Part II

DESCRIBING AND ABSTRACTING PICTORIAL DATA

Section 1
INTRODUCTION

To index and abstract pictorial data, much of the information concerning metric information must be discarded and only the objects, their properties, their relationships to other objects, and their locations are retained. Then, by using only a limited number of object names, modifiers, and relations, a wide class of pictures can be represented for future retrieval in response to general queries. Thus, one important purpose of picture description is that of generalization, and in our investigation the retrieval process consists of trying to match the query structure to general description structures.

It is possible to approach the study of pictorial data in either a formal or a nonformal manner. That is, one can attempt to present the descriptions in a formal, rule-bound, and analytic manner, or one can approach the problem in an analogical, intuitive manner. Much recent work has tended to be formal in nature (Refs. 1, 2) and has produced disappointingly few results relevant to our interests. In our study, we have taken a nonformal approach, studying pictorial description from the point of view of language in use.

Over the past several years, we have examined the variety of ways in which human observers describe and abstract pictorial structures. After a review of the field (Ref. 3), we carried out experiments in which descriptions of pictorial data prepared by human observers were examined. These experiments, reported in Ref. 4, gave us a data base of descriptive material and allowed us to examine the terminology and structure of such descriptions. In the first part of the present contract, we examined the problem of description of earth resources imagery* and indicated in Ref. 5 how a "dynamic" data entry system, i.e., one which allows for the entry of more information concerning a given picture obtained from later observers, could be used to obtain an

---

*See Table 1-1 for typical descriptions of the various earth resources disciplines.

LOCKHEED PALO ALTO RESEARCH LABORATORY
LOCKHEED MISSILES & SPACE COMPANY
A GROUP DIVISION OF LOCKHEED AIRCRAFT CORPORATION

Table 1-1

TYPICAL DESCRIPTIONS FOR THE VARIOUS EARTH RESOURCES DISCIPLINES

Geography

The Nile River and its distributaries within the delta appear as sinuous buff-colored ribbons. The main stream lies against the east (right) side of the valley to the south. The Rosetta Distributary is seen winding along the west margin of the delta, while the Damietta Distributary roughly bisects the irrigated area. Meanders and islands are everywhere in evidence. The major canal system appears as a series of straight buff-colored lines radiating outward from the settlement nodes. Such straightline segments are nearly always canals rather than roads.

Agriculture

Infrared photo taken from an airplane at Weslaco,Texas, shows crop vigor, growth, and soil salinity: (1) healthy cotton, (2) unhealthy cotton, (3) bare soil, (4) pig weeds in wet area, (5) pig weeds above short sorghum, (6) dry topsoil between rows of sorghum, and (7) bare soil between sorghum in moist areas.

Forestry

Figure 13 is an Ektachrome infrared picture taken by the Apollo 9 astronauts of the test site near Vicksburgh, Mississippi.... The dark blotches within this area are uncut patches of hardwood, and a large area of hardwood forest can be seen toward the lower right corner of the photo. Although the seasonal state of deciduous forest affects the appearance of such areas, little difficulty is encountered in distinguishing between cultivated land and forests in space photography.

Geology

(A Gemini 4 photograph of southwest Saudi Arabia.) Most of the upper half of the photograph is occupied by a range of rugged mountains composed of igneous rocks. Toward the base is an expanse of sedimentary rocks that dip very gently toward the upper right of the photograph. The distinction between igneous and sedimentary rocks is made by noting the presence of stratification, or layering, in sedimentary rocks. These two units appear to be separated by a major fault.

Hydrology

On the far side of the Tigris River shown in Fig. 22 is one end of a large lake that, judging from the pattern of dark tones around it, varies in size and is now at a low stage. Between the lake and the river are distributaries – subsidiary streams that carry water away from the main stream and which branch downstream. Such streams deposit large amounts of fine-grained sediment.

Oceanography

(A Gemini photo of Campeche, Yucatan.) Some of the prominent features here are longshore and offshore currents and a sandbar development. An easterly longshore current is made visible by the sediments discharged into the Gulf of Mexico's clear water by a river immediately to the left of the photo margin. The sediments are light-colored marls, silts, and sands. To the right of the laguna, the sandbar development is clearly visible.

Cartography

The valley of the Rio Grand cuts across the picture from the upper left to the bottom center. The Franklin and Organ Mountains are near the center. Below them are El Paso, Texas, and Jaurez, Mexico. These cities are in a strategic location in a pass in the range of mountains. There are irrigated areas on both sides of the international borderline. The Rio Grande Valley is slightly depressed and our highways pass through El Paso.

"encyclopedic" data base. The use of a relational net for representing the content of picture description was described, and methods for merging and searching such structures were indicated. A summary of this paper is given in Appendix A. The second portion of this contract period was devoted to the study of representations of the content of descriptions of terrain photography written in natural language. This report is devoted to that problem.

There are three basic problem areas in representing the content of a description expressed in natural language:

(1) The problem of finding canonical forms, i.e., how to represent paraphrases of the same concept
(2) How to translate from natural language to this representation automatically
(3) How to organize the data base so as to keep the representations simple, and still capture the full conceptual intent of the description

- The representation problem. Given a set of sentences such as:

  Many wharves line the river.
  On the sides of the river are a lot of wharves.
  Along the river are a number of wharves.

There are many variations of these sentences using "docks," "river," and "many" in place of "wharves," "stream," and "lot of," respectively, and it is important that these variations be given the same conceptual representation. The question of selection of the primitive terms used in the conceptual representation arises, e.g., should "side of" be considered a primitive, or should it be represented as "on the outside edge of an object, "using the primitives "on," "outside," and "edge." Finally, the conceptual representation must indicate the semantic role played by the primitives.

- The translation problem. The problem of translating from natural language expressions to a conceptual representation is very difficult. Because there are often many allowable parsings of a given sentence if only syntactic criteria are used, it is necessary to use additional aids. As indicated in Part I, section 2.1, one possible tool that can be used is the word government table, developed earlier in this study by Robison (Ref. 6), a dictionary of words and allowable adjacent syntactic structures for each word meaning. These associated syntactic structures pinpoint a given word's semantic meaning, and can therefore be used as an aid in deriving the conceptual representation of the sentence.

- Data base organization. If we have the description, "an airport is to the left of the freeway," and represent this as left of (freeway, airport), then a query such as "what is to the right of the airport?" cannot be answered unless the data base has stored the property of the relationship, "left of (a, b) implies right of (b, a)." Similarly, the system must contain axioms which indicate that if an object is inside another object, then it cannot also be outside that object, and that if an object is far from an object, then it is outside that object. If such properties of relations are not included in the system as general rules, then either the representation of each description becomes encumbered with a large number of relationships, or, if such relationships are not present in the representation, then the ability to respond to queries is diminished.

Section 2
REPRESENTATION OF MEANING

A parsing of a sentence indicates how the words interact, and which words or groups of words are related on various levels of aggregation. However, to represent the content or meaning of a sentence, one must indicate the conceptual roles played by the words or word groupings. For example, a representation of meaning should indicate whether a term relates to another so as to indicate quantity ("three lakes"), location ("at the river"), or attribute of an object ("waterway has docks"), etc. It is therefore important in any investigation of representation of meaning to identify these classes, to study natural language expressions representative of their use, and to determine effective techniques for representing content as a function of these classes.

