 5:4). The finding by Nachmias et al. (1969) indicates that at threshold the phase of visual or auditory signals is not detected by the perceptual system. However, for perception above the threshold level, the phase information is used in higher processors. After all, both the impulse function and white noise have the same flat amplitude spectrum, but very different phase spectra. The fact that they are heard and seen as being very different shows that ultimately the phase information is utilized.

1.3.3 Psychological Studies in Pattern Grouping

The topic of this section is a discussion of the psychological literature on texture grouping. The grouping process depends heavily on criteria of similarity of items. Although it has been known for some time that similarity is one of the most important features of perceptual grouping, only recently, in the work of Julesz (1971), Beck (1967), and Attneave and Olson (1970), has it been made clear explicitly what kinds of similarities are effective in this respect.

Beck (1967) has studied <u>perceptual groupings</u> produced by line figures. He showed that the overall orientation was essential for cluster formation, while more complex properties such as rated similarity or familiarity of figures were irrelevant. Example: T and tilted T are more similar than T and ⊣ . However, as a texture, T and tilted T form a more distinguishable texture than T and ⊣ . This has been confirmed by Attneave and Olson (1970) who have done similar and more extensive study, with different shapes such as L, ⅃ , ∧, ∨, lines of different lengths, and orientations. Directionality was important in grouping. We might expect curvature to be important also, but curved lines were grouped with straight lines which had the same direction.

Grouping was dependent also on orientation of the whole image. In general, grouping and complementary segregation is based on certain descriptors, some of which represent relationships of elements of the stimulus array to an internal Cartesian reference system.

Julesz (1962) has studied the clustering problem on random dot textures (stereograms). He described textures and predicted their properties by specifying their higher order statistics.

The usual joint probability distribution is an inadequate descriptor in perception, since it does not describe the shape of clusters. There are at least two ways to handle this difficulty. [One way is to define certain information rules for single clusters and parametrize them (orientation, compactness, etc.). The trouble with this solution is the arbitrariness of the selection of cluster parameters.) The second way is to use constructs of random geometry. Novikoff (1962) was the first to suggest such a solution.

The clustering process is dependent on the similarity and the proximity of elements. The similarity relation is relativized to brightness, color, geometrical descriptors and other parameters. The proximity relation is based on a distance measure. Nonmetric multi-dimensional scaling techniques (Shepard (1962), Kruskal (1964)) and hierarchial cluster-seeking algorithms are useful tools for handling similarity problems. The methods proposed by Shepard and Kruskal, however, are appropriate only for linear or multilinear cases. In a nonlinear situation an iterative algorithm is applied on small local regions in order to find an intrinsic dimensionality (Shepard and Carroll (1966), Bennet (1969), Fukunaga and Olsen (1971)). Applying multi-dimensional scaling to discrimination of textures composed of random 2 x 2 arrays, Julesz (1971) found that the most important factors for texture discrimination were <u>brightness and orientation</u>.

1.3.4 Texture in Machine Recognition

We have seen several aspects of visual textures, mostly from the point of view of psychology and psychophysiology. Now, we shall examine the characteristics and parameters of a texture with respect to some of the approaches that have been used in machine recognition of textures.

The best review paper about the current state of texture extraction technology is that of Hawkins (1970). According to him, there are four types of approaches that have been taken to texture classification: (1) Spatial frequency content, (2) Gray level content, (3) Local shape content, and (4) Higher order measures. Most of the early machine texture recognition was related to analysis of aerial photographs. As an example the work of Lendaris (1970) can be mentioned. Lendaris analyzed pictures of aerial photographs of agricultural landscapes as well as of urban areas. For recognition purposes he used the power spectrum of the brightness function over some windows of constant size. The power spectrum is analyzed in the following way: first, two functions are formed; one the energies along different directions; then the energies along different frequencies. Then, from these two functions, feature vectors are created. The features are the number of peaks and the strength of the peaks in both functions. These feature vectors are used for classification of the area. He distinguishes four classes:

man-made areas (cities)

agricultural areas

a single road

intersection of two roads.

Another way of describing repetitive patterns is to use some statistical features of the brightness function forming the pattern.

This has been used with some advantage in analyzing biological material (Lipkin et al. 1966, Prewitt and Mendelson (1963)), cloud pattern classification (Darling and Joseph (1968)), and the discrimination of strategic and tactical targets and terrain classification. The statistical features, though sometimes useful, have some limits. Thus the variance of a salt and pepper scene is the same as that of a white scene with a uniform dark area. The size of connected areas (think of clouds, for instance) can take a wide spectrum. The number of changes (zero crossings) is informative again only within a certain context, when combined with other features such as direction, etc. Histograms are useful in estimating light distribution in the picture and setting up the threshold values for measurements.

Shape measures used in texture analysis have involved applying a particular local "matched filter" to every point in the image area, and counting the number of points that match above some threshold. This has been applied to the previously noted examples of classifying biological material (Prewitt and Mendelson (1968)), to cloud classification, and to targets and terrain classifications (Hawkins (1970)). A more analytic approach to shape description of chromosomes is taken in terms of conic sections. An individual chromosome is defined as a non-negative function on the real plane, subject to certain constraints on position, size, orientation, etc. Ledley et al. (1965) suggested a simple method of measuring concavity and convexity. Integral geometry measures (Julesz (1971)) and their extensions amount to calculating the number of occurrences of n-tuples of specially arranged local points in all orientations over the image area.

Matched filters allow one to describe practically any shape. However, the matching process, due to the computation of a large number of correlations and the need of hundreds of patterns, is rather slow. The similarity relation can be defined in a straightforward fashion in terms of the threshold values.

Simple descriptors such as convexity, length of the boundary/area, etc., require small computation time, but similarity relations based on these simple descriptors are not usually sufficient for sharp decisions. Another set of simple descriptors has been suggested and implemented by Rosenfeld and Thurston (1971). They use, in parallel, several local averaging operators applied in different directions and on various sizes of windows. All results obtained from these local operators are evaluated and eventually a texture boundary is found. Though this method finds some texture boundaries, the operators are too trivial for handling a wide class of real textures. Besides, they do not provide any description of a texture, they only detect the texture differences.

All the approaches discussed above are pattern classification techniques. These techniques are not satisfactory for a description of real textures for the following reasons:

(1) Pattern classification techniques have concentrated on linear decision procedures, and domain independent formulations. Context appears as a set of numberical coefficients in a linear function, and in the choice of features. We have better models in terms of context dependent decision trees which provide a better basis for generalization and learning.

(2) Structural relationships and segmentation are part of the

desired analysis. We discuss this further in our analysis. The point has been made repeatedly by picture linguists.

PICTURE LINGUISTIC FORMALISM

In what follows we shall review the so-called linguistic approach; "picture linguists" take as their principal aim to analyze discrete pictures such as bubble chamber photographs, biomedical pictures of neurons, blood cells, and chromosomes, machine-printed, and hand-printed characters, fingerprints and the like. They argue rather convincingly that such pictures cannot be identified by means of classical receptor/categorizer devices. What one is after in this situation is not just a classification, but rather an articulated (discursive) description or explication, capturing the structured subparts of a picture and the relations between them (Miller and Shaw (1968), Narasimhan (1970), Clowes (1970)).

One has to assume that certain pieces of information have already been extracted from the picture by means of nonlinguistic techniques (texture elements and their possible structuring is known). We combine this prior knowledge with the data about the analyzed picture and then "deduce" its structural description. The "deduction" is accomplished by a grammar. Due to the fact that we cannot describe a picture in terms of strings of subpictures, phrase-structure grammars cannot be used directly. The rewriting rules must act on more general entities such as arrays, drawings, labeled graphs (webs), multigraphs, etc. For example, Kirsch (1964) and Dacey (1967) designed a grammar for two-dimensional languages, where the generating rules act on arrays. Pfaltz and Rosenfeld (1969) used for picture description the so-called web grammars in which the rules act on labeled directed graphs. Simply, in picture grammars one tries to replace the total ordering of strings by a partial ordering of graph structures so that the parsing can still work.

The language of the graph grammar is nothing but a collection of graphs that can be derived from initial graphs by iterated application of the rewriting rules.

For instance, one can construct a grammar for directed two-terminal series-parallel networks or neural networks (Pfaltz (1970)). It is believed that the organization of textured regions in scenes would be another promising field of application, particularly, when the number of different textured regions occurring in a scene is small and when their organization is such that a moderate set of rewriting rules can do the job.

Methodical scanning of the picture with a prescribed system of rules, which may be feasible when the variety of possible texture elements and their interconnections is small, becomes rapidly uneconomical where many varieties of wanted textures may exist, embedded in a background containing many similar forms which do not belong precisely to the required category.

The intricacy of textured picture recognition is associated not only with the presence of an incredibly large number of elementary texture elements,but also with the placement rules which seem to have extremely complicated grammatical structure.

To sum up, the linguistic method is suitable for such classes of pictures that contain a small number of primitive objects. The primitives have to be found with accuracy, otherwise the parsing process will terminate in misrecognition. The picture must be recursive in nature, so that a small number of rewriting rules can be used.

Picture languages are inappropriate in situations where the number of primitives is large and the geometrical relationships between these primitives are random. This is the case of most of the scenes such as fabrics, aerial photographs (large number of primitives), cloud covers, grass, bushes (random relationships), shading of smooth objects, textured surface of three dimensional objects (continuity), and similar natural or artificial scenes with strong aspects of repetitiveness, continuity, and regioning, and with intricate changes in gray levels and colors. On the other hand, descriptions of modes and scene elements are graphs, and there is a broad analogy to picture language in other approaches.

ANALYTIC FUNCTION APPROACH

A (real two-dimensional discrete rectangular) picture is represented by a pair $< I_1 \times I_2, p >$, where I_1 and I_2 are non-empty finite intervals of integers and p is an arbitrary real-valued function $p: I_1 \times I_2 \rightarrow$ Reals. If $X = I_1 \times I_2$ is fixed, one can identify the picture with p.

The definition itself is empty. We may proceed to try to approximate the picture function by analytic functions defined on subsets of the image plane. About the only useful analytic properties are those based on periodicity. These constraints are inspired by pictures in which the regions are generated from texture elements by a more or less straightforward family of analytic rules.

A sample of special cases of deterministically textured pictures is listed below:

(a) If p is <u>spatially periodic</u> on a connected region $S \subseteq X$, i.e.,

$$p(x + v, y) = p(x,y)$$

$$p(x, y + w) = p(x,y),$$

where $x,y \in S$ and $\langle y,w \rangle$ is the <u>spatial period</u>,

and if $p|S$ cannot be extended to a larger connected region S' $(S \subset S')$ without losing the periodicity of $p|S'$, then $\langle S,p \rangle$ is called a <u>periodically textured region</u>.

A picture decomposable into a family of periodically textured regions $\{R_i | 1 \leq i \leq k\}$ with spatial periods $\langle v_i, w_i \rangle$ ($X = U_i R_i$ and $R_i \cap R_j = 0$, when $i \neq j$), is called a <u>periodically textured picture.</u>

Simple visual patterns, such as a rectangle covered by a mosaic of squares, triangles, circles, etc., are examples of periodically textured pictures. Brick wall, honey-comb herring bone and many other ornamental or mosaic patterns also belong to this class of pictures.

Note that in this case only two <u>texture elements</u> are involved (black and white squares, etc.) and the whole picture is described by a finite group of translations in two directions (direct product of two translation groups of integers modulo $I_1 \times I_2$). Thus textured pictures of this kind can be defined in terms of their texture elements and an appropriate finite group of translations. Only a small degree of complication arises when the textured picture is decomposable into periodically textured regions.

(b) If p is <u>partially periodic</u> (periodic in one of its arguments) on a connected region $S \subseteq X$, i.e.,

(i) $p(x + v,y) = p(x,y)$ (Periodic in the first coordinate)

(ii) $p(x, y + w) = p(x,y)$ (Periodic in the second coordinate)

where $x,y \in S$ and v (w) is the <u>period</u>, and if $p|S$ cannot be extended to a larger connected region S' $(S \subset S')$ without losing the partial periodicity of $p|S'$, then $\langle S,p \rangle$ is called a <u>partially periodic textured region</u>. A picture decomposable into a family of spatially periodic textured regions $\{R_i | 1 \leq i \leq k\}$ with periods $\{v_i\}$ or $\{w_i\}$, is called a <u>partially periodic textured</u> picture.

(c) If p is <u>partially almost periodic</u> on a maximal connected region S, we obtain a new class of analytically characterized textured pictures. Here "almost periodic" means: For any $\epsilon > 0$ there exists a function $p': I_1 \times I_2 \rightarrow R$ of the form

$$p'(x,y) = \text{Sum}_i \text{Sum}_j (C_{i,j} e^{im(a_i x + b_j y)}),$$

such that $|p(x,y)-p'(x,y)| < \epsilon$, where a_i, $a_j \in R$, $0 \leq i \leq n$, and $0 \leq j \leq m$.

The function p' is not periodic in general, though it has periodic components $C_{i,j}e^{im(a_i x + b_j y)}$ and the absolute difference $|p'(x + v,y + w)-p'(x,y)|$ is an arbitrarily small number for a suitable pair $\langle v,w \rangle$.

Assuming that the pictures under consideration are composed of periodic, partially periodic, or almost periodic textured regions, we can utilize features like expansion of the picture function p over a textured region S into periodic <u>orthogonal series</u> such as Fourier, Hadamard-Walsh etc. In fortunate cases the orthogonal series-features compress the information hidden in the textured region into a few dominant components. This is followed by pattern classification (Rosenfeld (1962), Julesz (1962), and Bajcsy (1970)). If the periodicity or repetitiveness is not the most relevant aspect of the picture in question, an orthogonal expansion may scramble the information content so that no simplification occurs.

A typical case appears when the phase spectrum happens to be relevant in a Fourier expansion of a "nonperiodic" picture and we restrict ourselves only to the power spectrum. Here the information content is not only degraded, but also mixed in such a way that the Fourier features are no more relevant (Lendaris and Stanley (1970)).

1.4 The Contribution of This Research

We feel that theoretical and experimental advances have been made in programs for understanding textured scenes. These are:

1. A sheaf-theoretic formalism for describing textured and colored regions.

2. Symbolic structured description of textures.

3. Implementation of descriptors in terms of Fourier descriptors. Analytic expression of spacing, size and contrast of texture elements, and their approximate location.

4. Forming of color regions.

5. Forming of textured regions.

6. Spatial interpretation of regions in terms of texture gradient.

7. Description of a higher level procedure and world model for outdoor scenes.

2. TEXTURE DESCRIPTIONS

In this chapter we discuss qualitative descriptions of visual textures in order to suggest the corresponding implementation in procedures. Our aim will not be detailed descriptions; in a Borgese story, a project to make perfect maps lead to maps the full size of the countries mapped. Instead, we want to characterize textures in a compact symbolic representation which suggests correspondences with our models, and simplifies human communication and debugging. We feel that everyday texture descriptions are good models for these purposes. At a low level, we want to work with those descriptions to propose plausible colored and textured regions. At a higher level, our aim is a description in object space, not an image space map. Many interpretations and hypotheses should be in terms of objects and properties of the object space. An example is the interpretation of texture gradient in the image as distance gradient in space. Another interpretation is that overlapping regions correspond to foreground and background.

2.1 Examples of Outdoor Scenes

In the scene shown in Figure 1, we find three elements: grass, water, and rocks. The grass lies on an approximately level surface. The rock is in front of the water and behind the grass. We do not describe the image itself, but its interpretations as objects. In describing this scene, we emphasize its segmentation into elements which are objects and regions in object space. This structural description characterizes the relationships among objects and regions. For example, a tree stands above the ground and in front of the sky. The structure allows us to talk about complex scenes in terms of simple elements. To move about, we must know where the grass extends, where to walk around rocks, and where the water is. These spatial relations are essential; even if we were able to store and recognize whole scenes, we would need a mechanism to discover where we walk and what we can pick up.

Grass, rocks and water correspond roughly to three regions in the image. But these simple elements are not directly the sort of regions which come from existing edge or region finding programs. The elements we see are high level abstractions which do not coincide with color or texture regions. In the first approximation, color is the most relevant feature that distinguishes these regions. However, a closer look at the picture suggests that the color boundaries do not correspond exactly to the regions we see. Consider the white waves near the rocks or the dark areas inside the grass region. Our texture region growing also defines a set of regions. Directionality is important in the grass region, yet that property is not uniform over the region. Thus the regions defined by our texture descriptors do not coincide with the grass region we see.

Fig.

Yet there is a continuity over the region in some of the properties of color, size and density of grass stalks. The fact that we have similar stalks of grass over the whole field (sometimes with different direction or with different color) makes it possible to propose the field as an element. This complexity makes it impractical to attempt to identify local elements with local prototypes for grass, sky, or water, or to attempt to identify the low level regions from our programs.

In a second example in Figure 2 we have four elements: grass, trees, clouds, and sky. Again, color separates the sky, clouds, grass, trunks of trees, and in some areas, separates the crowns of trees. Texture, on the other hand, separates the trees from grass.

In the object space description, the sky and trees are distinct. We could arbitrarily define image regions as disjoint. Proximity of regions of like color is one basis for proposing a connectivity among tree branches and among fragments of sky. Those connectivities reflect the object space descriptions of trees as connected and sky as connected. The regions based on proximity in the image are unconnected and overlapping. That description allows an inference (which may not always be valid) that the trees are in front of the sky. Arbitrarily defining disjoint regions rejects these hypotheses of object space connectivity and the conclusion of interposition from overlap.

Although the trees are approximately of the same height, and the grass stalks are also roughly of constant height, their apparent size in the image decreases toward the center rear of the picture. The size of the grass stalks nearest us is the same as that of the trees farthest from us. Gibson [1950] has emphasized that perception relies heavily on the interpretation of systematic variation of apparent size with image position (texture gradient) as a variation of distance from the observer. For most purposes, the relative depth of elements in the world is sufficient. Assuming that we know the position of the observer, the gradient allows us to determine the absolute distance of objects. The measurement of observer or camera position and angles, and calibration of the image device (Sobel (1970)) are essential.

2.2 Textured Regions and Textured Elements

The examples from the previous section demonstrated a structural description of images by segmentation into elements of object space. We further structure these textured regions in terms of texture elements and their spatial relationships. In Table 1 we show some examples of texture elements and their relationships as they appear in object space and image space.

Texture elements cannot be determined in isolation. A single element may be unrelated to the texture. The relationships are frequently only approximate. In a texture of pebbles, the size similarity may be important even though the sizes vary significantly; still, there is a uniformity within a factor of 10 or so. Similarities of other properties such as contrast, shape and spatial distributions may also be only approximate.

In practical implementations we can describe only simple relationships: linear, periodic, regular but aperiodic, continuous, symmetric, and the like. Likewise, shape descriptors must be relatively simple. One may question the effectiveness of simple relationships and their descriptors; it is reasonable to think that a more complex description of texture elements and their relationships is necessary for adequate description of textures. The psychological experiments cited in Chapter 1 indicate that human differentiation of textures depends heavily on a few simple descriptors such as contrast and directionality, and ignores even curvature in making texture groupings. Although we cannot estimate the computational complexity of descriptors, we have an intuitive feeling that in terms of time, or in terms of complexity of wiring for parallel systems, that simple descriptors such as directionality are clearly preferred.

Table 1

	Name of Region	Grass	Water	Forest
Object Space	Texture elements	leaves, blades of grass	water waves	(a) Evergreen: trees (b) Deciduous: fruit trees
	Texture element size	width: 1/4" length: 2-10 in.	widely variable	width/height: 1/2 length: 5-20 ft.
	Spatial relationships between elements	dense, roughly parallel and vertical, and partial covering	quite parallel waves or con centric circular waves	(a) vertical and parallel (b) vertical and parallel partial covering
	Color	green, yellow or brown	blue, dark blue, dark green, silver gray	(a) crown of trees is green and the trunk of trees is dark brown. (b) crown: green, brown, yellow or red; trunk: light brown
	Boundaries of elements	fuzzy, smooth	fuzzy, smooth	sharp, not smooth
Image Space	Geometric description of elements	linear and directional	linear, directional, concentric circles	trunks of trees: linear texture crowns of trees: blob-like texture
	Expected contrast	very low	very low	high (trees with sky)

Table 1 (Continued)

	Name of Region	Sky	Clouds	Brick wall	Pebbles
Object Space	Texture elements	homogeneous	a cloud	bricks	pebbles
	Texture element size		1/4-2 miles	width: 3-4 length: 8-15 inches width/length 1/2	diameter: 1-3 inches
	Spatial relationships between elements	homogeneous	pattern depends on weather	horizontal rows	randomly distributed
	Color	blue	white, gray red	gray, red, brown, yellow	any color
	Boundaries of elements	sharp irregular (horizon)	fuzzy but contrasting	sharp and smooth	sharp and smooth
Image Space	Geometric description of elements	homogeneous	blob-like or directional	bidirectional	blob-like
	Expected contrast	high	low	depends on the back-ground low or high	low or high, depending on the back-ground.

.5 Textured Regions and Their Organization

In the previous section we discussed images in a part/whole structure: scene-regions-elements. The regions and elements were primarily in object space. A texture may have a few layers of heirarchical structure; in Fig. 5, the surfaces of the bricks have a rough texture. The regions formed by the bricks are elements of the brick wall texture.

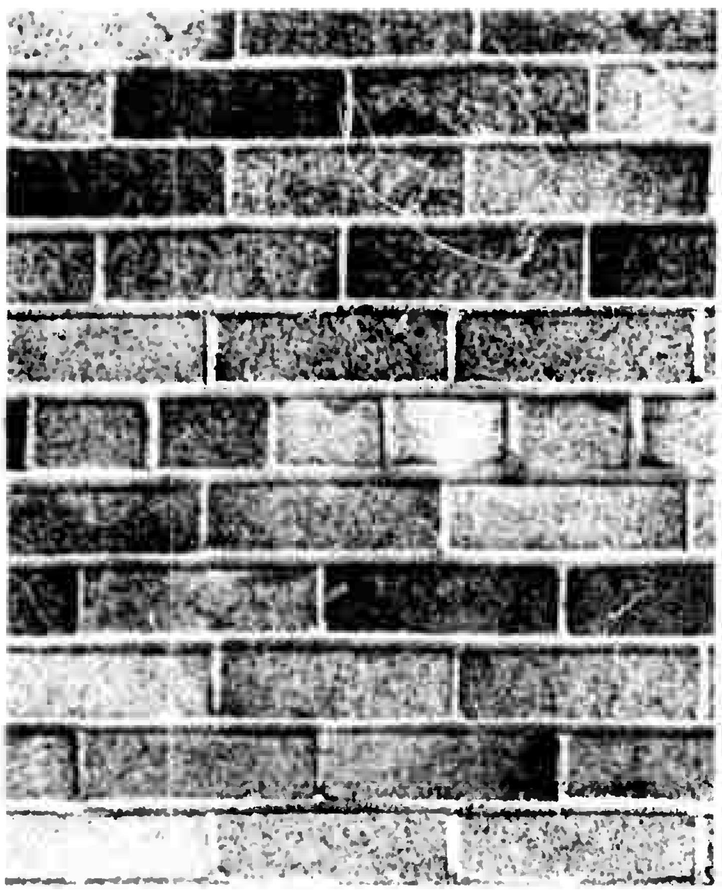


Fig.

The textured region thus can be a texture element in a textured super-region, or the texture element may have micro-texture.

Region and edge finders have some success with large homogeneous regions. Grouping of texture elements can be viewed as a generalization of homogeneity, the regions of a common property correspond to the regions of homogeneity from a region growing operation. The related operation of finding discontinuity in texture properties is analogous to edge-finding between homogeneous regions. Rosenfeld (1970) has discussed the problem of finding texture boundaries as that of finding gradients in the average values of statistical measures (which are assumed to be any suitable operator). While this is suggestive, it unnecessarily emphasizes statistical measures as opposed to structured descriptions which would be more suitable for patterned textures.

Let us approach the question of the organization of textured regions. In the simplest case, the picture to be described is partitioned into a disjoint covering of textured regions.

A somewhat more complex system of regions can be described by a tree structure. It may be used to represent the topological organization of brightness contours (Krakauer (1970)). While this may seem a great generalization, a tree does not well describe the system of regions from a number of descriptors. Even for a single descriptor, the tree is rigidly heirarchical. The nodes of the representing network are used for regions and the arrows correspond to the spatial relationships between the regions. Systems of features lead to several networks of regions. A single feature may give rise to a non-disjoint network of regions. For an operator to give disjoint regions (a partition) one must assume an equivalence relation (reflexive, symmetric, and transitive). Quantization would be an example leading to an equivalence relation. Gradient thresholding would be another example. Selection of typical values, followed by thresholding within an interval, would not lead to equivalences, so that it would not lead to a partition.

It is not necessary to fully expand the whole network or family of networks. Rather, instead of thinking of comparing several networks derived from different features, we use some simple hypotheses derived from a subnetwork of some particular network and supported by evidence from features which might imply another network (which may never exist as such).

We must deal with texture boundaries as well as textured regions. The boundary problem is dual to the grouping problem. Therefore the difficulties encountered in a grouping have their analogs in boundary detection. Take as an example the scene in Fig. 1. The objects in this scene (grass, water, and rocks) are separated by <u>physical</u> or <u>virtual</u> boundaries. Some of them are <u>visible</u> while others are <u>hidden</u> (grass covers the boundary between water and rocks). In the identification process it is not clear whether one should follow the boundaries defined by individual texture elements (look at the individual straws near the rocks) or whether one should look for some kind of average boundary or perhaps keep a spatial gap between two different textures.

Region growing operators use certain similarity criteria. These are applied in patching local structures into global ones. Whenever we meet a dissimilarity, a boundary point or segment is proposed. In the first approximation, a region is formed by patching continuous structures over connected areas. In this case the corresponding boundaries are also connected. There may also be internal, unclosed boundaries. When local discontinuities occur within a region, proximity criteria are used for bridging the gaps. The proximity here is used as an extension of continuity.

The same is true with interrupted boundaries. Proximity and continuity of boundary segments suggest continuation.

In the past it has been customary to think of regions as a disjoint covering of the image. The examples in Fig. 1 and Fig. 2 have shown that this conception is too simple to be useful. An equally simplistic point of view is that boundaries of regions are always closed curves.

3. PROCEDURES FOR TEXTURE DESCRIPTORS

In the previous chapter we discussed the description of texture in object and image space. In this chapter we shall specify the implementation of these descriptions. Specifically, we shall study texture descriptions in the spatial domain and in the Fourier domain. Algorithms for concrete descriptors will also be presented. Although the descriptors will be derived in the Fourier domain from the power spectrum, they actually refer to textural properties in the spatial domain.

We will find it useful to distinguish scalar, topological, and geometric features (shape, area, size, boundary, connectivity, thinness ratio) from relational features (spatial distribution, organization, gradient).

3.1 Texture Descriptors Derived in the Spatial Domain

Since descriptors refer to properties of objects represented in the image space, it is natural to look for operators acting directly in the spatial domain. The skeleton of this section is this: Procedures isolating the image elements, geometric description of image elements, and clustering of elements based on proximity and their spatial organization.

In the process of isolating the image elements the most important features are the following topological properties: connectivity, continuity, and proximity. These properties, applied to brightness or color, are used in all region finders (Fenema and Brice (1970)). Discontinuity is the basic property to be used in edge and line operators (Binford (1970), Hueckel (1971)). Current edge and line operators are designed for detecting discontinuities between two large homogeneous regions and they do not operate satisfactorily on small regions. The textured elements that one finds in outdoor scenes are too small in size and too large in number and therefore cannot be processed usefully by any of the above operators. However, under poor resolution conditions in the image, where the texture elements are smeared (so that the homogeneity stands out more than usual), one may be successful even with the above mentioned operators.

After completing the isolation of image elements - figures, we shall describe them. We select those descriptors which enable clustering, i.e., based on proximity those which will find the nearby elements. We had already a chance to note that color and brightness are among the most important descriptors in natural scenes. Image elements cannot be taken separately from their background. In fact, the common background of the elements is a strong clue for their clustering. The relationship between the background and color is expressed in terms of contrast, and therefore it can be used as another descriptor.

The descriptors corresponding to spatial relations depend on proximity relations just as cluster processes depend on proximity. Typically, we want to define colored regions by proximity, rather than only connectivity. Grass and trees are regions broken into many fragments defined by connectivity. But other like regions are nearby. This proximity in space and color can be phrased as a problem of proximity in 4 dimensions, using the multi-entry technique outlined by Binford in the Stanford Progress Report of January 1971. Likewise, super regions can be defined by brightness, contrast, size and shape descriptors clustered on the basis of proximity. Spatial relations, the intervals between elements and directions of these intervals, can be defined also among elements linked by proximity.

As an expedient which is suitable for linear textures, one can project the elements into several directions. Each projection will actually be a one-dimensional function of gray levels or color. Since this function is still too complicated for practical implementation, it is simplified by using a square wave approximation. The square waves are described either by edge detection operators or by magnitude and the distance between two consecutive zero crossings. Since the distances between zero crossings are intervals in which the approximating gray levels are constant, the method is called interval analysis. That technique has been used with some success to describe regular linear textures in an MIT term paper by Peter Wolfe, (1970).

Since the shape of a two or three-dimensional object in a general situation could be extremely complicated, we cannot hope and, in fact, we do not want to describe it in detail. Instead, complex shapes are decomposed into simpler ones which are (hopefully) easier to describe. A typical example is a tree which may be decomposed into its trunk and crown, where the trunk is geometrically linear while the crown is blob-like. In shape analysis of outdoor scenes we find directionality among the most useful features. One can see this immediately in Table 1. Directionality, combined with length/width ratio and length along the preferred directionality make up a linear element description of shapes or parts of shapes. These are all directly implementable descriptors. In ourdoor scenes, the shapes of texture element are quite important, while the shapes of the important regions of object space (sky, grass, trees, water) are not very important.

