

1. Introduction to Computer Vision

In this section we will introduce some terminology related to computer vision. Terms such as computational vision, image processing, computer vision, image understanding, low-level vision, intermediate-level vision, high-level vision, top-down analysis, bottom-up analysis, and other terms are often used to describe some aspect of computer vision. The concepts are not well defined and often overlap. Nevertheless they do have some ability to describe approaches to the analysis of images.

Computer vision (CV) is concerned with processing towards a goal of having a symbolic representation of the image [Fischler and Firschein, 1987, pp. 242]. The symbolic representation should correspond to the representation that humans obtain when they process an image. The objects in the image should be correctly labeled. Computer vision is concerned with scene interpretation. An interpretation is basically a mapping between a symbolic description of the scene and the structure of the image.

Another term is computational vision. Computational vision examines issues involved in extracting information from images [Wechsler, 1990]. One often is involved with changing the form of the representation. The transformation from one representation of the image to another should make the solution to the analysis problem easier. Computational vision is concerned with issues of discrete data, resolution, representation, numerical methods, and computational complexity.

Image processing is a term with some correspondence to computational vision. Image processing usually refers to methods for transforming an image into another image with desirable properties. These properties might be that certain features in the image are enhanced or that the output image is more suitable for subsequent processing.

Image Understanding (IU) is another term often used. This term has overlap with the term computer vision. It basically refers to processing methods that build symbolic descriptions of the image which are consistent with human interpretation. The symbolic description should accurately reflect the objects and their relationships in the scene.

1.1 Computer Vision Basics

A CV system analyzes images by forming an interpretation of the scene based upon the scene model in the vision system. An interpretation is a mapping between the object models in the scene domain to the features in the image domain such as regions, lines, and pixels [Matsuyama, 1989; Matsuyama, 1987]. Predictions about the occurrences of object in an image are called hypothesis [Hwang, Davis, and Matsuyama, 1986]. An instance is a specific occurrence of an object in an image. For example, one may have a general model of a road as a linear object constructed of concrete or asphalt. An instance would be a specific occurrence such as Interstate 10. When an instance occurs the structure of the general model are copied and the values are filled in for each instance of the object. The instance is linked to the general model. The object road is said to be "instantiated". See the following figure. A model might be of the form

```
name road
type linear
identifier
lanes
surface material
```

while a specific instances of the object might be

```
instance 1
object
name - road
type - linear
identifier - Interstate 10
lanes - 4
surface material - concrete
```

```
instance 2
object
name - road
type - linear
identifier - State Hwy 19
lanes - 2
surface material - asphalt
```