2.1 CONCEPTUAL CLASSES FOR PICTURE DESCRIPTION

Consider the description of an aerial photograph given in Table 2-1. In the table below the description, we have indicated the "concept kernels" – phrases which capture a basic idea or concept. On the right, we have noted the general concept class (or classes) of each kernel. If we represent the meaning of the description by a network in which terms are shown as nodes and the relationships by links, then we obtain a diagram such as is shown in Fig. 2-1. In that figure, the concept classes have been used to label the links with the conceptual role played by that link.

Such conceptual classes serve several roles. First of all, they help to organize the study of terms and expressions used in description, since we can examine terms and expressions representative of each class, rather than having to consider all terms and expressions. Second, they can be used in the conceptual net to indicate the semantic role played by relationships and thus simplify the search and analysis tasks when extracting information from the net.

Table 2-1

BASIC CONCEPT CLASSES OF A DESCRIPTION

Description by human observer

It is an aerial photograph, whose most arresting feature is a waterway, possibly an estuary or part of a harbor, with many ship docks. The waterway stretches across the upper part of the photograph with a single stream on the left breaking into three arms at not quite midphoto. The upper arm divides again into inlets cut off from view; at the top, the middle arm is a small, straight-sided, closed inlet or docking area in a peninsula or island; the lower arm passes out of the picture on the right, possibly dividing again at the right edge. Across the waterway just to the left of the division is a bridge.

Concept Kernels of the Description	Concept Classes
It is an aerial photograph	Photograph as physical object
Most arresting feature is a waterway	Attribute of object
Waterway is possibly an estuary or part of a harbor	Classification of object
Waterway has many ship docks	Set membership (has)
Waterway stretches across upper part of the photograph	Operative (stretches across)
Waterway breaks into three arms	Operative (partitions)
Break is at not quite midphoto	Location of object
The upper arm divides again into inlets	Operative (partitions)
Inlets are cut off from view	Existence
At the top of the middle arm is a docking area or inlet	Existence location
Inlet/docking area is straight-sided and closed	Attribute of object
Inlet/docking area is in a peninsula or island	Location of object
The lower arm passes out of the picture on the right	Existence (viewability)
The lower arm possibly divides again at the right side	Operative (partitions)
Across the waterway is a bridge	Operative (crosses)
Bridge is located just to the left of the division of the waterway	Location of object

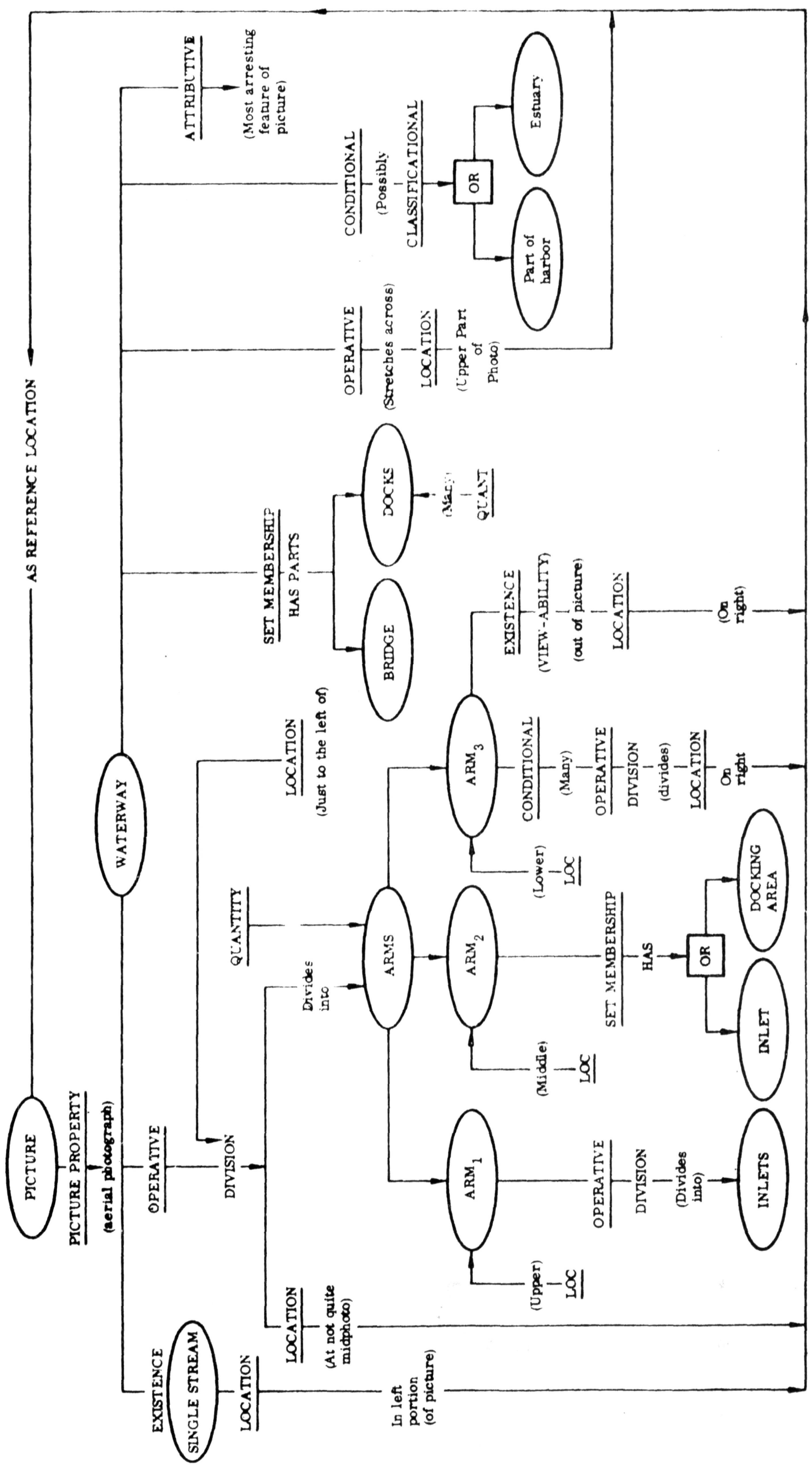


Fig. 2-1 Example of a Conceptual Net for a Picture Description

Finally, many verbs are associated with certain conceptual classes and allowable word forms, and such classes lead to constraints on the structure of the sentence. These constraints can be analyzed and tabulated for use as a guide in performing both the syntactic and semantic analysis. Recently, several approaches to linguistic analysis based on concept classes or "cases" have been presented by investigators such as Fillmore, Schank, and Tesler. We have examined case grammars which identify the underlying syntactic-semantic relationships, and give a brief summary of this work in Appendix A.

In examining descriptions of aerial photographs, we have chosen the following set of concept classes for representing meaning of picture description.

