A STUDY IN DESCRIPTIVE REPRESENTATION OF PICTORIAL DATA*

O. Firschein and M. A. Fischler
Information Sciences
Lockheed Palo Alto Research Laboratory
Palo Alto, California 94304, U.S.A.

ABSTRACT

Although much effort has been expended on automatic scene description, especially in the various robot and "hand-eye" projects, these efforts have usually been directed toward description for immediate use, i.e., description of a scene for the purpose of stacking blocks using a manipulator arm, or for allowing a robot to move through an environment. This paper concerns a somewhat different type of description in which a scene is described in general terms to be stored for an unspecified future use. This type of description has application in (1) advanced robot systems, where the robot, similar to the human, will build up an "encyclopedia" of descriptions for possible use, and (2) in question-answering systems for image data bases in which the descriptions are to represent an "encyclopedic" knowledge of the image. Experiments in non-goal-directed description using human subjects are described, experiments which seek to determine how general descriptions are generated and the nature of such descriptions.

1. INTRODUCTION

Much work has been done in describing a scene for specific use, such as description of an image data base for definite retrieval goals, or description of a scene for predetermined use by a manipulator arm or for specified robot tasks. In contrast to these efforts, the main theme of the present investigation is that of non-goal specific description of imagery, i.e., preparation of a scene description for a broad range of possible use. This is the type of description a human stores as he moves through an environment, when he has no specific goal in mind. He is later able to recall the stored image and to use it for various purposes. Our investigation is concerned with how such descriptions can be generated (the prior knowledge, "set," and deductive capability required), and the nature of such descriptions (the vocabulary, relationships used, and structural forms).

To clarify what we mean by goal-specific description, consider the types of pictorial description described below.

- Descriptions for Reconstruction. These can be used to reconstruct a picture and can employ relatively simple vocabulary. They are concerned largely with providing detailed information on location, size, shape, color, texture, etc., and even for a simple picture can be quite extensive.
- Descriptions for Classification. These are used in distinguishing one scene from another, or to place a scene in one or more sets of distinct categories. In this case, the person preparing the description must have considerable knowledge of what is typical and what is atypical for the subject matter of the picture. He must also know the possible classification categories.
- Descriptions for Retrieval of Pictures. These must take into account possible user queries and must capture the content or meaning of the picture using some labeling scheme which indicates these aspects of the picture to the user.
- Descriptions for Picture Comprehension. These are used to aid the observer in understanding a picture.

In contrast to the above descriptive types, we are interested in more general descriptions such as those defined in the next paragraphs.

- Descriptions That "Paint Mental Images." These are descriptions in literature that are used to "paint a mental image" of the scene in the reader's mind.
- Descriptions for Answering Queries. These are descriptions of an image for use in answering general questions about the image directly, without the user's having to see to the actual imagery.

Some relevant questions concering the preparation of general image descriptions are:

- What strategies do human subjects use in preparing such descriptions ?
- How do instructions and constraints on the subject affect these strategies?
- What is the relation of training and background to such strategies?

As far as the deductive aspects used in description are concerned:

- How deep is the reasoning required ? In other words, what are the logical structures employed ?
- What information included and not included in the photo is required ?
- How relevant are the deductions to the overall impression of the photo ?
- What is the validity of the reasoning and how were the conclusions drawn ?

And, finally, as far as normalization is concerned:

- How can an unconstrained description be formalized without losing important information ?

*Work sponsored by The Lockheed Independent Research Program and by Office of Naval Research Contract N00014-70-C-0239.

- What transformations are necessary in going from the free natural language description to the canonical form ?
- What mechanisms are required to combine individual canonical descriptions so as to form an overall "complete" description?

1.1 QUESTION ANSWERING BASED ON SYMBOLIC DESCRIPTION OF PICTORIAL DATA: THE ENCYCLOPEDIA CONCEPT

At first glance, an attempt to create informal descriptions of imagery for a future unspecified use seems overly ambitious. There 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. Consider, however, an analogous situation that arose in the attempt to summarize all human knowledge in such a way that questions could be answered using this summary. Although this task also seemed to be insurmountable, the resultant collections of knowledge, called encyclopedias, have been used for many years, and they have proved to be a useful tool for answering a diversity of questions.

Our work has similarities to that of Schwarzlander (1), who is concerned with "repackaging" knowledge, as represented in written form, into an encyclopedic storage arranged in such a way that users can have convenient access to the information. The investigation that he describes is concerned with the extraction of "information items," basic concepts which capture the important units of knowledge from the written material. A data base consisting of these information items is then used to answer user queries.

Similarly, it should be possible to develop an image data base using the concept of an encyclopedia for the body of knowledge residing in either a single scene or a set of scenes. 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 one for 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.

1.2 THE STRUCTURE OF AN ENCYCLOPEDIC ENTRY FOR PICTORIAL DATA

Because the user cannot adequately evaluate a description without knowing the context in which it was generated, i.e., the emotional set, background, and deduction procedures of the person or process producing the description, we have considered an encylopedic entry as consisting of two parts, the context of the description and the description itself.

1.2.1 The Context of the Description

The contextual portion of the description indicates the emotional impact of the scenes on the observer, his doubts or expressions of alternatives, and indications of how he made his deductions. Typical entries are:

Emotional. "The picture is rather beautiful. "

Doubts. "I guess," "perhaps, "appear to be"

Alternatives. "Water or gas storage tank"

Deductions. "This area is developing fast, because at the opposite end of the runway that the subdivision is on, there are building sites on both sides. "

1.2.2 The Description

The descriptive portion of an encyclopedic entry could consist of one or more of the following content-indicating elements:

Descriptor Data. Words or phrases concerning the acquisition of the photograph Itself, such as the sensor used, date, sensor parameters, and altitude, as well as the subject matter in the photograph. Firschein and Fischler (2) have reviewed such descriptor systems in applications ranging from art libraries to indexing of psychiatric videotape.

Natural Language Descriptions. Informal descriptions which capture the theme, general layout and arrangement of the photograph, the objects pictured, and their relationships with other objects. Descriptions from different points of view (e.g., for the case of question-answering systems, geology, hydrology, military intelligence) are often necessary for adequate description of a picture for encyclopedic purposes.

Semantic Maps. The contents of a picture indicated by a two-dimensional structure which shows objects as nodes and the relationships between objects as links. The semantic map is discussed in more detail in section 3.2.