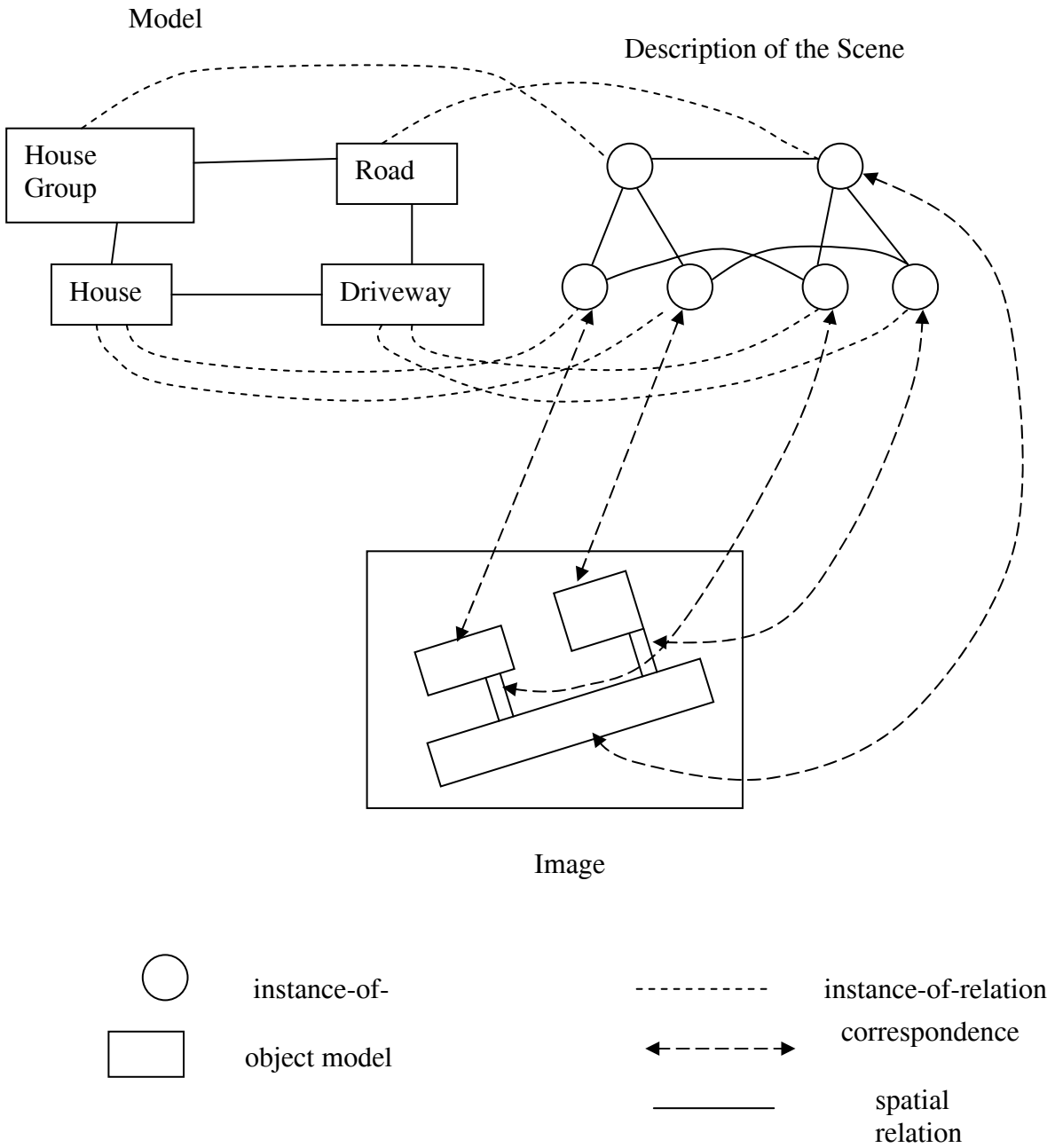


Figure 1. Interpretation of Scene

Objects usually have functions. For example, a glass holds a liquid and sits on a table. Sometimes the function of the object can be an asset in characterizing an object. There are many different sizes and shapes of glasses but they all fulfill the same basic function. Characterizing a glass using shape properties would be difficult. It might be easier to determine if the glass function is achieved by a given object. Understanding a scene requires knowledge about relationships between objects as well as knowledge about the intrinsic properties of an object. Intrinsic properties would be such properties as an object's shape, color, reflectance, or texture.

Another term often used in describing vision processes is the term bottom-up analysis [Ballard and Brown, 1982; Palmer, 1999, pp.84]. This form of analysis proceeds from lower level details and groups these entities to form more complex relationships. One would start with low-level analysis methods such as edge detection and then group edge pixels to form lines. Lines might then be grouped to form more complex objects. This is also referred to as data driven analysis. The term top-down analysis is also used [Ballard and Brown, 1982, pp. 343; Harlow and Eisenbeis, 1973]. This form of analysis is more model or hypothesis driven. High-level models in the scene domain generate expectations of image structure. These hypotheses are then verified from the object data. This is a goal directed process. A high-level hypothesis generates sub hypothesis until the hypothesis are simple enough to be verified from the image data. The analysis procedure is hypothesize and verify.

1.2 Computer Vision Levels

Computer vision approaches are often discussed in terms of levels. The first level is called low-level vision (LLV). LLV is related to local properties concerned with continuity or discontinuity of intensity, texture, or color [Fischler and Firschein, 1987, pp.242]. Noise reduction, smoothing, contrast enhancement, and edge detection methods fall into the category of low-level vision. These image processing methods are often general and can be applied to a variety of different image classes. Low-level vision is sometimes called early vision [Forsyth and Ponce, 2003, TOC]. Some other terminology related to low-level vision are the 2.5 dimensional sketch or the primal sketch [Winston, 1992, pp. 337,338]. This processing would determine features at each point such as the edge properties, texture properties, or surface normals for 3-D imagery. The grouping of points into regions corresponding to physical properties is often called forming the intrinsic image [Ballard and Brown, 1982, pp. 63][Chan, pp. 95??]. The intrinsic properties of an object are properties such as shape, color, reflectance, 3-D surface, or texture. The intrinsic image would have values for these quantities at each point. This processing derives meaning quantities at a level above the raw pixel values.

Another level of vision processing is call intermediate-level vision (ILV) [Fischler and Firschein, 1987, pp. 262, 87, 279]. Intermediate level vision is concerned with integrating local or point features into global constructs or regions. Examples would be edge point aggregation into lines and point aggregation into regions. This relates to image partitioning or segmentation into regions which perceptually relate to objects in the image. Typical models here are lines, parallel lines, perpendicular lines or polygons which represent simple geometric objects such as roads, doors, and buildings that commonly occur in scenes.

High-level vision (HLV) utilizes higher levels of processing and more complex object relationships in the modeling of objects [Fischler and Firschein, 1987, pp. 281, 87]. For example, processing methods might involve formal logic, constraint programming, rules, frames or other artificial intelligence programming methodologies. There is a more complex knowledge representation. Inferencing is involved in scene interpretation. The vision system has more semantic content and therefore the software is more specifically tailored to the specific image class. This corresponds to the more cognitive levels of image interpretation by humans. This methodology is sometimes called late vision.

These different levels of vision may overlap in any practical vision system. The concepts of the levels are useful mainly in characterizing the complexity of the vision system. It is also useful in relating to the computational processing in the human vision system.

1.3 Domains in Computer Vision

Now let us consider different aspects of computer vision analysis. Developing a computer vision system involves an interplay between the real 3-D world, the camera data, models of the world, software systems that implement a vision system, and human perception. These different aspects are referred to as domains [Argialas and Harlow, 1990]. This terminology is useful in describing different aspects of computer vision systems. First let us consider the world domain.