The apparent size of an object in an image is not relevant if considered in isolation. This fact was already noted in Fig. 2. There the apparent size of grass was the same as the apparent size of trees, located further from the observer. However, the size of region could be relevant, particularly in the initial stage of a scene analysis when one is searching for large connected regions. Despite the importance of descriptors derived in the spatial domain, we shall not use them in this work. Currently available edge finders and region finders are tailored for large homogeneous regions. In natural scenes, textured areas are composed of small texture elements. Even to the extent that the boundaries of small regions are determined, the data structures require unreasonably large memory, since the boundary descriptions are no longer economical. The next steps of description of elements and clustering elements of similar direction, size, color, or brightness, seem prohibitively time consuming and difficult for grass, pebbles, sand, etc. The one-dimensional interval analysis might have some utility but is very limited; combined with other methods such as Fourier description, interval analysis is potentially useful.

3.2 Texture Descriptors Derived in the Fourier Domain

In what follows we shall need some elementary and well-known notions of Fourier analysis. They will be reviewed presently.

Consider a real picture function of two variables in a matrix form $g(x,y)$, where x and y are variables from fixed intervals of natural numbers $I_1 = \{0, 1, \ldots, p_2-1\}$. The two-dimensional discrete finite Fourier transform of the function $g(x,y)$ is then given by

$$F(n,m) = \frac{1}{p^2} \sum_{x=0}^{p-1} \sum_{y=0}^{p-1} g(x,y)\exp(-2\pi i(xn+ym)/p), \quad (1)$$

where $p = p_1 = p_2$ and i is the usual imaginary unit.

In general, $F(n,m)$ is a complex function, given uniquely by its power spectrum $P(n,m)$ and phase spectrum $PSI(n,m)$:

$$P(n,m) = SQRT(F_{re}^2(n,m) + F_{im}^2(n,m)),$$

$$PSI(n,m) = ARCTAN(F_{im}(n,m)/F_{re}(n,m)).$$

From the elementary properties of the Fourier operator it follows that any real periodic function has a symmetric Fourier image with respect to the origin. An equally well-known but somewhat more interesting fact is that the power spectrum is invariant with respect to translation in the spatial domain, but not with respect to rotation. A trivial consequence of this property is that the directionality of a pattern in the picture is preserved in the power spectrum but the phase of the transform is not.

If a function is periodic, partially periodic, or almost periodic, then its Fourier transform compresses the data considerably without great loss of information and the relational features derived from the Fourier image form a good description of periodic or almost periodic patterns.

As we have pointed out above, the power spectrum contains the information about the form of a periodic picture function restricted to a window. The phase spectrum, on the other hand, represents by and large the locational (positional) information in a window.

We said also that directionality is preserved in the power spectrum. This fact allows us to infer some gross shape properties. We are able to distinguish directional and non-directional components of texture. For this reason, it is useful to transform the power spectrum from a cartesian coordinate system $\langle n,m \rangle$ into a polar coordinate system $\langle r, \varphi \rangle$. Then in each direction φ, one can regard $P(r, \varphi)$ as a one-dimensional function $P_{\varphi}(r)$. Similarly, for each frequency r, function $P_r(\varphi)$ is a one-dimensional function. Thus, the description of the texture depends in this method on the form of the pair of functions $\langle P_{\varphi}(r), P_r(\varphi)\rangle$.

Function $P_r(\varphi)$ determines whether there is a directional or non-directional component. If function $P_r(\varphi)$ is flat then the corresponding texture is nondirectional. If it has few distinguished peaks, the texture is directional. One peak leads to a monodirectional texture. Two peaks under certain constraints lead to a bidirectional texture.

The nondirectional texture could be homogeneous, noisy or blob-like.

Function $P_{\varphi}(r)$ distinguishes between noisy and blob-like texture. The noisy texture corresponds to a flat nonzero function $P_{\varphi}(r)$. Whereas in the case of the blob-like texture, function $P_{\varphi}(r)$ will have some peaks. The homogeneous texture corresponds to an almost constant function $P_{\varphi}(r)$ for $r>0$ and with a large value for $P_{\varphi}(0)$. In the case of a directional texture function $P_{\varphi}(r)$ will have peaks similar to the case of blob-like texture. The frequency in the maximum of $P_{\varphi}(r_{max})$ will roughly correspond to the distance between two parallel stripes (in the case of directional texture) and to the distance between two blobs in the case of a blob-like texture.

We have shown the interpretation of function $P(\varphi)$. Now we want to analyze a further possible interpretation of function $P(r)$. Consider a monodirectional pattern that appears as a one-dimensional (in the particular direction) square wave function shown in Fig. 4 .

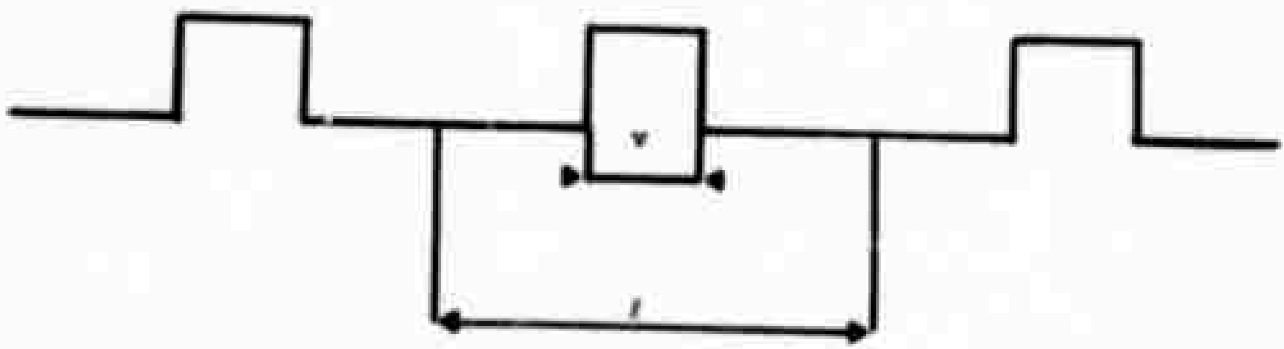


Fig. 4

Denote the replicative symbol $\text{ш}(x)$ and the wave form by $f(x)$. The periodic function in Fig. 4 is expressed as a convolution of $f(x)$ and $\text{ш}(x)$, thus

$$F(x) = f(x) * \text{ш}(x).$$

The Fourier transform of $F(x)$

$$\mathcal{F}(F(x)) = \mathcal{F}(f(x.v)) \cdot \mathcal{F}(\text{ш}(x.l))$$

$$= \operatorname{sinc} \frac{x}{v} \cdot \text{ш}\left(\frac{x}{l}\right).$$

Applying the window function of the width w, appears as a convolution in the Fourier domain.

$$\mathcal{F}(W(x.w) \cdot F(x)) = \operatorname{sinc} \frac{x}{w} * \left(\operatorname{sinc} \frac{x}{v} \cdot \text{ш}\left(\frac{x}{l}\right)\right).$$

This function displayed graphically is seen in Fig. 5 .

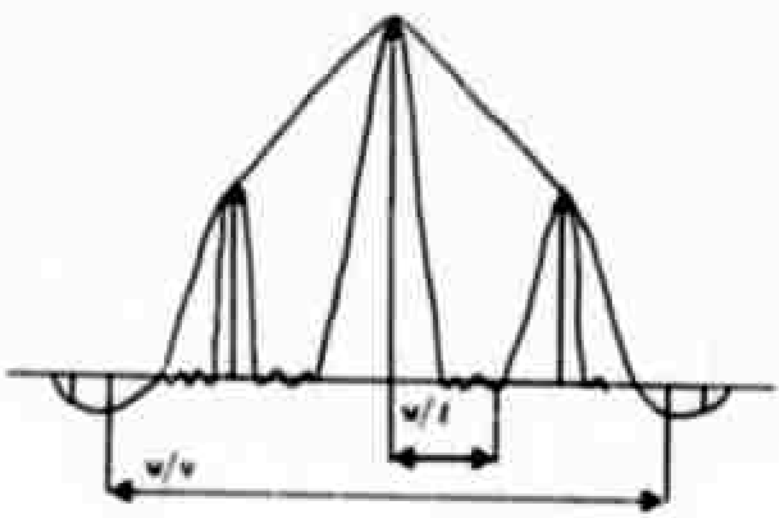


Fig. 5

It is clear that we can measure $\frac{w}{v}$, $\frac{w}{l}$ in the power spectrum from the function P(r), for every directionality and window size w. Consequently we can estimate (how well, depends on the brightness function) the wavelength l, as before and, in addition, the size of the smallest element, v. v and l will be parameters associated with each description. Examples of functions P(φ), P(r) of texture samples will be presented next. The size of samples is 32 x 32 points. The points on the y axis have the corresponding values of the functions P(φ) and P(r) respectively. The points φ (on the x axis) in the graph for function P(φ) represent the value $(x-1) \cdot \frac{\pi}{16}$, for x = 1, 2, ..., 16. The points (on the x axis) in the graph for function of P(r) have just the actual values of frequency r = 1, ..., 16.

Each pair of functions $<P(\varphi), P(r)>$ will be described by some parameters, listed in a table. Below is the list of the parameters and their description.

NAME: The natural language names of the texture samples.

DESCRIPTOR: A hypothetical description of the sample according to some criteria (thresholds) applied on functions $<P(\varphi), P(r)>$.

MAX $P(\varphi)$: The maximal value of $P(\varphi)$.

φ_{max}: Is such φ that $P(\varphi_{max}) = \max P(\varphi)$.

WIDTH: The distance between φ_1, φ_2, where $\varphi_1 \leq \varphi_{max} \leq \varphi_2$ and $P(\varphi_1) = \text{MIN } P(\varphi)$, the left side with respect to $P(\varphi_{max})$. $P(\varphi_2) = \text{MIN } P(\varphi_2)$ (the right side with respect to $P(\varphi_{max})$.

DIR: If the descriptor is directional, first perform a fan filtering in such a way that the fan filter is centered in φ_{max} and then find $\text{MAX } P(n,m) = P(n_{max}, M_{max})$ and thus compute $\text{DIR} = \text{arctg } \frac{m_{max}}{n_{max}}$. If the descriptor is nondirectional then just find

$\text{MAX } P(n,m) = P(n_{max}, m_{max})$

and compute DIR as above.

RO: Is the wavelength computed from the maximal point energy. $\text{RO} = \text{window size}/\sqrt{n_{max}^2 + m_{max}^2}$.

M_φ: is the mean value of function $P(\varphi)$.

V_φ: is the variance of $P(\varphi)$.

MAX $P(r)$: is the maximal value of $P(r)$.

r_{max}: is such r that $P(r_{max}) = \text{MAX } P(r)$.

WIDTH r: Is the distance between the center of $P(r)$ and the threshold value of the envelope of $P(r)$.

M_r: is the mean value of $P(r)$.

v_r: is the variance of $P(r)$.

v: is the element size, = window size/width r of the envelope.

ℓ: is the spacing between elements.

= window size/frequency of the first peak.

In the case of bidirectional texture a pair of values is listed for the following parameters:

MAX $P(\varphi)$, φ_{max}, width φ, DIR and RO.

The texture names are on the top of each picture displaying the corresponding function $P(\varphi)$ and $P(r)$. The actual samples of texture-lines, wood, circle, and sand - are in Figures 14, 15, 20, and 21. The texture water is a sample from the upper left corner of the picture in Fig. 1.

Fig. 6 and Fig. 7a display functions $P(\varphi)$ and $P(r)$ of textures, parallel lines and water, followed by Table of Parameters. For the identification of parameter v we have used the directional part of the water picture. The filtered alternative of functions $P(r)$ and $P(\varphi)$ for water is in Fig. 7b.

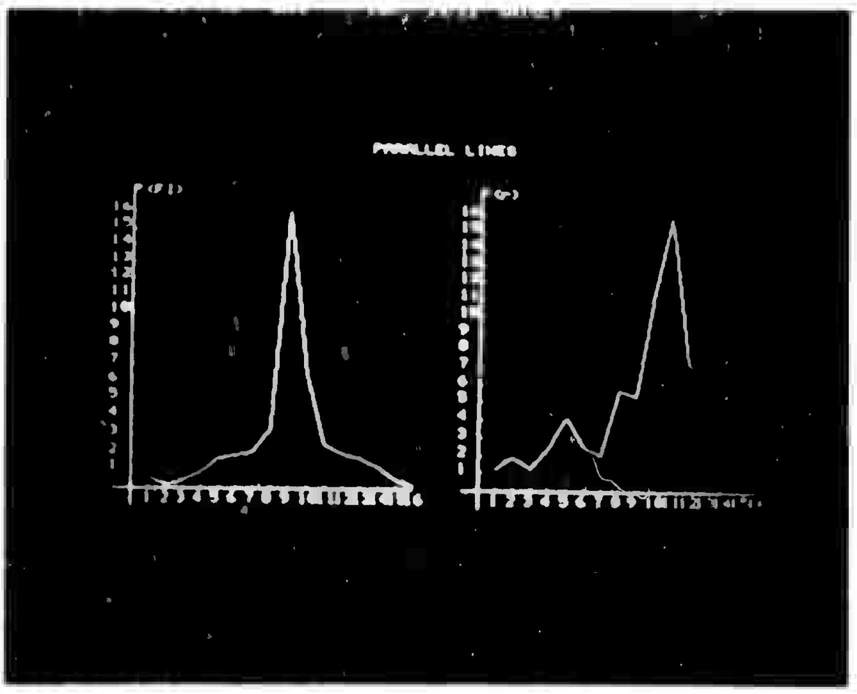


Fig.

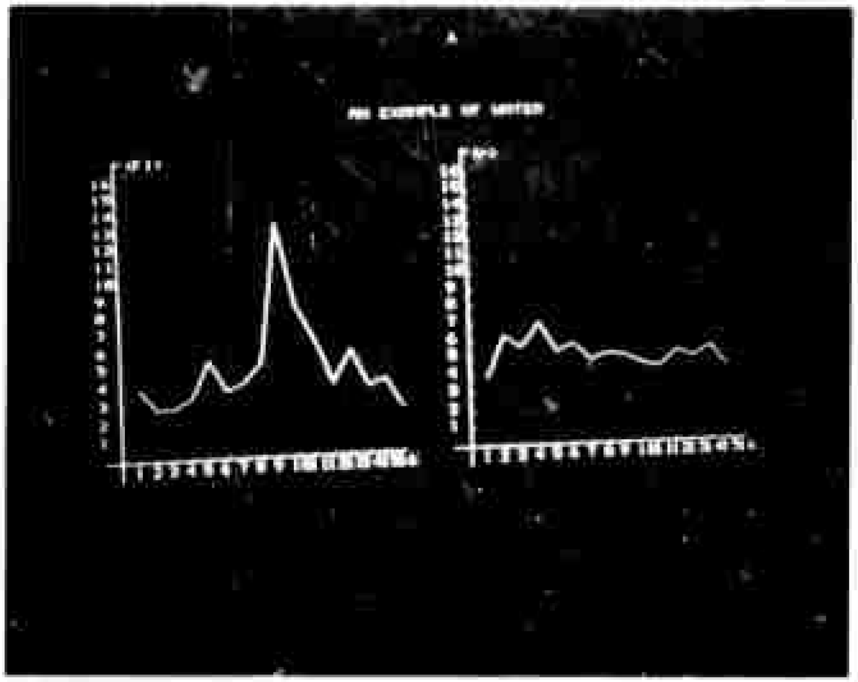


Fig. 7a

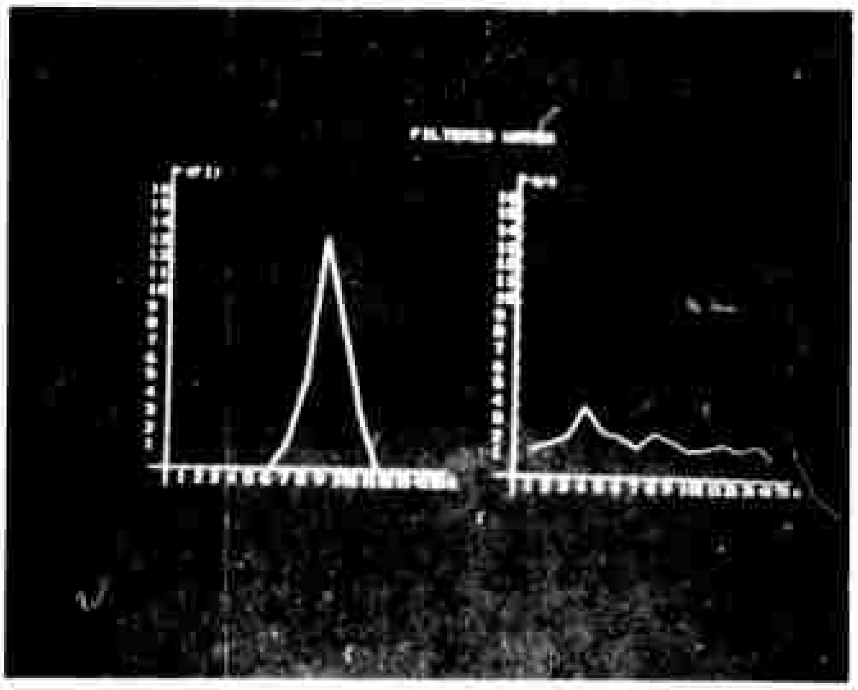


Fig. 7b

TABLE 2

NAME DESCRIPTION	LINES MONODIRECTIONAL	WATER MONODIRECTIONAL
MAX $P(\varphi)$	242.12	13.5
φ_{max}	9	9
WIDTH φ	4	6
DIR	1.57	1.57
RO	2.9	16
M_φ	39.8	5.08
V_φ	14.22	0.64
MAX $P(r)$	105.2	6.96
r_{max}	11	4
WIDTH r	16	16
M_r	37.9	4.86
v_r	7.42	0.364
ℓ	3	8
V	1	1

COMMENTS: Both textures, lines and water are described by the program as monodirectional (they have one signficant peak in $P(\varphi)$ form, geometrically speaking, parallel vertical lines). That is why φ_{max}, and DIR in both cases are the same, e.g. $\pi/2$. The contrast in the picture of lines is much higher than in the picture of water as indicated by the values of $P(\varphi)$ and $P(r)$. The regular pattern of lines shows higher values of the directional component vs the nondirectional component than the texture of water. (Compare for instance the values of MAX $P(\varphi)$ and M_φ). The water waves are broken and thus they form parallel broken lines organized in random fashion. This shows up in the function $P(r)$ of water texture. That is rather flat in comparison with $P(r)$ of the texture of lines.

In Fig. 8 we display a sample of grass from the scene in Fig. 1. The upper left window in Fig. 8 is the original sample, the upper right window is its corresponding power spectrum, the lower left window is the power spectrum after a high pass filter and the lower right window is the resynthesized original picture after the high pass filter.

This example is presented in order to demonstrate the necessity for separating the slow changes from the real texture pattern. The rationale for this is that most of the objects (texture elements) tend to have the same reflectivity and the lighting varies smoothly, thus shading in the Fourier domain generates a low frequency component.

Functions $P(\varphi)$ and $P(r)$ of textures grass, wood and canvas are displayed in Figs. 9a, 10a, and 11a respectively. The analyzed samples from grass are in Fig. 8, from wood in Fig. 15, and on canvas in Fig. 18. For the sake of considering the main directionality and thus to be able to determine ℓ and v we display the filtered alternatives in Fig. 9b for grass, Fig. 10b for wood, and Fig. 11b, and 11c for canvas (for one directionality). The table of their corresponding parameters is below:

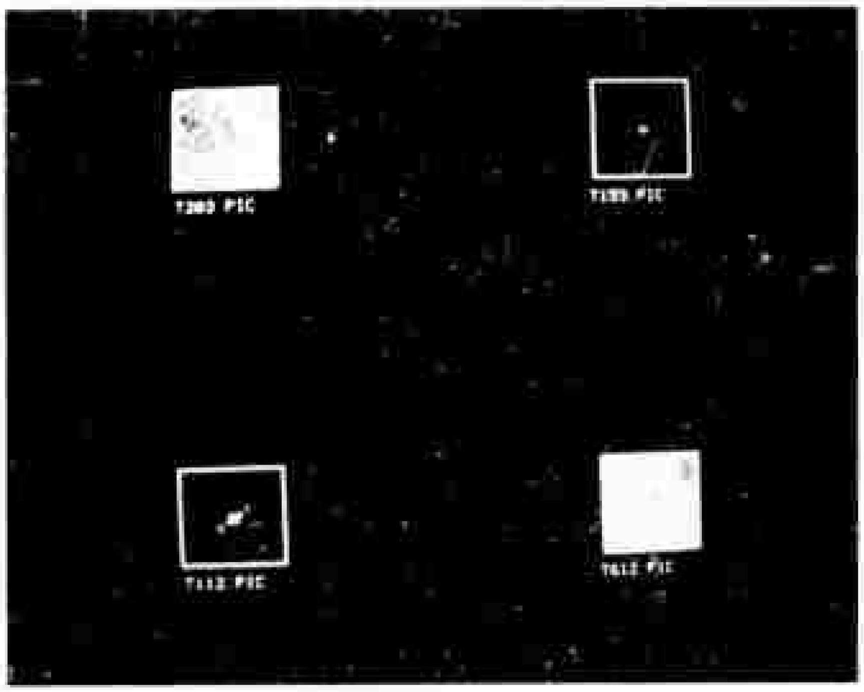


Fig.

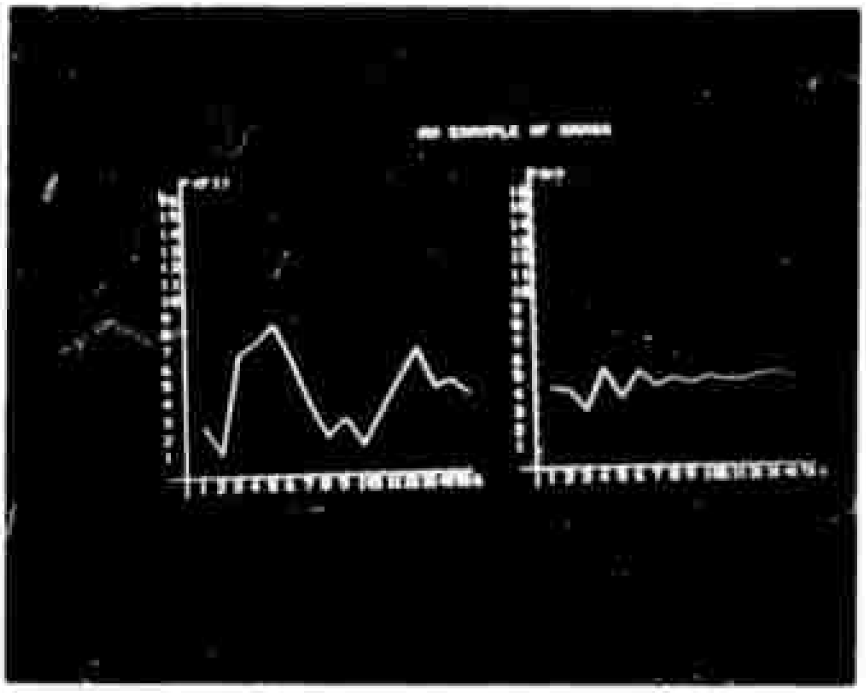


Fig. [illegible]a

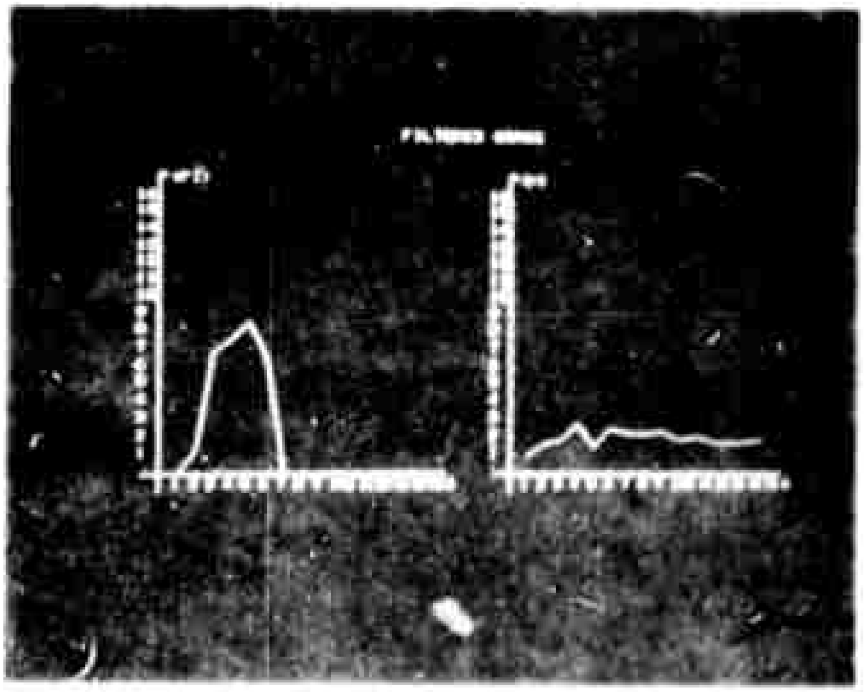


Fig. [illegible]b

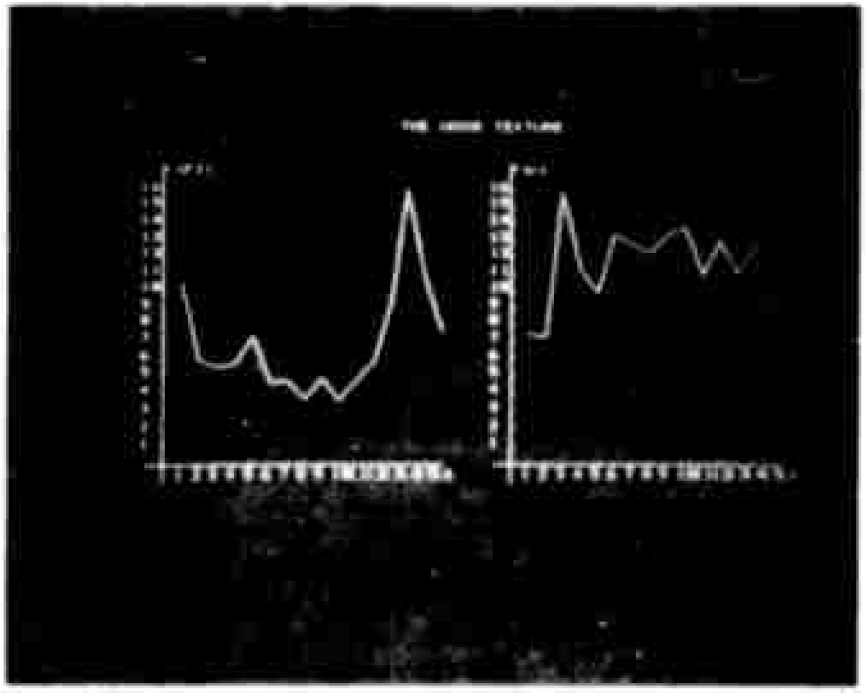
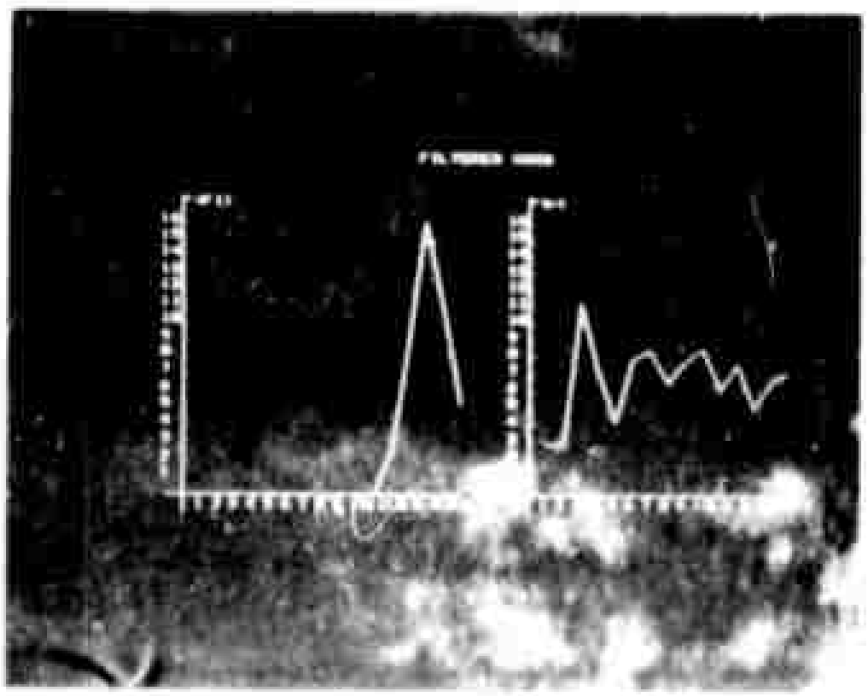