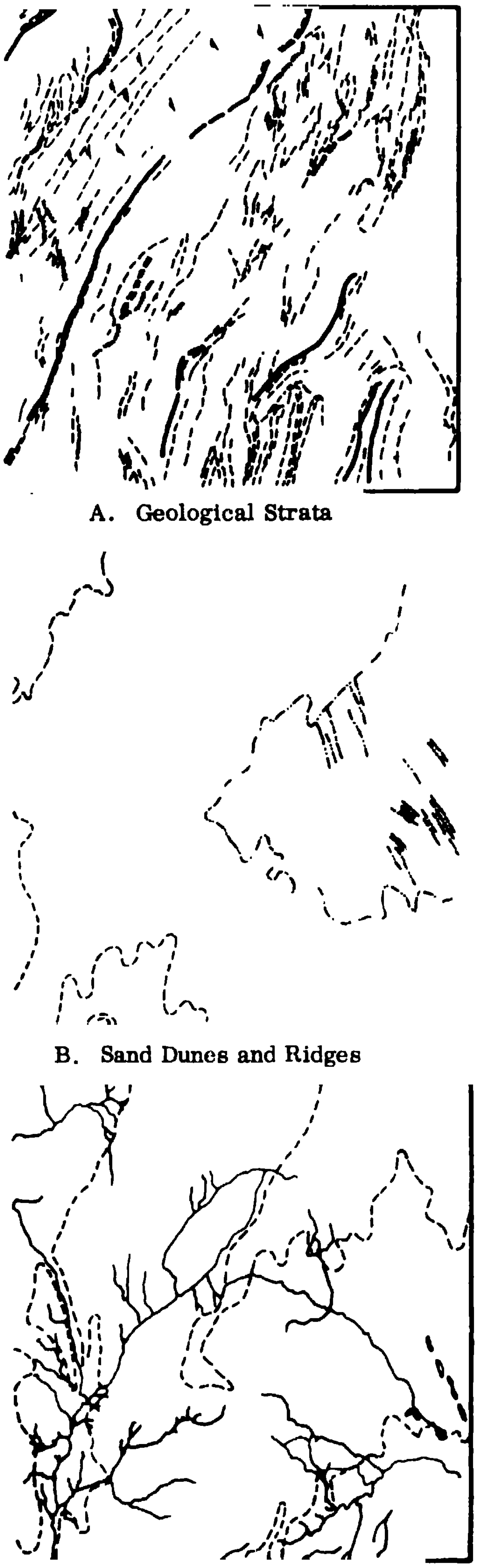
Line Drawing Extracts. Line drawings extracted from a photograph for expressing specialized relationships such as transportation networks, geologic boundaries, and water courses. Figure 1 indicates three line drawings that were extracted from a single Gemini 5 space photograph (3), one showing geologic features, one showing sand dunes and ridges, and the third showing stream channels and vegetation. The structure of such line drawings can be expressed in one of the formal notations using grammar-based techniques.

1.3 RELATION TO PREVIOUS WORK IN PICTURE DESCRIPTION

Descriptions can be either "formal, "i.e., use notation, procedures, or rules to express the content or meaning of a picture, or be "informal, "i.e., use natural language in an unconstrained manner for the same purpose. Work using these two approaches to description is reviewed briefly below.

1.3.1 Formal Descriptive Techniques

Tne general field of formal pictorial description that relates to the present work is reviewed in Ref. 1. We considered three classes of formal descriptive technique: (1) the grammar-based approach, which



A. Geological Strata

B. Sand Dunes and Ridges

C. Vegetation and Intermittent Streams

Fig. 1 Three Different Line Drawings De rived From the Same Space Photograph

uses symbolic notation to describe relationships among primitive pictorial units (in practice, these units were invariably points are line segments); (2) the descriptor-based approach, which attempts to capture the content of the picture by using a number of terms or phrases; and (3) the procedure-based approach, in which a high-level control mechanism causes one description to be generated from many possible descriptions.

It was noted that the grammar-based approaches (4, 5,6,7) can be used to answer questions concerning certain types of line drawings but that these methods had not yet been extended to true photographic material. It would seem also that, because such approaches are currently dependent on parsing an expression consisting of basic picture elements, they would tend to be overwhelmed by gray-level imagery, in which the picture elements are difficult to detect, and by extremely complicated expressions which become impractical to parse, ft is also not clear whether these methods could provide useful "hlgh level" descriptions in the more complicated environment of gray-level photography.

The descriptor-based approaches (8,9,10) are used to locate pictures that might satisfy a query; the set of index terms cannot, in general, be used to answer questions directly. Such descriptor approaches are attractive today because of the availability of computers which can efficiently perform searches based on required combinations of descriptors.

Finally, there are the procedure-based approaches to description (11,12,13). These are systems, possibly grammar-based, capable of preparing a large number of descriptions of a given picture, coupled to a control mechanism which selects procedures and the order of procedures to be used, so as to produce only a single desired description. As in the case of the grammar-based approach, the procedure-based methods have not yet been extended to photographic material.

1.3.2 Informal Descriptions

Two investigations concerned with informal description of pictures by human subjects are of direct interest to the present study. In both cases, the descriptions were used for pattern recognition purposes.

Fischler (13) used a set of four samples of each of 14 handwritten Sanskrit characters on four human subjects totally unfamiliar with the Sanskrit alphabet. The nature of the experiment was described to one of the subjects, who was asked to prepare a written description for each character in one of the character sets (14 distinct items). These written descriptions were then used by the other three subjects to identify the characters in the three remaining (but randomly intermixed) character sets. It was found that surprisingly high classification performance could be achieved with very simple and seemingly crude descriptions. However, this efficiency of communication can be accomplished only when the communicants share a common model of the universe of discourse and have similar visual and perceptual abilities.

A similar result was obtained by Krauss (14), who used children as subjects in an experiment concerning communication about patterns. The subjects, separated by a screen so that they could not see each other, were provided with identical sets of six blocks, each block imprinted with a different design. One child removed a block from a special dispenser and stacked it on a peg. He described the pattern on the block to the other child who had to find the corresponding block from the blocks spread out in front of him and stack it on his peg. For the subjects, the object of the game was to build identical stacks of blocks. For them to perform their task successfully, the first subject had to describe the figures well enough so that his partner could select the correspond ing figure, a task that was not simple because the figures imprinted on the blocks were complex and were not easy to relate to known pictures. It was found that children 4-1/2 years old had difficulty in performing the description task because they used a "private encoding"-a description that was not socially shared. The children performed the task successfully when their own encodings were used as the description. Thus, while the aware adult formulates his description with regard to the result that it will produce in the listener or reader, the young child (and the unsophisticated machine) lacks this talent.