1.3.1 World Domain

The world domain is the actual physical world of 3-dimensional objects. In every situation, one has an understanding of the objects. Examples in aerial scenes would be cars, roads, houses, etc. Objects in turn may be composed of surfaces. An example would be a car with the different surfaces which appear at the fenders, hood, etc. A surface is characterized by a change in the 3-d representation of an object. e.g. plane to sphere. An object may be considered to be composed of sub objects which correspond to the different surfaces. Another example would be a residential object which might have sub objects houses, streets, yards.

1.3.2 Image Domain

Another domain is the image domain. The image domain consists of the image data obtained from the camera and data acquisition system. The light source, the world domain, and the camera system affect these data. These are the data our processing systems utilize. Some of the features present in the image domain are:

pixel - a point with a gray-level,

patch - this is a connected set of pixels with a uniform property such as gray level, and

region , this is a set of pixels which represent an object in the 3-d world domain.

Observe that regions representing objects in the image domain may be adjacent while the objects in the world domain (3-D) may not be adjacent. This is called occlusion. See the following figure that shows the relationship between the world domain and the image domain for a simple camera. Note also that the image does not uniquely determine the world as shown in the figure. Different scenes may produce the same image. Often one may loosely refer to the region associated with an object as an object or image object. One may also refer to world domain features such as junctions, points or lines in the image domain where one means the pixels or regions associated with the corner, junction, or line. Regions associated with linear objects such as roads, rivers, or edges of buildings have some special properties and are called elongated regions. One will not always clearly distinguish between the different domains when the meaning is clear.

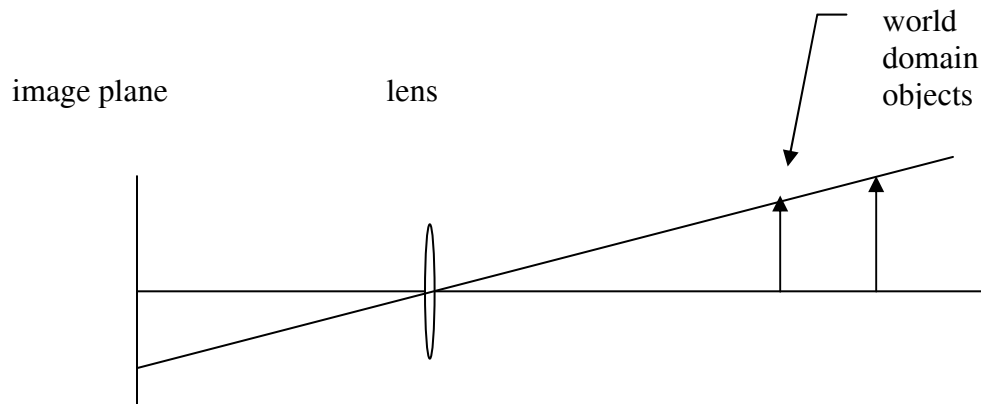


Figure 2. Image Domain and World Domain

The image domain data will be 2-dimensional unless one has made special arrangements to collect or 3-dimensional image data. There may be many different types of data available such as reflectance data from different spectral bands; thermal data, range data, and radar data. We make measurements in the image domain on pixels, patches or regions. Examples would be the gray level of a pixel, the average gray level of a region, or the area of a region. Extracted properties may be called image cues. Each pixel in the image corresponds to a location and intensity in the world domain. The basic problem in computer vision is to determine the structure in the world domain from the data in the image domain.

1.3.3 Scene Domain

The scene domain consists of the representations or models of the physical objects in the world domain [Maysuyama, 1989]. Examples would be a road modeled as a linear element which is white or black in color or has a given spectral response. Another example would be a house modeled as composed of rectangles. A basketball might be modeled as a brown sphere. We again call these entities objects in the scene domain. Each object must have a description of the 3-d object in the world domain it represents. Objects may be composed of sub objects in the scene domain. The object models in the scene domain are often called a world model. One interprets the image data according to the objects in the scene domain.

1.3.4 Software Domain

Another domain is the processing or software domain. This domain consists of the collection and organization of the software which performs the analysis and interpretation of the image data. Examples would be edge detectors, region formation methods, line followers, or texture analysis methods. This domain reflects the manner in which the software is structured and interacts with the image domain and scene domain.

A method of programming called object oriented programming is prevalent. The concept of objects occurs in the processing domain. These objects may be only loosely related to the objects in the scene domain since they are designed to optimize the software structure.

1.3.5 Perceptual Domain

The manner in which humans perceive the world using their visual system is called the perceptual domain. This refers to human interpretation of his visual information. Objects and features such as corners and lines are perceived. There is organization to the objects and processing done in this domain. Perception is directed and under the control of the cognitive system. Visual tasks are solved with a distributed computational system which forces constraints upon the solution. We have active perception strategies that reduce the computational complexity and makes the vision system robust and not sensitive to noise or extraneous data.