Fig. 1a



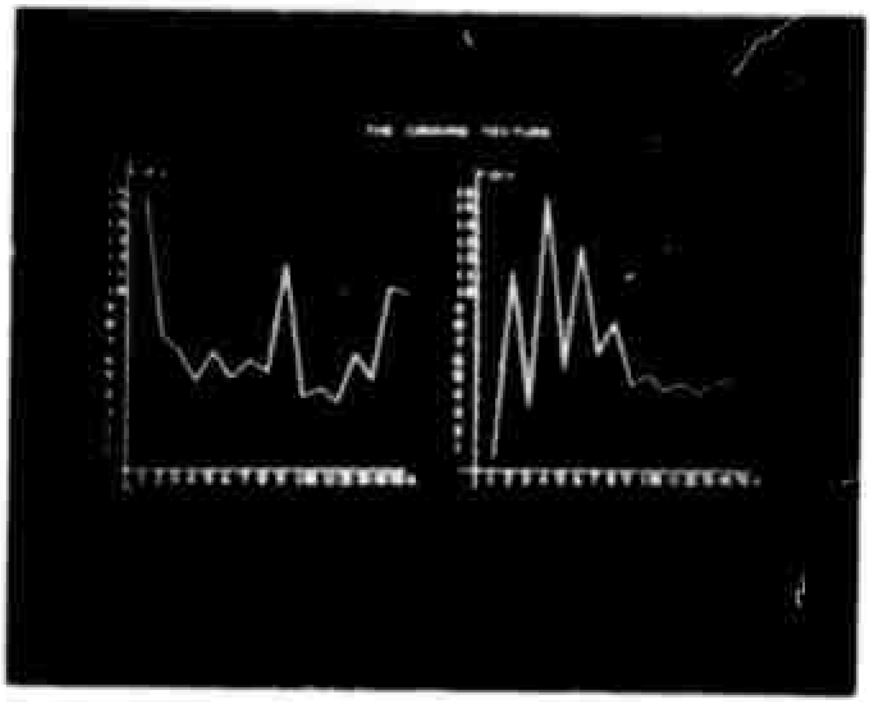


Fig. 11a

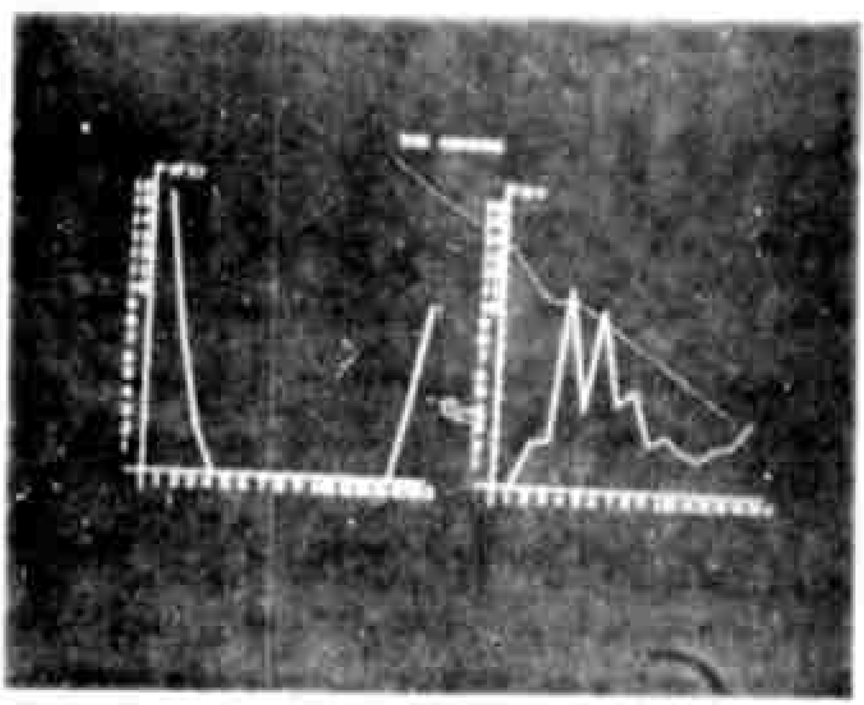


Fig. 11b

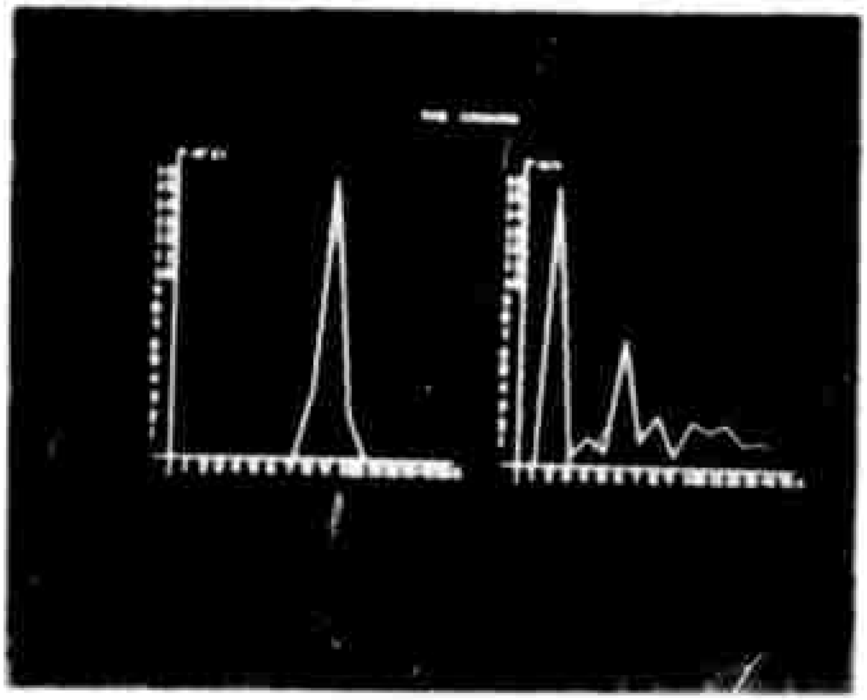


Fig. 11

TABLE 3

NAME DESCRIPTOR	GRASS BIDIRECTIONAL	WOOD MONODIRECTIONAL	CANVAS BIDIRECTIONAL
MAX $P(\theta)$	<8.35, 7.5>	64	<108, 80>
θ_{max}	<5, 13>	14	<1, 9>
WIDTH θ	<6, 4>	5	<4, 2>
DIR	<0.463, 2.03>	2.55	<1.57, 0>
RO	<14.31, 16>	8.81	<16, 8>
M_θ	4.76	32.84	49.3
V_θ	0.534	3.76	5.46
MAX $P(r)$	5.62	44.8	120.94
V	4	3	4
WIDTH r	16	16	9
M_r	4.52	31.46	47.14
V_r	0.324	2.54	7.64
ℓ	<8,16>	10	<16, 8>
ℓ_{max}	8		8
V for MAX DIR	1	1	1.8

COMMENTS: First of all, notice that grass is described as bidirectional, contrary to what would be expected. The reason is that even after high pass filtering, there is still signficant slow change left (wavelength = 16) which forms the second peak. One needs to know more about the scene (its illumination, continuity, context) in order to remove this kind of slow change. It is impossible without further knowledge about the area to handle this situation appropriately.

because the same component (wavelength = 16) which in the case of grass is undesirable, in the case of the canvas texture is an essential part of its description.

Function P(r) in case of grass and wood shows similarities which suggests that both of these textures have some noisy, irregular backgrounds. On the other hand the canvas texture displays signficant peaks in low frequency and decreasing power in higher frequencies.

For more detailed analyses of P(r), one has to separate the different directionalities. This is what we have followed up in Figures 9b, 10b and 11b and 11c.

The last two examples of texture of blobs and sand demonstrate the differences between nondirectional textures. In Fig. 12 and 13 are functions P(Φ) and P(r) of samples of texture recorded in Fig. 20 and Fig. 21 respectively. Table 6 contains their corresponding parameters.

The P(Φ) is a flat function in both textures as to be expected. P(r) in the case of blobs has one signficant peak, whereas in the case of sand P(r) is approximately flat.

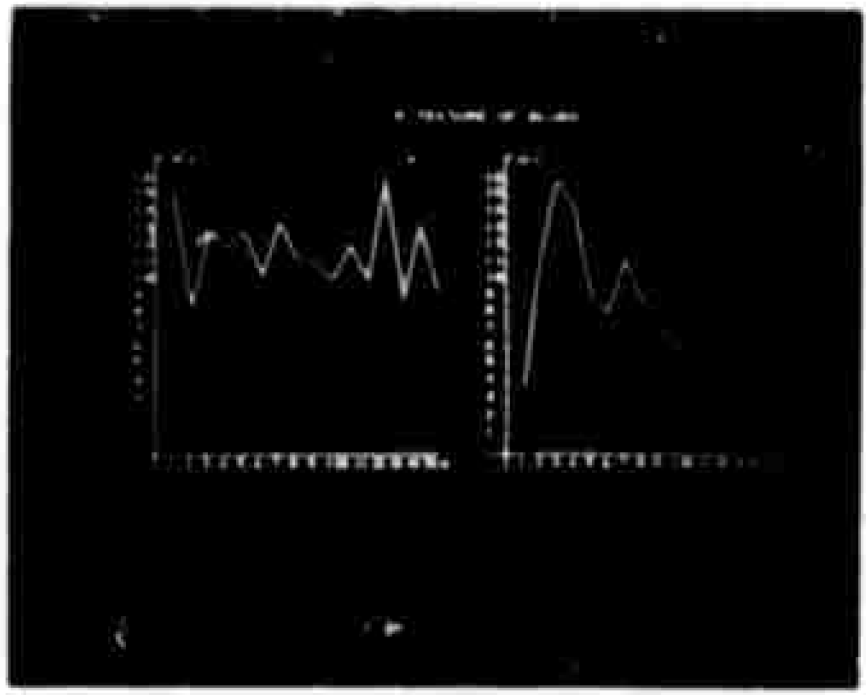


Fig. 12

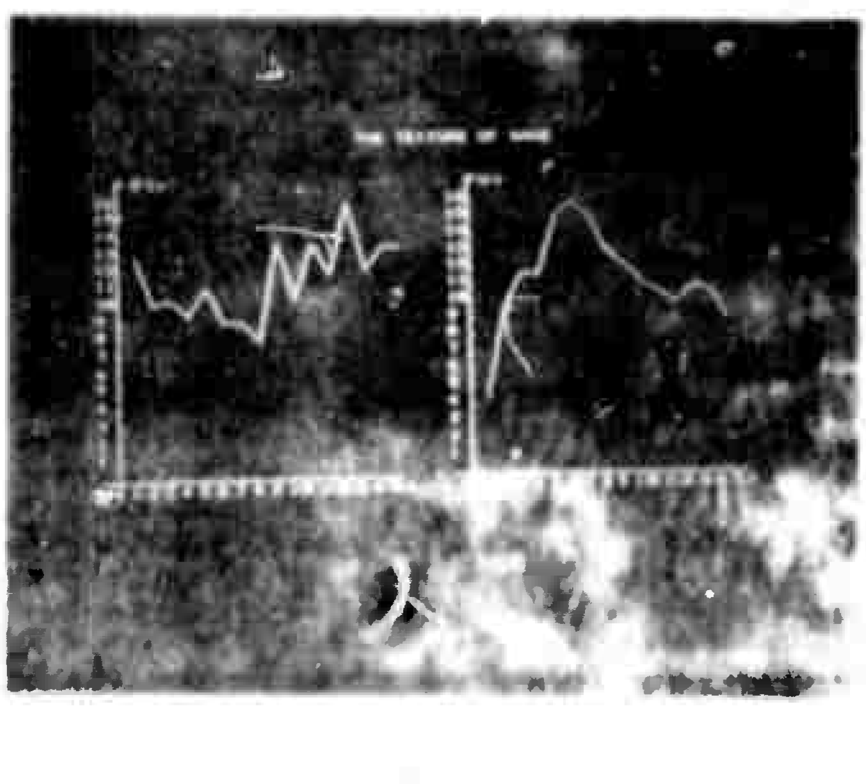


TABLE 4

NAME DESCRIPTOR	BLOBS BLOB - LIKE	SAND NOISY
MAX $P(\varphi)$	82.26	73.74
φ_{max}	13	13
WIDTH φ	3	3
DIR	2.35	2.35
RO	11.31	5
M_φ	60.2	52.8
V_φ	2.72	2.48
MAX $P(r)$	120.46	75.8
V_{max}	3	6
WIDTH r	6	12
Mr	61.70	54.4
Vr	6.52	3.18
φ	10	5
V	2.6	1.3

We must make some comments about the differences between continuous and finite discrete Fourier transforms. The continuous Fourier transform exists for every function with finite energy, while the finite discrete Fourier transform exists for any function. Our interpretations will be based on the continuous transform and the actual computations on the discrete transform (fast Fourier transform). The discrete transform is really a Fourier series. A continuous Fourier transform is rotationally invariant (except for windowing effects) while a discrete transform has distinguished axes along the coordinate axis and the diagonals. Thus a directional image has a continuous Fourier transform in a very narrow band, while the discrete transform has a narrow band transform only for directions along the preferred axis. There is a corresponding difficulty in defining fan filters which we have not succeeded in solving. The difficulty with narrow fan filters is demonstrated in the following example, a line with directionality $\theta = 22\ 1/2^{\circ}$ in digitized form, with a window size of 8×8 points. Due to the sampling problem the line is represented by only four points instead of the desired 8 points.

The values of the corresponding power spectrum are in matrix 2. From inspecting the values in matrix 2 it is clear that there is a spread of energies in different directions besides the expected direction $\theta' = 112\ 1/2^{\circ}$. This effect is due to poor sampling. For more details see Huang (1970).

MATRIX 1 f(x,y)

0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1
0	0	0	0	0	1	0	0
0	0	0	1	0	0	0	0
0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

MATRIX 2 P(n,m)

4	0	0	4	0	0	0	0	0
3	1	2.6	1	1	1	1	2.6	2.6
2	1	0	0	4	0	0	0	4
1	2.6	1	1	2.6	2.6	1	1	2.6
0	0	0	0	0	4	0	0	0
-1	2.6	2.6	1	1	2.6	2.6	1	1
-2	0	4	0	0	0	4	0	0
-3	1	2.6	2.6	1	1	1	1	2.6
	-4	-3	-2	-1	0	1	2	3

All the real values in F(n,m) have to be divided by coefficient 64.

One should make a note of a fairly important though elementary mathematical fact, namely that the Fourier transform does not preserve functional restriction. More specifically, if $g(x,y)|W$ denotes the restriction of the image function $g(x,y)$ to a window W (so that $g(x,y)$ is truncated outside W), then

$$F[g(x,y)|W] = F[g(x,y)]|W$$

is true for every W only when $g(x,y)$ is periodic with period equal to the size of W. Thus a Fourier image of a truncated function, truncated outside a window, will in general depend also on the part of the function $g(x,y)$ whose domain is outside W. What this means practically is that certain texture elements could be split in half by windowing and as a consequence, an improper interpretation would be derived. This problem can be partly compensated for by overlapping windowing.

Human perception allows us to discount smooth changes in shading. This fact allows us to separate shading from edges. The Fourier transform, on the contrary, reflects not only edges, but also slow changes which are ignored in human visual perception. Perhaps the simplest way of demonstrating this is by recalling the basic dictionary of the Fourier transform. We find that a rectangular impulse is transformed into a sinc function, a triangular impulse into a $sinc^2$ function, and a cosine signal is transformed into two impulses. We are accustomed to regarding images in terms of homogeneous regions with sharp boundaries, and to describe elements by brightness and color contrast and outline shape. In the Fourier domain, these become jumbled in a way that is only approximately resolved by our heuristics; thus they are not always usefully described.

In addition, the texture elements (their shape) and their organization are also jumbled together in the Fourier domain. So, for instance, dots and small segments of lines organized in parallel lined fashion, will be described equally as monodirectional texture. Thus they are not described in full details. As we said before, for more detail, one has to apply the spatial, local operators.

For areas of a scene for which homogeneous regions are too small for use of usual edge-finding and region-growing techniques, the Fourier transform provides useful and compact descriptors. Many of the examples of textured regions showed linear texture elements, crudely aligned, and with roughly uniform size and spacing. These shape descriptors have natural counterparts in the Fourier domain. Directionality in the spatial domain corresponds to a directional transform, and uniform spacing corresponds roughly to dominant frequencies in the Fourier domain. However, a much more common uniformity in the spatial domain, cons' nt size elements randomly distributed, does not have a clear counterpart. There has been much oversimplification of the use of the frequency spectrum. In reality, it appears as though it has very restricted utility, however that utility corresponds to a few descriptors which have primary importance in human perception. Since most descriptors are spatial domain descriptors not directly related to the transform, frequency domain techniques are quite limited. Nevertheless, the proper combination of the Fourier descriptor together with the spatial domain technique is the suggested approach for a texture identifier. The Fourier technique will compactly describe large areas with repetitive features. The description will contain some characteristics of the shape of the elements and their organization as directional opposed to nondirectional. It will fail to detect some detailed description of the shapes of elements; as well the Fourier technique cannot be very local. So the spatial technique can complement the Fourier technique, being more local and therefore more accurate in some sense.

All this concerns black and white pictures. In colored pictures, each point is represented by at least a three-dimensional real vector; the coordinates could represent either the brightness through red, green and blue filters (possible other filters), or their normalized values (R/R + G + B, G/R + G + B, B/R + G + B), or perhaps the chomatic triple (hue, brightness, saturation).

It appears that color is a local property, meaning that the color is determined by local contrast (with global constancy judgment). The Fourier transofrm is an integral operator, that mixes up different local properties consequently. Direct application of the Fourier texture operator on an area is not useful for color in the general case, however, under certain constraints, one can suggest some applications of the Fourier operator on colored textures.

The simplest case is when the color is constant and the texture is encoded in the brightness function. Examples are grass, water, brick wall, etc. In this case the Fourier operator is used in the same way as in the black and white picture.

The second case is when the texture is formed by only two alternating colors. Here, let us assume the representation of the color of a point as a vector whose coordinates will contain the brightness of the point taken through red, green and blue filters respectively.

Since we have only two colors, clearly, the brightness functions will be correlated or anti-correlated with each other. Fourier analysis of functions could give a reasonably good description of the texture in terms of the contrast of the color components. This is analogous to spatial domain analysis.

This discussion points out a crucial weakness of Fourier transform techniques in dealing with color.

3.3 Concrete Texture Descriptors: Local Descriptors

According to our theme, a texture is characterized by a structure of texture elements and their spatial distribution. Each descriptor is associated with a procedure and a set of geometric measures.

The descriptors may be derived from parameters that come from the spatial and/or Fourier domain. In fact, often we will have to deal with two different measurements of the same parameters (e.g., length, width, direction), one performed in the spatial domain and the other in the Fourier domain. Here we seek a common interpretation of these measurements.

The input data from which we derive the local Fourier parameters is the power spectrum of the picture over every window. Since we are able to describe only what we measure, the technique that we implement will determine the system of descriptors we can use. In particular, the technique of Fourier analysis leads to the following system of descriptors: monodirectional, bidirectional, blob-like, homogeneous, and random. Using the input data of a local area one may expect to have more than just one descriptor.

Next we discuss the particular types of local descriptors we shall be using in our work.

(a) Monodirectional Texture

In the spatial domain, a monodirectional texture is approximately invariant along some direction. An example of monodirectional texture is a system of parallel stripes. In the Fourier domain the spectrum is approximately zero along the direction of near invariance, and is concentrated along the direction normal to that. We take this description to be adequate for spatial domain elements with some curvature or superimposed on a non-directional background. It makes sense to describe as directional a spectrum in which the dominant energy is along one direction, and where the directional peak is narrow.

Next we proceed to give a qualitative description of an algorithm that provides monodirectional descriptors. As alluded to above, this algorithm is based on the assumption that the texture will show concentration of energy in a certain direction of the Fourier domain. Thus we want to find a peak in the function of energy vs angle. This function is a sum of energies over a fan with a certain angle γ and direction φ. Remember that the data structure is a matrix, and thus only four directionalities (horizontal, vertical, and the diagonals) coincide with the matrix unit invariant direction. The fan technique permits one to include also the points near the investigated direction. The peaks of the function are defined as its local maxima, greater than the average value of the directional energy function. The width of a peak is defined as the distance between two consecutive zero crossings of the directional energy function minus the average value. The algorithm used two additional parameters, namely, γ and $\frac{E_{dir}}{E - E_{dir}}$, where the latter must be greater or equal to 2, while the former should not be greater than $\frac{\pi}{10}$. The angle γ is the measure of width of the peak and its threshold value corresponds to the limit of a useful directional description. E_{dir} is the energy in the angular stripe (fan) and E is the total energy. Its threshold corresponds to the condition that the ratio length/width must be at least two.

The algorithm determining the descriptor derived in the Fourier domain is given below:

Algorithm Monodirectional:

(1) Form a function $P_r(\varphi) = \sum_{r=1}^{WD/2} P(r,\varphi)$, where WD is the window size.

(2) Find the number n of peaks of the function $P_r(\varphi)$.

(3) If $n = 1$, then check the magnitude of the peak

$$\max_{\varphi} (P_r(\varphi)) = E_{dir},$$

and go to step (4) else mark the window by message: "There is more than one direction, do further analysis", and go to the end.

(4) If $E_{dir}/(E-E_{dir}) \geq 2$, then check the width of the peak which corresponds to the angular strip γ and continue in step (5) else mark the window by the message: "There could be blob-like or a noisy texture here, do further analysis", and go to the end.

(5) If $\gamma \leq \pi/10$, then mark the window: "Monodirectional texture" and go to the end else mark the window: "It is a monodirectional texture with nondirectional components" and go to the end.

(6) End.

The power spectrum along the direction of maximum power is the power spectrum normal to the invariant direction. In the spatial domain humans characterize these profiles by step functions.

In the Fourier domain we can find the approximate wave length of parallel strips (distance between two neighboring stripes), and the width of stripes from our previous analysis. One way of identifying width in the spatial domain would be to use one-dimensional interval analysis along a direction. This technique could be used also for more precise localization of monodirectional textured edges than one can achieve in the Fourier domain. The interval analysis method has not yet been implemented.

The above algorithm has been implemented and tested on examples. A sample is shown in Fig. 14 and Fig. 15.

In Fig. 14 we have a texture of parallel lines and in Fig. 15 we have a texture of parallel strips (wood grain). In both figures the upper left pictures show the original textures, divided into four windows (each window is of size 32 by 32 points). The pictures in the upper right corner are resynthesized textures, produced according to the description. The pictures in the lower left corner show the power spectrum of the original textures. Note the two different directionalities in the lower quadrants of the picture. Here the diagonal directionality corresponds to wood grain pattern and the vertical directionality represents the shading effect (slow changes in brightness).

(b) <u>Bidirectional Texture</u>

The descriptor "bidirectional" is associated with two sets of monodirectional stripes, described in the monodirectional texture. This description belongs to the spatial domain and does not have a unique Fourier counterpart. In terms of the power function vs. angle $P_r(\varphi)$ it corresponds to two distinguished peaks of $P_r(\varphi)$, while the converse is not true.

If function $P_r(\varphi)$ for φ from $<0,\pi>$ has two distinguished peaks, then it could represent at least one of the following two cases in the picture (its window):

(α) two different directional textured subregions are adjacent (are next to each other) in one window, or

(β) two different directional regions are superimposed (one is on the top of the other).

The problems discussed above are shown in Fig. 16 and Fig. 17 and 18 where the pictures in Fig. 16 show the case (α) and the picture in

Fig. 17 and 18 exhibits the case (β).

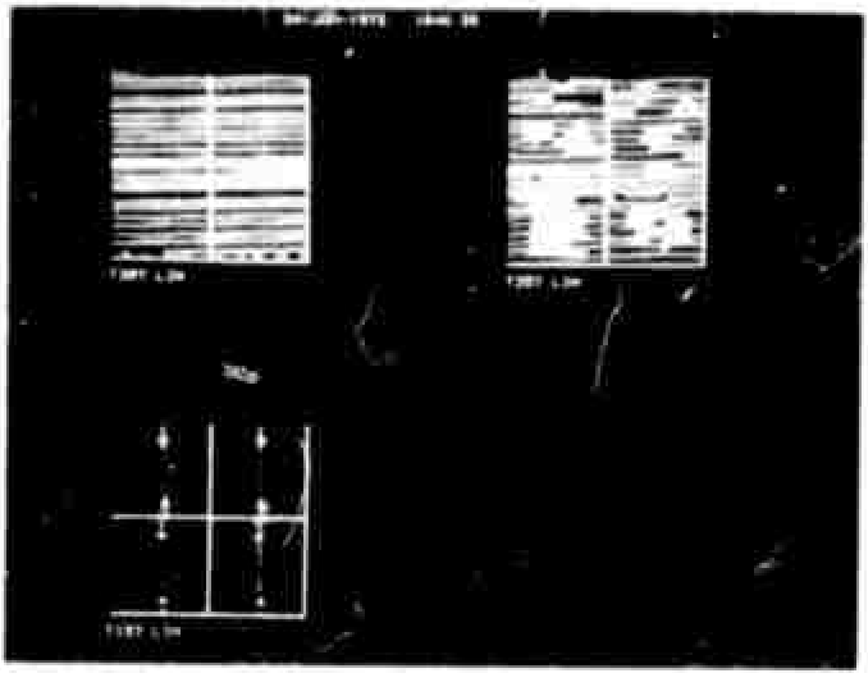


Fig. 1

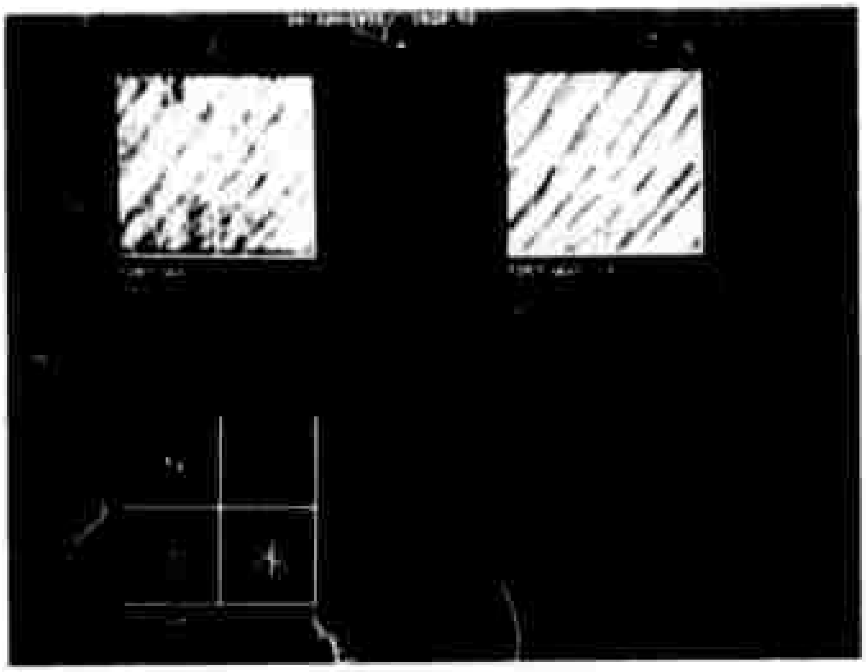


Fig. 1

Each of the figures displays nine pictures whose meaning is explained in the table below, where the row and column numbers refer to particular pictures in Fig. 1 .

Row	Column	Description of Pictures in Fig. 16
1	1	Four windows, where each contains horizontal and/or vertical stripes.
1	2	Picture <1,1> after directional filtering process, performed in every window separately.
1	3	The "complement" of <1,2>.
2	1	The power spectrum of <1,1>.
2	2	The phase spectrum of <1,1>. Here the phase is transformed from the range $<-\pi,\pi>$ to $<0,2\pi>$.
2	3	The absolute values of the phase spectrum of the picture <1,1>.
3	1	The power spectrum of picture <1,1> parametrized by the absolute value of the phase in range $<0,\pi/3>$.
3	2	The same as in <3,1> but this time with range $<\pi/3, 2\pi/3>$.
3	3	The same as in picture <3,2> but with range $<2\pi/3,\pi>$.