2. EXPERIMENTAL SETTING AND INITIAL OBSERVATIONS

With the idea in mind of an image data base for unspecified future use, experiments were performed to determine:

- The extent of utility of a general symbolic description of a photographic scene
- The extent to which contextual factors influence description
- A set of primitive vocabulary terms, linguistic forms, and pictorial relationships that can be chosen as the basis for semi-formal language for describing scenes -in particular, aerial terrain photographs

As noted in the Fischler experiments in human communication about pictorial objects (13), an observer is often guided in his description by the expected use of the description. This predisposition to interpret, i.e., to concentrate on different aspects of a picture, is what the psychologist thinks of as "set" or "expectation." We attempted, therefore, to give instructions that would motivate the subjects to tell as much as they could about a picture without providing any information that might result in a limiting "set."

A preliminary set of experiments was conducted using 13 subjects who had a variety of backgrounds. We were interested in observing the deductive and meta-descriptive aspects of description, as well as the scope, format, and forms of descriptions produced.

Aerial photographs in both color and black and white were selected for use in these experiments because of the many practical applications employing such data, e.g., earth resources studies. *The* following photographs were used:

(1) Black and white, 10,000-ft view of a suburban area having much open space

(2) Color, 10,000-ftview of a highly urbanized area

(3) Black and white, 150-mile-altitude view of the San Francisco Bay area

(4) Color, 10,000-ft view of river and industrial area (see Fig. 2)

(5) View of the moon from 40-mile altitude, black and white

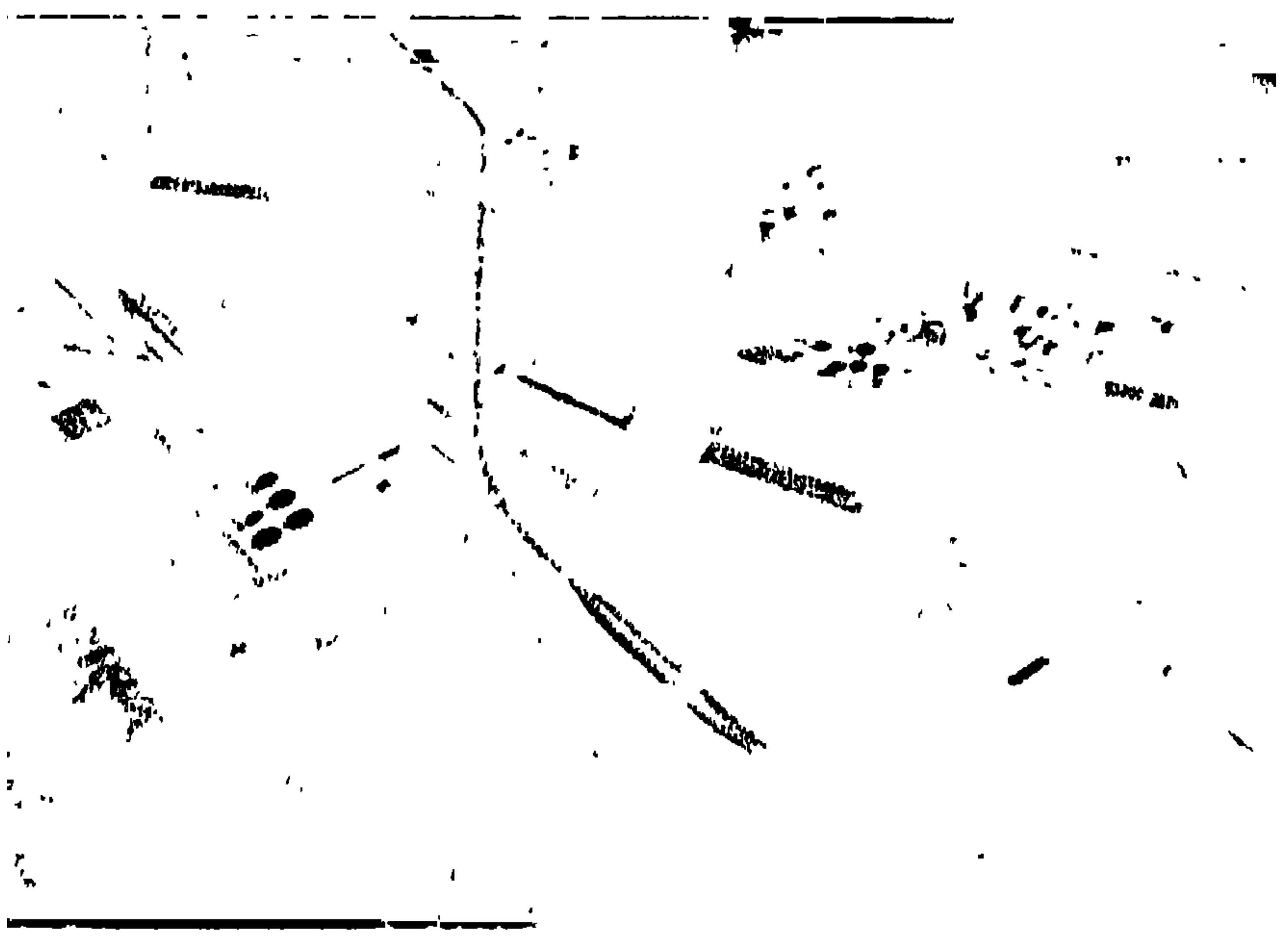


Fig. 2 Aerial Photograph of Industrial Area

The instructions used were of three types:

(1) Please write a description of the attached photo. The description will be used in an information retrieval experiment.

(2) Please write a description of the attached photo. The description will be part of a data base dealing with city planning.

(3) Please answer the following questions concerning the picture....