- 1-8 Case relations. Relations such as (1) location, (2) localizing, (3) qualifying (4) quantifying, (6) operative, (7) use, (8) existence, which are often strongly related to the verb of the sentence.
- 9 Attributes. Attributes of an object, such as color, texture, shape, size. Many of the attributes have a value, e.g., "red," "rough," etc.
- 10-Comparison. Comparisons can be made on the values of attributes, e.g., "redder than," "rougher than," "bigger than."
- 11-Logical linking. Linking of concepts using logical relationships such as "and," "or," "not," "if....then"
- 12-Set membership. Indication that an object is a subset or part of another object.
- 13-Doubt, approximation, or uncertainty. Indications of doubt, approximation, or uncertainty on the part of the observer.
- 14-Picture properties. Expressions which deal with the physical properties of the picture, such as size, resolution, and picture quality.

Some of these conceptual classes are described more completely below.

2.1.1 Case Relations

The first eight conceptual classes grouped together under the term "case relations" are listed in Table 2-2 which shows the role played by each case and examples of each case. It is expected that the list will be modified as the analysis proceeds.

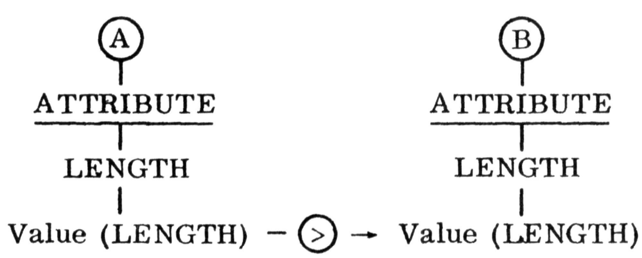
Because location tends to figure strongly in observer descriptions of pictures, we investigated expressions for location in some detail. In Table 2-3 (a tabulation of expressions used to denote location), expressions appearing in the same box have roughly the same meaning, while those appearing in adjacent columns are either antonymns, or express somewhat related ideas. Similar concepts have been grouped together, and the column on the left indicates the name of the grouping whenever a suitable general name can be found. The question of suitable primitives for these expressions will be discussed in section 2.2.2.

2.1.2 The Remaining Conceptual Classes

The remaining six conceptual classes are described briefly below.

- Attributes. Various attributes that appear in pictorial description are indicated in Table 2-4. Each attribute has a value, e.g., the value of "color" may be "red," and often the values of the attributes can be compared, as described below.

- Comparison. Comparison of attribute values can be indicated in the conceptual net as follows. Suppose we wish to indicate, "object A is longer than object B." We use the net,



to indicate that the value of the LENGTH attribute of A is bigger than that of B.

Table 2-2

CASE RELATIONS USED IN PICTURE DESCRIPTION

Case	Role Played	Examples
1. Location/extent	Location of object with respect to frame of picture, or with respect to another object. Also extent of object	X is near Y, X is west of Y (see Table 2-4)
2. Localizing	Portions of objects (can be thought of as location with respect to object itself)	Top of X, edge of X middle of X
3. Qualifying	Which of an object is meant	The upper arm (to distinguish this arm from other arms)
4. Quantifying	How much or how many	Three arms, few, many
5. Classificational	An entity in the photo is identified as to class or name	X is a Y X appears to be a Y Y seems to be a Y X is similar to a Y X resembles a Y X is classified as a Y X is related to a Y
6. Operative	An entity does something to another entity (usually intrusion, crossing, or cutting)	X crosses Y X cuts Y X disrupts Y X intertwines Y X invades Y X divides Y X stretches across Y
7. Use or application	Purpose of object	X developed for Y X used by Y
8. Existence	An entity can be seen	There is an X an X can be seen an X is visible an X can be found an X is observed an X will be noted an X appears

Table 2-3

EXPRESSIONS FOR LOCATION

Concept	Expressions Used		
Distance to	X is at Y X is in the vicinity of Y X is by Y X is with Y X is close to Y X is near Y X is toward Y X is in place Y	X is far from Y X is away from Y X is distant from	X is (distance) from Y
	X is on Y X is above Y X is over Y	X is under Y X is underneath Y X is below Y X is beneath Y	X is off Y
	X is in Y X is inside Y X is within Y	X is out of Y X is outside of Y	X is along Y
	X is here	X is here	X is everywhere
	X is between Y and Z		
	X is around Y X is alone Y		
	X is among Y		
Relative position	X is to the left of Y	X is to the right of Y	to the side of
	X is East of etc.	X is West of etc.	
	X is in front of Y	X is in back of Y X is behind Y	
Orientation	X is perpendicular to Y X is a diagonal	X is parallel to Y X runs north (south....)	X makes an angle of Q with Y

Table 2-4

ATTRIBUTES OF OBJECTS

Attributes	Values of Attributes	Comparatives of Values
Shape	curved straight round	more curved, curvier, straighter rounder same shape as
Size Length Area	big, little short, long, 5 in. big, small, one acre	greater than, smaller than, equal
Color	red, blue	redder than, bluer than, same color as
Grey Level	light, dark bright, dull	lighter than, darker than, same grey level as, brighter, duller
Texture	grainy, sandy, mottled smooth sprinkled with scattering of clumps of	rougher than, grainier than, rough as smoother than smooth as

- Logical linking. Objects are often linked using logical relationships such as "and," "or," "not," "since x, then y," etc. Such linking can be indicated using the standard logical notation. Some of the expressions used are shown below.

Logic Expression	Symbol Used	Natural Language Form
and	*	roads and mountains
or	+	oil or water
not	~	does not cross
if X then Y	X → Y	X shows that Y X indicates that Y X confirms that Y X gives evidence that Y judging from X, then Y because of X, then Y

- Set membership. Indications that an entity is a subset of another entity are given by such natural language expressions as:

A is part of B
A is an element of B
A belongs to B

- Doubt/uncertainty/approximation. It is important that the links of the conceptual net retain an indication of doubt, uncertainty, or approximation concerning observations, size, and quantity. Thus, a value of an attribute should be labeled as "approximately," and that of an observation "possibly," when so indicated by the observer. It may be desirable to use a scale of approximation or of possibility, but this should offer no difficulty.

- Picture properties. Because of the importance of such picture properties as clarity, resolution, scale, etc., we have separated this topic area out as a conceptual class.

2.2 PICTURE PRIMITIVES

The terms appearing in the relational net should be selections from a basic set of primitives to enable matching to be performed between queries in canonical form and the data base. Primitives appropriate for the various conceptual classes are described below.

2.2.1 Attribute Primitives

Some of the attributes concerning size, quantity, and grey level (given previously in Table 2-3) permit representation in terms of a small set of primitives. For example, if one defines BIG as a general primitive denoting a large value of an attribute, then we can define

QUANTITY:		LENGTH:	
few	= ~ BIG	short	= ~ BIG
a lot, many	= BIG	long	= BIG
several, some	= ~ (BIG* ~ BIG)	not long and not short	= ~ (BIG* ~ BIG)
SIZE:		GREY LEVEL:	
little, small	= ~ BIG	light	= ~ BIG
large, big	= BIG	dark	= BIG
moderate sized	= ~ (BIG* ~ BIG)	greyish, not dark and not light	= ~ (BIG* ~ BIG)

However, the primitives for shape, direction, color, and texture are not as straightforward. In the case of shape, for example, one could define as primitives a few selected geometric shapes. Human description of shape, however, often uses the names of objects to indicate shape, e.g., "S-shaped," "hook-shaped," and it is not clear whether any set of primitives of reasonable size can be found to cover the spectrum of such descriptions.