1.3.6 Domain Properties

There is interaction between these domains. The camera inputs data from the world domain into the image domain as data. The software system interacts with the image domain reading data and forming regions under direction of the models in the scene domain. The human viewer views the same world domain and forms his own models of the object relationships in the scene using the human visions system.

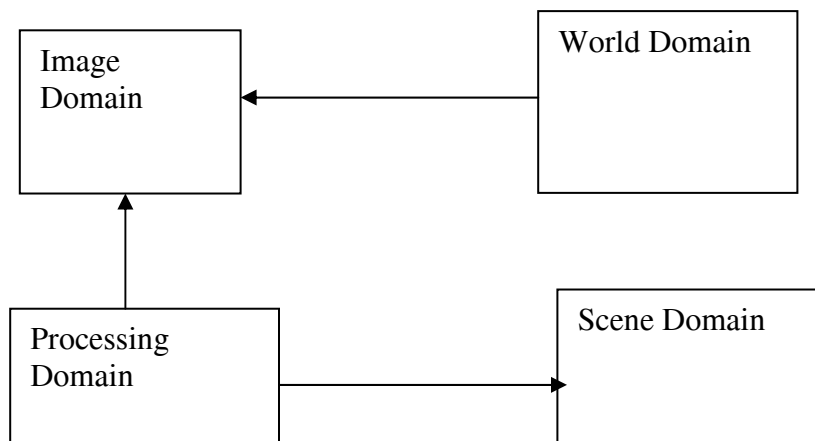


Figure 3 . Block Diagram of Domain Interaction

Each of these domains will have structures such as hierarchies. The structures may be closely related. For example, one may have a hierarchy of regions in the image domain, a hierarchy of objects in the world domain, a hierarchy of objects in the scene domain, and a hierarchy of processing modules in the processing domain.

These domains are related and there must be close interaction in the vision system. This interpretation should be similar to human interpretation of the world

domain. The next figure shows a scene domain object hierarchy [Harlow, Trivedi, and Connors, 1986].

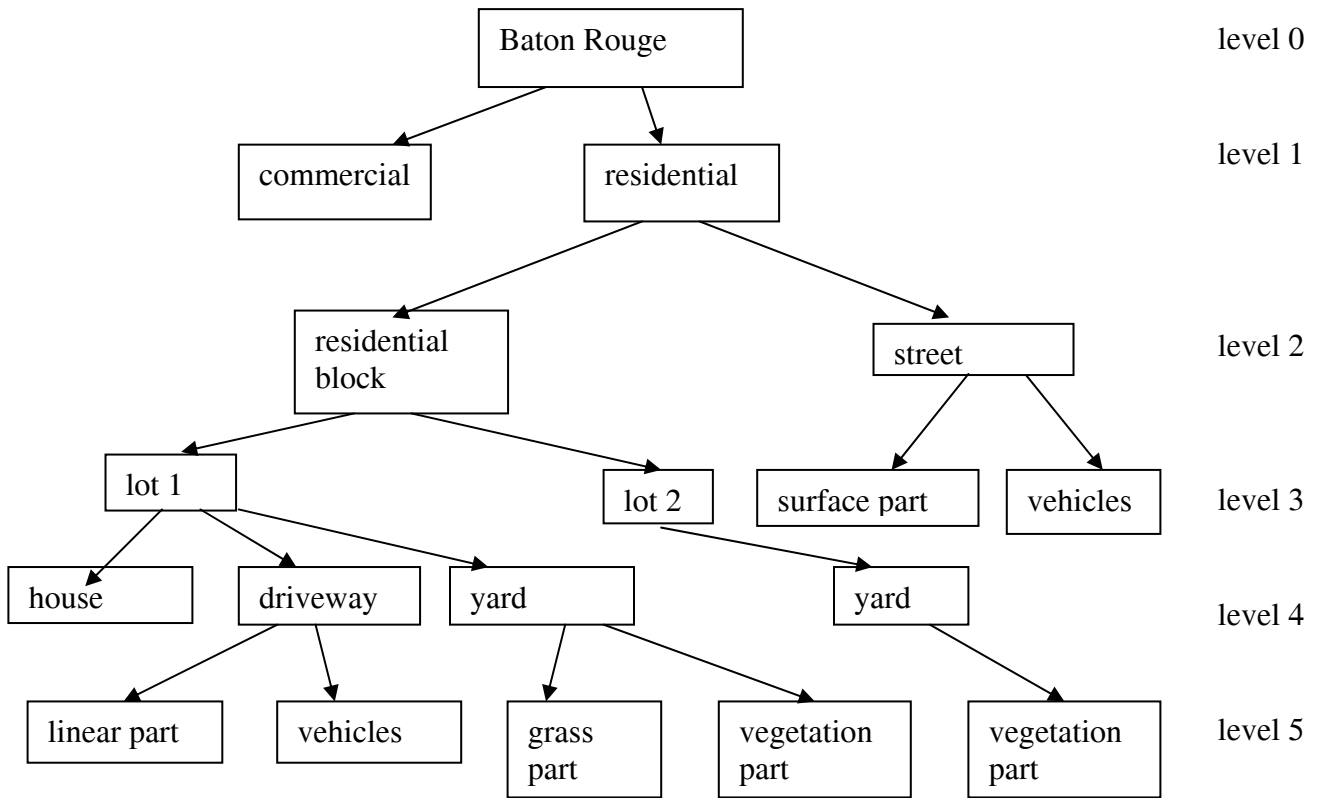


Figure 4. Scene Domain Object Hierarchy

The following figures shows some of the different types of structures that can occur in an urban scene. The variety of different structures is high.