The type of photograph used-i.e., simple or complex, uniform or varied— had a strong effect on the subjects. In the preliminary experiments, the subjects had the following general responses:

- Subjects find highly detailed scenes having little or no centralizing theme or focus very difficult to describe. For example, subjects asked to describe an aerial photograph of a suburban sprawl, or a photograph of the moon showing uniform small cratering, are not merely unenthusiastic but even hostile toward the task.
- Subjects respond to pictures that have a centralizing theme and do relate other portions of the picture to the theme. For

example, for an aerial photograph of a harbor area, the subjects will describe the river shape and then the various refinery, port areas, etc., surrounding the river.

- If the picture has no centralizing theme but is not totally unfocused, subjects will partition the picture into regions and describe each region separately, or may partition the picture according to whether the sub region 8 contain items that can be focused upon.
- Untrained subjects will come to conclusions that are unwarranted by the pictoral material that they are viewing. Thus, conclusions concerning the depth of a crater or the height of a mountain may be made even though there is no scale that would enable reliable estimates.
- Subjects are uncertain about how to respond when they have a large amount of information about an area. For example, a subject who recognizes a photograph as being of an area with which he is intimately acquainted does not know which of the many items to include in his description. He is not sure which of his conclusions are really warranted by the imagery and which are purely a product of prior knowledge.

Since the first two sets of instructions did not indicate how to organize the descriptions, the approaches that were used by the subjects (approaches that were often determined by the nature of the photographs) were of interest. The approaches noted were as follows:

- Central focusing subject. The subject first gave the central focus of the picture and then related all other objects to it.
- Descriptor approach. The subject provided a list of descriptor terms that he felt captured the salient points.
- Partition approach. The subject partitioned the picture, often into quadrants (e.g., upper right, upper left) and described the objects and their relationships within the quadrants. Sometimes the nature of the photo was such that the partitioning was upper third, middle third, lower third, or some other top-to-bottom divisional scheme.
- Classification approach. The subject established categories (e.g., lakes, shopping centers, docking areas) and then described the characteristics and positions of the items within the categories.
- Unorganized listing. The subject scanned the picture in some manner, listing the objects and their attributes as he scanned.

3. ANALYSIS OF EXPERIMENTAL DESCRIPTIONS

As indicated previously, we are interested in both the factors that influence the approach to generating a pictorial description and the nature of the descriptions. It is therefore pertinent to determine the emotional or mental set, the background of the observer, and the kinds of deductions made, as opposed to the observations made when one views scenes directly. For this reason, each description was analyzed in terms of two aspects: context and content.

3.1 CONTEXTUAL ITEMS

The contextual analysis included the objective impression, the emotional impression, the metadescriptions, and the deductions made, as defined below.

Objective impression. Statements related to the subject or title, e.g., "a color aerial photograph of a land area invaded by a trlton-shaped waterway," that the subject feels are objective evaluations. (In fact, however, these statements may be quite subjective.)

Emotional impression. Statements concerning emotions aroused in the observer, e.g., "Picture is rather beautiful."

Metadescriptions. Statements concerning the indexer, the descriptive process itself, or the picture as a physical entity, that reveal aspects of the emotional or men tal set, the contextual knowledge, and the certainty of the subject; e.g., respectively, "I am not sure about ... " "It is probably not of interest to detail the locations of buildings. " "The picture is 6 inches wide and 4 inches high. " For human-gene rated description, metadescriptive Information is of consequence because it provides an indication of the validity of the description - an important consideration if the description is to be included in a data base.

Deductions. Conclusions drawn by combining several elements of the picture with the subjects prior knowledge; e.g., "expensive house" is a deduction, while "house with swimming pool and tennis court" is an observation. It is not always possible to determine whether a deduction has been made; e.g., the description "auto bridge" could be based either on the shape of the bridge or on the presence of a highway leading up to the bridge and the absence of railroad tracks.

Figure 3 presents a typical description and an analysis of the contextual aspects.

3.1.1 Deductions

Subjects describing pictures often perform rather subtle deductions, frequently without being aware of the elements that they employed. It is of interest to examine the deductive process for three reasons: (1) to determine whether higher level deductions are crucial to the descriptive process, (2) to determine whether better deductive procedures can be formally instituted, and (3) to determine what the requirements are for automatic deduction using computers.

Some of the deductions made by the subject can be considered as "lower level," involving direct pattern recognition of objects in the photo; other deductions are "higher level," requiring relating various clues in the photo to information known to the observer.

The picture is a color aerial photograph of a land area invaded by a triton-shaped or a three-pronged, shaped waterway. Wharves line the sides of the waterway; a bridge, probably for auto traffic but possibly for rail traffic, crosses the "handle" of the triton. The land is used primarily by industry; many large, low buildings and fluid storage tanks, such as those used to store oil and water, are on the land. The land depicted has dimensions of perhaps 1-1/2 by 1-1/2 miles; the photo exhibits a significant parallex effect. The waterway, perhaps a river or canal, was perpendicular to the line-of-sight of the camera; its average width is about 1/4 mile.

a. A Typical Description of Aerial Photograph Shown in Fig. 2

Objective Impressions	A color aerial photograph of a land area invaded ty a triton-shaped or a three-pronged, fork-shaped waterway.
Emotional Impressions	None.
Metadescriptions	
Of Indexer	None.
Of Descriptive Process	Probably for auto traffic, but possibly for rail traffic; perhaps a river or canal; about 1/4 mile.
Of Photo	Area of perhaps 1-1/2 by 1-1/2 miles; color; aerial photograph, waterway perpendicular to line-of-sight of camera; significant parallax effect.
Deductions	Area of perhaps 1-1/2 by 1-1/2 miles; average width of waterway about 1/4 mile.

b. Contextual Aspects

Fig. 3 A Typical Description and an Analysis of the Contextual Items It Contains

Some of the mechanisms used to make deductions are:

- Observation of use of object, e.g., "auto bridge," "runway for airplanes"
- Observation of adjacent object, e.g., "buildings near sports area probably gyms"
- Analysis of texture, e.g., "sandy ground"
- Analysis of shadows, e.g., "River is flowing from the top of the picture to bottom of the picture judging by the shadow cast by the dam."
- Analysis of color change, e.g., "Channel must have been dredged. "
- Distance and height estimation, based on comparison with known objects shown

3.1.2 Validity Checking and Consensus

To verify conjectures concerning deductions made and the mechanisms used in obtaining them, the deductive statements made concerning a picture, along with the picture, can be presented to a panel. The panel is asked whether the assertions are true or false, and the reasons for their answers. In this way, it is possible to determine whether the deductions were valid, at least by a consensus of other observers, and it is possible to investigate more precisely the deductive mechanisms used by subjects.