For color, one can use the primary colors, indicating the other colors in terms of primaries, e.g., "orange" would be "red/yellow," and shades could be indicated using BIG. Thus, an orange that was on the reddish side could be indicated "BIG-red/yellow." Texture is a difficult class to deal with in terms of a small set of primitives, and it is not clear as to what an acceptable set of such primitives might be.

2.2.2 Location Primitives

A tentative set of 15 to 20 location primitives chosen from Table 2-3 is currently being evaluated. The number of such location primitives can be reduced by defining a set of primitives based on the idea of an object being inside or outside a suitably defined space or region. Tesler (Ref. 8) has used such a set of primitives, as follows:

space(Y)	=	the space defined by the boundaries of object Y
nearspace(Y)	=	the space near object Y, where "near" would depend on the particular problem area
betweenspace(Y, Z)	=	the space between Y and Z
amongspace(Y, Z, . . .)	=	the space in a region of a set of objects

Each of the spaces or regions is assumed to have an edge, as does an object. One then can define as primitives "inside," "outside," and "on," having their usual meanings. We can then define many of the location terms using these primitives as follows:

Expression	In Terms of Primitives	Paraphrase of the Notation
X is inside Y	inside(X, space(Y))	X is inside the space of Y
X is far from Y	outside(X, nearspace(Y))	X is outside the nearspace of Y
X is along Y	on(X, edge (nearspace(Y))	X is on the edge of the nearspace of Y
X is between y and z	inside(X, betweenspace(Y, Z))	X is inside the betweenspace of Y and Z

Localizing terms such as "top of," "bottom of," "right side of," which indicate portions of an object are thought of as defining a region. Thus, the outside of "top" is "above," while its inside is "below."

2.2.3 Localizing Primitives

Localizing terms such as "top," "bottom," and "side" will be expressed in terms of the following primitives.

LOCKHEED PALO ALTO RESEARCH LABORATORY
LOCKHEED MISSILES & SPACE COMPANY
A GROUP DIVISION OF LOCKHEED AIRCRAFT CORPORATION

Primitives	Examples of Use of Primitives
top	bottom = ~ top; center = 1/2 (top)
right	left = ~ right
front	back = ~ front
edge	edge
side portion	} possible combining into one term

2.2.4 Qualifying Primitives

Qualifying terms, such as "the upper arm," are used in natural language expressions to denote which of several objects is being referred to. In terms of the conceptual net, they indicate the proper link connections, but do not appear as a high level identifier of a link.

2.2.5 Classificational Primitives

The classificational primitives depend on the subject area or discipline for which the description is being prepared. A set of such primitives would therefore have to be prepared for each such discipline.

2.2.6 Operative Primitives

To deal with terms such as "cuts," "crosses," etc., our initial selection is one term which indicates a crossing over, one term which indicates a cutting off, and one term which indicates entry into, as shown below:

cuts across = crosses, spans, divides, partitions
cuts off = disrupts
cuts into = invades, enters

2.3 STRUCTURING THE DATA BASE

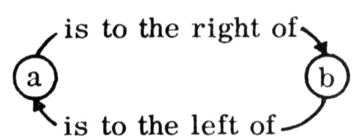
We have considered several strategies for structuring the data base so as to allow efficient search of the pictorial descriptions. Two approaches are the use of the "metanet," and the use of relationship properties.

2.3.1 The Metanet

The metanet is a high-level network which uses only conceptual classes and not the relationships themselves to label links. In addition, only the "upper" levels of the original conceptual net are retained. Search in response to a query is first to be made on the ordered set of metanets, each of which indicates the general theme or structure of the actual conceptual net.

2.3.2 Relationship Properties

To keep the number of links appearing in a conceptual net to a minimum, it is desirable to make available to the retrieval system certain "world knowledge" concerning relationship properties. For example, instead of using the net



it is preferrable to use the net (a)— is to the right of →(b), and to store the rule: right of (a, b) = left of (b, a).

Additional properties of relationships are required because mismatches can occur between query and data base even when relationships are reduced to a primitive form. For example, "at" and "near" cannot be expressed by the same primitive, yet something which is "at" a location is also "near" that location, and therefore the

LOCKHEED PALO ALTO RESEARCH LABORATORY
LOCKHEED MISSILES & SPACE COMPANY
A GROUP DIVISION OF LOCKHEED AIRCRAFT CORPORATION

property, at(m, n) implies near(m, n) should be part of the system. An example of properties of relations concerned with location is given in Table 2-5, derived with the aid of Ref. 9.

If an extensive body of world knowledge is to be provided to the system, some efficient method of searching this collection of facts must be provided. That is, the system must somehow locate relationship properties that are relevant to a query without performing an exhaustive search of the collection of its world knowledge. Although it may be possible to develop a suitable organization and indexing scheme for this collection, the alternative of portraying this information using a two-dimensional map seems more attractive. Using this approach, various concepts are defined individually in this space, and the interrelationships are determined by performing "experiments" in the space. For example, if the relationship, "X is between Y and Z" is depicted in this space, (where Y and Z are parallel lines and X is a point), then, as X approaches Z so that "X is nearer to Z," the system can note experimentally that "X is farther from Y."

We plan to explore the two-dimensional portrayal of two-dimensional relationships during the next contract period.

Table 2-5

PROPERTIES OF RELATIONS CONCERNED WITH LOCATIONS

<u>Reciprocal relations</u>

Near(m, n)	=	Near(n, m)
Far from(m, n)	=	Far from(n, m)
Beside(m, n)	=	Beside(n, m)

<u>Opposite relationships</u>

Upon(m, n)	=	Under(n, m)
On(m, n)	=	Under(n, m)
Above(m, n)	=	Below(n, m)

<u>Mutually exclusive relationships</u>

Inside(m, n)	excludes	Outside(m, n)
Near(m, n)	excludes	Far from(m, n)
Below(m, n)	excludes	Above(m, n)
Below(m, n)	excludes	Beside(m, n)
Above(m, n)	excludes	Beside(m, n)
At(m, n)	excludes	Between(k, m, n)[a]
Inside(m, n)	excludes	Between(k, m, n)

<u>Implications</u>

Near(m, n)	implies	Outside(m, n)
Far from (m, n)	implies	Outside(m, n)
At(m, n)	implies	Near(m, n)
Upon(m, n)	implies	At(m, n)
Upon(m, n)	implies	Above(m, n)
Under(m, n)	implies	Below(m, n)

(a) i.e., k cannot be between m and n.

LOCKHEED PALO ALTO RESEARCH LABORATORY
LOCKHEED MISSILES & SPACE COMPANY
A GROUP DIVISION OF LOCKHEED AIRCRAFT CORPORATION

Section 3
NATURAL LANGUAGE ASPECTS OF CONCEPTUAL MAPPING

The previous section dealt with problems of representation of meaning; this section deals with problems in determination of meaning from natural language expressions.