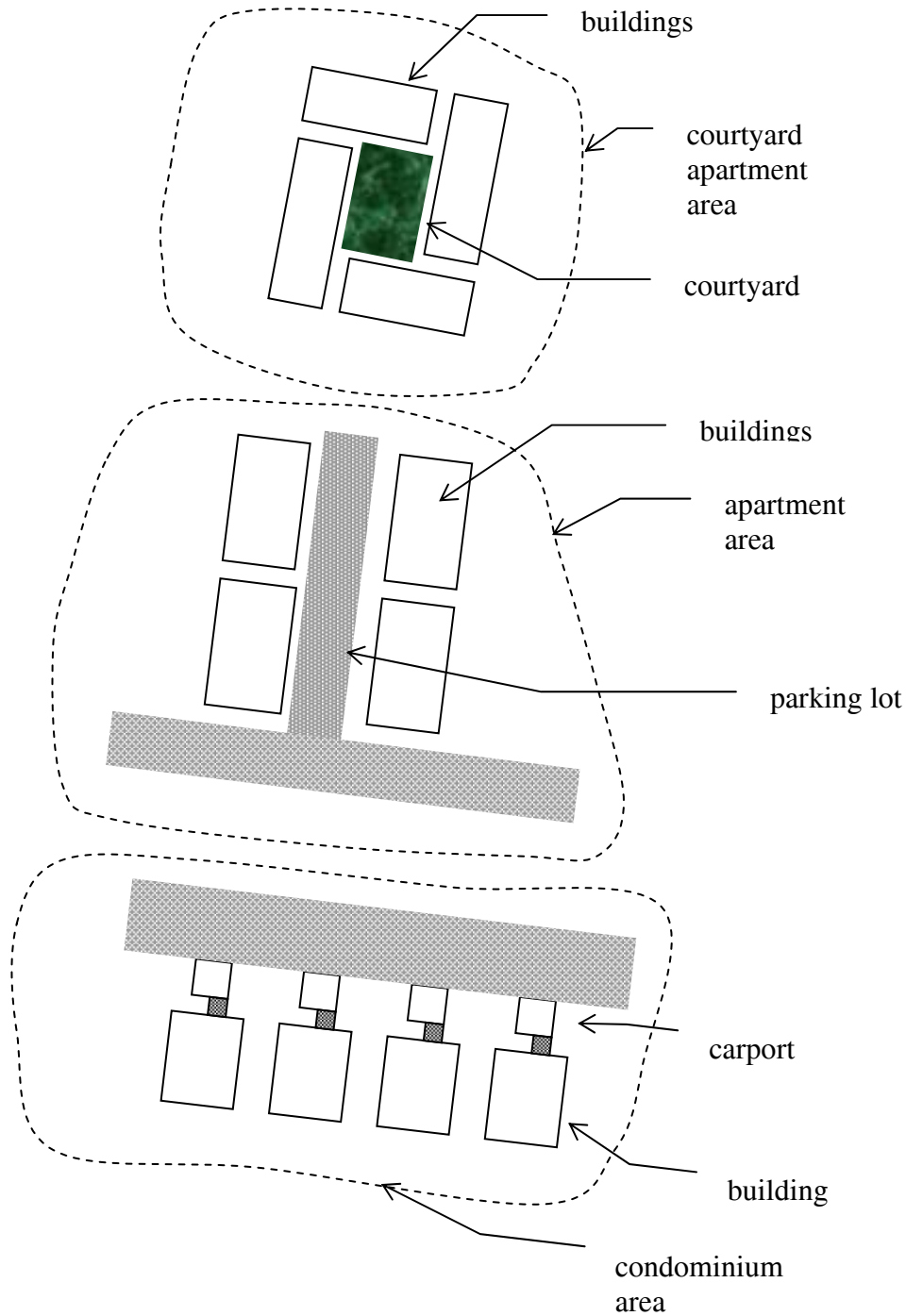


Figure 5. Structures in an Urban Scene

The following figure shows a possible interpretation of a scene with the interplay between the image and the scene domain.