It should be noted that consensus is an operation that can be used by a robot in observing a scene at a later time. Winston's thesis (15) shows how an automated system can improve its original description after seeing several examples of a class of objects, which obviously relates to the idea of a consensus.

3.2 CONTENT OF THE DESCRIPTION

Each description was transformed to a canonical form to allow comparison of descriptions to be made, and also to allow descriptions of the same picture to be combined to obtain a composite description. To be useful in computer data bases, such canonical forms should utilize a small number of primitives and relationships. We therefore analyzed the relationships contained in the descriptions to determine the relationship categories used. Examination of the categories not only aids in the selection of basic relationships to be used in the canonical forms but also provides insight into" the analysis required for an automated system to provide general image descriptions.

3.2.1 Canonical Forms

To use informal descriptions in a mechanized data base, the representation must be normalized because of the wide variation in narrative style used by the subjects. A variety of canonical forms could be used; we have chosen a tabular form and a network representation. It should be noted that in transforming from the informal description to the canonical forms, at times we have had to make "creative" alterations. In so doing, there is always some danger that the investigator will warp the original data presented by the subjects. We have not, as yet, investigated the consistency of analysts in transforming a given description to canonical form; this warrants study.

Tabular form. The descriptions were reduced to the general form: (modifier, object, location), (relationship, location), (modifier, object, location), namely, a relationship between two objects, each having a location and descriptive modifiers. The relationship often has a location, e. g., "crosses at." Using this form, such a description as "waterway, possibly an estuary or part of a harbor, with many ship docks" becomes "(possibly estuary or part of harbor, waterway,_____) (with_____) (many ship docks,_____), " where______indicates that no information was given. Note that location can be given with respect either to other objects or to the photograph itself. A complete tabular canonical form for the description given in Fig. 3a is shown in Table 1.

Table 1

TABULAR FORM OF CANONICAL DESCRIPTION

(Capitals indicate that item is part of overall theme of picture.)

Modifier	Object	Location	Relationship	Modifier	Object
	LAND AREA		INVADED BY	THREE-PRONGED, FORK-SHAPED	WATER-WAY
	wharves		line	the sides of the	water-way
auto, possibly rail traffic	bridge	handle of triton	crosses		triton
	land		used by		industry
many, large, low	buildings		are on		the land
fluid storage	tanks		are on		the land
possibly river or canal	water-way		was perpendicular to		line-of-sight of camera

Network representation. A network representation, sometimes called a "relational graph," is shown in Fig. 4, using the same informal description as for the tabular analysis. Objects are circled, and the modifiers and the relationships to the objects are indicated by links. This representation is related both to the list structures used in computer graphics systems such as Sketchpad (16) and to the semantic networks as used by Quillian (17). Portions of the informal description that are higher level are closely related to the semantic network concept; lower level portions of the description are closer to the computer graphics area. As noted by Winograd (18), what is important is not the network representation but rather what the representation is made to depict. For example, it is possible to link concepts on the basis of their closeness in the sense of meaning, or by the roles played by the concepts (actor, action, object, modifier) within the description. In our recent work using relational nets, it appears that the simple linked structure presented in this paper does not always capture the conceptual structure appropriately. Recent work on conceptual nets by Schank et al. (19) describes a labeled link structure in which the link symbol denotes the role played by the concepts. This auxiliary information adds an additional dimension of semantic content to the network representation.

3.2.2 Description Completeness

For a data base consisting of picture descriptions to be useful, it is necessary that the descriptions be "complete"[1] in some sense. This problem of completeness of an "information item file (IF)" is faced by Bottle and Schwartz lander (20) in their encyclopedia information system, and in the field of information science in the preparation of information abstracts for information retrieval. In the present case of a description to be used in an encyclopedic data base, we have approached the problem of completeness as follows. Several different descriptions of a given picture are combined to obtain a composite. The composite description is then used as a standard of completeness.

An example of a composite description is depicted in Fig. 5, which was formed as follows. An additional informal description obtained from a subject was converted into the network form. TMs network was compared with the original network of Fig. 4, and additional links and nodes were then added. *The* phrases used for these additional links and nodes are underlined in the description of Fig. 6.

For a set of subjects having no special technical qualifications and no highly specialized set, it was found that convergence was obtained quite rapidly - after two or three descriptions were combined. On the other hand, an individual with a speciality can introduce a whole new set of links and nodes into the semantic net, and therefore convergence to a certain representation does not guarantee completeness; it merely indicates completeness for a particular set of subjects.

3.2.3 Deriving the Composite Representation

It is difficult to formulate rules for obtaining the network representation or for forming the composite network, since these transformations depend on the analysts understanding of the meaning of portions of the description. However, some general guidelines are possible: First, identify the theme sentence of the description; then draw a network using the objects and relationships found in the theme sentence. This provides the nucleus for the rest of the network.

The composite network is formed by selecting the most complete network for a given picture and then examining the other network representations of descriptions of that picture, noting whether any new aspects arise. Even though other terminology has been used, e.g., "waterway" for "canal," it is not difficult to identify such paraphrasing, even without recourse to the picture.

For composite descriptions to be useful in practice, automatic conversion from natural language to conceptual structure must be available, and composite descriptions must be obtained automatically using structure-matching techniques. For matching of queries to the data base, a true normal form must be derived in which paraphrases of the same concepts are structured identically.

One problem that arises in a practical system based on composite descriptions is conflict between the descriptions of a given image that are provided by different observers. When this problem occurs, the system must be able to detect the conflict so that a review mechanism can determine which description was inappropriate.