3.1 SEMANTIC AMBIGUITY

To indicate the problems of ambiguity in picture description, let us examine three descriptions from Table 1-1; often sentences that offer no problem to the human are quite difficult for an automated process to handle. Consider the following sentence:

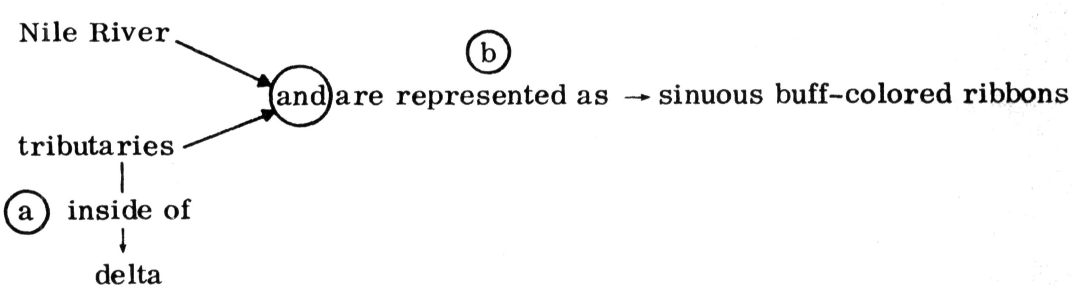
(1) The Nile River and its tributaries within the delta appear as sinuous buff-colored ribbons.

Note that "appear" can be appear$_1$ = are represented as or appear$_2$ = come to sight. The program must, from a government table, identify the appear as form as indicating appear$_1$, and the parsing program must then be able to identify the two subjects the Nile River, and its tributaries and that within the delta is a prepositional phrase modifying tributaries. If we use the primitive terms,

a inside of b are represented as c approximately in

c consisting of e is visible in

then the conceptual net for sentence 1 would be mapped as,



LOCKHEED PALO ALTO RESEARCH LABORATORY
LOCKHEED MISSILES & SPACE COMPANY
A GROUP DIVISION OF LOCKHEED AIRCRAFT CORPORATION

(2) Such straightline segments <u>are</u> nearly always canals <u>rather than</u> roads.

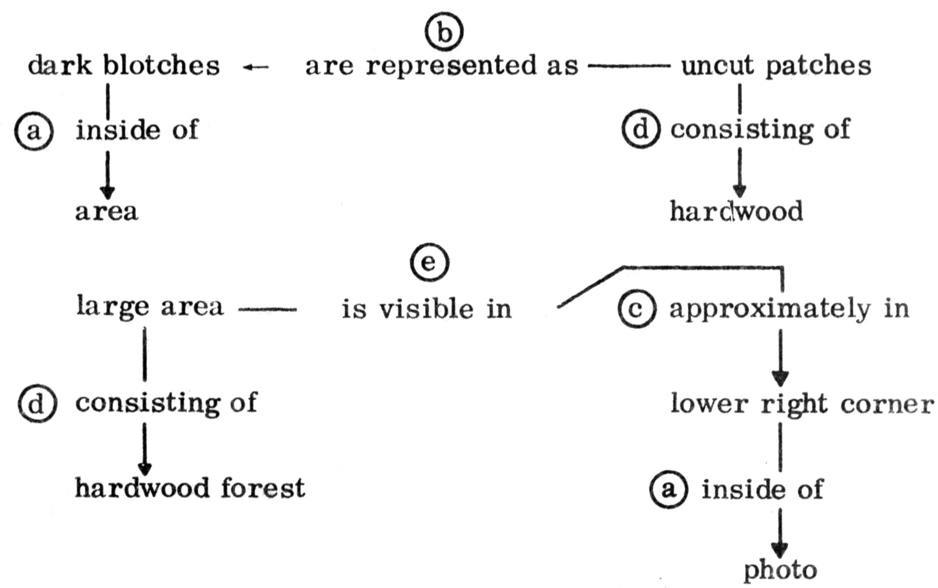
Here we are faced with the ambiguity of $\underline{are_1 = identity}$, and $\underline{are_2 = represent}$. The program would have to identify the $\underline{are_2}$ meaning and the <u>rather than</u> would have to be identified as <u>not</u> to obtain the representation,

straightline segments ← are represented as (b) — canals

straightline segments ← not (are represented as) ~(b) — roads

(3) The dark blotches within this area are uncut patches of hardwood, and a large area of hardwood forest <u>can be seen</u> toward the lower right corner of the photo.

The <u>can be seen</u> has the alternative meanings, $\underline{\text{can be seen}_1 = \text{is visible}}$, and $\underline{\text{can be seen}_2 = \text{are represented as}}$. The program would have to handle the ambiguity of both <u>are</u>, and <u>can be seen</u> to obtain the conceptual net,



The importance of the government tables in resolving ambiguity can be seen from this brief examination of mapping from natural language to concepts. (It should also be noted that not all ambiguities are resolved by such tables.) A word government table suitable for pictorial description is being prepared by extracting and expanding upon the entries from the Robison tables. An example of the entries for the letter a as extracted from the Robison tables is shown in Table 3-1. The tables to be constructed will be in the form described in Part I, section 2.2 (which includes word meanings), and is expected to have less than 1000 entries.

3.2 PARSING PICTORIAL DESCRIPTIONS

The PHRASE parsing program of Lois Earl was used to analyze four descriptions as follows: (1) an earth resources geological description, (2) a description of an aerial photography by a lay person, (3) description of a surgical scene by a surgeon, and (4) a botanical description. The resultant parsing are shown in Fig. 3-1. It will be noted that there were nine sentences in the descriptions, and there were seven phrase-constructions whose function was incorrectly labeled. Two of these errors were due to incorrect part-of-speech determination of technical words, and two were due to incorrect interpretation of conjunctions.