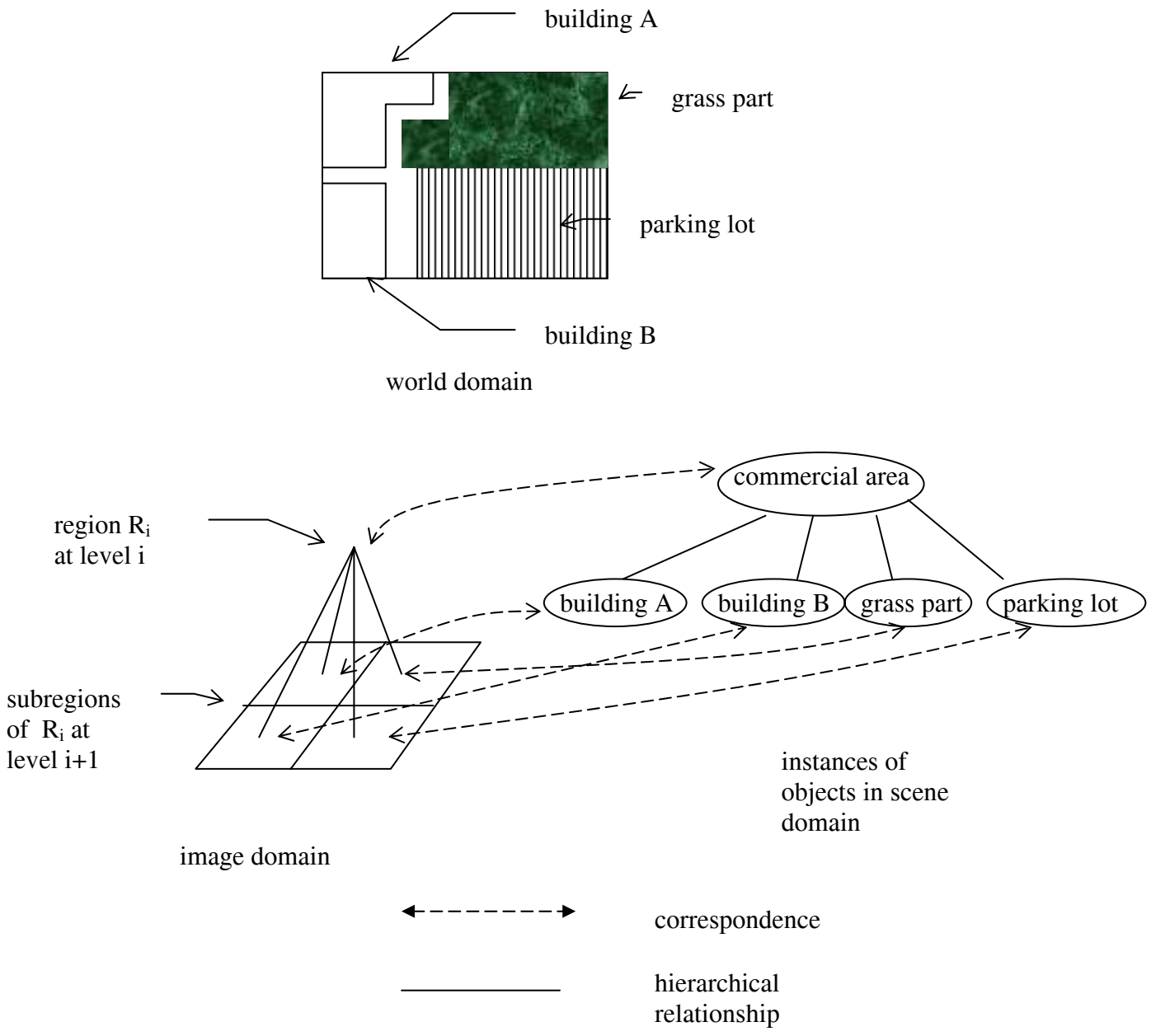


Figure 6. Interaction between the Image and Scene Domain

The next figures show some possible operators that might occur in the processing domain. They would use information from the scene domain to determine the expected relationships between objects [Harlow, Trivedi, and Connors, 1986].