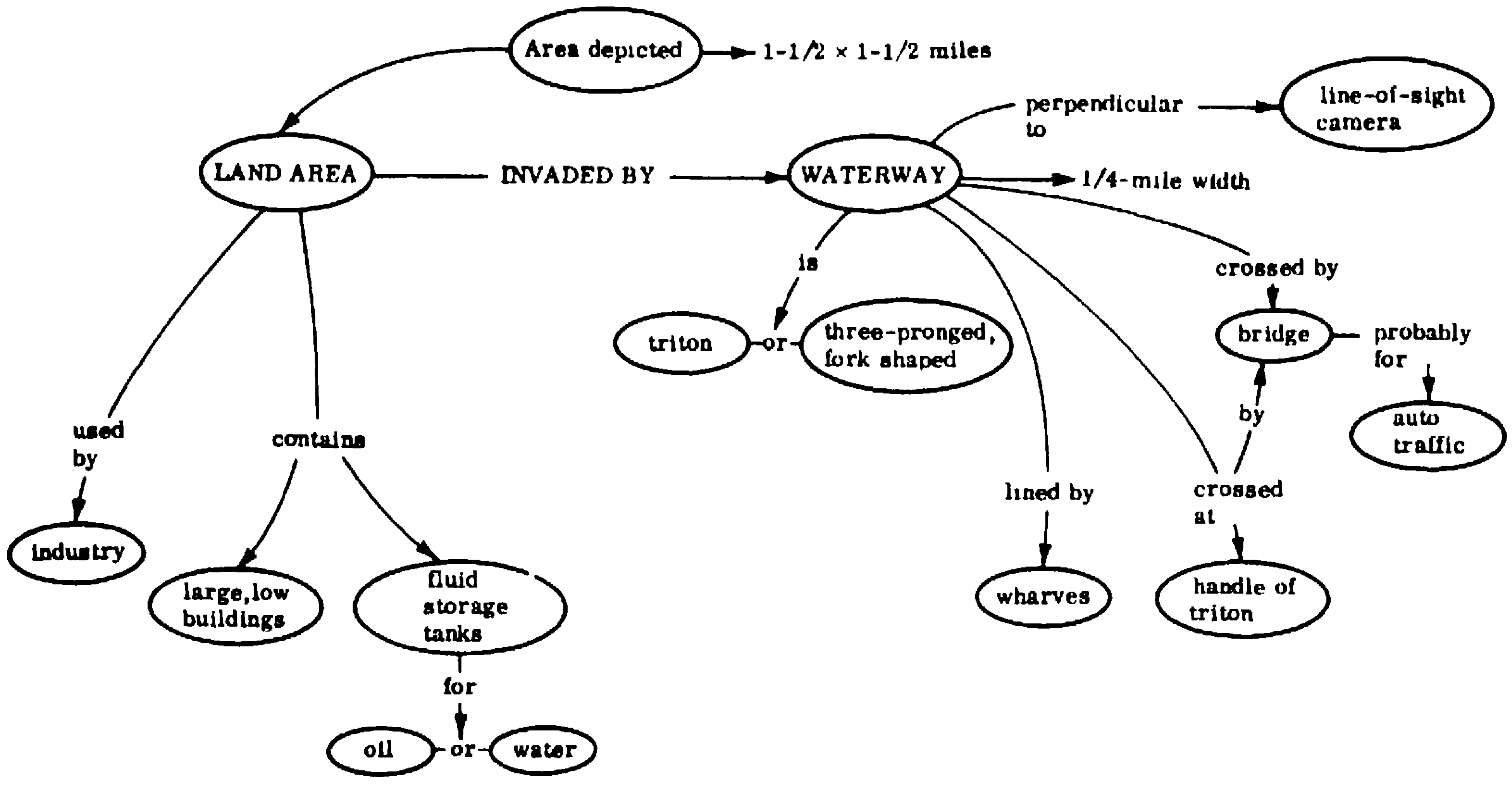


Fig. 4 Network Representation of Informal Description

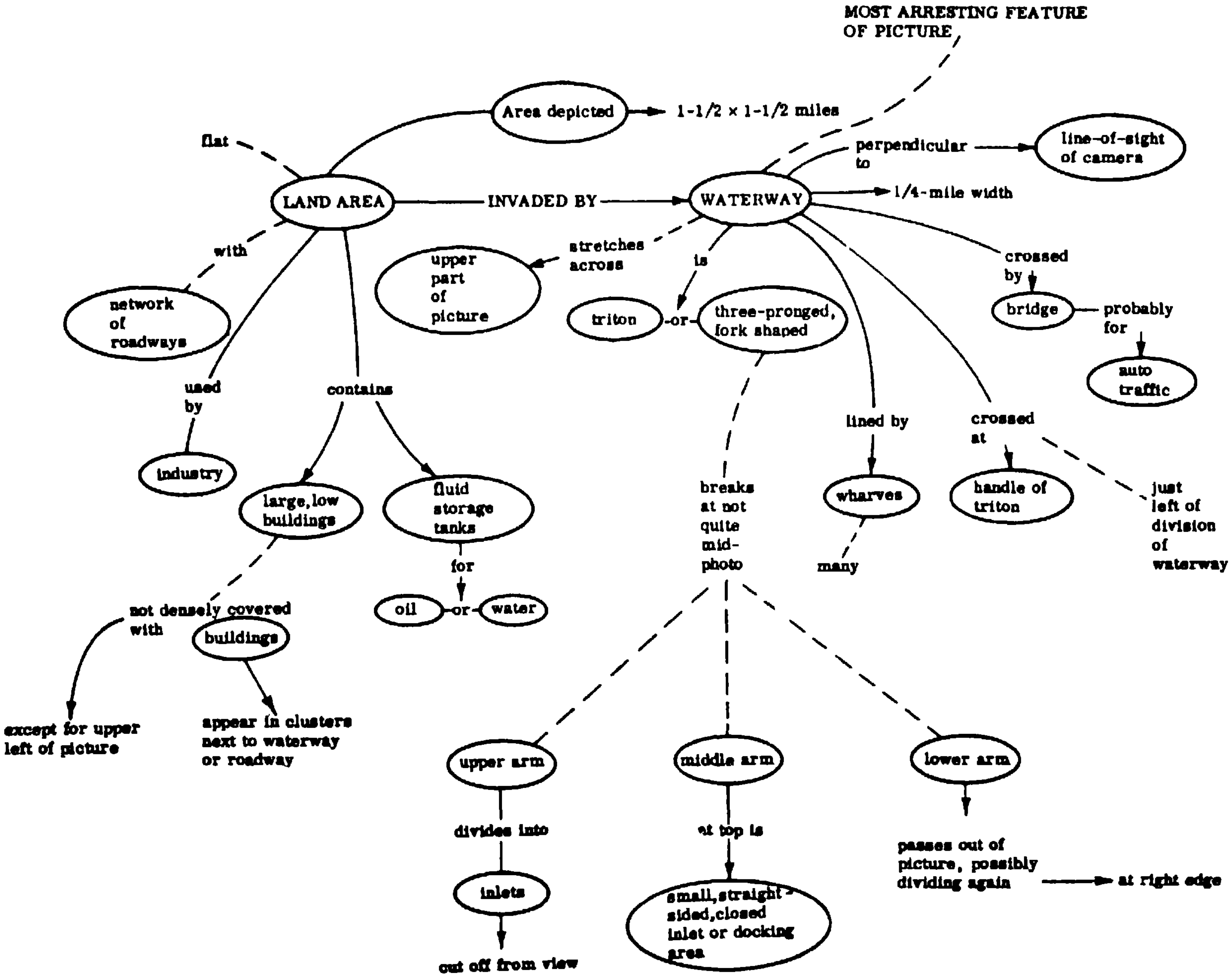


Fig. 5 Composite Description

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 3 arms at not quite mid-photo. 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. The land areas are flat and apparently developed for industrial use, with a network of roadways. There are many storage tanks and warehouse-like buildings. Except on the upper left, the land is not very densely covered with buildings; rather they appear in clusters next to the waterway or a roadway.

Fig. 6 Additional Description Used To Make Composite Description. Underlining indicates additions to original representation.

4. EXPERIMENTAL RESULTS

This section presents summary tables of the results obtained in the experiments to date with respect to the metadescriptive terms, the relationships used, and the deductions made.

4.1 METADESCRIPTIVE TERMS

Typical metadescriptive terms concerning the observer and comments concerning the description itself are given in Table 2. Most subjects were frank to admit uncertainty and inability to analyze a picture, using terms such as "I don't have the experience to," as indicated in Column (a). Moreover, they often included a rough estimate of the accuracy of their observations, using terms such as those shown in (b), ranging from low-probability terms such as "may" to high-probability terms such as "probably."

Table 2

TYPICAL METADESCRIPTIVE TERMS FOUND IN DESCRIPTIONS

(a) Concerning the Observer	(b) Concerning the Description
I guess	may
I am not sure	perhaps
I can't see well	possibly
I cannot tell	might be
I do not know	seems
I don't have the experience to	looks like
I had expected	would appear to be
There is little more I can say	probably
	about
	approximately
	I would say
	If I'm correct

4.2 RELATIONSHIPS USED

The relationships used in the descriptions are shown in Table 3. A partitioning has been made to indicate the general category that the relationship belongs to, such as the "joining, intrusion, division" category. Relationships in this particular category were found to be numerous in the descriptions, partly because of several of the photos used but also, we feel, because the human observer tends to use global relationships quite readily. Interestingly enough, such global relationships are probably the most difficult for automated analysis.

Table 3

RELATIONSHIPS USED IN THE DESCRIPTIONS

Conceptual Class	Role Played	Examples
1. Attributive	Attributes of object(s)	Straight-sided, closed inlet
2. Locational	Location of object(s) with respect to other object(s) or frame of picture	Highways are in the lower left
3. Containment	Inclusion of object(s) in other object(s)	Swimming pools in backyards
4. Localizing	Portion of object(s)	The sides of the waterway
5. Qualifying	Which of a set of objects is meant	The upper arm
6. Quantifying	How much of any object or how many	Three arms, very large, many docks
7. Logical connectives	Logical connection of objects	Oil and water
8. Comparison	Comparison of objects	Area is deeper than others; boats are larger than
9. Geometrical relationships	Geometric orientation of object(s)	Waterway is perpendicular to the line-of-sight
10. Joining, intrusion, division	Connection of objects or splitting apart of an object	Cuts into, leads to, invades; stretches across; breaks into; spreads out; branches into