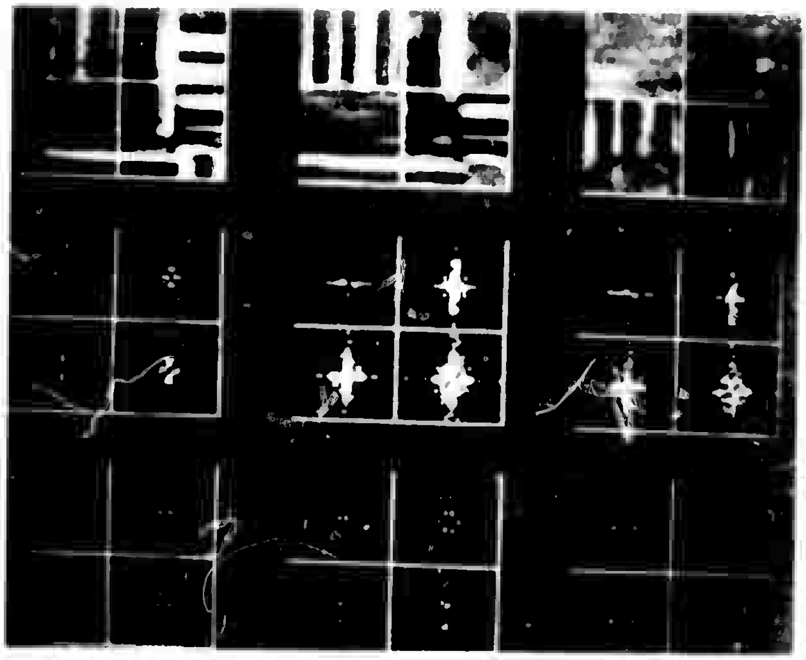


Fig. 1

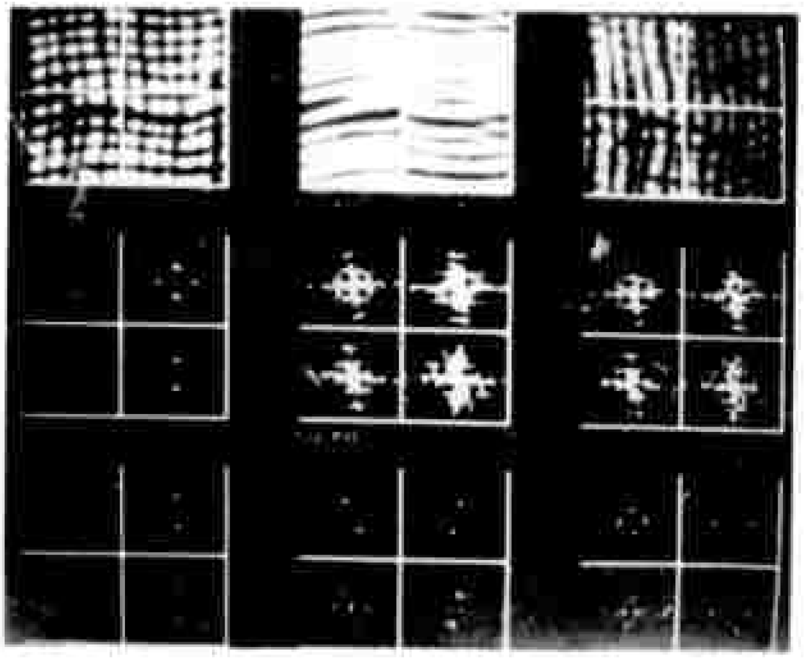


Fig. 17

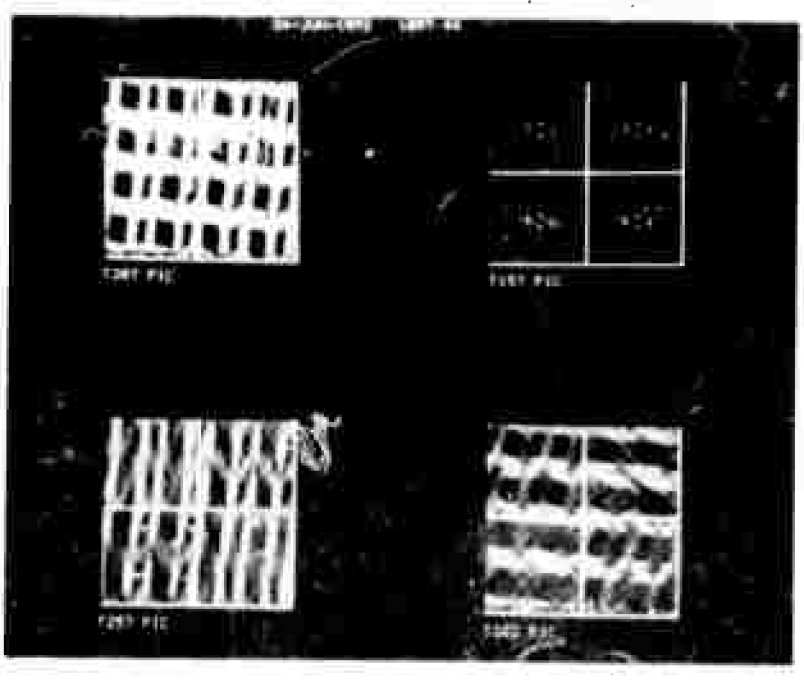


Fig. 1

The description of pictures in Fig. 17 is the same as that of pictures in Fig. 18, except in Row 1 and Column 1, where we have four windows, each containing a superposition of horizontal and vertical lines.

Let us concentrate for a moment on the windows of the first and third quadrant of picture <1,1> in Fig. 17. Each of the windows is a composition of horizontal line textures and vertical line textures. It is impossible to distinguish the cases of separate (α) from overlap (β) in the power spectrum. Using the phase spectrum one would hope to separate the region containing the horizontal lines from the region containing the vertical lines, or one would at least hope to be able to identify their positional relationships. Unfortunately, it is not known at present how to carry out the separation. To our knowledge, no one has yet used the phase spectrum in a meaningful way.

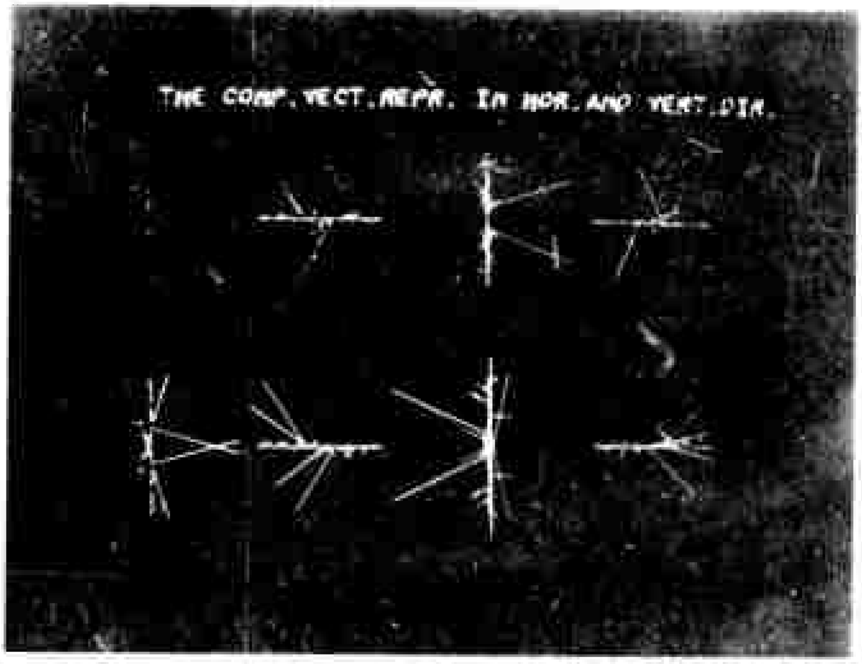


Fig. 19

The figures in Fig. 19 show the vector display of the complex function $F(n,m)$ in a direction λ; in our case it is in horizontal and vertical direction for each $\langle n,m \rangle$. The direction of the vector is equal to the phase, and the length of the vector corresponds to the value of the power. As one can see from the pictures, there is no evident distinctive feature which would describe the relationship Left-Right or Right-Left.

We have shown above that in spite of the nonuniqueness of the representation of bidirectional textures in the power spectrum, using decomposition techniques, one can construct a suitable algorithm for identification purposes.

We shall soon give such an algorithm. However, before we do that, we want to point out that the domain of validity of the parameters associated with this descriptor is given by the domain of validity of the parameters used for monodirectional textures, except that the lower and upper boundary of d is now changed from $\langle 0,\pi \rangle$ to $\langle \gamma,\pi - \gamma \rangle$. Moreover, the peaks are defined in the same way as was done in the monodirectional algorithm.

<u>Bidirectional Algorithm</u>:

(1) Find the number of peaks (n) of function $P_r(\varphi)$. If $n = 2$, then find the corresponding directions φ_1 and φ_2 for each peak, else write the measage: "This is not a bidirectional texture, do further analysis", and go to the end.

(2) If $9\pi/10 > \mathrm{Abs}(\varphi_1 - \varphi_2) > \pi/10$, then go to step (3), else write the message: "This is a deformed monodirectional texture" and go to the end.

(3) Partition region R into four equally large subwindows R_1, R_2, R_3, and R_4.

(4) Check each subwindow R_i for $i=1,\ldots,4$, whether it is a bidirectional textured region or not, using the algorithm "Bidirectional". If the answer is yes, then set MR_i On, otherwise set MR_i Off.

(5) If MR_i are On for all $i=1,\ldots,4$ then describe the given window (region R) as a "bidirectional texture" and go to the end, else go to step (6).

(6) If MR_i are Off for all $i=1,\ldots,4$ and all the subwindows are monodirectional, then describe the corresponding region as "Two monodirectional textures with different directions are adjacent" and go to the end, else issue the message: "Further texture localization is necessary" and go to the end.

END.

(c) Blob-like Texture

This descriptor is associated with blobs and nonlinear distribution. It should be noted that these two components go together to the effect that it is not sufficient to have blobs as texture elements for inferring a blob-like description. For instance, blobs on a grid would show a directional texture, and the 'bloblikeness' will be very weak.

In the Fourier domain, blobs are represented by a concentric energy distribution. An annulus with the greatest energy value is the peak annulus. In the implementation of the description (see the algorithm below) we approximate areas of the transform by circles. The radius of the approximating circle is inversely proportional to the radius of the approximating circle in the spatial domain.

It would seem logical to pass from mono- and bidirectional textures to tri-, tetra-, ..., n-directional textures, before turning to blobs. However, it is very hard to interpret these higher order directionalities in the spatial domain.

The blob-like algorithm describes blobs and their nonlinear distribution. It is based on the assumption that patterns which do not have directionality, noise, nor homogeneity, are some sort of blob-like textures. In the Fourier domain this assumption corresponds to two conditions. First, $P_r(\varphi)$ is constant and, second, $P_\varphi(r)$ is not constant.

<u>Algorithm Blob-like</u>:

1. Form functions $P_r(\varphi) = \sum_{r=1}^{WD/2} P(r,\varphi)$ and

$$P_\varphi(r) = \sum_{\varphi=0}^{\pi} P(r,\varphi)$$

and then compute their respective mean values

$$M_r = \frac{2}{WD} \sum_{\varphi=0}^{\pi} P_r(\varphi) \quad \text{and,}$$

$$M_\varphi = \frac{2}{WD} \sum_{r=1}^{WD/2} P_\varphi(r) \quad ,$$

where WD is the window size.

Next compute their variations

$$v_r = \frac{2}{WD} \sum_{\varphi=0}^{\pi} (P_r(\varphi) - M_r)^2 \quad \text{and}$$

$$v_\varphi \quad \frac{2}{WD} \sum_{r=1}^{WD/2} (P_\varphi(r) - M_\varphi)^2$$

2. If $M_r > CN$ and $M_\varphi > CN$, then go to step 3 else print the message: "The structure is on the level of the camera noise" and go to the END.

3. If $v_r > CN$, then go to step 7 else print the message: "All energies are equally distributed in every direction", and go to step 4.

4. If $v_\varphi > CN$, then go to step 5 else print the message: "It is a noisy texture" and go to the END.

5. Find Max $P_\varphi(r) = P_\varphi(r_{max})$.

If $r_{max} \leq 2$, then print the message: "There is only one texture element"; go to step (6).

6. Form a new discrete function $I(i)$ from $P_\varphi(r)$ in the following way:

Assume that $P_\varphi(r)$ is a combination of sinc(r,φ) type functions. Find all the local maxima and all the minima of the function $P_\varphi(r)$. For every local maximum $r_{max\ i}$, there are two surrounding minima r_{1i}, r_{2i} such that

$$r_{1i} < r_{max\ i} < r_{2i}, \text{ where}$$

$$I(i) = \sum_{r = r_{12}}^{r_{2i}} P_\varphi(r).$$

If $I(i)$ is a convex function, then print the message: "Texture elements are blob-like" and go to step 7, else print the message: "There is an unidentifiable texture" and go to END.

7. Assume that $P_r(\varphi)$ is a combination of sinc functions. Find all the local maxima of the function $P_r(\varphi)$ and if their number is greater

than 2 , issue the message: "The blob-like texture has some directional features". else print the message: "There is no blob-like texture".

8. END.

The above algorithm has been tested on an example shown in Fig. 20 which should be self-explanatory.

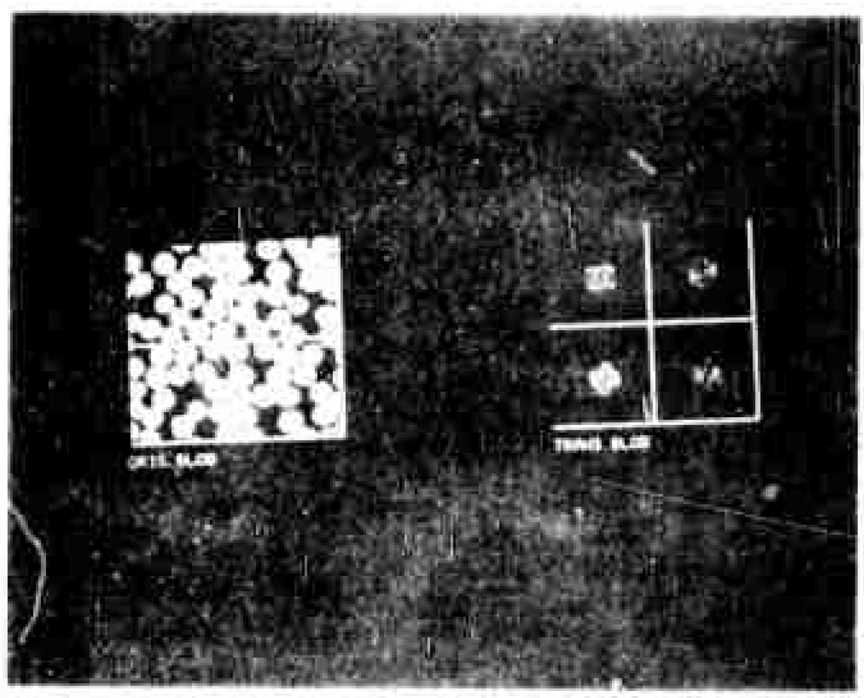


Fig.

(d) Noisy (Random) Texture

The spatial case:

A random distribution of dots (pepper and salt pattern) forms a model of noisy texture. This model describes the random spatial organization of dot - texture elements versus periodic or regular distribution of texture elements.

Fourier Case:

The texture in this model corresponds to a homogeneous distribution of energies in the power spectrum.

Descriptor "noisy" is associated with certain parameters (obeying some threshold constraints), explained below:

EL will denote the ratio of the size of the one-dot-texture element and the size of a real texture element. The inequality $EL \leq WD/4 + CN$ means that a texture element with area of two dots and WD (window size) of 8 x 8 points will still be a dot-texture element.

ED is the parameter of random distribution. ED is the ratio of

EM and M_r , where

$$EM = \max_{\varphi} \text{Abs}(P_r(\varphi) - M_r) \quad \text{and} \quad M_r = \frac{2}{WD} \sum_{\varphi = 0}^{\pi} P_r(\varphi).$$

The value of ED is set to be $ED \leq 0.1 + ON$.

CN is the noise of the TV camera.

Algorithm Random:

1. Form functions $P_r(\varphi)$, $P_\varphi(r)$, M_r, M_φ, v_r, and v_φ as they were described in the algorithm blob-like.

2. If $M_r \leq CN$ or $M_\varphi \leq CN$, then write the message: "The texture

structure is on the TV camera noise level, check if it is a homogeneous texture" and go to END, else go to step 3.

3. If $v_r > ED$, then write the message: "There is no random distribution" and go to END, else go to step 4.

4. If $v_\varphi > EL$, then write the message: "There might be a blob-like texture" and go to END, else write the message: "The texture is a randomly distributed dot pattern".

END.

The algorithm has been tested by an example shown in Fig. 21 . The picture in the left upper corner is a texture of sand; the picture in the right upper corner is the resynthesized image of the original, according to the description. Finally, the left lower corner shows the power spectrum of the original.

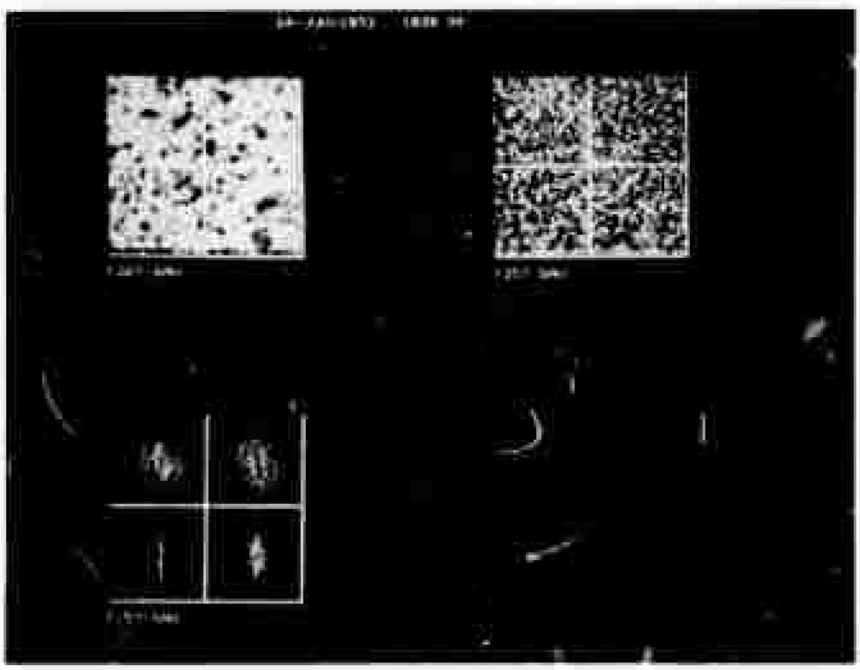


Fig. 21

(e) Homogeneous Texture.

A homogeneous region of uniform brightness (in black and white picture), color (in colored picture) forms a model of a homogeneous texture.

The Fourier counterpart is represented by a Dirac function with its center in the zero point of the coordinate system.

The only threshold parameter in this model is the level of the TV camera noise (CN).

Algorithm Homogeneous

1. Form a function

$$\text{noise}(r,\varphi) = 2 \sum_{\varphi=0}^{\pi} \sum_{r=0}^{WD/2} P(r,\varphi) - P(0,0)$$

P(0,0) is also called the DC value.

If noise $(r,\varphi) \leq$ CN then write the message: "The texture is homogeneous" else write the message: "The texture is not homogeneous".

END.

An additional parameter - the average value of the intensity of light over a particular window is associated with every description of a homogeneous texture.

The following table summarizes the texture descriptors we have implemented:

Table 5

DESCRIPTORS	PARAMETERS
Monodirectional	DIR - direction of lines ww - distance between two parallel lines γ - a measure of straightness of a line $E_{DIR}/(E-E_{DIR}) = t$ t - a measure of "thickness" of a line
Bidirectional	DIR_1, DIR_2 - directions of lines l_1, l_2, respectively. ww_1, ww_2 - distances between two parallel lines in two different directions DIR_1 and DIR_2. γ_1, γ_2 - measures of straightness of lines l_1, l_2. t_1, t_2 - measure of "thickness" of lines l_1, l_2, respectively. $d = DIR_1 - DIR_2$ Comment: lines l_1 and l_2 are assumed to be nonparallel.
Bloblike	R - the distance between two texture elements in direction DIR.

Table 5 (Continued)

DESCRIPTORS	PARAMETERS
Random (Noisy)	EL - a measure of "dotness" of the pattern. ED - a measure of random distribution $(M_r + M_\varphi)/2$ - the mean value of "constantly" distributed energies.
Homogeneous	noise (r,φ) - a measure of the degree of variation of the "homogeneous" area. DC - the average value of the intensity of light over window W.

4. COLORED AND TEXTURED REGIONS

In the previous chapter, we discussed procedures for texture descriptors. This chapter describes the determination of textured and colored regions and introduces a mathematical description, topological sheafs, to formalize the region-forming process. The texture descriptors are used to form regions with similar descriptors. The region-growing is low-level in that it does not use the context of a world model. It is intended as a tool for higher level routines. The proposed regions function as initial guesses about important areas of the image. Thus, the routines favor large regions at the expense of smaller regions, a sort of "law of the fishes", the big ones eat the smaller. Since there are few useful texture descriptors and organization procedures, this attention to low level modules was a necessary focus for our research.

The informal distinction between low-level and high-level processes refers to the context which the process takes into account. Roughly, we mean low-level when the context is local and based on the image, and by high-level we mean an object space interpretation which depends on several levels of abstraction and relations (global). We would like a larger armament of texture descriptors and low-level organization mechanisms. However, it is important to have a balance between the low-level and higher-level systems, and to design for their communication.

We emphasize that we use a technique where we start from large windows and take smaller windows at boundaries. This approach has a limitation of missing substructure. The microscopic approach of starting from small windows and trying to piece together global structure has a complementary weakness of missing global order. Overall, we think this points up the need for a range of sizes for local organization. For texture, we prefer our approach to the microscopic one.

As the scanner traverses across the picture in a television-like raster scan, the local texture descriptors (these descriptors might be spatial, or histogram, etc., in addition to those we use)over each window are sent to the program which detects the appearance of similarities or dissimilarities of the structures, over the given pair of windows. The knowledge of the existence of similarities is retained together with locations. All the windows with similar structures are joined together by a two-way list which is constructed during the scanning process. The program also detects the break of similarities between two structures and gives a command to the scanner to scan with windows of smaller size.

When two windows are joined or split apart, different texture names are assigned to them. Each structure associated with a window is tested to determine its similarity with other structures or its proper association with the existing similarity classes.

Region boundaries do not usually coincide with the grid windows, and hence there occurs both merging of two adjacent areas, and the splitting of an area into at least two portions.

In this work a set of real life and artificial pictures was scanned and processed by our program to demonstrate the capability of the implemented Fourier method. The results of testing indicate that our method is capable of decomposing pictures into regions, where each region corresponds to a different texture or color.

In our implementation of region growers, the emphasis was on testing some of the ideas and not on the efficiency of programming. However, for illustration we present in Table 6 the average time and memory load for our programs. The programs have been implemented on a PDP-1(, at the Artificial Intelligence Project, Stanford University.

Table 6

NAME AND FUNCTION OF THE PROGRAM	SIZE OF THE PICTURE	CPU TIME (min)	CORE (k)
FANAL.SAI FAST FOURIER TRANSFORM AND SEGMENTATION	256 x 128	4.15	39
TEXTUR.SAI TEXTURE ANALYSES ON WINDOWS (32 x 32) POINTS	256 x 128	2.07	34
MIKROA.SAI LOCALIZATION TEXTURE ANALYSIS OF WINDOWS (8 x 8)	256 x 128	12.8	27
TREE.SAI TEXTURE REGION GROWER	256 x 128	8.0	36
COLOR.SAI COLOR REGION GROWER	192 x 128	2.0	22

4.1 An Algorithm for Finding Regions

The process of localization of structures was described in detail in Section 3.3. Here we shall focus our attention on finding the connections between local structures in terms of continuity, discontinuity, and proximity. The actual job to this effect is carried out by a region grower that we shall describe momentarily. The region grower can be used both for continuous textured regions and continuous colored regions. The algorithms for our region grower use the principle of local constancy whose content is summarized in the phrase: "Unite connected locally similar areas into one global one." Our algorithm uses the notion of a cell which is nothing but an arbitrary window of the smallest possible size, carrying meaningful information.

Algorithm "Region Finder"

1. Set regional index i to 1 and produce a mark R_i.

2. Take the first untested cell and call it the first pilot cell (which thereby is also a pilot cell).

3. Set XSIDE to be RIGHT SIDE, YSIDE to be LEFT SIDE, and XADJ to be RIGHT ADJACENT.

4. If the pilot cell has been tested for its XADJ cell, then go to step 8, otherwise mark the pilot cell by a mark signifying the fact that it has been tested on its XSIDE, and continue in step 5.

5. Find the next XADJ cell. Ask whether this new cell does not exceed the size of the picture and has not been tested on its YSIDE. If the answer is NO, continue in step 6, else go to step 8.

6. If the pilot cell and the adjacent cell are similar, then continue in step 7, else mark the pilot cell on its XSIDE, indicating that it has been tested, and go to step 8.

7. Join the two cells (pilot cell and the new cell), mark the new cell by a mark R_i and indicate the fact that it has been joined on its YSIDE. Store the new cell in an array of new cells. Make the cell a pilot cell and go to step 8.

8. If XSIDE is the RIGHT SIDE, then set XSIDE to be LEFT SIDE, YSIDE to be RIGHT SIDE, and XADJ to be LEFT ADJACENT and go to step 9. If XSIDE is the LEFT SIDE, then XSIDE is set to be the UPPER SIDE, YSIDE is set to be LOWER SIDE, and XADJ is set to be UPPER ADJACENT, and go to step 9. If XSIDE is the UPPER SIDE, then set XSIDE to be LOWER SIDE, YSIDE to be UPPER SIDE, and XADJ to LOWER ADJACENT, and go to step 10.

9. Set the pilot cell to be the first pilot cell and go to step 4.

10. Take the array of new cells. Take the index j (initially j = 0) and increase it by 1. If j exceeds the number of all new cells, then go to step 11, else take the element n_j from the array of new cells and make it the first pilot cell and go to step 3.

11. Zero the array of new cells. If there is any cell in the picture that has not been yet tested, then increase the index of regions i by 1, make a new mark R_i and go to step 2, else go to the end.

END.

4.2 Texture Regions

This algorithm has been tested on textured regions as well as on colored regions. The scanning process is, for instance, shown in Fig. 22 with white squares, each representing windows of 32 x 32 points. Fig. 22 displays the boundaries of different textured regions of the picture shown in Fig. 23, after the first pass. One can see the different sizes of windows.

Over every window there are several descriptors and parameters. Since we used several window sizes (32, 16, 8) and some of the parameters are size dependent, we reduced all descriptors and parameters to the smallest window size (8). Then the criteria of similarity had to be set.

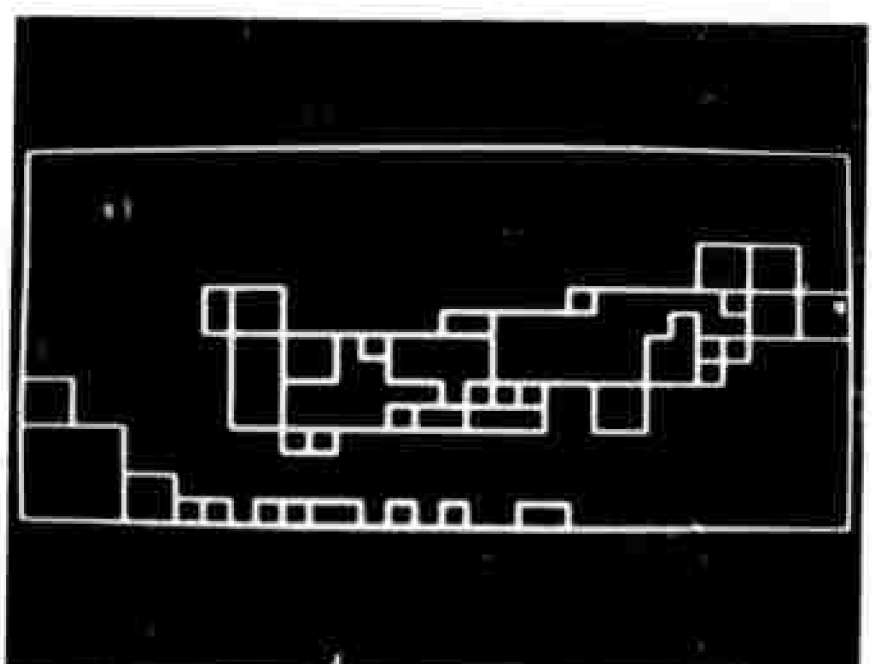


Fig. 22

The criteria of similarity are set by the higher level program. In our work we used two approaches, not exclusive but rather complementary.

One approach used only black and white pictures and did not assume any previous knowledge about the scene. The similarity criteria were determined by the camera noise and expected error of the method. The whole region growing was based only on the similarities of certain geometric properties described by the Fourier texture operator. The results of this approach are displayed in Fig. 24 and 26, where one can see that while this approach is sufficient for separating regions on simple, more or less artificial scenes (the rastered cube on Fig. 23, the cube on a grid surface in Fig. 25, it is not adequate for finding boundaries of regions of real outdoor scenes. In the latter case one needs to know more about the scene and thus conduct a directed texture region growing or texture boundary detection.

The directed texture region growing and/or boundary detection is the other approach that we used. It uses information gained through a color region grower. This information directs the application of the textured operator for two purposes:

One is to look for a common texture where the colors are the same or proximal. The other is to look for texture differences where there are colored boundaries.

This approach identifies more efficiently the real regions and their boundaries. The example in Figure 27 shows the different textured regions of the original picture displayed in Fig. 1. Most of the grass region came out as directional texture. Only two areas (one on the left side and the other on the right side) within the grass region were identified as noisy texture, though with the same direction as the directional textures. It requires further verification of the continuity in those two textures in order to remove the boundaries.

The main difference between the two approaches is that in the latter we use the texture operator in a directed way. This means that as well as applying the texture operator only in certain areas (not all over the picture), we also have the choice of asking for continuity and proximity in several descriptors and parameters independently.