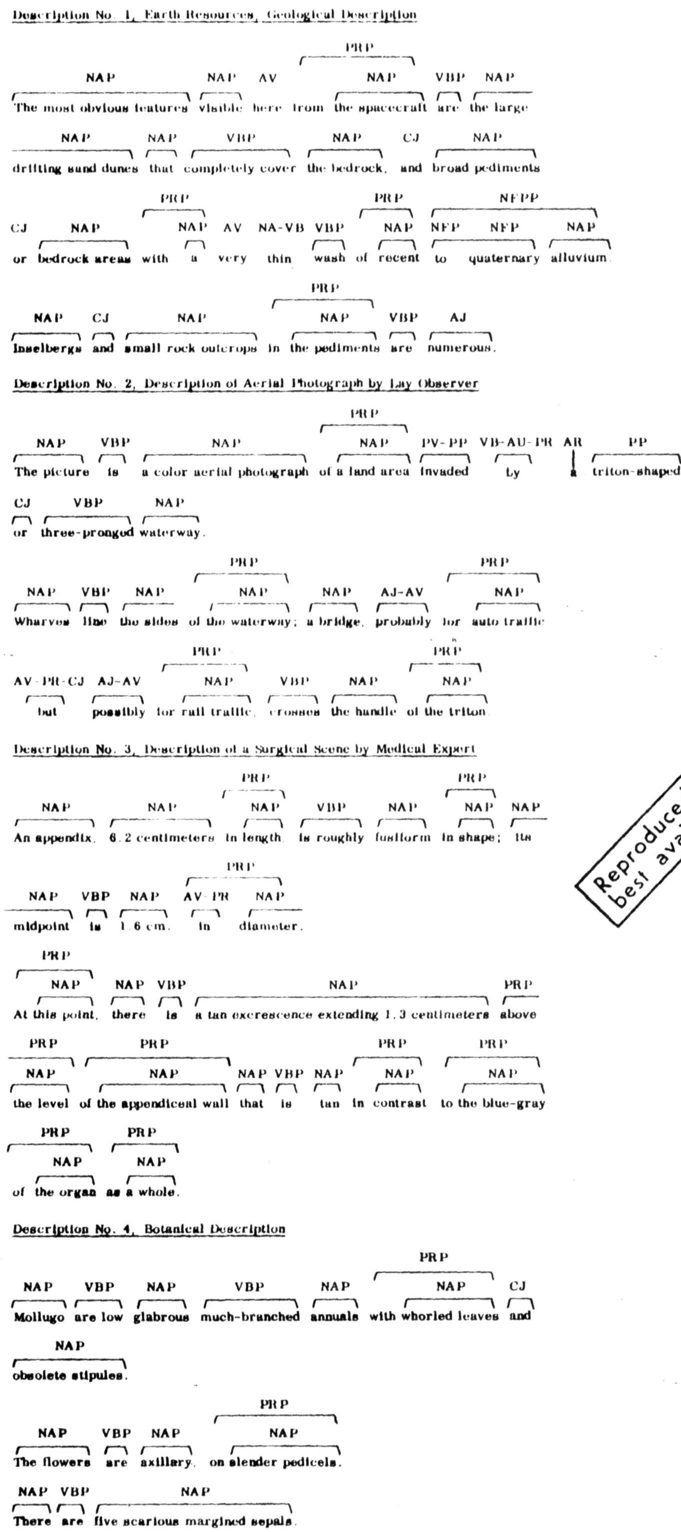
These results were obtained using both Level I and Level II parsing, as defined in Section 1. The Level II parsing is necessary for the identification of word government patterns in text, while for full semantic analysis, all four levels of parsing will be needed. During the next contract period, the Level III parser will be available for parsing these descriptions, and it will be interesting to see what improvement in parsing results.

Table 3-1

ENTRIES FOR LETTER "A" FROM GOVERNMENT TABLES PERTINENT TO PICTURE DESCRIPTION

absence	1	n	of S/from S
	2	n	of S/in S
absent		vi2	Px/from S
		aj	from S
access		n	of S/to S/by S
accessible		aj	to S/by S
accompany	1	vt	S/to S
	2	vt	S/with S
add	1	vt	S/onto S
	1	vt	S/to S
	2	vi	to S
	3	vi	up to S
adjacent		aj	to S
adjoin		vt	S/to S
adjunct		n	to S
advance	1	vt	S/to S
	1	vt	S/toward S
	1	vt	S/into S
	1	vt	S/on S/from S
	1	vt	S/upon S
	2	vt	S/from S/to S/over S
	1	n	of S/to S
	1	n	of S/toward S
	1	n	of S/into S
	1	n	of S/on S/from S
	1	n	of S/upon S
	2	n	of S/to S/over S
affix		vt	S/to S
		vt	S/onto S
		vt	S/upon S
ahead		aj	of S/by S/in S
aim		vi	for S
		vi	at S
align		vt	S/with S

alignment		n	of S/with S
amount		n	of S/in S
angle		vt	S/at S
		vt	S/toward S
		n	of S/with S
		n	between S/and S
anterior		aj	to S
apart		av	from S
append		vt	S/to S
		vt	S/onto S
appendage		n	to S
		n	of S
apportionment	1	n	of S/into S
	1	n	of S/to S
	2	n	of S/between S/and S/ according to S
	2	n	of S/among S/according to S
approach		vi	to S
		n	of S/to S
arise	1	vi	from S
	1	vi	out of S
	2	vi	at S
arm		n	of S
arrive		vi	at S/by S
association	1	n	of S/with S
	1	n	between S/and S
attachment	1	n	of S/to S/with S
	1	n	of S/to S/by S
	2	n	between S/and S
	2	n	of S/for S
attribute		n	of S
augmentation		n	of S/with S
		n	of S/by S



Reproduced from best available copy.

Fig. 3-1 Description Parsings Using BPHRAS

Section 4
REFERENCES

1. A. C. Shaw, "Parsing of Graph-Representable Pictures," ACM Journal, Vol. 17, No. 3, 1970

2. M. B. Clowes, "Pictorial Relationship — A Syntactic Approach," Machine Intelligence 4, B. Meltzer, and D. Michie, eds., New American Elsevier Publishing Co., Inc., 1969, N. Y.

3. O. Firschein and M. A. Fischler, "Describing and Abstracting Pictorial Structures," Pattern Recognition Journal, Vol. 3, No. 4, Nov 1971

4. O. Firschein and M. A. Fischler, "A Study in Descriptive Representation of Pictorial Data," Second Int. Joint Conf. on Artificial Intelligence, Sep 1971, London, England

5. O. Firschein and M. A. Fischler, "Descriptive Representation of Remotely Sensed Image Data," 1971 IEEE Systems, Man and Cybernetics Group, Annual Symposium Record, IEEE Catalog No. 71 C46-SMC, Oct 1971, Institute of Electrical and Electronic Engineers, N. Y.

6. H. R. Robison, "Computer-Detectable Semantic Structures," Information Storage and Retrieval, Vol. 6, Pergamon Press, 1970, Oxford, New York, London, and Paris

7. L. Tesler, "New Approaches to Conceptual Dependancy Analysis," in Spinoza II: Conceptual Case-Based Natural Language Analysis, by R. C. Schank, L. Tesler, and S. Weber, Stanford Artificial Intelligence Project Memo AIM-109, Jan 1970

8. E. Sandewall, "Representing Natural Language Information in Predicate Calculus," in Machine-Intelligence 6, B. Meltzer and D. Michie, eds. American Elsevier Publishing Co., Inc., New York

9. C. J. Fillmore, "The Case for Case," in Universals in Linguistic Theory, E. Bach and R. T. Harms, Holt, Rinehart and Winston, Inc., Chicago, 1968

10. R. F. Simmons, "Some Relations Between Predicate Calculus and Semantic Net Representations in Discourse," Second Int. Joint Conf. on Artificial Intelligence, Sep 1971, London, England

11. R. C. Schank, L. Tesler, and S. Weber, "Spinoza II: Conceptual Case-Based Natural Language Analysis," Stanford Artificial Intelligence Project, Memo AIM-109, Jan 1970

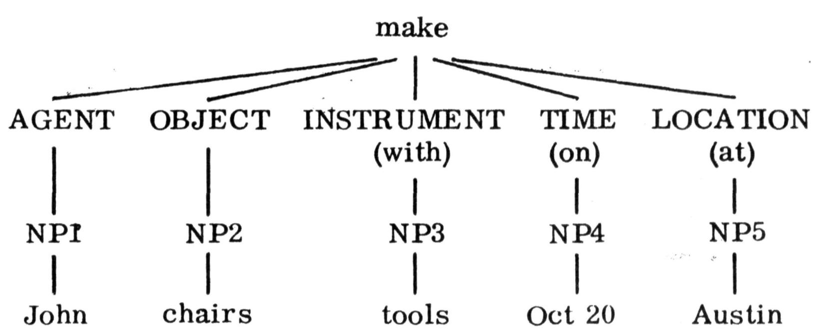
12. R. C. Schank, "Finding the Conceptual Content and Intention in an Utterance in Natural Language Conversation," Second Int. Joint Conf. on Artificial Intelligence, Sep 1971, London, England

Appendix A
THE USE OF CASE STRUCTURES IN SEMANTIC MAPPING

The "case" notions comprise a set of universal concepts which identify certain types of judgments human beings are capable of making about the events that are going on around them; judgments about such matters as who performed an action, who did it happen to, with what was it done, and what was changed. The present appendix reviews some of the work in the use of case concepts to obtain a representation of the meaning of a sentence.