11. Use or application	Purpose of object(s)	Developed for rail traffic; used by shipping

4.3 DEDUCTIONS MADE

The deductions made are given in Table 4. Again, a partitioning has been used to indicate the mechanisms used in the deduction. (When more than one mechanism is probably involved, the number of additional categories is indicated in parentheses.) In addition, we have partitioned the table so as to indicate whether the reasoning used was stated in the description or whether we made assumptions regarding the nature of the reasoning.

Table 4

DEDUCTIONS MADE

Reasoning Category	Our Assumption of Reasoning Used	Reasoning Used Stated in Description
(1) Use of objects	Shipping depot Auto bridge (2) Ship channel Runway for airplanes (2) Land used primarily by industry (2) Educational facility Area developed for recreation (2) Parking facility	
(2) Deductions based on adjacency or components	Recreation complex Track field Undeveloped land Freeway under construction Oil storage Shopping center Areas of industry	Might be a train because object is composed of long objects too close together to be cars or trucks. Building near sports area is probably a gym. Lumber yard, judging from the stacks of lumber. A church, judging from pointed roof.
(3) Analysis of texture	Sandy ground Dry, uncultivated ground Sand-lot yard Wooded area (5)	
(4) Analysis of shadow		The river flows from top to bottom of the picture, judging from the shadow cast by the dam.
(5) Analysis of color or gray level	Land looks fertile and green	Channel is darker, so it must have been dredged. Area is deeper than others; may lead in from ocean.
(6) Distance, area, or height estimation	Area of 1-1/2 × 1-1/2 miles Width of waterway, 1/4 mile Slight elevation in ground 2,000 feet in height 1/4 acre lots 100 houses in 70,000 square yards	
(7) Combination of many factors	Early morning affluence High density suburban sprawl Photo was taken on a clear day. The sun is reflected on the water. Does not appear to have been created by a single geological event	Structure seems to be more than a bridge – it might be a dam. Decent middle-aged community because there are a lot of trees and new communities do not have trees. Houses have no small front and back yards, indicating that the area is not a slum. This is eastern U.S. because there are a lot of trees, large lots, older homes.

Most of the deductions were found to be based on the use of objects or the presence of adjacent objects, e.g., "parking facility" and "Building near sports area is probably a gym." Analyses of texture, shadow, and color/gray level, although extremely important in specialized description, such as analysis of space photography (3), were not used frequently by the lay observers who constituted the majority of our subjects.

Only a representative set of distance, area, and height deductions has been entered in Table 4. *The* subjects were fairly free with such estimates, and often their estimates appeared without any probabilistic modifiers.

The final category, "combination of many factors," was most evident in responses to the instruction "Provide a title for this picture." The titles supplied, such as "Early Morning Affluence," and "High-Density Suburban Sprawl," represented an extremely high level of deduction, based on many factors.