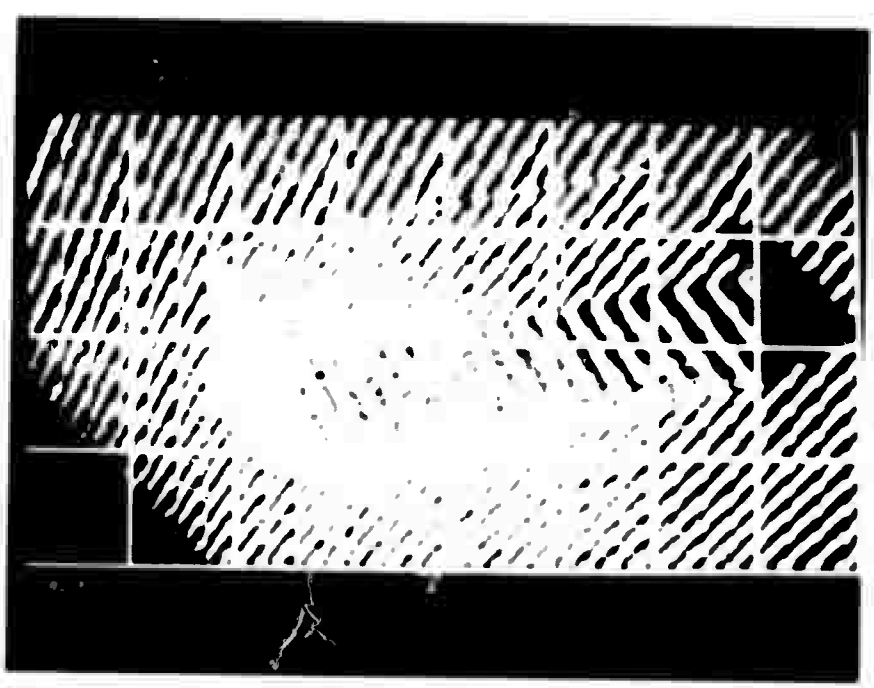


Fig. 23

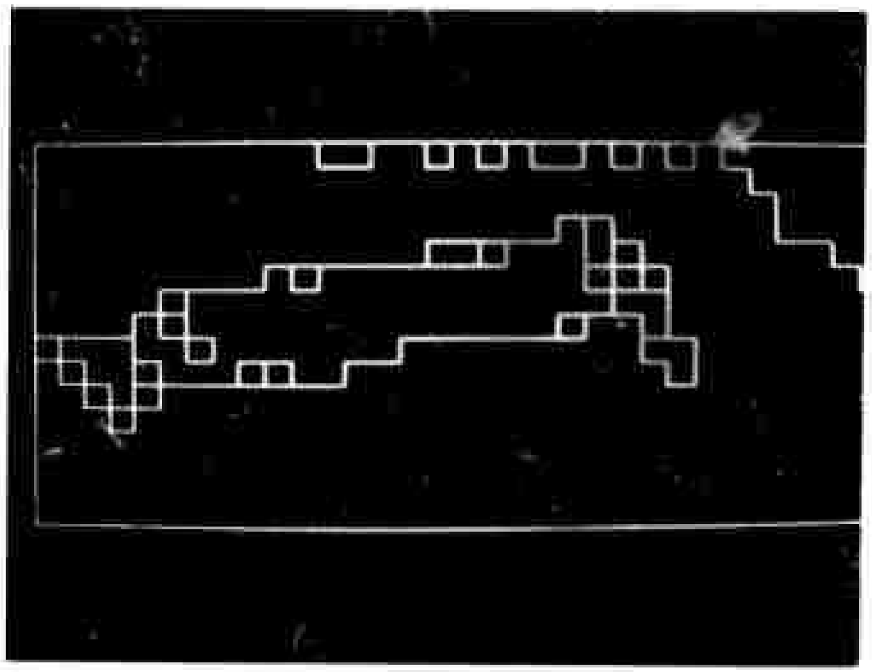


Fig. 24

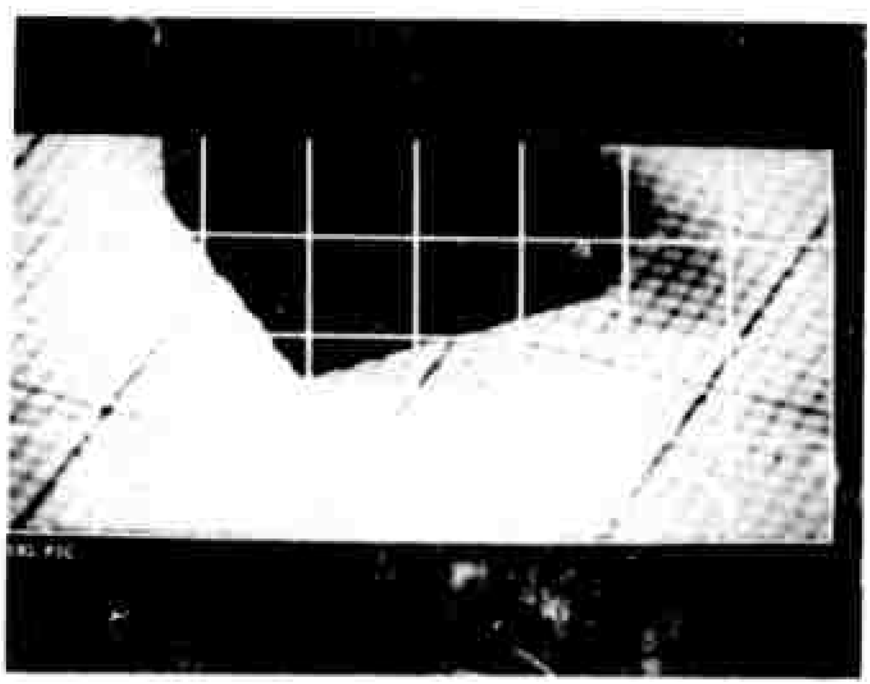


Fig. [illegible]

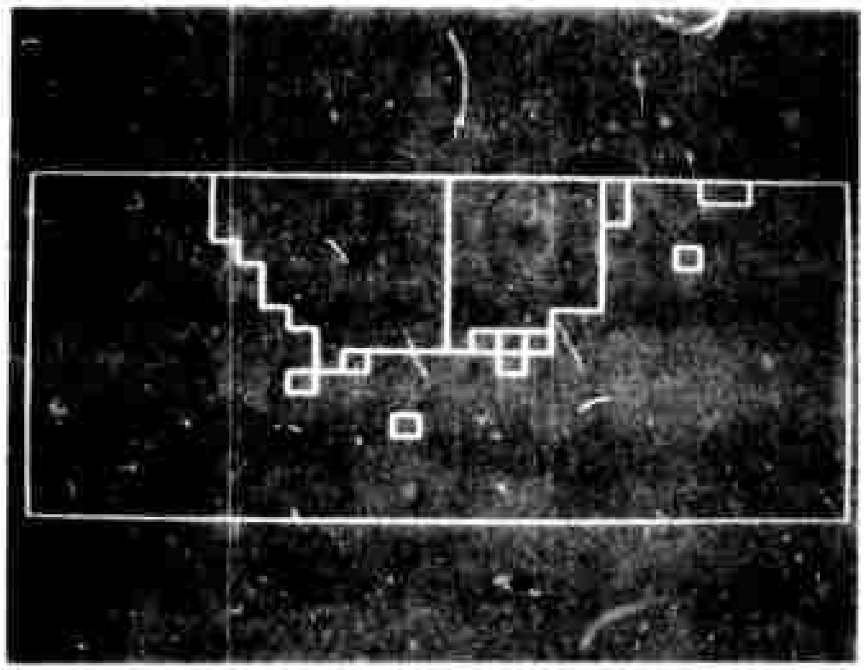
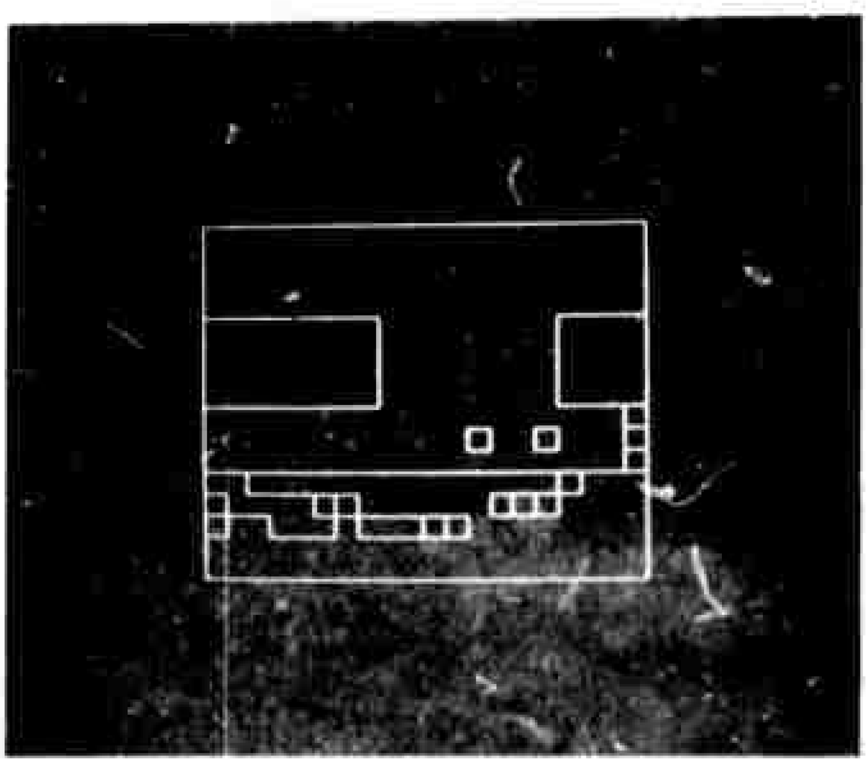


Fig.



4.3 Color Regions

Similarly, as for textured regions, the region growing algorithm has been used for colored regions. The colored picture consists of three files, each representing the brightness through red, green and blue filters. We use the normalized values of color for each point (e.g. $R/R + G + B$, $B/R + G + B$) where R,B,G are the intensities through the red, blue and green filters respectively. As in the texture region grower, here we use again windows over which the average values of $R/R + G + B$ and $B/R + G + B$ are computed. The size of the windows depends on the structure of the picture we have chosen (8 x 8). The windows are overlapped, so that continuity is checked strictly. The threshold value that determines the similarity criterion depends on the resolution of the picture as well as on the window size. In our case, it is set to 2, provided that we deal with 6 bit pictures. The example in Fig. 44 shows the result of the above described color region grower, applied on the picture in Fig. 1. The original picture is only 4 bits resolution, so the threshold has to be different (0.75). Otherwise every thing is the same.

4.4 A Sheaf-Theoretic Point of View of Finding Regions

The geometric analysis of pictures, in particular, partition of a picture into regions, can be neatly presented in the language of sheaves (for details see the APPENDIX). From a sheaf-theoretic point of view, the region identification process is based on an assignment of structures to windows (the local structure) and on passing from LOCAL STRUCTURES (over windows) to GLOBAL ONES (regions). Thus, each region is specified by one sheaf. Over every window, we can have several different descriptors, thereby different structures. Each of these structures will partition the picture in a different way. These different partitionings of the picture, described by different sheaves, correspond to the different layers of description of the picture. Naturally, the sheaves could be interconnected through some connecting mappings. The difficulty in making use of the structure of sheaves in scene analysis is that we usually do not know the connecting mappings between two different sheaves.

The sheaves constitute a vehicle for checking the continuity and proximity of structures with respect to some well defined connected mapping. In a concrete application of a texture region grower, this mathematical tool has the following limitations:

(i) If the structure is a texture, then it will find the continuity in the texture, but it will find discontinuity in the texture element. Thus the smallest window size must be restricted to the size of the texture elements.

(ii) The sheaf-theory assumes that the structures over every two windows, which are in inclusion relationship, are related by a connected mapping. However, in reality the different positions of windows may cause false continuities or discontinuities. One has to do several different overlapped windowing in order to overcome this error.

The contribution of the sheaf point of view to region growing is that it defines precisely the conditions for continuity and discontinuity of a structure with respect to some connected mapping. The sheaf theory shows that if the structures from two (overlapped) windows and their overlapped part are connected by the mapping, then the union of these two windows is continuous with respect to the structure and the mapping. It is interesting that the sheaf conditions are similar to natural continuity conditions for use of the Fourier power spectrum.

In most of our applications (texture or color region grower), the connected mapping is the local similarity relationship (it must be an equivalence relation). Naturally, the theory allows much more complicated mappings as well as structures.

After this discussion let us present the sheaf-theory more formally. The topology we shall use is discrete and is induced by certain norms, taken from the structure to integers. Once the topology is fixed, we introduce a convenient system of neighborhoods, called windows. We think of windows as a system partially ordered by inclusion. Procedures which evaluate the data over the windows assign to every window a structure of descriptors. When two windows, say v and w, are in inclusion relationship $v \subset w$, the corresponding networks of descriptors N_w and N_v are related by a connecting mapping

$$\beta_v^w : N_w \to N_v$$

which essentially restricts the network over the bigger window to a network on the smaller window. Since the process of restriction is transitive, one obtains by this formalization a PRESHEAF associated with the image function

$$N = \langle N_{V}, \beta^{V}_{W} \rangle$$

Sheaves are presheaves satisfying additional axioms. A definition of a sheaf in its full generality requires several additional technicalities. A more direct definition of a sheaf with a fairly clear picture-theoretic interpretation is given below.

Thus, loosely speaking, a sheaf is a system of structures over a lattice of windows, where each structure represents one particular texture.

Consider a presheaf $S = \{S_V; \beta^{W}_{V}\}$ of structures over a cellular space X, i.e., on the lattice of subsets $\langle Sub(X), \subseteq \rangle$. Then S is a <u>sheaf</u> over X precisely when for any family $\{V_i | i \in I\}$ of subsets of X with $V = \bigcup_i V_i$, the following two conditions are satisfied:

(1) <u>Uniqueness axiom</u>: $\forall i[\beta^{V}_{V_i}(s') = \beta^{V}_{V_i}(s'')] \Rightarrow s' = s''$;

(2) <u>Existence axiom</u>: $\forall i,j[\beta^{V_i}_{V_i \cap V_j}(s_i) = \beta^{V_j}_{V_i \cap V_j}(s_j)] \Rightarrow$

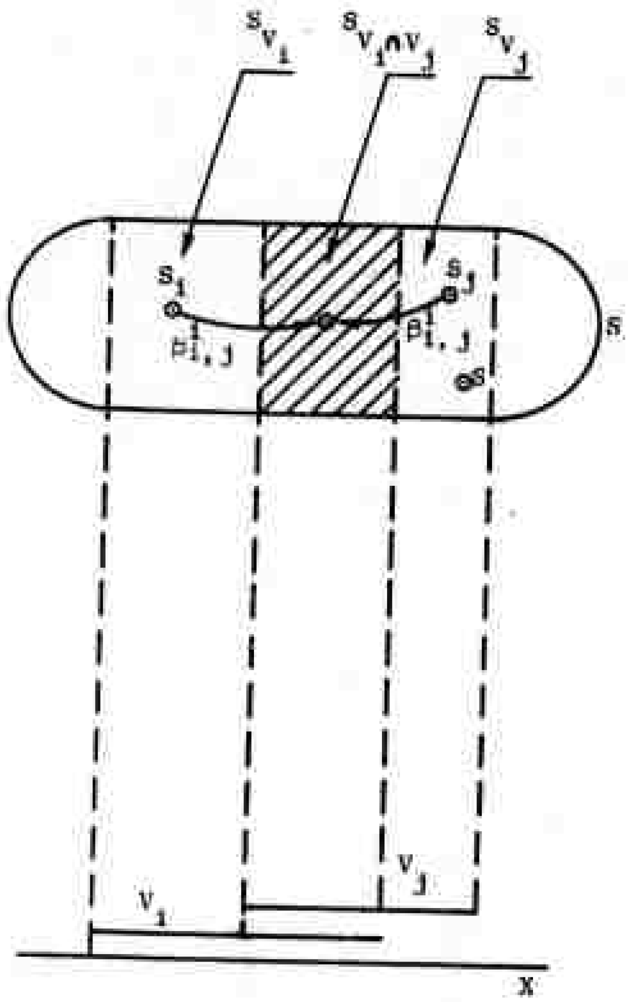
$s \forall k\ [\beta^{V}_{V_k}(s) = s_k]$,

where $s, s', s'' \in S_V, s_i \in S_{V_i}, s_j \in S_{V_j}, s_k \in S_{V_k}$, and $i,j,k \in I$.

The condition (1) says that if the structure elements s are locally identical, then they are also <u>globally</u> identical. That is elements are uniquely determined by <u>local data</u>.

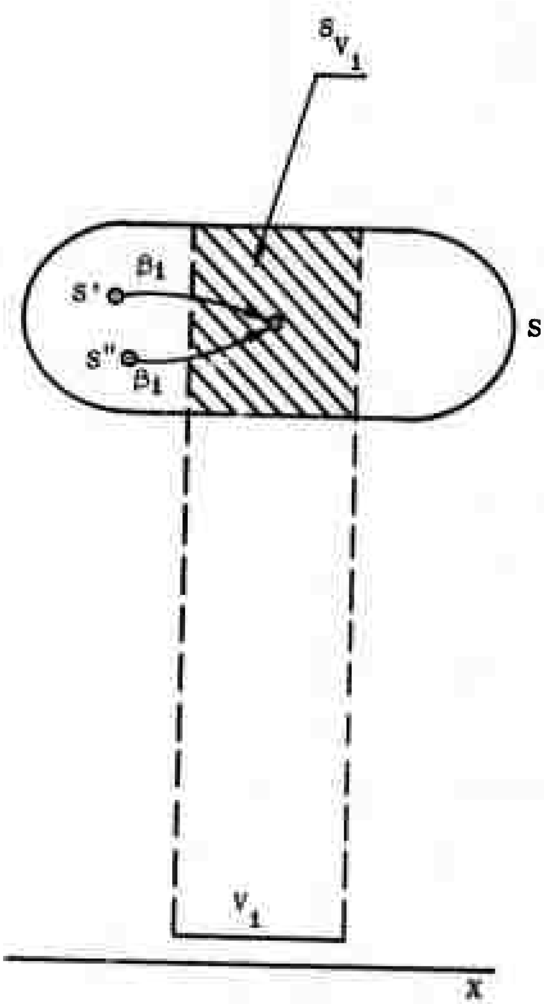
The condition (2) says that if we have <u>local</u> data which are compatible, they actually "patch together" to form <u>global</u> data.

The geometric meaning of axioms (1) and (2) is displayed below in Fig. 28 and Fig. 29.



S_{V_i}
$S_{V_i \cap V_j}$
S_{V_j}
s_i
s_j
$\beta^1_{i,j}$
$\beta^1_{i,j}$
S
s
V_i
V_j
X

Fig. 28



S_{V_i}
β_i
s'
s"
β_i
S
V_i
X

Fig. 29

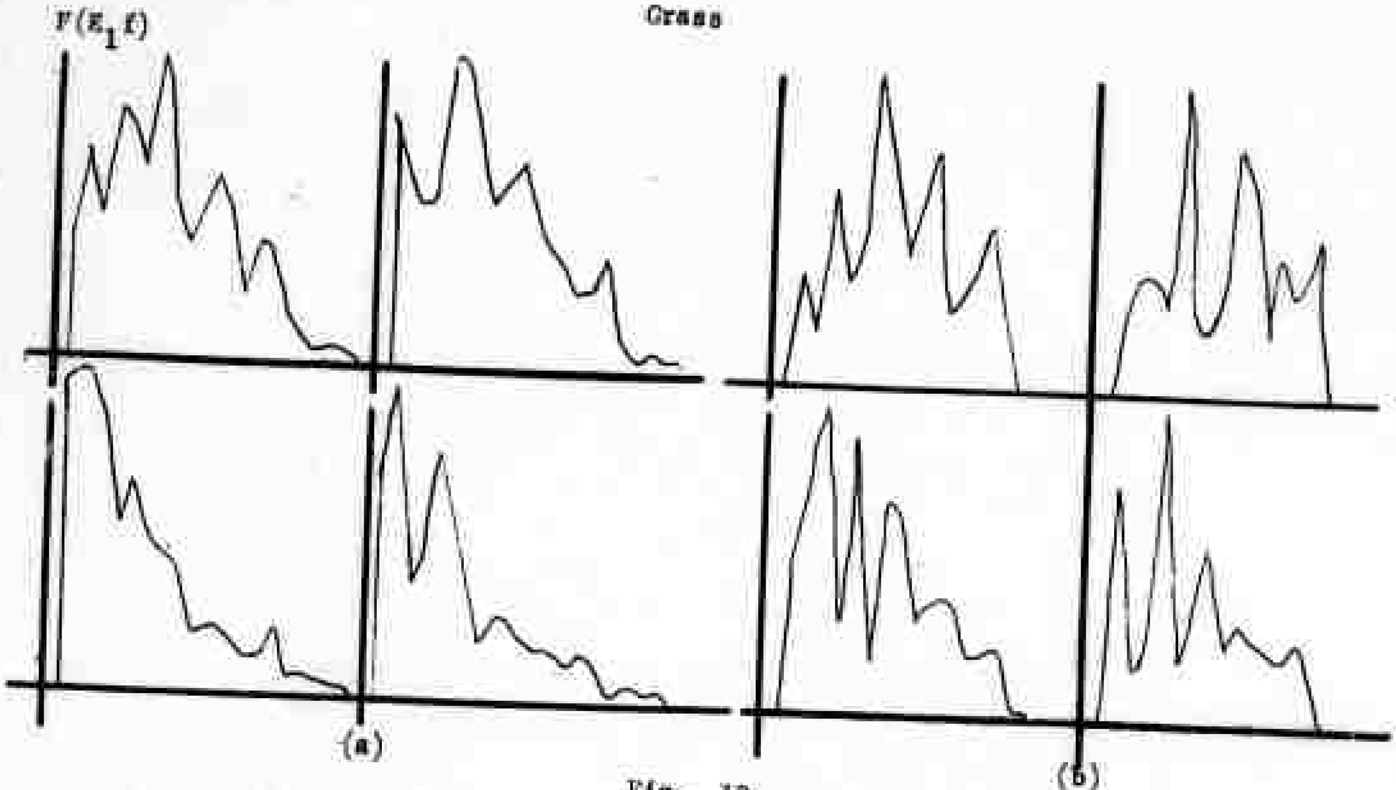


Fig. 30

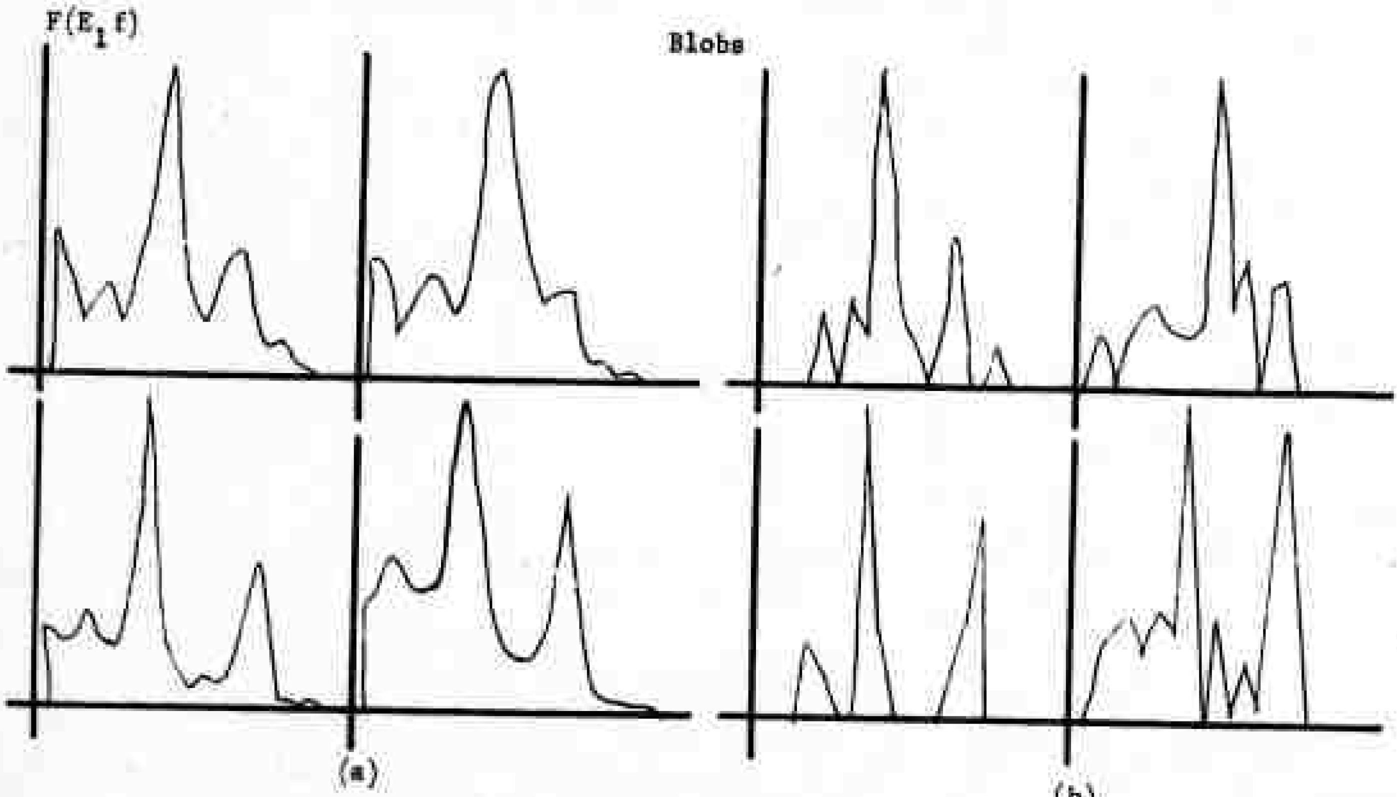


Fig. 31

100-a

5. INTERPRETATION OF OUTDOOR SCENES.

The main issue in this chapter is how to recognize and interpret real outdoor scenes of grass, water, sky , etc.

5.1. Pattern Recognition Approach.

In an early stage of our research, we tried to recognize texture using a pattern recognition method (Bajcsy, 1970). We computed a function of energy (E) along the frequencies (f) and derived a feature vector from this function. The features were the number of peaks, their energies, their width and their corresponding frequencies. In addition, we characterized the function as flat or with peaks. These features were used for classification of the texture into classes: grass, water, regular pattern (like blobs, brick wall) and unidentified. As an example, the grass and water had more flat function than the regular patterns. Samples of the function of the energy and the frequency of textures of grass, water, brick wall and blobs is displayed in Figs. 30 - 33.

Each picture consists of two graphs. One is the function (energy, frequency) computed in the window without any preprocessing (indexed by (a)), and the other is the same function as above computed from the data, which was preprocessed (indexed by (b)). Preprocessing, in the case of grass and water, was a high pass filtering. The purpose of the preprocessing was to eliminate the effects of shadows on grass or water.

For the regular patterns, the preprocessing consisted of a low pass filtering. The purpose of this filtering was to enhance the main frequency components of a regular pattern and suppress the noise.

By this method we could distinguish well the regular patterns (or man made patterns) from the natural textures encountered in outdoor scenes. It was more difficult to distinguish the water from the grass unless the main frequency component was sufficiently different. The training feature vector was extremely sensitive to differences in how the picture was taken, in particular, the distance between the observer and the scene, and the orientation of the observer (whether he is on the ground or in an airplane) with respect to the scene. This method did not consider any corrections for texture gradient. It simply classified some areas of a scene into some given classes of textures.