Fillmore, Ref. 10, views the sentence as consisting of a verb and one or more noun phrases, each associated with the verb in a particular case relationship. Some of the cases and their definitions are given in Table A-1. He gives a slight indication of how one can go from the surface structure of the sentence to the underlying cases by noting the characteristics of the verbs and the prepositions. For example, the Agentive preposition is by, the Instrumental preposition is by if there is no Agentive, otherwise it is with; the Objective and Facitive prepositions are typically deleted, and the Dative preposition is typically to.

Simmons and Bruce, Ref. 11, use the Fillmore case concept to obtain an "attribute-value" representation for the meaning of a sentence. For example, given the sentence, "John made chairs with tools on October 20th in Austin," they get a diagram based on the cases as follows:



LOCKHEED PALO ALTO RESEARCH LABORATORY
LOCKHEED MISSILES & SPACE COMPANY
A GROUP DIVISION OF LOCKHEED AIRCRAFT CORPORATION

Table A-1

CASES USED BY FILLMORE

Agentive (A). Instigator of the action identified by the verb.

Instrumental (I). The inanimate force or object causally involved in the action or state identified by the verb.

Dative (D). The case of the animate being affected by the state or action identified by the verb.

Factitive (F). The case of the object or being resulting from the action or state identified by the verb.

Locative (L). The case which identifies the location or spatial orientation of the state or action identified by the verb.

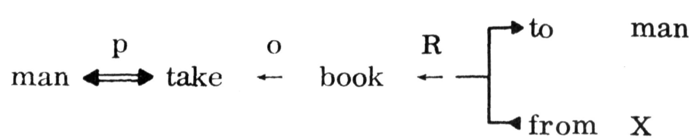
Objective (O). The semantically most neutral case, the case of anything representable by a noun whose role in the action or state identified by the verb is identified by the semantic interpretation of the verb itself (conceivably, the concept should be limited to things which are affected by the action or state identified by the verb).

This diagram can also be represented as a series of triples:

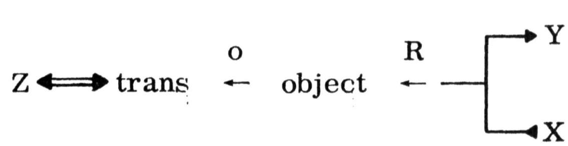
make	AGENT	NP1
make	OBJECT	NP2
make	INSTRUMENT	NP3
make	TIME	NP4
make	LOCATION	NP5
NP1	TOKEN	John
NP2	TOKEN	chairs
NP3	TOKEN	tools
NP4	TOKEN	Oct 20
NP5	TOKEN	Austin

Since this set of triples represents the meaning of the sentence, Simmons and Bruce suggest that to answer a question one simply finds the set of triples in the data base corresponding to the triples in the question. They do not deal with the problem of paraphrase, a problem that keeps such matching from being a straightforward procedure.

Schank et al., Refs. 12, 13, use case concepts in their parser, Spinoza II. The parser makes use of relations between conceptual actions and the implications of these actions. This enables the conceptual analyzer to discover not only the conceptual content of an utterance, but also the intention of that utterance in context. The conceptual dependency analysis which they use is directly related to the case concept, but they use an interesting generalization to handle the problem of paraphrase. For example, the sentence, "the man took the book," would have the network



where p = actor or performer, o = object, R is the recipient, and the X denotes the fact that the entity from which the book was taken was not specified in the sentence. However, this network is typical of an entire class of actions involving transfer of an object from one entity to another, e.g., "given," "send," "steal," which have the general network



The verb dictionary used in Spinoza II would indicate that for "give" the "trans" network above is relevant, and that Z = X, i.e., the person performing the action is the same as the person from whom the object comes.

The cases used by Schank are shown in Table A-2, and those of Tesler, a coworker, Ref. 8, in Table A-3.

Table A-2
CASES USED BY SCHANK

Conceptual cases	Prepositions Used	Relationship
Objective	(none)	Object of the action
Recipient	to, from	Receiver or transmitter
Instrumental	with, by	Instrument used
Directive	to, from, toward	Direction of action
Attributive Cases	Prepositions Used	Relationship
Possession	of, with	Has
Location-near	near, at, by, in before	Located
Containment	in, of	Contained

Table A-3
CASES USED BY TESLER

Conceptualization Class	Relationship
Attributive	Existence, has attribute, in a state
Classificational	Something which is the same as something else
Behavioral	Behavior without change of state, victim, goal, or direction, e.g. rotate, dance
Motive	Something in motion from one station to another
Operative	Something active does something to something passive
Mutual	Two or more actors interacting to perform an action, which would not be observed to be performed just by observing one of them, e.g. fight
Hypothetical	An attitude toward an issue, e.g., say, claim, know, fear, hear
Causal	Something makes something happen, either intensionally or unintentionally

Appendix B

DESCRIPTIVE REPRESENTATIONS OF REMOTELY SENSED IMAGE DATA*

O. Firschein
M. A. Fischler
Lockheed Palo Alto Research Laboratory
Palo Alto, California 94304

There is a growing interest in high-altitude image data obtained from various sensors for use in earth resources applications. The lines, patterns, and colors in such imagery reveal what both nature and man have done to the face of the earth. In photos taken from high altitudes, highways appear as networks of fine lines, farm fields produce recognizable patterns, and the areas from which forests have been cleared stand out brightly. Colors are indicative of the vegetation, soil conditions, depth of water, and many other matters of vital concern to mankind.

The present paper deals with the nature of the image descriptions used to express the content or meaning of earth resources satellite data, rather than the mechanics of handling such data, such as is given in Ref. 1.

Natural Language Descriptions

Considerable variation occurs in the scope of natural language description of imagery from the various earth resources disciplines.[2] Geology, geography, hydrology, and oceanography tend to more global descriptions, while agriculture, forestry, and cartography tend to focus on particular attributes of the image, often relying on overlays or line drawing representations. In addition, we note that the grammatical structures for describing imaged data tend to be much simpler than conventional samples of literary or spoken language – the verb forms and relationships are simpler, and ambiguity is easier to resolve.

Several questions arise in dealing with such descriptions: (1) How can the descriptions be put into some normal or canonical form to regularize variations in description from observer to observer, and also for use in computer based systems? (2) How can a number of different descriptions of a given image be combined to obtain a more comprehensive description? (3) How can the entities and relations between entities in a description be converted to a small set of basic primitives to simplify the canonical forms?