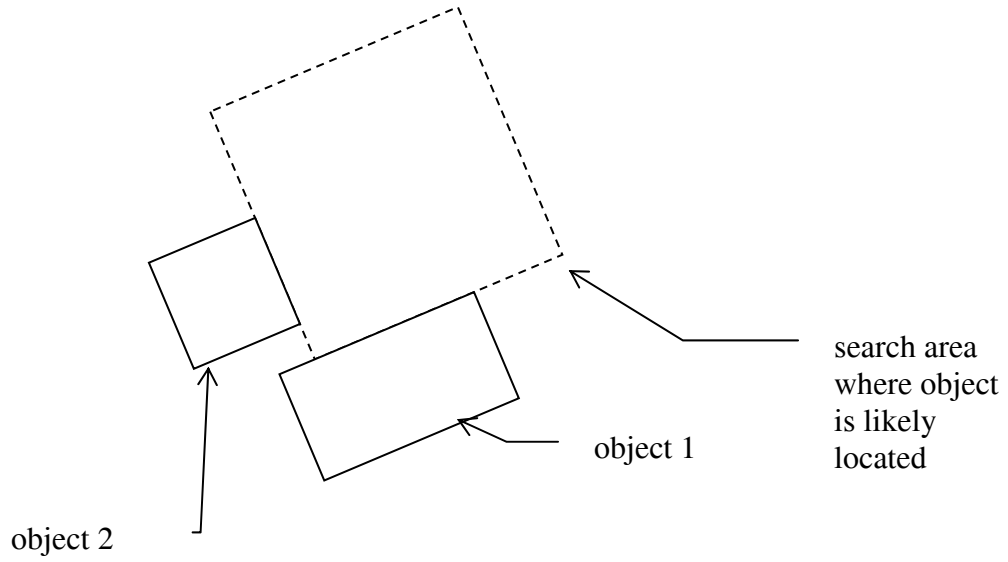


Figure 7. Corner Search Operator

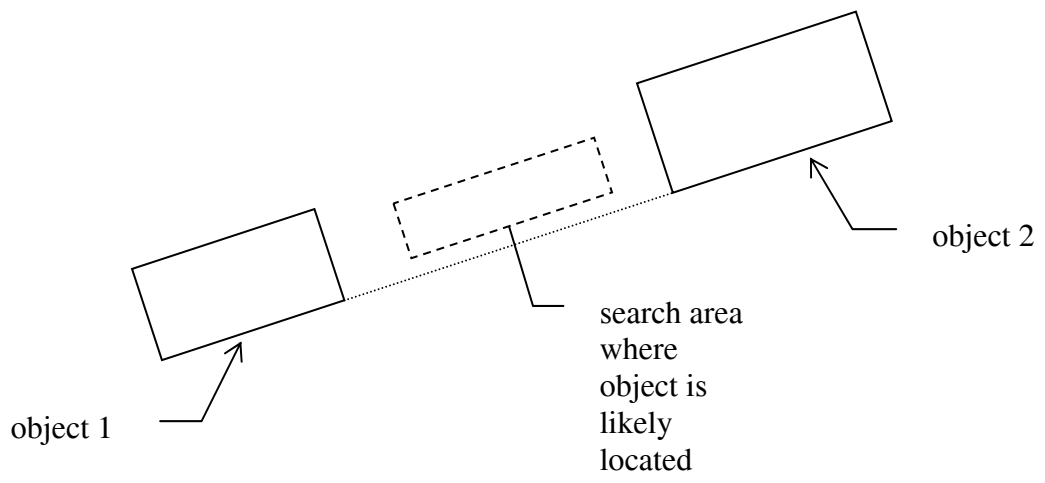


Figure 8. Between Operator

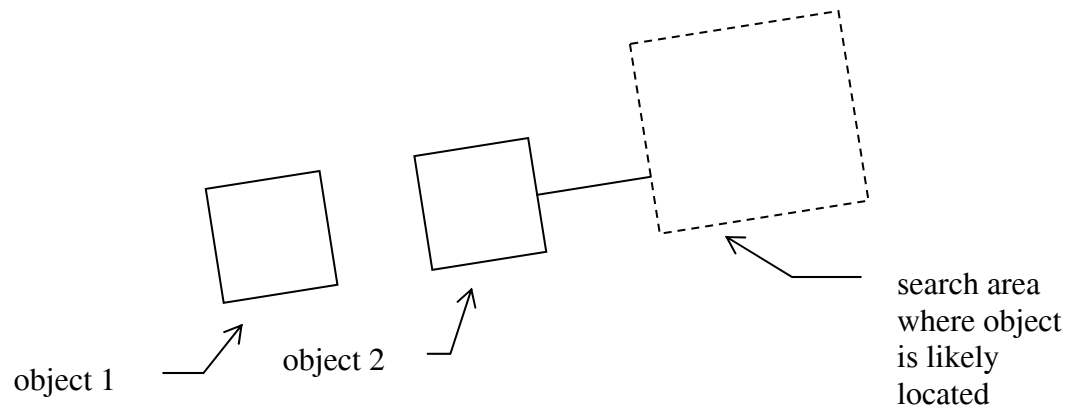


Figure 9. Directed next-to Operator

1.4 Approaches to Computer Vision

A computer vision system can be conceptualized as being composed of vision operators, an inferencing system, and a knowledge base of models and facts. The vision system should label and describe the objects, sub objects and relationships in the scene. The vision operators are used to extract image cues from the image. The role of the inferencing system is to sequence or select the models in the knowledge base, match the extracted image cues against the models, resolve conflicts and track the inferences. Sophisticated control structures will be required to implement vision systems for aerial scenes.

In creating a vision system, a number of problems must be addressed. These are: the interaction between the vision operators and the inferencing system, the selection and adequacy of the vision operators, the method of forming an optimal global interpretation of the scene, and the integration of system modules and data structures so that the system can be improved as additional knowledge about the scene becomes known.

Knowledge must be expressed in terms natural to the objects so that the extraction of image cues and the descriptive and explanation capabilities of the vision system are enhanced. It might be noted, that knowledge replaces the need for training samples that are often not available. One would like to develop analytic models of appropriate descriptive power with the proper invariant properties relative to imaging transformations that describe the scene and imaging process. This should help to systemize the input of knowledge and reduce the reliance upon heuristic input of knowledge. It is difficult in general cases to express spatial knowledge and make reliable spatial inferences.

Most of the early vision systems focused upon edge detection and strategies for connecting edge points into lines. This approach has intuitive appeal because the eye is sensitive to edges and humans can recognize objects from the object's outline. These approaches have encountered difficulty in practice because the operators often confuse scene detail and noise. This is especially true in images of natural scenes. It is also, at times, difficult to correctly link the edge pixels together to form the correct outline of objects in complex scenes. Reviews of edge operators are given in [Peli and Malah, 1982; Brooks 1978].