We could have improved the feature vector using further features similar to Lendaris' and thus enlarged and refined the classification procedure of texture. We did not do it for the following reasons:

(i) Feature vectors offer very specific and rigid description of a texture, which is an obstacle in finding continuity of textured regions unless the texture is a very regular pattern without any features such as texture gradient. Naturally, one

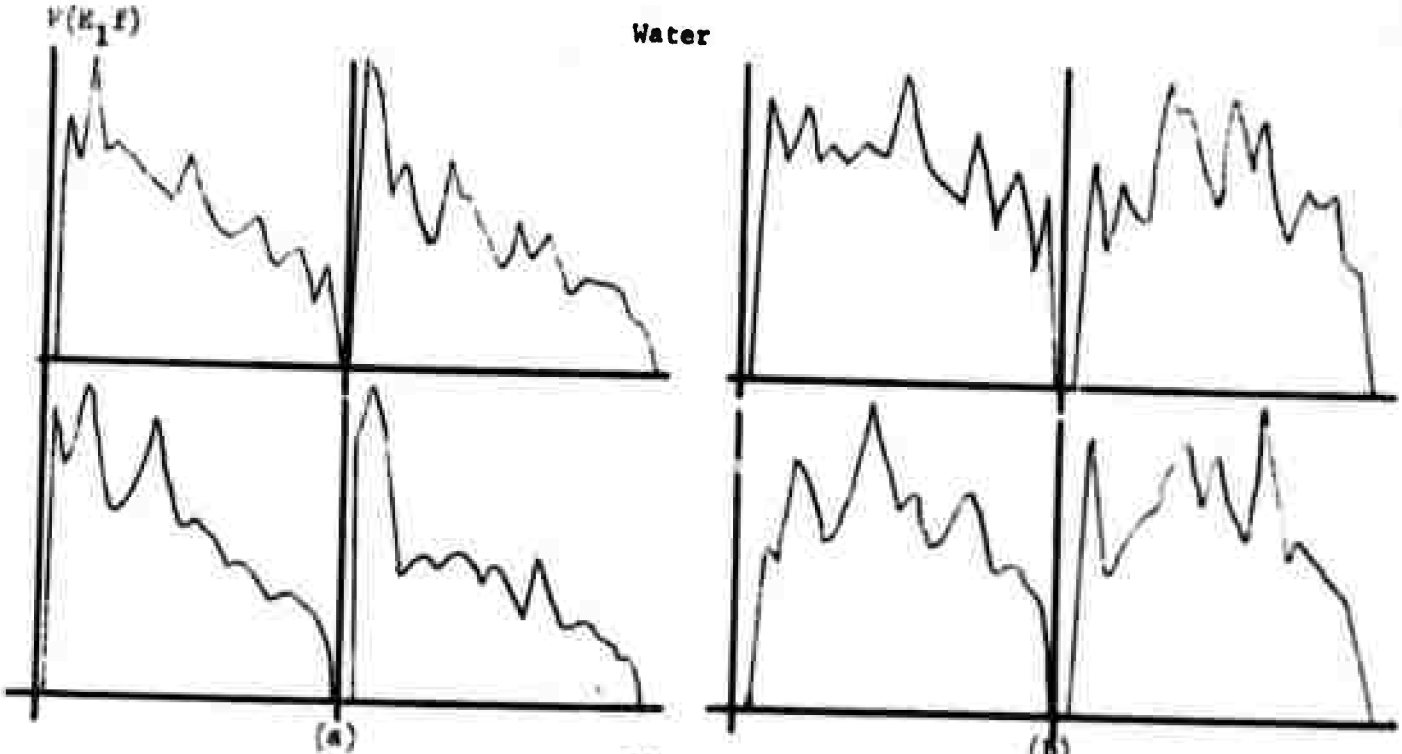


Fig. 32

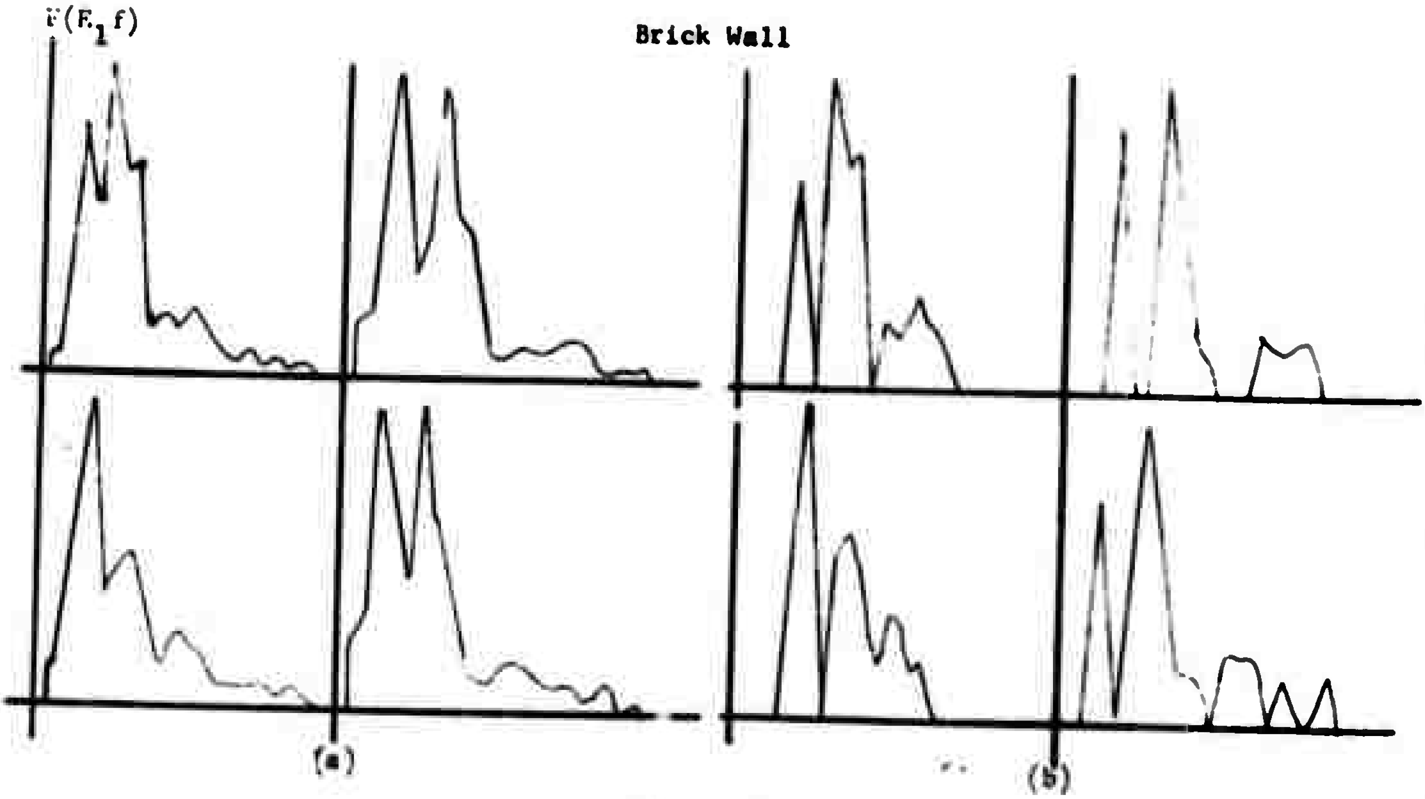


Fig. 33

can construct feature vectors less specific, but then the sensitivity for the differences between two different textures will be lessened, which in general is not desirable. To sum up, in the texture region finder one needs to have a flexibility in choosing features for grouping or discriminatory purposes. One also wants to have symbolic descriptions with some parameters as opposed to only numeric description (as in the feature vector). The symbolic description (if properly chosen) is invariant with respect to several metric (scalar) features and thus it represents a certain abstraction which is useful for recognition purposes.

(ii) The classification process of textures into some classes besides feature vectors uses some distance measurements between the training feature vector and the sample feature vector. This process does not consider any topological properties of windows, namely connectivity, continuity and proximity. Furthermore, metric description of a real texture is not sufficient for identification purposes. For instance, grass is identified as grass not only because of its color or the geometry of its texture but also through its spatial relationship with other objects on the scene (e.g. grass is always on the ground, below a sky, etc.).

A different approach had to be sought for describing textures; an approach that would give symbolic descriptions of a texture together with some parameters and would find continuous regions with respect to their descriptions. In Chapter 3 we have described the texture operator that produces such a description. This operator can function on different window sizes. The large windows capture the global textures, whereas the small windows are used for recognition of fine texture that in the large window is not noticed. The continuity and proximity of some structures are the basic properties used in a region grower. So far, we talked mostly about the texture structure. However, the structure that forms a region could depend on many properties, such as color, shape, size, and others.

5.2. Texture Gradient.

Many elements of the world are made up of texture elements of a constant size, (grass, brick walls, wheat, water waves). The apparent size of texture elements depends upon distance. Although there is a chance for mistake, it is natural to interpret consistant variation in apparent size of texture elements as a measure of relative distance. If there is little variation, the interpretation is that the surface is everywhere approximately at the same distance from the observer. Such surfaces are nearly perpendicular to the line of sight and are called frontal surfaces. If there is a systematic variation of apparent size of texture elements, smaller elements are assumed further away. Such a texture gradient suggests that the surface is longitudinal, that is, along the line of sight. The presence or absence of a systematic texture gradient gives a rough indication of the angle, curvature, and relative distance of objects. The role of texture gradient in human perception of depth has been described by Gibson (1950).

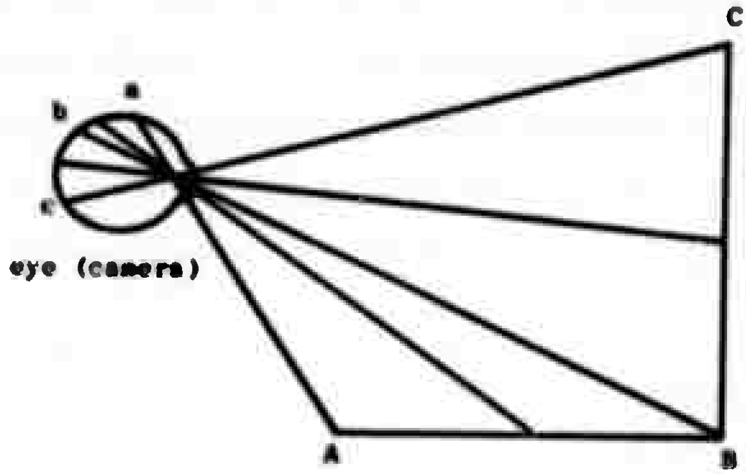


Fig. 34

In Figure 34, surface AB is a longitudinal surface and surface BC is a frontal surface. In the image there exists a gradient of texture, from coarse to fine along ab, whereas in the image, no such gradient occurs along bc, and the texture is uniform throughout.

The texture gradient can be used as a measuring stick whose scale we don't know, but which gives us relative depth estimates:

B is twice as far as A.

For familiar surfaces for which we know the texture element size, the scale of the measuring stick is known, and we have an estimate of absolute distance (provided we have an estimate of the surface angle with regard to the observer's image plane - we shall show soon that we can determine that angle). Since the observer knows his orientation with regard to gravity, by assuming a level ground plane, he can estimate the distance of areas near his feet with reasonable accuracy. This helps in establishing absolute size of grass and other textures on the ground.

There is one reasonableness condition on texture gradients. The apparent size of texture elements should decrease toward the horizon. That is, we don't expect large nearly level overhangs, above us, and for opaque surfaces below the horizon, we must see decreasing apparent element size toward the horizon.

The projection of a longitudinal or slanted surface on a picture plane is obtained by perspective geometry. The principles governing such a projection are as follows (See Fig. 35).

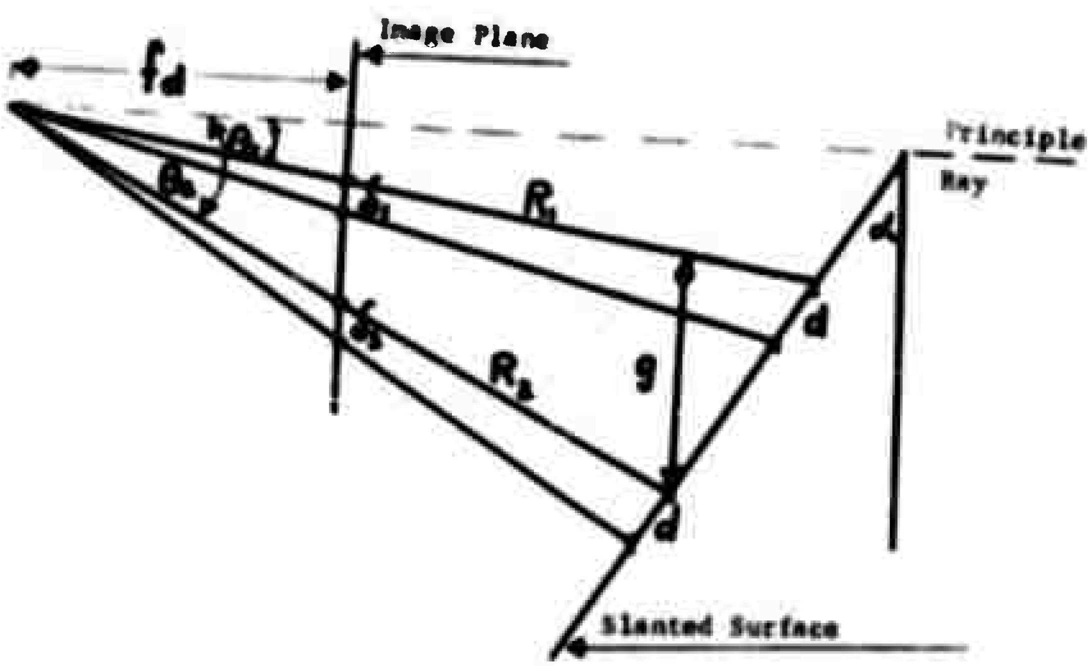


Fig. 35

$$\frac{d\cdot\cos\alpha}{R_1} = \frac{\delta_1}{\frac{fd}{\cos\theta_1}}$$

$$\frac{d\cdot\cos\alpha}{R_2} = \frac{\delta_2}{fd/\cos\theta_2}$$

for small θ_1, θ_2,

$\cos\theta_1 \cong \cos\theta_2 \cong 1$

$$\frac{\delta_1}{\delta_2} \cong \frac{R_2}{R_1}$$

From the similarity of two triangles follows

$$\frac{fd\cdot(\tan\theta_2 - \tan\theta_1)}{g} = \frac{fd\cdot\cos\theta_2}{R_2}$$

For small $(\theta_2 - \theta_1)$, $R_1 = R_2 + dR$

$$dR = g\cdot\tan\alpha = (\tan\theta_2 - \tan\theta_1)\, R_2 \tan\alpha$$

Now let us define a fractional (Gradient)

$$\epsilon = \frac{\text{Fractional change in element size (image)}}{\text{Baseline in image}}$$

$$\epsilon = \frac{\delta_1 - \delta_2}{1/2(\delta_1 + \delta_2)\ fd\ (\tan\theta_2 - \tan\theta_1)},$$

where δ_1, δ_2 are texture element sizes in the image.

After some approximation we obtain formula:

$$\epsilon = -\frac{\tan\alpha}{fd}.$$

Rephrasing the formula in terms of angles (on retina) instead of length on retina, we get:

$$\epsilon^1 = -\tan\alpha.$$

Thus, we can calculate the angle with respect to observer.

Since the observer knows his angles with respect to gravity and he knows the angle with respect to the observer, he thus knows the angle of the surface with respect to gravity.

Then, the texture element in the object space can be computed as follows:

$$d = \frac{R \cdot \delta_1}{fd \cdot \cos \alpha}$$

How sensitive are estimates of the distance to the assumption that the ground is level?

Consider Fig. 36 .

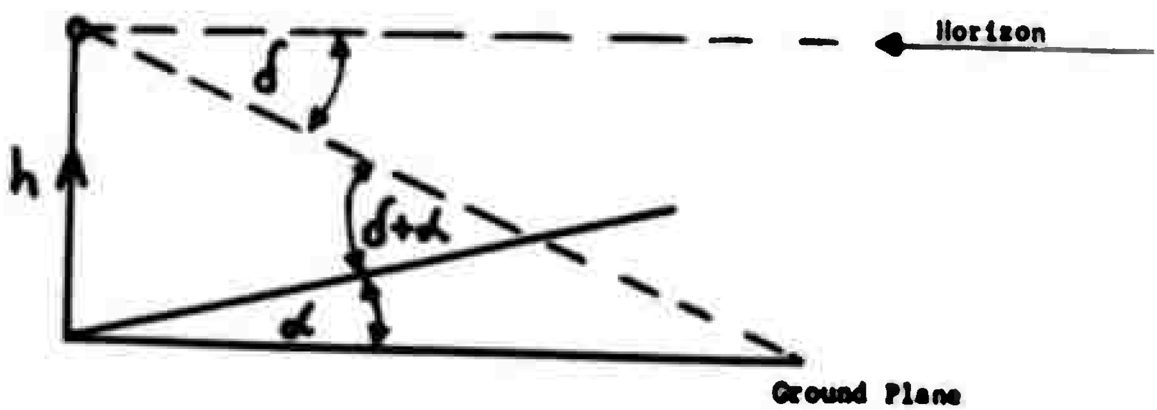


Fig. 36

We want to calculate the distance from observer to the ground for level and non-level cases. δ is the angle between the horizon and the observer view. And α is the angle of the slanted surface.

Then $\frac{S_1}{S_2} = \frac{\delta}{\delta + \alpha}$

is the ratio between the distance S_1 and the distance S_2, which is the distance to the level surface. The formula shows that there is a fairly strong dependence on α, except for small distances.

As an example of the texture gradient and its recognition, we present a picture of the ocean (See Fig. 37); without recording the texture gradient we find a partition of the picture into several regions (See Fig. 38). All regions are described as monodirectional textured regions, with the same directionality but with different wave lengths. However the wave length changes linearly in a vertical direction across the picture. (From the bottom of the picture the wavelength = 32 to the top of the picture where the wavelength = 8). Thus, by recognizing the texture gradient, we recognize the whole picture as one textured region, displayed in Fig. 39.

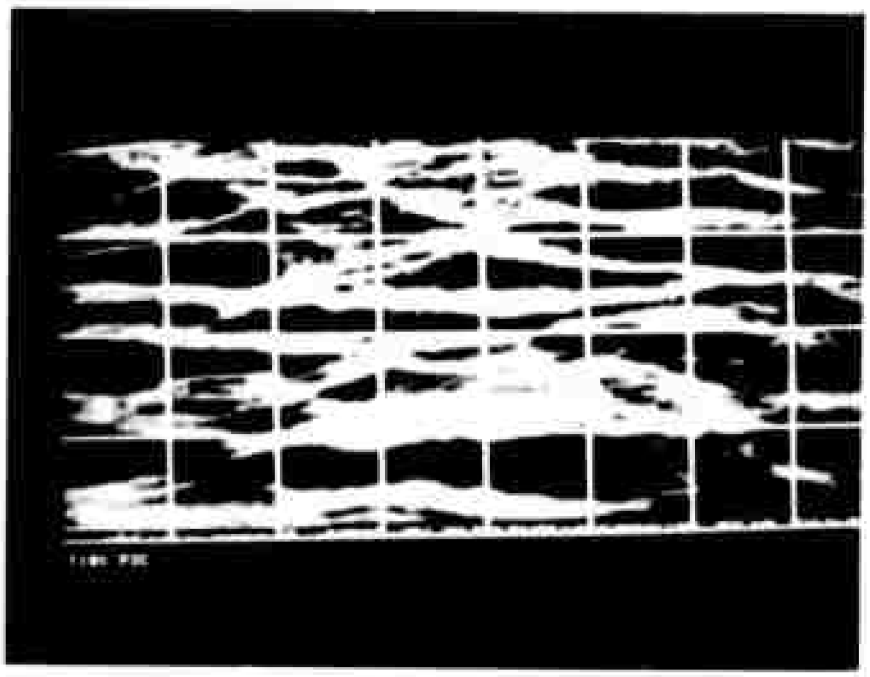


Fig. [illegible]

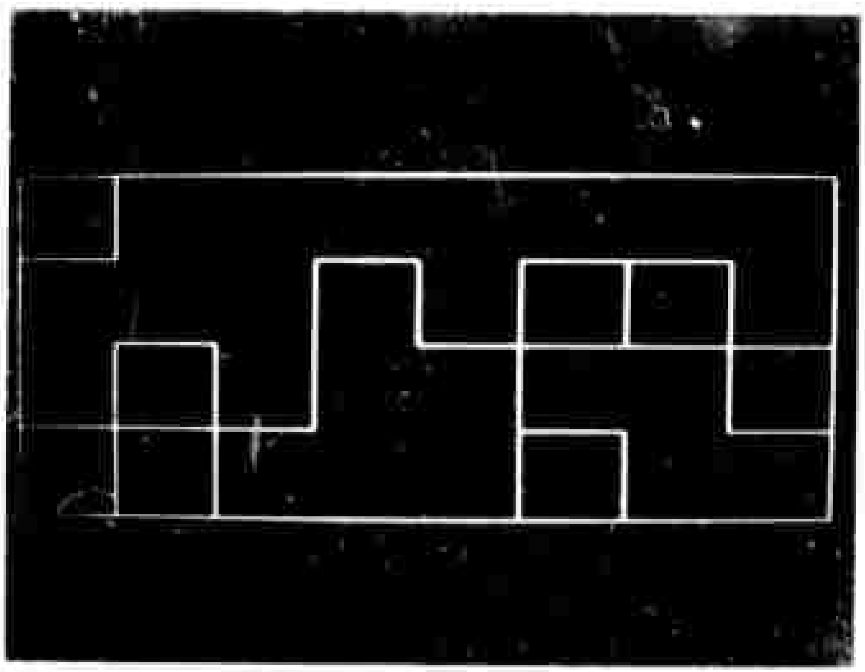


Fig.

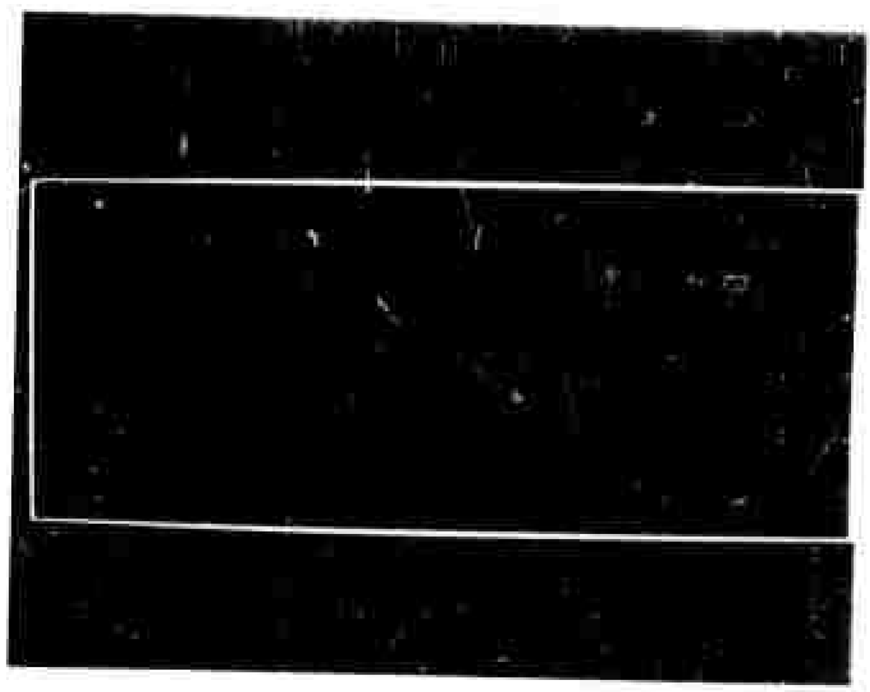


Fig.

5.3. The World Model

In chapter 2 we looked closely at elements of an outdoor scene, involving grass, sky, clouds, water, and trees. On of our purposes was to introduce the sort of texture descriptors whch we have implemented. The other was to lead the way into a discussion of our world model. We saw a great range of variation for sizes, colors and other properties of texture elements in these outdoor scenes. Grass ranges in color through greens, browns, and yellows. Trees range from a few feet to a few hundred feet in height. Because of this variation and the variation of apparent size of objects at different distances from the observer, it appears that no immediate identification of image textures with elements of the world is reliable. In some cases, the understandings depend on perhaps unconscious reasoning: the spray on rocks is not very similar in appearance to the ocean around it. In many cases, the identifications are simply resolved by considering relations between image regions; motion obscuration identifies trees in front of clouds, shadows identify trees as standing above ground, obscuration implies a background. Relative depth determines that the ground is roughly level and that trees stand above the ground.

It is reasonable to question whether a model which must allow as much flexibility as to allow the range of sizes for objects, and the variation in relations, is of any use at all. There are several ways in which it is useful. The first is that certain relations are reasonably stable. The sky is bright against the horizon. The proportions of grass and trees are roughly independent of size. Certain regular shapes are usually man-made. The second is that much of the variation is connected with subsidiary conditions. If trees appear different colors, they are different species, and have other identifiable properties. If the grass is yellow, then it must be dry. An apparent size gradient probably means a distance gradient.

However, the usual mode of perception is continuous perception. In scene analysis, we often think of showing a single picture with no context and expect the observer to understand it. Indeed, humans can do just that usually. But the bulk of perceptual activity is involved in moving in a world in which changes happen slowly and locally. Most of the world is nearly unchanged from one moment to the next. Most of the recent perceptual understanding are useful at any instant; the system knows a great deal about the environment and makes incremental changes to its model. The making of the changes to the model is aided by the detail of the knowledge already available.

That does not mean that we can do without the ability to actually build up the model, either from the picture shown out of context, or guided by an already detailed model. But it does mean that a large part of perceptual activity is guided by detailed models.

Another aspect of the world model is that it contains the information about the observer's point of view. The observer's motion provides a depth sense equivalent to stereo, but much more useful for distant objects. Distance estimates using motion parallax depend on the observer's estimate of his motion. Stereo distance measurements depend upon a model for the convergence position of the two eyes, the eye separation, and correspondence of the coordinate systems of the two eyes. An equivalent observer model has been implemented at this laboratory in the work of

Sobel (1970), Tenenbaum (1970), and the use of observer's motion for depth perception has been implemented by Nevatia (unpublished). Formally, the model has two levels:

(a) Regions in the object space, objects and collections of objects, called the elements of the model;

(b) Structured description of the elements in the object space. These descriptions are almost directly interpretable in a program as procedures.

A world model is a dynamic structure that changes during the identification process. The description of the elements of the model is carried out in the object space and, wherever it is possible, with counterparts in the image space. Not all descriptors in the object space have a meaningful counterpart in image space. An example is the size of objects, which can be interpreted from distance estimates and apparent sizes.

The properties of grass, sky, water and trees have been described in Table 1. All these descriptions are included in the world model and some new ones are included in Table 7.

All objects in the model, except rocks and unnamed objects, have broken boundaries.

One may wonder what other descriptors (besides texture descriptors and color descriptors) could be relevant in the model. Unlike in the case of grass blades, water waves, and trees, where their size plays an important role, the size of rocks varies so much that it is hardly a useful feature for them. On the other hand, the shape of rocks (bloblike), is signficant because it can be contrasted with the linear shapes of grass leaves or water waves. However, the only rocks of interest are those which are big enough to stick far out of the ground (might impede navigation). Here we worry about the relevance of size and shape of texture elements. What about the size and shape of regions? Size and shape is not significant for regions of grass, ocean, forest, and ensembles of rocks.

Tab. (

Regions in Object Space	Color Attributes	Spatial Relationships
grass	Usually green, sometimes yellow or light brown, never blue.	Located on ground, under the sky and trees.
water	Blue or green, sometimes gray with silver waves, never red	Located at the ground plane below sky and trees. In the image space, ocean and grass, trees or rocks could form overlapping regions.
sky	Light blue, the brightest area in the scene.	Sky is the farthest region in the scene and it is always above any element of the world model. In the image space it can form overlapping regions with grass, trees, and rocks.
clouds	Objects in the sky.	
tree	The crown is usually green, sometimes yellow, brown or red. The trunk if dark brown.	Trees are below the sky, and above grass or ocean.
rock	All shades of gray, brown or red.	Rocks are always below sky and on the ground. They could be scattered in grass and water.
unnamed objects	any color	On ground, below sky.

5.4 <u>The Higher Level Program</u>

We discuss briefly a suggested higher level program which we simulate. We concentrate on the two scenes in figures 40 and 42. Fig. 40 contains three pictures taken of the scene in Fig. 1 through 3 filters, red, green and blue. The names SCENE1, SCENE2 and SCENE3 correspond to red, green and blue filtered scenes. Figure 42 contains four pictures of a scene in Fig. 41. The top two and the bottom left pictures correspond to the red, green, and blue filtered pictures, respectively. The bottom right picture is the brightness function of the scene in Fig. 41.

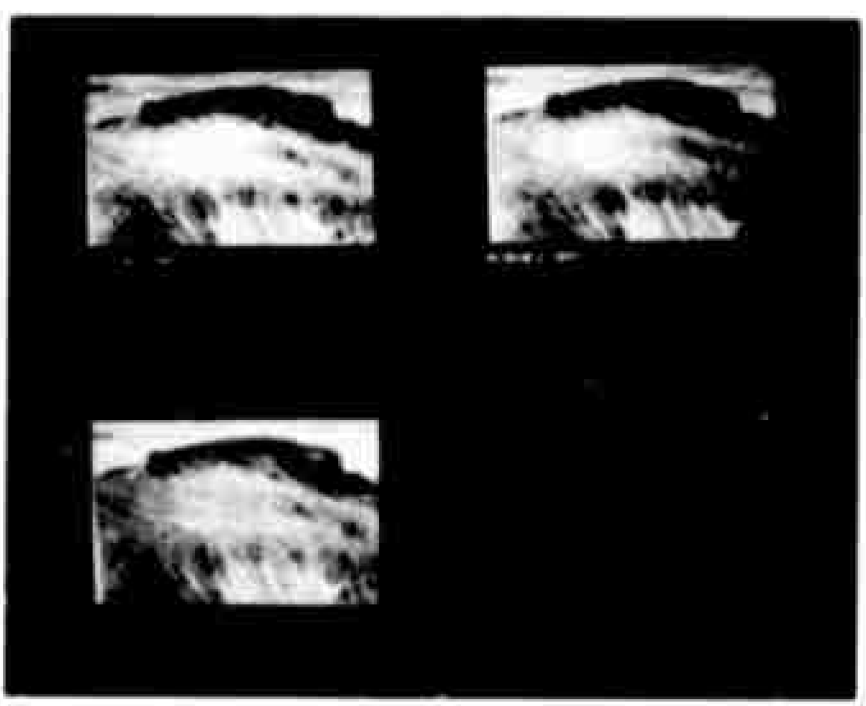
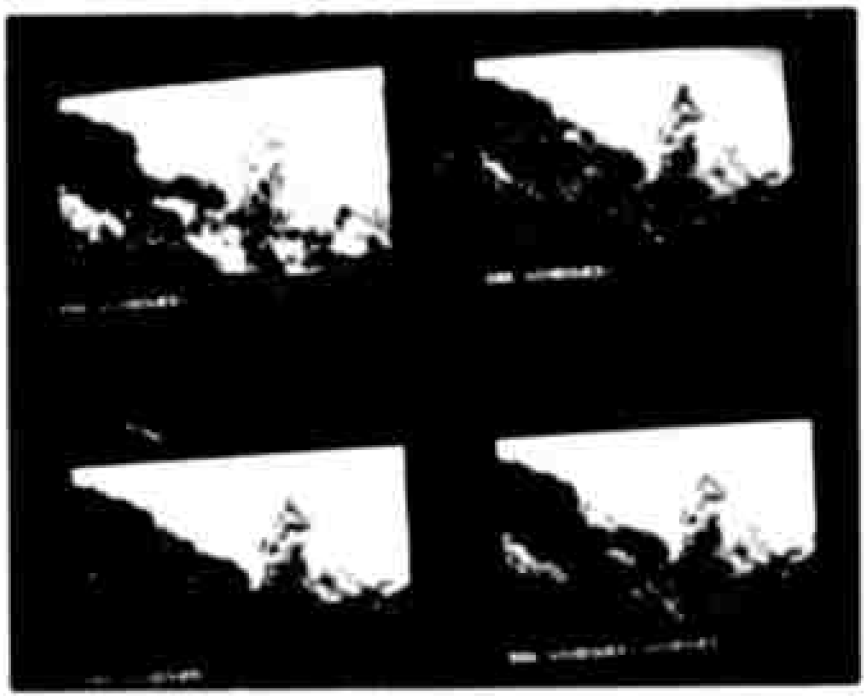


Fig. 40


SCENE



Let us call the scenes in Fig. 40 and 4? WATER and ROSES respectively. There are three gross parts of this process of interpreting the structure of the scenes, these might be called:

organization of regions having continuous properties

determination of spatial relations

identification of elements in object space;

They are not strictly heirarchical, since identification determines new spatial relations, and suggests other low level organizations.