5. ANALYSIS OF RESULTS

The number of sentences, and of sentences containing statements concerning the observer, expressing uncertainty of observation, and containing significant deductive content are given for each description in Table 5.

Table 5

SUMMARY OF METADESCRIPTIVE TERMS USED

Type of Instructions	Number of Sentences	Number of Sentences		
		Concerning Observer	Expressing Uncertainty	Containing Deductions
Minimal instructions given to subjects	8	0	1	1
	6	0	1	5
	22	2	6	6
		Used descriptors		4 terms
	13	2	5	2
	5	0	2	2
	25	2	4	3
	8	3	3	0
	7	0	1	2
Intended purpose of description given to subjects	30	0	4	8
	10	0	2	3
Specific questions asked of subjects	37	4	4	8
	15	2	1	3
Overall average	20	1 (approx. 5%)	3 (approx. 15%)	4 (approx. 20%)

It will be noted in Table 4 that there is quite a variation in the items from description to description, and we have used averages just to obtain a rough indication of the importance of some of the factors. It was found that sentences containing expressions concerning the observer were not used by most observers; that all observers expressed some form of uncertainty; and that almost all observers used deductive expressions.

Other results of interest are:

- In the few composite descriptions prepared, convergence was obtained within two or three descriptions.
- In the descriptions, we identified five different global approaches, as indicated in Section 2: central focusing subject, descriptor, partition approach, classification approach, and unorganized listing.
- Eleven categories of relationships and the roles they play were identified, as indicated in Table 3.
- The metadescriptive elements provide clues to the validity of the descriptions. Since the number of terms used by observers to indicate doubt and uncertainty was fairly limited, there iB a reasonable potential for formalizing or mechanizing the process of using such information to aid in assessment of validity.

6. DISCUSSION

The question of whether a symbolic representation can be used to capture the "meaning" or content of a picture has been raised. * In particular, the concept of whether such representations could be complete enough to act as an "encyclopedia," a data base for answering questions about the picture, has been the focus of this study. It is felt that this question is of increasing importance for machine perception as well as for information retrieval using pictorial material.

Due to the generality of the questions posed, we feel that it is not practical to conduct an investigation using large-scale experimentation. That is, the large number of viewpoints (i.e., different sets) that the subject can bring to an aerial photograph makes it difficult to obtain general results. Instead, we feel that completeness and validity should be examined by the use of a number of different observers, and an attempt should be made to obtain consensus on the various aspects of the description.

The utility of such representations will depend on the completeness of the representation and the accessibility of the data. This paper has dealt with the first topic; currently, we are studying the question of accessibility.

A basic requirement that arises in dealing with informal description is that of transformation of descriptions to a canonical form so as to decrease the variation in nomenclature, structure, and length of the descriptions. A tabular form and a network form for use as canonical representations were indicated, and a method of forming composite descriptions from the canonical forms was given. Our initial evaluation of the innovation of forming a composite semantic network from individual networks Looks very promising. The composites formed in this manner were found to give a clear representation of the original pictorial objects.

7. REFERENCES

(1) H. Schwarzlander, "Encyclopedic Storage of Scientific and Technical Knowledge," IEEE Trans. Eng. Writing & Speech, Vol. EWS-13, No. 2, Sep 1970

(2) O. Firschein and M. A. Fischler, 'Describing and Abstracting Pictorial Structures," Pattern Recognition J. (in press)

(3) Ecological Surveys From Space, NASA SP-230, National Aeronautics and Space Administration, Washington, D.C., 1970

(4) W. F. Miller and A. C. Shaw, "Linguistic Methods in Picture Processing - A Survey, " Fall Joint Computer Conference, 1968, pp. 279-290

(5) A. C. Shaw, The Formal Description and Parsing of Pictures, Ph.D. Thesis, Computer Science Department, Stanford University, 1967

(6) T. G. Evans, "A Grammar-Controlled Pattern Analyzer," Info. Proc., Vol. 68, Amsterdam, North-Holland Publishing Co., 1969

(7) R. Narasimhan, "On the Description, Generation, and Recognition of Classes of Pictures," Automatic Interpretation and Classification of Images, A. Grasselli, ed., Academic Press, 1969

(8) M. Ogi, "Pattern-Matching Techniques Applied to Indexing and Retrieving Films for Television Use," Proc. ASIS Annual Meeting, 1968

(9) B. L. Kenney et al., "An Automated Index and Search System for Psychiatric Videotapes," Proc. ASIS. 1970

(10) S. Rice, "Picture Retrieval by Concept Coordination, " Special Libraries, Dec 1969

(11) M. B. Clowes, "Transformational Grammars and the Organization Pictures," Automatic Interpretation and Classification of Images [seeRef. (7)]

(12) J. F. O'Callaghan and P. C. Maxwell, "On Describing Line Drawings," Seminar Paper No. 18, C. S. I. R. O., Division of Computer Research, Jan 1970

(13) M. A. Fischler, "Machine Perception and Description of Pictorial Data," Proc. Intemat. Joint Conf. on Artificial Intelligence, Washington, D.C., May 1969

(14) R. M. Krauss, "Language as a Symbolic Process in Communication," Am. Scientist, Vol. 56, No. 3, 1968, pp. 265-278

(15) P. H. Winston, "Learning Structural Descriptions From Examples," MAC TR-76, Project MAC, Massachusetts Institute of Technology, Cambridge, Mass., Sep 1970

(16) I. E. Sutherland, Sketchpad - A Man-Machine Graphical Communication System, Tech. Report No. 296, Massachusetts Institute of Technology, Lincoln, Lab., 1963

(17) M. Ross Quillian, "The Teachable Language Comprehender," Comm. ACM. Vol. 12, No. 8, Aug 1969

(18) T. Winograd, "Procedures as a Representation for Data in a Computer Program for Understanding Natural Language," MAC TR-84, Project MAC, Massachusetts Institute of Technology, Cambridge, Mass., Feb 1971

(19) R. C. Schank, L. Tesler, and S. Weber, "Spinoza II: Conceptual Case-Based Natural Language Analysis," Stanford Artificial Intelligence Project, Memo AIM-109, Computer Science Department, Stanford University, Jan 1970

(20) R. T. Bottle and H. Schwarz lander, "Variations in the Assessment of the Information Content of Documents," Proc. ASIS Annual Meeting, 1970