A basic philosophical question also arises in designing retrieval systems: Should one determine the best query structure for a particular user population and then design a suitable data base structure, or should one use a data base structure having certain desirable properties and design the most appropriate query structure?

We are taking the latter course. By examining various representations of natural language description, we hope to determine structures that have certain desirable properties such as allowing automated structure determination from text, and allowing automatic combination of separate descriptions of the same image. Query and retrieval systems would then be designed around such representations.

Canonical Forms and Description Merging

We have examined two types of canonical forms,[3] the descriptor approach in which a set of words or phrases is selected to represent the image, (see our review in 5) and the relational graph approach in which a conceptual representation is obtained by indicating relationships between entities in the picture. If two descriptor representations have been prepared for the same image, a combined descriptor set can be obtained by forming a single combined list of descriptors. If two network representations for the same image are available, the problem is more complex but, abstractly, the networks are combined based on detection of common nodes.

In the experiments described in [5], using a set of subjects having no special technical background, convergence to a merged form was obtained quickly; the relational net remained relatively unchanged when additional descriptions of the same image were combined with it. However, in the case of experts dealing with earth resources data, an individual with a specialty can introduce a whole new set of links and nodes into the net, and therefore convergence to a particular representation does not guarantee completeness; it merely indicates completeness for a particular set of specialties.

The Concept of an Encyclopedic Entry

This study is concerned with the problem of whether a representation of a picture could be used to answer general questions about the picture. There would seem to be too many diverse questions that could be asked concerning the picture, and too many different requirements in terms of the level of response required to allow such a procedure to be successful. Consider, however, an analogous situation which arises in the attempt to summarize all human knowledge in such a way that questions can be answered using this summary. Although this task would also seem to be overwhelming, collections of knowledge, called "encyclopedias," have been used for many years, and turn out to be a useful means of answering a diversity of questions.

Similarly, it should be possible to develop a question-answering system based on the concept of an encyclopedia for the body of knowledge residing in either a single photograph or a set of photographs. The same problems of level of detail and use of technical terminology arise in the preparation of both an encyclopedia based on knowledge expressed in natural language and

*The work described here was supported partially by the Office of Naval Research Contract N00014-70-C-0239, and partially by the Lockheed Independent Research Program.

Paper presented at Joint National Conference on Major Systems, sponsored by the IEEE and ORSA, Anaheim, California, October 1971.

one in which knowledge is expressed in pictorial form. In both cases, one has to consider how the material is to be organized, the nature of the potential user, and the amount of material to be gathered.

Relation of Queries to Descriptions

We can gain some insight into the requirements for an encyclopedic entry by examining the types of query that are likely to occur. It should be kept in mind that an encyclopedia contains material suitable for answering many questions, but not all questions on a topic. In addition, while several diverse items of information from the encyclopedia may be pulled together to answer a question, there is no guarantee that this can be done. Similarly, for an encyclopedic entry for an image, we should expect to be able to answer many questions, to combine elements of the entry to answer other queries, and to fail completely for the remaining queries. In our examination of queries concerning earth resources imagery, we have found that the following elements of description are often required.

Discontinuity and contrast. An indication of discontinuity and contrast by identifying regions homogenous in such attributes as texture, color, pattern, and temperature (on IR) imagery.

Terraine structure identification. Queries will often ask for the existence of flow patterns, surface film patterns, silting and sediment, circulation, thermal structure, man-made features, etc.

Change detection. Queries often ask for changes in cultural features, stream flow, reflectivity, crustal structure, snow areas, sand bars, etc. Such questions can be answered if previous and present descriptions of an image are available and contain line drawing extracts or equivalent. Such line drawing extracts can be readily compared and the differences noted.

Boundary area, and volume analysis. The boundaries of snow and ice, lakes, tectonic features, crop areas, etc., are often required for area and volume computations.

Counting of items. Queries will often ask for a count of the number of items, or some indication of what percent of the image is taken up with some items.

Content in a broad sense. These are queries that concern the global aspects of the image, such as: is the area predominately rural? ; what are the general characteristics of the mountain ranges?; what are the transportation networks?; describe the shoal formations.

The natural language descriptions are most appropriate for the last type of query.

The Nature of an Encyclopedic Entry

An encyclopedic entry for an image capable of dealing with the above queries would consist of one or more of the following content-indicating elements.

Image citation data. Identification information concerning the source of the imagery, the equipment used, time, date, geographical location, image quality, etc. The image citation data usually constitute one of the main content-bearing elements used for indexing image collections today.

Descriptor data. Words or phrases concerning the subject matter in the photograph.

Natural language descriptions. Descriptions which capture the theme, general layout and arrangement of the photograph, objects pictured, and their relationships with other objects. This type of description has been the main concern of the present paper.

Analysis of imagery. Measurements and delineation of aspects of the image, such as identifying contours of bodies of water, determining distance between objects, etc.

Conceptual maps. Indication of the content of the picture by means of linked structures. Such relational nets are used in present day question-answering systems having limited data base sizes.

Line drawing extracts. Line drawings extracted from an image for expressing specialized relationships such as transportation networks, geologic boundaries, and water courses.

These disparate elements must be tied together in such a way that queries are suitably responded to, with direct answers, extracts of imagery, measurements, or discussion provided to the user.

A Dynamic Data Bank for Earth Resources Imagery

We have examined [6] the automated processing of descriptions to obtain encyclopedic entries for an image. Such automation can lead to what we have termed "dynamic data banks" for earth resources imagery. Experts using the imagery would return to the archival center their analysis and description of the imagery developed as part of their investigation. Such descriptions would be processed and merged, so that with time each image would have associated with it a combined or encyclopedic description.

Of course, such descriptions would have to be tagged so that data on the background of the person making the description (his name, set, or field of specialty, analysis used, comments regarding the analysis, and reports produced using this image data) could be retrieved if desired. This "metadescriptive" material gives the context of the individual descriptions, the confidence that a user can have in the description, and links to related work.

References

1. H. M. Gurk, C. R. Smith, and P. Wood, "Data Handling for Earth Resources Satellite Data," Proceedings of the 6th International Symposium on Remote Sensing of Environment, Center for Remote Sensing Information and Analysis, University of Michigan, Ann Arbor, Oct 1969

2. Ecological Surveys From Space, NASA SP-230,

National Aeronautics and Space Administration, Washington, D. C. 1970

3. O. Firschein and M. A. Fischler, "Describing and Abstracting Pictorial Structures," Pattern Recognition Journal (in press, 1971)

4. M. A. Fischler, "Machine Perception and Description of Pictorial Data," Proceedings of the International Joint Conference on Artificial Intelligence, Washington, D.C., May 1969

5. O. Firschein, and M. A. Fischler, A Study in Descriptive Representation of Pictorial Data, 6-83-71-3, Lockheed Palo Alto Research Laboratory, Palo Alto, California, Feb 1971 (also, Second International Joint Conference on Artificial Intelligence, Sept 1971, London, England)

6. O. Firschein and M. A. Fischler, Descriptive Representations of Remotely Sensed Image Data, 6-83-71-5, Lockheed Palo Alto Research Laboratory, Palo Alto, California, May 1971 (complete version of the present short paper)