Some of the mechanisms for organization of continuous regions were previously discussed. The regions based on continuity in color and some texture descriptors are natural starting places. Proximity provides the basis for the suggestion of texture super-regions which are disconnected, but may be usefully considered as a unit. In this operation, we group together nearby regions of like color or like textural properties. The next mechanism is that of hypothesis-verification: a particular color or texture region is a hypothesis of continuity. If the region has some physical continuity, we should find that other properties are continuous over the region. If the boundaries are false, then there should be textural properties which continue across the boundary, from which we would assume a continuity which would be tested by looking for other continuity. If the boundaries correspond to physical boundaries, we will usually be able to find a discontinuity in some textural property.

We can infer a few spatial relations from the texture gradient, from guesses about interposition (which object is in front of which) and from the observer orientation and position, combined with the ground plane hypothesis. In a complete system, we could call on depth perception by

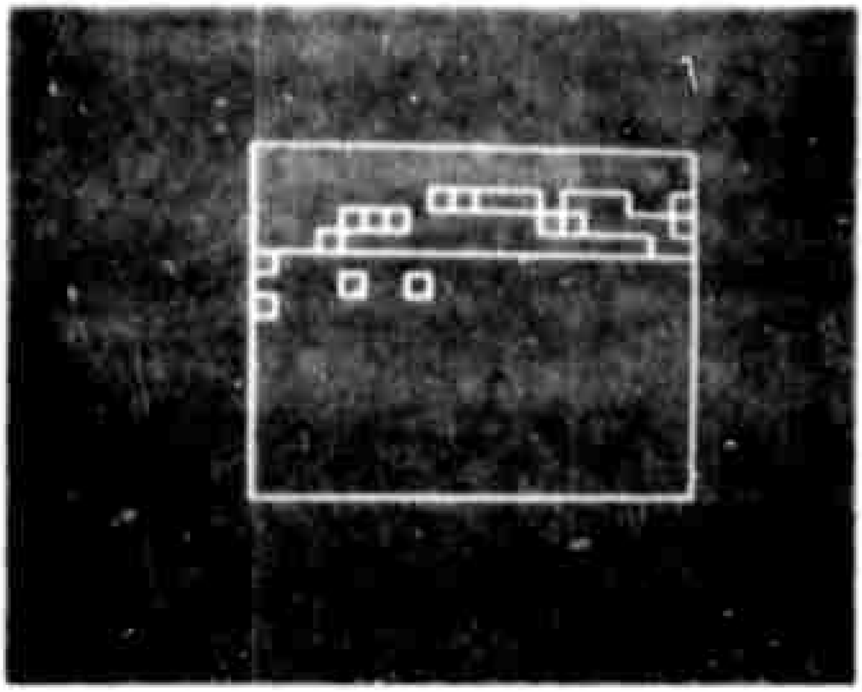


Fig. [illegible]

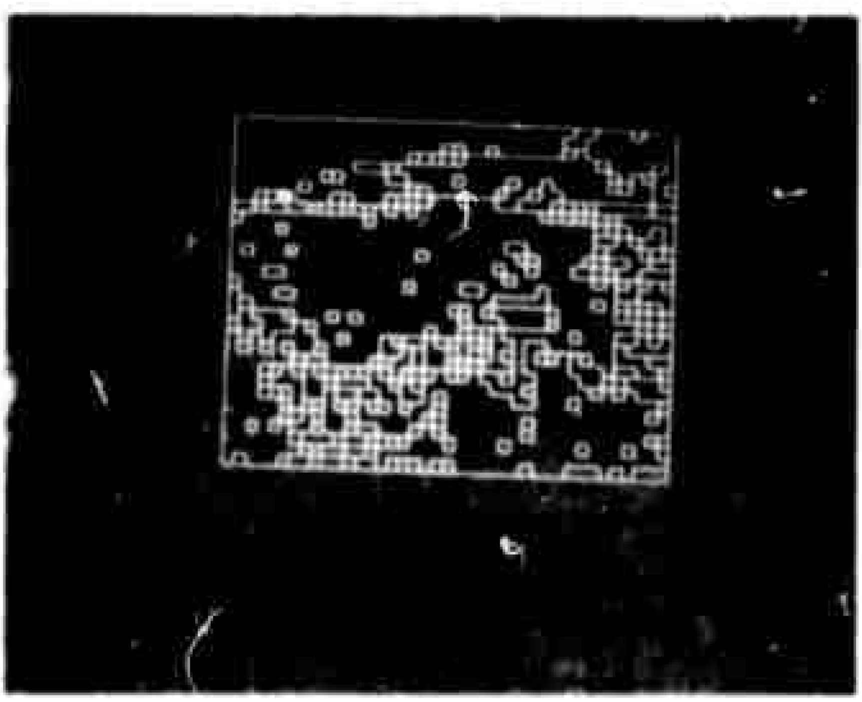


Fig. [illegible]

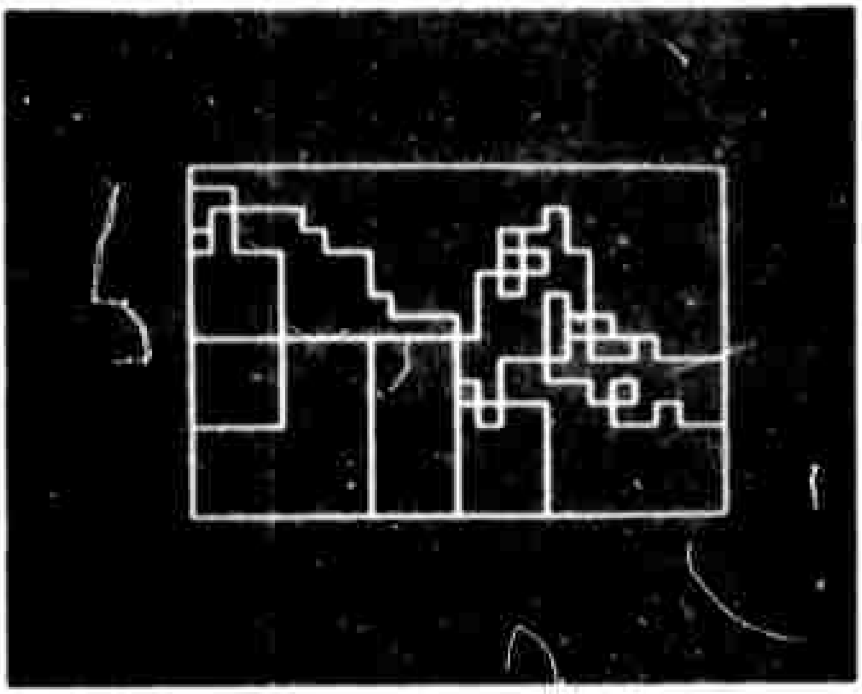
stereo and motion parallax. Our inferences would be a good guide to economical use of these modules.

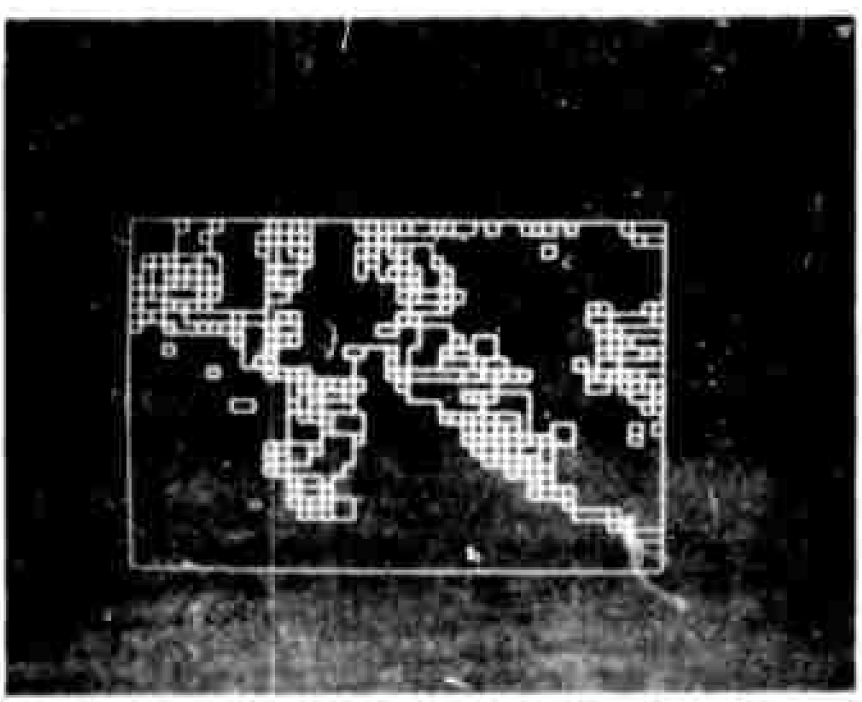
Identification, in our suggested system, proceeds both from the world model and from the data. Some elements of the world model are better starting places than others. We assume that the sky would be easily established in most cases. Other image elements should be approached after finding the important structural elements in object space, i.e., sky, ground plane, and trees. We are assuming that a full variety of properties and relations aid us in making initial and tentative identifications of sky, etc.

In Figure 40, which we call WATER, we have two major regions which correspond to grass, a region which corresponds to the rock, and two regions which correspond to the water. In the scene called ROSES, the sky appears as one large region and several small regions; there are three regions of the bush, and several small regions which correspond to roses.

In the analysis of these two scenes based only on texture analysis without guidance, the scene WATER is described partly adequately. After the texture gradient suggests further continuities, the grassy regions merge, and there remain three main image elements which correspond to grass, rock, and water; see Fig. 43.

This texture organization is suitable for providing hypotheses of continuity for regions which are broken by color organization for WATER; see Fig. 44. The similarity in color of the joined color regions confirms the texture continuity. In ROSES, the sky is adequately described by texture, while the bush and flower regions are chaotic, as one can see in Fig. 45. This reflects one of the inadequacies of the Fourier transform,





the weakness with feature sizes approaching the window size. This is normally accomplished by subdividing regions with slow changes which correspond to probable region boundaries. That was suppressed in this version of the program.

The region-growing does not succeed in isolating the flowers by cutting up windows containing flowers to partition off smaller cells of adjoining areas of the bush. This is the worst performance of the texture region finding process, but it is instructive. On the whole, the unaugmented texture region analysis is unable to aid in proposing useful alternative hypotheses for organization. However, texture boundaries for the sky coincide with color boundaries (the color boundaries of the roses are displayed in Fig. 46), and a slight relaxation of the criteria for continuity, verified by continuity in contrast among the color components, does provide a set of larger texture regions among the bush and flowers. Even in that worst case, the jumbled areas of color correspond to regions of moderate size under texture, so that there are no large regions of the picture which appear entirely chaotic under both aspects. The texture descriptors are useful for analyzing the color regions, and have more utility used in that directed mode. The element size and contrast are meaningful when restricted to the bush; in the unaided texture analysis, these descriptors mix the flowers and bush.

In evaluating our higher level procedures, it is usual that we re-evaluate the quality of the lower level modules. We find significant ways in which they could be improved, and in ways which would best be done at that level. In general, it is better to proceed to a fully developed system then to put disproportionate work at the low level. We will later specify what improvements we would make in the low level modules.

Let us make the preliminary organization of the two scenes. With ROSES, we begin with proximity of color regions. The bush regions and the flower regions are alike in color; for example in two areas of the bush, the color coordinates $r/(r+g+b)$ and $g/(r+g+b)$ are:

Sample 1	(.46, .43)
Sample 2	(.47, .40)

thus we can conjecture these as a super-region. Let us compare contrast and dominant wavelength for these two color regions which we conjecture to be similar. Compare the two color regions in Fig. 46 with the Tables 8, 9, and 10 of average intensity, wavelength, and contrast, over 8 x 8 windows. We see that the dominant wavelength is short over much of these two color regions. In fact, if we define a region from the small wavelengths (< 4) the region spreads over most of the bush. In the scene, the sky is a region under color and all texture descriptors. The sky boundary in color is reinforced by the existence of texture boundaries. As we have indicated, textural properties are probably adequate to confirm continuity of the regions suggested by color for bushes and flowers, and to show discontinuities of frequency. In WATER, the two regions corresponding to water are joined by proximity in color and continuity in texture. The water boundary shows up strongly as a change in color and in texture, directional to homogeneous for the water-rock boundary, and different directionalities with distinctly different color at the water-grass boundary. The grass is continous in directionality, size, and color. We can now make correspondence with the world model. Since the sky is often prominent in outdoor scenes, we attempt to find the sky. We look at white and blue regions which are near or above the horizon. In WATER, we might try the region which is really water. The color is acceptable, but the directionality is very unlikely for sky, and the contrast and size of texture elements is also unlikely. (This estimate is based on a few months of sporadic sky watching. (Of course, there are directional clouds, "mackerel sky", but it seems quite infrequent. Also, the clouds seem to have much lower frequency.) The water region is below the horizon. If there were a significant view, we could see a texture gradient and thus substantiate that the surface is flat. Also, in continuous perception, we would find that the water motion is very different from cloud motion. Motion would also allow interpretation of the breakers around the rock as part of the water. The region corresponding to grass is directional, low contrast, and has a texture gradient, implying that it is horizontal. The color is consistent with grass, which lies on the ground plane. From the ground plane assumption, we can estimate the size of the elements of the grass:

image size*angular resolution*distance

= 2*(1/666/*300 cm = .9cm

where we estimated the image size previously, the camera parameters are known, and the distance is obtained from a crude guess, but is known in principle from the observer position and orientation. The size is also consistent with grass blades. For the rock, neither color nor homogeneity tell us very much. Since the rock is convex downward on its boundary with water, we assume that the rock is in front of the horizontal water surface. We assume thus that it is an object which sticks up from the surface, and calculate the vertical height and length along the ground. From the image, the texture gradient tells us that the distance at the rock is about 4 times that at the front of the picture. Thus, the above expression gives:

$$18*(1/666)*1200cm = 36cm$$

while the width of the rock is approximately 300cm. These are only approximate values which depend on our guesses about the ground plane and texture gradient. On the other hand, the conclusions depend most strongly on relative size conclusions. Grass elements are small; rocks are often big compared to grass. We can make the comparison between the rock and grass near the base of the rock. In the image, the rock is big, and from all assumptions about objects in the image being further away as they recede in apparent position toward the horizon, the rock is much bigger than the elements of the grass. These give some strength to the assumption.

In ROSES, we begin by attempting to find the sky. The only region of acceptable color is the sky itself. The color is white, indicating clouds, with low contrast as seen in Table 8. The texture is homogeneous. As a verification, we might find blue patches, find motion, and find that the distance of this region is very great. The region is far above the horizon, and is very bright; see the brightness in Table 6. From the concave downward boundary with the other regions, we assume that it is behind the green elements. With the identification sky; we find that the green elements are in front of the sky, thus probably approximately vertical and are frontal (they show no systematic texture gradient , also indicating that they are vertical). The texture is blob-like; the blob size is interesting. If we can guess that these are leaves rather than leaf clusters or branches, then we can estimate the distance to the bush. Finding the stems would aid in that. Because the bush is probably vertical, it is not grass. If we include leaf elements, fruits and flowers in our descriptions of trees and bushes, then by guessing that the flowers are really associated with the bush which surrounds them, we can guess the scale of the leaves relative to the flowers, and thus establish that the texture elements are leaves and establish approximate distance. Of course, at any level, we could establish relatively unique elements to correspond in two views, and determine distance by stereo or motion.

6. CONCLUSIONS

We presented a representation of textured scenes which was not a two-dimensional representation of the projected image, but a three-dimensional representation of the elements and spatial relations in object space. We feel that the representation of spatial relations such as "grass is found on the ground plane" and the nearly infinite distance of the sky are characteristic of these elements, and help more than any other properties to identify them and to orient the observer. Our representation is effective, also, in that it is segmented into distinct elements, which are described by a heirarchy of texture regions and textured elements. Textured regions may be texture elements of a super texture, or texture elements may be textured regions of a sub-texture. This is not only a formal nicety, but a usual part of our description of ourdoor scenes; for example, in trees, the leaves are texture elements of leaf clusters or branches, which are texture elements of a tree, which is a texture element of tree clusters. The description of shape of texture elements depends heavily on a linear approximation to shape, and describes directionality, width of texture elements, and spacings. We argue from psychological evidence that these are the most important of descriptors, and further, that they are natural for computer implementation. These descriptors are most useful for directional textures. The representation is included in a world model for which an example was given, but which awaits implementation with the high level procedures.

A simple color region analysis was very useful in texture analysis. The normalized color coordinates, $r/(r+g+b)$ and $g/(r+g+b)$ were compared for continuity, and regions were defined by neighbors of continuous color. There are some potential problems in an analysis of color in outdoor scenes, since many of the texture elements are very small. Typically, the regions are leaves or blades of grass. As an expedient to find larger regions, we have averaged over a small window, and compared colors of adjacent windows. As a consequence, we sacrificed localization of edges. The expedient is only partially successful however. We notice that the averaging works best among clusters of leaves where the color is uniform to begin with. Where the leaves are isolated against the sky, the color contrast is large, and continuity of the averages are only by chance. Thus, the averaging is not very successful. A better mechanism to define larger regions of color is perhaps to go to the computationally more difficult operation of finding like colors within windows, that is to implement color regions based on proximity rather than continuity. The obvious value of defining color regions which ignore the brightness fluctuations of individual leaves should not lead us to ignore brightness edges and consider only color edges. We also make use of the size of the brightness regions, the leaves in this case.

A sheaf-theoretic description formalizes the process of region-growing and gives an exact account of the shift from local to global and vice versa. One must be cautious about interpreting the sheaf-theoretic notions in the context of color and texture regions. Due to sampling we have a finite scale of window sizes and the definition must include a least window size, the size of texture elements. No such discreteness conditions are embedded in sheaf theory.

In the implementation of texture descriptors, we were able to translate those spatial domain descriptors that we found important from Fourier transforms over windows of various sizes. Directional and non-directional components were separable to a useful extent. These reflected shapes of texture elements and their spatial relations, or statistical properties of irregular textures. We argued that the human description of an image makes use of a step function approximation; we often describe in terms of regions of constant intensity. We obtain analytic expressions for the contrast, the element size and spacing, and some location information; these were based on a spatial domain model of pulses of equal amplitude, width and spacing, on a uniform background level. The implementation of Fourier transform descriptors had some minor implementation difficulties which are usually overlooked. They are consequences of the fact that the Fast Fourier Transform is really a Fourier series and not a Fourier transform. A Fourier transform has no preferred directions, while the Fourier series has preferred axes along the x and y coordinate axes. This introduces a non-isotropy in the fan filter. We have used only a straightforward fan filter and have not attempted to compensate for the peculiarities of the Fourier series. There are also spurious broadening of peaks which are consequences of the Fourier series. Despite these difficulties, the spectra show useful directionality properties, and we have been able to work with them.

The more serious problems with the Fourier transforms are conceptual difficulties which are true of any orthogonal expansion and of the true Fourier transform. Interpretation is based on the power spectrum and phase information is ignored. These transforms are non-local, and give very poor edge and position information. To an extent, we have tried to get around this by an expedient of using local windows. This provides a crude localization, which was not adequate for many purposes. The usefulness of the transforms was very dependent on the scale of the windows. The descriptors were most useful when the window size was such as to exclude other regions and to include some repetition of texture elements within the window. This meant that a range of window sizes was necessary, and that we could not always have windows small enough or large enough. To a certain extent, we could probably get positional information from our analytic expression using the phase of the transform. However, that is useful only for uniformly spaced texture elements. In most useful cases, the spacing is quite irregular. The non-local nature of the transform is a severe disadvantage in using color information. As in other cases, we can make examples where the averaged quantities are useful in smearing out and bridging local boundaries. This has limited utility. There are a few simple cases for which the color contrasts can be extracted from the Fourier transform of separate intensities through three filters. We have not yet incorporated these into our descriptors.

There are some incremental improvements to be made to our Fourier descriptor scheme. One of these is to use interval analysis for determining widths and locations of discontinuities in directional textures. In some cases, overlapped windowing would be useful near boundaries. Including Fourier color components was mentioned above. We should include edge and region operators (perhaps on filtered directional components) to find the extent to linear elements and to find boundaries of quasi-homogeneous regions.

Textured regions were obtained on the basis of the texture descriptors. We found that a range of window sizes was necessary; our strategy was not adequate to use the window sizes as well as it might have. The choice of window sizes would be better left to higher level choice among suggested regions. The actual program which gave textured regions classified the descriptors according to mono-directional, bi-directional, blob-like, homogeneous, and noisy. Then adjacent regions were merged into continuous regions based on continuity of the descriptors. A change of scale was used if there was a strong frequency 1 component (that is, there probably was an edge in the window) or if there was a bi-directional texture which could arise from a boundary of two directional textures. A second region-growing pass relaxes the criteria and ignores classifications to extend large regions by including acceptable small neighboring regions under relaxed criteria.

A guided program determined textured regions based on information passed down from above. This program was guided by the user, but the advice could equally well have come from other programs. The guided programs determined texture regions within specified areas and according to specified similarity criteria, for example, frequency and contrast. Continuity could be explored on the basis of separate parameters.

We see some incremental improvements to the region-growing based on Fourier descriptors. We would add an alternate tactic for windows with long wavelengths; now we subdivide based on the possibility of an edge in the window. We would also increase window size to look for a repeated feature of a larger size. We would eliminate the classification step now used. There is some argument in favor of the classification. There are actually many natural objects which fall into one or another of these classes. But the classification in the early stages introduces artificial boundaries and ignores the multiple descriptors which are available. Only later does it relax these classes. A further and more significant addition would be an implementation of proximity of like descriptors on nearby windows for texture super-regions.

We presented an outline for a higher level procedure to make a correspondence between the image elements and the world model. This outline was not implemented but simulated in two examples. The striking conclusions from the simulation were that interpretation of the three-dimensional structure of the scene was necessary to make the identification. This contrasts with the work done in interpretation of aerial photographs, for which the world is effectively on a locally flat surface, and depth considerations are unimportant. A perceptual system begins to have interesting structure when it can work with more than a single property. Our model deals with multiple properties by a hypothesis-verification paradigm for proposing boundaries and regions. Continuity in a physical surface can be hypothesized on the basis of one of the properties. Often, the region description from some one property will be particularly simple and useful. It is not clear that we can always pick in advance which property it is; with sufficient context we could usually make a good choice. But we can try several different choices as hypotheses of meaningful surfaces, and test that continuity in that property corresponds to continuity in other properties; discontinuity in physical surfaces often reflects discontinuities in several properties. In these examples, color regions were joined into color super-regions of like color. Textural properties of dominant frequency and contrast showed continuity of these properties over the color super-region. At this point, the lower-level modules could function in a guided mode to expand the region corresponding to these typical descriptors and include nearby areas of the image which were not well-described in the earlier analysis with little context.

The regions which come forth do not make a neat image. They overlap and do not cover the whole image. Still, we are aiming to interpret those parts of it that are simple to understand. Inferences of spatial relations are important here. Texture gradient gave estimates of surface orientation. The ground plane assumption gave a local coordinate system when combined with the assumption that objects stand out from the ground. These relations depended primarily on relative distance and relative size estimates which were not greatly sensitive to the assumptions. In some cases it was possible to guess which object was in front of another, either because of concavity, or from identification of the sky or water.

In identifying elements and structure of the world model, our simulation attempted first to establish the sky. Based on color brightness, contrast with the horizon, and its position (above the horizon estimated by gravity), this is assumed a simple match. We can also then determine the sky line and guess which objects stand out from the ground plane. Sizes of texture elements were of considerable use; knowledge of size is much more useful here than in the blocks world where context is limited. The knowledge of the size of grass blades is more useful than the knowledge of the size of one particular block.

We have mentioned some incremental improvements to texture local description and to forming texture regions. The primary weaknesses at this level are the crude localization and limited use of proximity in establishing super-regions. Improvement of the color region-finder is of primary importance. Since must of the interpretation depended upon inference separate from strictly textural properties of areas of the image, we feel that the most significant next stage is to embed these elements in a complete visual system. This would involve more than just implementation of the simulated higher level program. The typical system would navigate in an outdoor or planetary exploration environment. The navigation goals make explicit which problems the system needs to solve at any time. The situation is one of continuous perception which allows the model built up at one instant to be used in subsequent problem-solving. Continuous perception also allows us to tell which objects are moving, which is of use in outdoor scenes. The complete system would have stereo and motion parallax for depth at small and great distances. To a certain extent, the system would avoid finding solutions which could be found purely from single projections, but such a system appears feasible within the current state of computer vision, while a system which ignores so much information does not appear to be achievable soon.

APPENDIX

Topological Models

In this section we give a brief account of a possible approach to the topology of pictures and then explain the sheaf-theoretic model of textured scenes, involving several different structure sheaves.

The topology we shall use is discrete and is induced by certain norms, transplanted from the structure of integers. Needless to say, the purpose of this topology is to make precise the use of such notions as continuity and proximity.

Once the topology is fixed, we introduce a convenient system of neighborhoods, called windows. These will be used throughout this work.

Given a textured picture with the discrete topology as indicated above, we assign to every window over the picture a structure of some sort, depending on the picture under the window and, perhaps, on some fragment of prior knowledge concerning the picture. The structure in question can be something very simple such as a set of descriptions or something more involved such as a vector space, generated by the attribute vectors of the picture under the window. We emphasize the degree of generality involved in the specific choice of structures. In the implementation of our picture identification program we use a structure induced by the Fourier image of the picture function.

After the species of structures has been selected, we reduce the degree of freedom in the set of structures by assuming that the structures over any pair of windows standing in an inclusion relationship are closely related. That is, one of the structures can be transformed into the other precisely when the picture function under the two windows is continuous. The transformation, which in a general situation is called a homomorphism, depends on the picture function and possibly on a prior knowledge relevant to the picture. If we imagine that the structures carry the local picture information, then the corresponding homomorphisms tell us how this information changes as we move from one window to another. Thus the question as to when and how to join two locations on the picture is answered by the homomorphisms, interrelating some of the structures.

Topology and Metric of Digitized Pictures

One of the most efficient ways of arriving at the topological model of a digitized picture is to consider the picture as a set of cells X and coordinatize or parametrize it by the finite normed two-dimensional space of integers modulo $\langle n,m \rangle$:

$$Z^2_{n,m} = \underline{Z \times Z} / \langle n,m \rangle$$

More specifically, if $\wedge : X \rightarrow Z^2$ is a selected coordinatization function, we put

(i) $x + y = z$ iff $\hat{x} + \hat{y} = \hat{z}$; (Vector addition)

(ii) $jx = y$ iff $j * \hat{x} = \hat{y}$; (Scalar multiplication)

(iii) $x \leq y$ iff $\hat{x} \leq \hat{y}$; (Partial ordering)

(iv) $||x|| = |i| + |j|$; (Sum norm)

(v) $\langle\langle x \rangle\rangle = \mathrm{Max}(|i|,|j|)$, (Max norm)

Where $x,y \in X$, $\hat{x} = \langle i,j \rangle$, and $i,j \in Z$.

The structure $\{X,+,*,\leq, ||\ ||, \langle\langle\ \rangle\rangle\}$ is called the cellular space, where the conceptual ingredients are, respectively, vector addition, scalar multiplication, partial ordering, sum-norm (city block norm), and max-norm.

Both norms induce a discrete topology in the cellular space X. The intended interpretation of the elements of X is the <u>retina point</u>, a geometric location of the point information, which is of interest in input data. Geometrically we can think of X as a finite, rectangular, two-dimensional array of congruent squares, whose coordinates are given by a grid of pairs of integers, located at their midpoints. The advantage of defining X in this way lies in the possibility of using a coordinate-free (topological) language, and when necessary, we can carry over the concepts of vector calculus to X.

When several coordinization functions (sampling) are given, one can order them partially by the <u>fineness</u> or <u>coarseness</u> relation. The finest coordinization is usually that which is suitable for capturing the ultimately relevant local information concerning the gray level shape or color change. Clearly, a finer coordinization function leads (at least potentially) to a more complete description of a picture.

The subsets of X are subjected to generalized vector operations such as

<u>Algebraic sum</u>:

$$A + B = \{a + b \mid a \in A \text{ and } b \in B\},$$

where $A, B \subseteq X$.

We shall not use these operations in this work since other operations will play a far more important role. The horizontal and vertical rectangular subsets of X (i.e., planar intervals) are called <u>windows</u>:

A is a <u>window</u> iff $A = \{a \in X \mid x \leq a \leq y\}$, where x and y are some cells in X. Thus, windows are essentially two-dimensional intervals.

The empty window is denoted by 0. The set of all windows Wind is partially ordered by inclusion. In fact, it forms a finite distributive lattice with zero element 0, unit element X and with operations:

<u>Intersection</u>:

$$A \wedge B = A \cap B;$$

<u>Union</u>:

$$A \vee B = \bigcap \{C \in \text{Wind} \mid A \subseteq C \wedge B \subseteq C\},$$

where $A, B \in$ Wind.

The window $A \vee B$ denotes the smallest rectangle containing A and B. The lattice $< \text{Wind}, 0, x, \vee, \wedge >$ will be the basic structure in picture identification. (A similar lattice is obtained by taking the <u>convex</u> subsets of X.)

The choice of norms in X induces a special system of neighborhoods, suitable for developing the basic properties of continuous and proximal functions on X.

For every natural number p we define the p- <u>neighborhood</u> (von Neuman template) of a cell x by

$$N(x;p) = \{y \mid \|x - y\| \leq p\}.$$

If we neglect the effect of the picture boundary, a p- neighborhood forms a diamond shape cluster of cells about x.

Another system of p-<u>neighborhoods</u> (Moore templates) is defined by the max-norm:

$$M(x;p) = \{y \mid \ll x - y \gg \leq p\}.$$

These neighborhoods form square windows about x, if we forget about the effect of the picture boundary.

As pointed out in the introduction, our main interest will be in pictorial relationships such as neighboring, inside near, equidistant, perpendicular, overlap, above, etc., and in pictorial objects such as figure boundaries, regions, and the like. These are certain metric-topological entities, definable in terms of the primitives of the cellular space X. Theoretically one may think of a broader class of geometric entities (projective, affine, metrical, and topological), but this is an auxilliary issue now. We shall totally disregard at present the semantic relationships and semantic objects, induced by a particular object-world model.

The starting point of a picture representation is a picture function $p: x \to R$, whose values are called gray levels. In the case of colored pictures, the values $p(x)$ for $x \in X$ are vectors, representing the intensity of light for a fixed system of colors.

The difficulty with the picture function lies in the fact that it is a point function, as opposed to an area or set function. We need a data structure, where the point information is usefully transformed into a local or areal information which is the only one we are interested in, now.

In order to achieve this, we associate with every restriction $p|A$ of the picture function p, where A is a window in X, a structure, carrying the desired local information. But before we explain how can this be done, we shall review some of the sheaf-theoretic notions.

Presheaves of Pictorial Structures of a Given Species

What are presheaves and what are they good for? These are the questions we intend to answer in this section. As for the theoretical details, the reader may consult Bredon (1967). First, we state the general definition of a presheaf and then we give a number of concrete examples of presheaves relevant to picture theory.

Let $\langle E, \leq \rangle$ be a partially ordered set. Then by a presheaf of structures of species SIGMA on E we shall mean a pair of sets

$$S = \langle \{S_a \mid a \in E\}, \{\beta_a^b \mid a \leq b\} \rangle$$

such that for all $a,b,c \in E$ the properties displayed below are valid:

(1) S_a is a structure of species SIGMA and β_a^b for $a \leq b$ is a SIGMA-homomorphism $\beta_a^b : S_b \to S_a$, called the connecting (transition) mapping from S_b to S_a.

(2) $\beta_b^a : S_a \to S_a$ is the identity SIGMA-homomorphism (automorphism) on S.

(3) $a \leq b \wedge b \leq c \Rightarrow \beta_a^c = \beta_a^b \circ \beta_b^c$.

A presheaf on E will conveniently be denoted by $S = \{S_a ; \beta_a^b\}$.

The species SIGMA refers to the type of a structure which could be just a plain set, a set endowed with certain relations and/or operations, or anything that resembles a mathematical structure (group, vector space, automaton, etc.). The only point to be realized is that the structures in question should be of the same sort or type. The SIGMA-homomorphism may be defined in various ways, the simplest, perhaps, being the structure preserving mapping from the domain of one structure into the domain of the other.

Before we launch ourselves into a more specialized study of sheaves, it seems useful to illustrate the definition of a presheaf by a couple of intended interpretations. This will hopefully help us to envisage the picture - theoretic applications.

One of the simplest presheaves is the <u>constant presheaf</u> of sets. Here $S_a = S$ is a fixed set of elements and β_a^b is the identity mapping on the fixed set S.

(a) <u>Presheaf of continuous functions</u>

Let X be a topological space. For each $V \subseteq X$ let S_V be the set of all continuous real-valued functions $f:V \rightarrow R$, and for $W \subseteq V$ let $\beta_W^V:S_V \rightarrow S_W$ be the mapping which assigns to each $f \in S_V$ its <u>domain restriction</u> $f|W$. Then, of course, $f|W \in S_W$, since a restriction of a continuous mapping is again continuous. This construction gives a presheaf $\{S_V;\beta_V^W\}$ on the set of all subsets of X, partially ordered by inclusion. We call it the <u>presheaf of continuous real-valued functions on</u> X.

If the topological attribute "continuous" is replaced by "uniformly continuous", "proximally continuous", "differentiable", "analytic", etc., we get a whole family of new presheaves on the same space. Moreover, we get some other presheaves on X when we consider only a system of neighborhoods, e.g., $N = \{N(x;p)|\ x \in X,\ p \geq 0\}$, rather than the set of all subsets of X.

(b) <u>Presheaf of Histograms</u>

Consider a picture function $p: X \rightarrow R$ together with the lattice of all windows $< Wind, \subseteq >$ of the cellular space X. Assign to every window W a set of <u>histograms</u> S_W or more precisely, a set of distribution functions corresponding to a family of random variables, characterizing certain features of the picture p.

As connecting mappings β_W^V, choose for $V \subseteq W$ an appropriate stochastic transformation, (stochastic matrix) transforming the elements S_W into the elements of S_V. The presheaf axioms are readily verified. The structure $\{S_W; \beta_W^V\}$ is a presheaf, called the presheaf of histograms. Again, we can take the system of square windows $M = \{M(x;p) \mid x \in X, p \geq 0\}$ and consider another presheaf of histograms.

(c) Presheaf of Geometric Models

Let $p: X \to R$ be a picture function together with the lattice of windows $\langle Wind, \subseteq \rangle$ of the space X. Assign to every window a geometric model S_W, induced by the picture over W. The geometric model is essentially a set of figures (lines, circles, etc.) together with the figure attributes and their placement rules.

The connecting transformations $\beta_W^V: S_W \to S_V$ are restrictions or in a more general situation, they are certain similarity functions, assigning to every element S_W a most similar element from S_V. In the case of pictures with local gradients, some other homomorphisms may be of interest.

(d) Presheaf of Feature Spaces

Let $p: X \to R$ be a picture function and let $\langle Con, \subseteq \rangle$ be the lattice of convex subsets of X. Define S_W for $W \in Con$ as the linear space generated by the feature vectors associated with p/W via measurement, and identify each β_W^V for $W \subseteq V$ with a linear transformation. Then $\{S_W; \beta_W^V\}$ is a model for the presheaf axioms.

Several other scene analysis concepts turn out to have a sheaf - theoretic interpretation.

Simply, a presheaf is a formal device which assigns to certain local areas of a topological space a specific structure in such a way that whenever two areas are in inclusion relationship, the assigned structures are in homomorphism relationship.

The reader should see by now the connection between presheaves and picture structure identification by windowing. Often a presheaf S on $< E, \leq >$ has a relatively simple structure "locally" about every point $a \in E$.

A suggestive picture of a presheaf over a window in a cellular space is given below, in Fig. 47. It is important not to confuse the presheaf structure with the picture function.

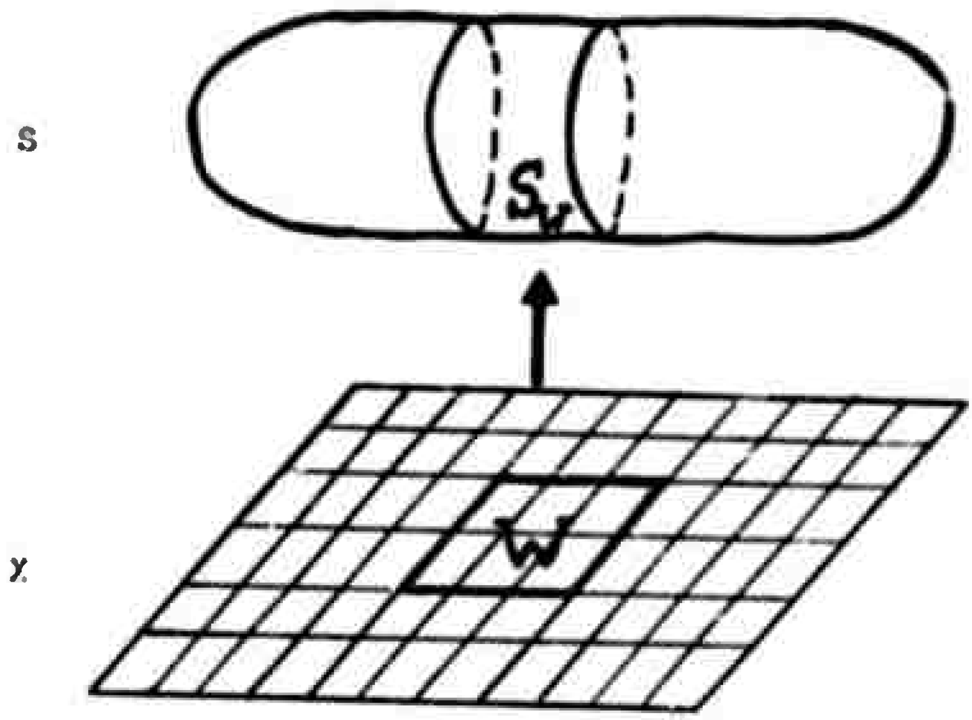


Fig. 47

Frequently we are not interested in all windows of a cellular space but rather in a subset of windows. This is the case when the texture elements are large enough and we do not want to enter into their structure. In situations like this, the following notion appears to be relevant.

Let $\mathbb{E}' \subseteq \mathbb{E}$ be a partially ordered subset of $< \mathbb{E}, \leq >$. Then the presheaf

$$< \{S_a | a \in \mathbb{E}'\}, \{\beta_a^b \mid a \leq b \wedge a, b \in \mathbb{E}'\} >$$

is called the __restriction of__ S __to__ $\mathbb{E}'$ and is denoted $S \restriction \mathbb{E}'$. Thus, the presheaf structure is considered only over some points of $\mathbb{E}$.

Often several presheaves are of interest on the same base $\mathbb{E}$. In situations like that we want to know how to relate them.

Let $S = \{S_a : \beta_a^b\}$ and $T = \{T_a : \lambda_a^b\}$ be two presheaves on $\mathbb{E}$ of the same species SIGMA.

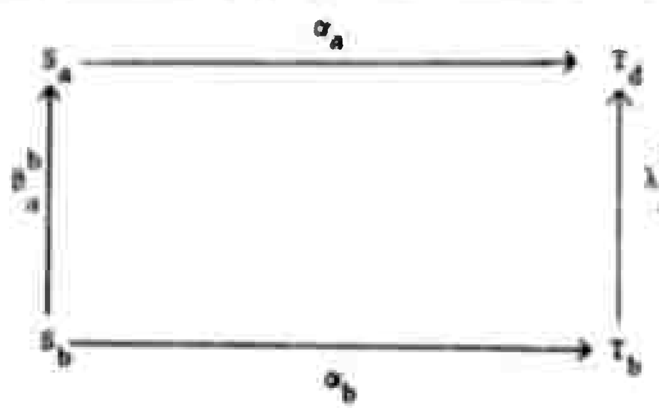
Then by a homomorphism

$$\alpha : S \to T$$

of one presheaf into another we shall mean a family of SIGMA homomorphisms $\alpha = \{\alpha_a | a \in \mathbb{E}\}$ such that for all $a, b \in \mathbb{E}$:

$$a \leq b \Rightarrow \alpha_a \circ \beta_a^b = \lambda_a^b \circ \alpha_b .$$

The condition above is explained suggestively by stating that the following diagram of functions is commutative for all $a \leq b$:



In less formal terms, the way the structures in S are related can be modeled in terms of the structures in T. This notion is important in the semantics of picture identification. We take S as a "geometric" presheaf and T as a "semantic" or "World-model" presheaf and translate the geometric information of S into a world-model information of T.

As pointed out, in picture-theoretic applications presheaves are essentially families of structures of certain species, interconnected by transformations, expressing continuity.

With every presheaf S of species SIGMA on a base $< E, \leq >$ we associate two important structures of the same species SIGMA, provided that certain existence conditions are met. Before we show how this is done, two auxiliary notions are in order.

Namely, the direct product $\mathrm{Prod}_{a \in E}\, S_a$ and the direct sum $\mathrm{Sum}_{a \in E}\, S_a$. If S_a are plain sets, then $\mathrm{Prod}_{a \in E}\, S_a$ is just the Cartesian product of the sets S_a and $\mathrm{Sum}_{a \in E}\, S_a$ is the disjoint union of these sets.

Given a presheaf S of species SIGMA then by the section (Projective or inverse limit) of S we mean the structure $\mathrm{Sect}(S) = \{s \in \mathrm{Prod}_{a \in E}\, S_a \mid \forall a,b\, [a \leq b \Rightarrow \beta_a^b (s_b) = s_a]\}$.

Thus, $\mathrm{Sect}(S)$ of a substructure of the direct product $\mathrm{Prod}_a\, S_a$ (provided that it exists).

Given a presheaf S of species SIGMA then by the cosection (inductive or direct limit) of S we understand the structure

$$\mathrm{Cosect}(S) = \mathrm{Sum}_a\, S_a / \equiv \, .$$

where $\operatorname{Sum}_a S_a = \{ \langle a,s \rangle \mid s \in S_a \}$ denotes the direct sum of the family (S_a) for $a \in E$ and $\equiv$ is the smallest equivalence relation on the direct sum, containing the binary relation $\#$;

$$\langle a,s \rangle \# \langle b,t \rangle \iff \exists\, c \le a,b \mid \beta^a_c(s) = \beta^b_c(t) \,],$$

with $s \in S_a$ and $t \in S_b$.

Thus, Cosect(S) is a <u>quotient structure</u> of the direct sum $\operatorname{Sum}_a S_a$ (provided that it exists).

Given a presheaf S, it can be shown that if "e" is a <u>first</u> element of E, then the isomorphism

$$S_e \cong \operatorname{Sect}(S)$$

holds and also

$$\sigma: S \to T \implies \sigma_{e'}: S_e \to T_e .$$

In other words, the structure Sect(S) can be identified with the structure S_e, assigned to the first element e in E. In applications, Sect(S) corresponds to the structure of particular texture elements.

Dually, if e′ is a <u>last</u> element of E, then

$$S_{e'} \cong \operatorname{Cosect}(S)$$

and

$$\sigma: S \to T \implies \sigma_{e'}: S_{e'} \to T_{e'} .$$

Again, Cosect(S) can be identified with the structure $S_{e'}$, assigned to the last element e′ in E. In applications, Cosect(S) corresponds to the structure of the textured region.

<u>Sheaves of Pictorial Structures of a Given Species</u>

Sheaves are presheaves satisfying additional axioms. A definition of a sheaf in its full generality requires several additional technicalities. We shall present therefore such a definition which is free of abstract conceptual constructions, general enough, and yet still relevant in picture analysis.

Let $< Wind, \subseteq >$ be the lattice of windows over the cellular space X. Consider a presheaf $\{S_W, \rho_W^V\}$ over the window system $< Wind, \subseteq >$. Then this presheaf is called a <u>sheaf</u> if for any family of windows $\{W_i\}_{i=1}^n$ such that $W_i \subseteq W \rightarrow W \in \{W_i\}_{i=1}^n$, the following isomorphism is valid:

$$S_{\bigcap_i W_i} \simeq \text{Sect}\,(S|\{W_i\}).$$

Thus, loosely speaking, a sheaf is a system of structures over a lattice of windows, where each structure represents one particular texture.

Dually, we call presheaf $\{S_W, \rho_W^V\}$ a <u>cosheaf</u> if for any family of windows $\{W_i\}_{i=1}^n$ such that $W \subseteq W_i \rightarrow \in \{W_i\}_{i=1}^n$ the following isomorphism is valid:

$$S_{\bigcup_i W_i} \simeq \text{Cosect}\,(S|\{W_i\}).$$

A more direct definition of a sheaf with a fairly clear picture-theoretic interpretation is given below.

Consider a presheaf $S = \{S_V; \rho_V^W\}$ of structures over a cellular space X, i.e., on the lattice of subsets $< Sub(X), \subseteq >$. Then S is a <u>sheaf</u> over X precisely when for any family $\{V_i \mid i \in I\}$ of subsets of X with $V = \bigcup_i V_i$, the following two conditions are satisfied:

(1) Uniqueness Axiom:

$$\forall i [\beta^{V}_{V_i}(s') = \beta^{V}_{V_i}(s'')] \Rightarrow s' = s'';$$

(2) Existence Axiom:

$$\forall i,j [\beta^{V_i}_{V_i \cap V_j}(s_i) = \beta^{V_j}_{V_i \cap V_j}(s_j)] \Rightarrow \quad \exists s \forall k [\beta^{V}_{V_k}(s) = s_k],$$

where $s, s', s'' \in S_V, s_i \in S_{V_i}, s_j \in S_V, s_j \in S_{V_j}, S_k \in S_{V_k}$.

The condition (1) says that if the structure elements s are locally identical, then they are also globally identical. That is, elements are uniquely determined by local data.

The condition (2) says that if we have local data which are compatible, they actually "patch together" to form global data.

This might appear as a perhaps unduly sophisticated way of looking at the windowing process in which by overlapping windowing we are capable to recover the unique structure of the picture from several local structures. The definition of a sheaf will turn out to be a test method for texture region identification.

The picture - theoretic substance of sheaves is this. A sheaf is essentially a system of "local coefficient". In picture-theoretic applications we start by assuming that a picture has certain local pictorial properties which are captured by a structure S_V of certain species. We then express these properties in terms of the properties of the structure sheaf S over a picture region. Finally, we apply the theory of sheaves to deduce certain global properties of the picture. Consequently, the importance of sheaves in scene analysis is simply in giving relations between local and global properties of a scene.

REFERENCES

[1] Andrews, H.C., Computer Techniques in Image Processing, New York Academic Press, 1970.

[2] Atteneave, F., and Arnoult, M.D., "The Quantitative Study of Shape and Pattern Perception", Psychol. Bull., 53 (1956), 452-71.

[3] Atteneave, F., and Olson, R.K., "What Variables Produce Similarity Grouping?", The American J. of Psychology, 83 (1970), 1-21.

[4] Bajcsy, R., "Computer Texture Analysis" in B.S.M. Gransberg (Ed.), Proceedings of the Third Hawaii International Conference on System Sciences, Western Periodical Co., 1970, 1010-1015.

[5] Beck, J., "Perceptual Grouping Produced by Line Figures", Perception and Psychophysics 2, (1967), 491-495.

[6] Bennet, R.S., "Intrinsic Dimensionality of Signal Collections", IEEE Trans. Inform. Theory, Vol. IT-15, (1969), 517-525.

[7] Brice, C.R., and Fennema, C.L., "Scene Analysis Using Regions", Artificial Intelligence, Vol. 1, No. 3, (Fall 1970).

[8] Binford, T., "A Visual Preprocessor", Internal Report, Cambridge, Mass., MAC, MIT, (1970).

[9] Blakemore, C. and Campbell, F.W., "On the Existence of Neurones in the Human Visual System Selectively Sensitive to the Orientation and Size of the Retinal Images", J. Psychol. 203, (1969), 327-260.

[10] Bredon, G.E., Sheaf Theory, New York, McGraw-Hill, 1967.

[11] Brill, E.L., "Character Recognition via Fourier Descriptors", WESCON Technical Papers, Session 25, Qualitative Pattern Recognition Through Image Shaping, Los Angeles, August, 1968.

[12] Brom, D.R., and Oewen, D.H., "The Metrics of Visual Form", Psychol. Bull., 68, (1968), 243-259.

[13] Campbell, F.W. and Kulikowski, J.J., "Oriental Selectivity of the Human Visual System", J. Psychol. 187, (1966), 437-445.

[14] Campbell, F.W., and Robson, J.G., "Application of Fourier Analysis to the Visibility of Gratings", J. Psychol. 197, (1968), 551-566.

[15] Clowes, M.B., "Picture Syntax", In: S. Kaneff (Ed.), Picture Language Machin., New York, Academic Press, 1970.

[16] Dacey, M.I., "Description of Line Patterns", Quantitative Geography, Northwestern Univ. Studies in Geog., No. 13, Evanston, 1967, 277-87.

[17] Darling, R.M., Joseph, R.D., "Pattern Recognition from Satellite Altitudes", IEEE Trans., SSC-4, (1968), 38-47.

[18] Duda, R.O. and Hart, P., Pattern Recognition and Picture Processing, (forthcoming), 1972.

[19] Enroth-Cugel, C., and Robson, J.G., "The Contrast Sensitivity of Retinal Ganglion Cells of the Cat", J. Psychol., 187, (1966), 517-552.

[20] Falk, G., "Computer Interpretation of Imperfect Line Data as a Three-dimensional Scene", AIM-132, Artificial Intelligence Project, Stanford University, Stanford, California (1970).

[21] Frish, H.L., and Julesz, B., "Figure-ground Perception and Random Geometry", Perception and Psychophysics, 1, (1966), 389-398.

[22] Flock, H.R., "Optical Texture and Linear Perspective as Stimuli For Slant Perception", Psychol. Review, 72, (1965), 505-514.

[23] Freeman, R.B., Jr., "Theory of Cues and the Psychophysics of Visual Space Perception", Psychol. Mongr. Supll., (1970), Vol. 3, No. 13 (Whole No. 45).

[24] Fukunaga, K., and Olson, D.R., "An Algorithm For Finding Intrinsic Dimenionality of Data", IEEE Trans. on Computers, Vol. C-20, No. 2, 1971.

[25] Gibson, J.J., "The Perception of the Visual World", Boston, Houghton Mifflin, 1950.

[26] Goodman, J.W., Introduction to Fourier Optics, San Francisco, McGraw-Hill Book Co., 1968.

[27] Guzman, A., "Decomposition of a Visual Scene Into Three-dimensional Bodies", Proc. Fall Joint Computer Conf., AFIPS 33, Washington, Thompson Book, 1968, 291-304.

[28] Hawkins, J.K., "Textural Properties for Pattern Recognition", In: B.S. Lipkin (Ed.), Picture Processing and Psychopictorics, New York, Acad. Press, 1970.

[29] Horn, B.K.P., "Shape From Shading: A Method for Obtaining the Shape of a Smooth Opaque Object From One View", MAC-TR-79, Project MAC, MIT, Cambridge, Mass. (November 1970).

[30] Huang, T.S., 'Digital Fourier Analysis.

[31] Hueckel, M.H., "An Operator Which Locates Edges in Digitized Pictures", J. of ACM, Vol. 18, No. 1 (Jan. 1971), 113-125.

[32] Hubel, D.H., "Single Unit Activity in Lateral Geniculate Body and Optic Tract of Unrestrained Cats", Psychol., 150, (1960), 91-104.

[33] Hubel, D.H., and Wiesel, T.N., "Perceptive Fields of Optic Nerve Fibres in the Spider Monkey", J. Psychol., 154, (1960), 572-580.

[34] Hubel, D.H., and Wiesel, T.N., "Perceptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex", J. Psychol., 160, (1962), 106-154.

[35] Hubel, D.H., and Wiesel, T.N., "Receptive Fields and Functional Architecture of Monkey Striate Cortex", J. Psychol. 195, (1968), 215-243.

[36] Johnson, S.C., "Hierarchical Clustering Schemes", Psychometrica 32, (1970), 241-54.

[37] Julesz, B., "Visual Pattern Discrimination", IRE Trans. on Information Theory, IT-8, (1962), 84-92.

[38] Julesz, B., "Some Recent Studies in Vision Relevant to Form Perception", In W. Shaten-Dunn, (Ed.), Models for the Perception of Speech and Visual Form, Cambridge, Mass., MIT Press, 1967, 136-154.

[39] Julesz, B., and Stromeyer, C.F., "Masking of Spatial Gratings by Filtered One-dimensional Visual Noise", Talk at the 10th Annual Meeting of the Psychonomic Society, San Antonio, Texas, Nov. 5-7, 1970.

[40] Julesz, B., Foundations of Cyclopean Perception, Chicago, The University of Chicago Press, 1971.

[41] Kirsch, R.A., "Computer Interpretation of English-Text and Picture Patterns", IEEE Trans. On Electric Computers, EC-13, No. 4, (1964) p. 63-76.

[42] Kovalevsky, V.A., "Present and Future of Pattern Recognition Theory", In Proc. IFIP Congress, Vol. I, Spartan Books, Inc., 1965, pp. 37-43.

[43] Krakauer, L.J., "Computer Analysis of Visual Properties of Curved Objects", MAC-TR-82, Project MAC, MIT, Cambridge, (May 1971).

[44] Kruskal, J.B., "Multi-dimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis", Psychometrica, 29 (1964), 1-27, 115-129.

[45] Kuffler, S.W., "Discharge Patterns and Functional Organization of Mamualim Retina", J. Neurophysiol., 16, (1953), 37-68.

[46] Ledley, R.S., Rotolo, L.S., Golab, I.J., Jacobson, J.D., Ginsberg, M.D., and Wilson, J.B., "FIDAC: Film Input to Digital Automatic Computer and Associated Syntax-directed Pattern Recognition Programming System",

In J.T. Tippett (Ed.) Optical and Electro-Optical Information Processing, Cambridge, Mass., MIT Press, 1965.

[47] Levine, M.D., "Feature Extraction: A Survey", Proc. IEEE, 57, (1969), 1391-1407.

[48] Lendaris, G.G., Stanley, G.L., "Diffraction Pattern Sampling for Automatic Pattern Recognition", Proc. IEEE, 58, (1970), 198-216.

[49] Lipkin, L.E., et al., "The Analysis, Synthesis, and Description of Biological Images", Am. N.Y. Acad. Science, 128, (1966), 984-1012.

[50] Miller, W.F., Shaw, A.C., "Linguistic Methods in Picture Processing: A Survey", Proc. AFIPS, 1968 FJCC, Vol. 33, Washington, D.C., Thompson Book Co., pp 279-290.

[51] Mountcastle, V.B., "Modality and Topographic Properties of Single Neurons of Cat's Somatic Sensory Cortex", J. Neurophysiology, 20, (1957), 408-434.

[52] Munson, J.H., "Experiments on the Recognition of Hand-Printed Text", Part I - Character Recognition, Proc. FJCC, 1968, 1125-1138.

[53] Nachmias, J., Sachs, M.B., and Robson, J.G., "Independent Spatial-Frequency Channels in Human Vision", J. Opt. Soc. Am., 59 (1969), 1538.

[54] Narasimhan, R., "Syntax-directed Interpretation of Classes of Pictures", Comm. ACM, 9, (1966), 166-173.

[55] Narasimhan, R., "Picture Languages", In (Ed.) S. Kaneff, Picture Language Machines, 1970, New York, Academic Press.

[56] Novikoff, A.E.J., "Integral Geometry as a Tool in Pattern Perception", In (Ed.) H. von Foerster and G.W. Fopf, Principles of Self-Organization, 1962, New York, Pergamon Press.

[57] Pantle, A., and Sekuler, R.W., "Contrast Response of Human Visual Mechanisms Sensitive to Orientation and Detection of Motion", Vision Res. 9, (1968), 397-406.

[58] Papoulis, A., Systems and Transforms with Application in Optics, New York, McGraw-Hill Book Co., 1968.

[59] Pfaltz, J.L., "Web Grammars and Picture Description", Techn. Report, 70-138, 6J-754, University of Maryland, Computer Science Center, 1970.

[60] Pingle, K.K., "Visual Perception by a Computer", In (Ed.) A. Grasseli, Interpretation and Classification of Images, 1969, New York, Academic Press.

[61] Pollehn, H., and Roehnig, H., "Effect of Noise on the Modulation

Transfer Function of the Visual Channel", J. Opt. Soc. Am., 60, (1970), 842-848.

[62] Powel, T.P.S., and Mountcastle, V.B., "The Cytoarchitecture of the Postcentral Gyrus of the Monkey Macaca Mulatta", Bull. John Hopkins Hosp., 105, (1959), 108-131.

[63] Prewit, J.M.S., and Mendelson, M.L., "The Analysis of Cell Images", Annals N.Y. Acad. Science, 128, (1968), 1035-1053.

[64] Quam, L.H., "Computer Comparison of Pictures", AIM-144, Artificial Intelligence Project, Stanford University, Stanford, California, (May 1971).

[65] Rosenfeld, A., "Automatic Recognition of Basic Terrain Types From Aerial Photographs", Photogram. Eng., 28, (1962), 115-132.

[66] Rosenfeld, A., "On Models for the Perception of Visual Texture", In Models for the Perception of Speech and Visual Form, Proceedings of a Symposium, Nov. 11-14, 1964.

[67] Rosenfeld, A., and Pfaltz, J.L., "Distance Functions on Digital Pictures", Pattern Recognition, Vol. 1 (1968), 33-61.

[68] Rosenfeld, A., Picture Processing by Computer, Academic Press, New York, 1969.

[69] Rosenfeld, A., Pfaltz, J.L., "Web Grammars", TR-69-84, Computer Sciences Center, University of Maryland, College Park, Maryland (1969).

[70] Rosenfeld, A., and Thurston, M., "Edge and Curve Detection For Visual Scene Analysis", IEEE Trans. On Computers, Vol. C-20, No. 5, May 1971.

[71] Shaw, A.C., "The Formal Description and Parsing of Pictures", SLAC Report No. 84, Stanford Linear Accelerator Center, Stanford University, Stanford, California, March 1968.

[72] Shepard, R.N., "The Analysis of Proximities: Multi-dimensional Scaling With an Unknown Distance Function, I and II", Psychometrica, 27, (1962), 125-140, 219-246.

[73] Shepard, R.N. Caroll, J.D., "Parametric Representation of Nonlinear Data Structures", In Multivariate Analysis, (Ed.) P.R. Krishnaiah, New York, Academic Press, 1966, pp. 561-592.

[74] Sobel, I., "Camera Models and Machine Perception", AIM-121, Artificial Intelligence Project, Stanford University, Stanford, California (1970).

[75] Tenenbaum, J.M., "Accommodation in Computer Vision", AIM-134, Artificial Intelligence Project, Stanford University, Stanford, California (1970).

[76] Ward, T.N., Jr., "Hierarchical Grouping to Optimize an Objective Function",

J. Am. Statistical Assoc., 58 (1963), 236-244.

[77] Wertheimer, M., "Experimentelle Studien Uber das Sehen von Bewegung", Z. Psychol., 61 (1972) pp. 161-265, Translated in large part in: T. Shipley (Ed.), Classics in Psychology, New York, Philosophical Library, 1961.

[78] Wiesel, T.N., and Hubel, D.H., "Spatial and Chromatic Interaction in the Lateral Geniculate Body of the Rhesus Monkey", J. Neurophysiol., 29, (1966), 1115-1156.

[79] Wolf, P.D., Term Project, MIT (1970).