Region formation methods were developed to overcome some of the problems with edge detection [Zucker, 1976]. In region formation one attempts to locate areas of the image that share some uniformity property. The most common property used is uniformity of gray level. The hope is that it is easier to locate regions with a uniform property than to measure dissimilarity with edge detectors. [Pavlidis, 1977] describes split and merge techniques for region formation using various methods to measure region similarity. All the various thresholding methods using histograms are a form of region formation. [Ohlander et. al. 1978; Trivedi et. al. 1985; Kapur et. al. 1985] indicate some thresholding methods. Color [Ohlander et. al., 1978], multispectral [Kettig and Landgrege, 1976] and texture properties [Chen and Pavlidis, 1979,1983] have been used to some extent in region formation. The term region formation or region growing is the term historically given to these processes.

The important conceptual distinction between the image domain and the scene domain was noted in [Clowes,1971]. The image domain is a domain of observable facts which can be obtained from the image while the scene domain is an abstract domain in which objects are modeled and relations are defined. [Kanade,1977] describes the

hierarchies which exist in a vision system. These hierarchies are a processing unit hierarchy and a detail hierarchy. Within the image domain the processing unit hierarchy refers to the size of the areas processed in an image such as pixels or regions. The detail hierarchy that is in the both the scene and image domains refers to the precision of description. The description may range from crude to very detailed. In the scene domain this might be the precision of shape descriptors. In the image domain this be the resolution of the data. There is also a composition hierarchy which includes part-of relations. The details of these hierarchies must be made explicit [Kanade, 1977].

1.5 Vision Control Strategies

When discussing vision systems three control strategies are often mentioned. These are top-down, bottom-up and a combination of the first two strategies. In a top-down approach one starts with a hypothesis about the scene. The hypothesis is then decomposed into sub hypothesis. Matches are made at lower levels of the hierarchy to determine if the hypothesis is valid. If the hypothesis is not valid another hypothesis is chosen. Top-down processing is goal directed processing. High-level goals generate sub goals until one can solve the goals and then back the results up the hierarchy. In vision systems one is making hypothesis about objects and their appearance. The hypothesis must then be verified by extracting image cues. On most vision systems the hypothesis are verified at the lowest levels in the hierarchy. A number of vision systems developed to analyze chest radiographs employ a top-down strategy [Harlow, 1973,1976]. A top-down approach works for this class of scenes because they are highly structured. In general a pure top-down approach is not suitable for scenes with a large amount of variability in their structure such as aerial scenes.

Most vision systems use a bottom-up strategy, a process which begins at the bottom of the hierarchy with the examination of fine detail, e.g. cars and houses, and attempts to identify more general entities such as residential areas by recognizing appropriate groupings of the smaller objects. One usually starts by segmenting the image to find homogeneous regions or lines that correspond to objects at the lowest level. Local properties are usually used to produce the segmentation. One should note that the analysis methods used are called low-level vision techniques since little information is utilized. Simple properties of scene geometry and gray level contrast between object and background typify the process. The segmentation process is in general difficult to perform correctly. Even if a good segmentation is achieved, the segmented image contains less information than the original image. The lost information might be important in obtaining the best interpretation of the scene. The analysis is driven by the data analysis at the lowest level. Bottom-up analysis is similar in spirit to forward chaining in an inferencing system which derives consequences from established results [Rich, 1981].

A bottom-up approach has a number of limitations. The segmentation of the scene is based entirely upon image cues obtained from operators that have relevance only to the information at the lowest levels of the hierarchy in the scene domain. These operators use little or no information from the scene domain that makes errors likely and also contributes to insufficient information being available for correct interpretation. Since little context is utilized, an extensive search process is required to locate regions which

might correspond to the low-level objects in the scene domain. Also, there is too much emphasis placed upon getting a good initial segmentation of the image.

The objects in the scene domain that can be explicitly verified in the image domain correspond to low-level objects. This means complex models in the scene cannot be directly verified and independent evidence cannot be gathered at intermediate levels of the hierarchy for a belief maintenance system.

Since explicit models for high-level objects and their relationships cannot be verified with image cues, the objects at high levels in the scene domain hierarchy must be verified by defining relationships on lower level objects. After one has obtained interpretations of low-level objects from image cues, one can then infer the interpretations of the objects at higher levels in the hierarchy. When errors are made they invoke incorrect models at the lower levels. These errors will be inferenced to the higher levels because of the dependence upon image cues obtained at the lowest levels [Kanade, 1977]. These inferences may proceed several levels in the process making interpretation at high levels in the hierarchy more error prone than necessary.

Only local constraints and not global constraints can be utilized since only low-level objects can be explicitly verified in the scene. Global constraints often are important in scene interpretation.

1.6 Summary of Some Vision Systems.

In general, it is important not to constrain a vision system to a pure top-down or bottom-up approach. One needs a vision system which incorporates both goal and data driven analysis. In [Havens and Mackworth, 1983] a number of the above limitations are described and a hierarchical scene analysis system is proposed to overcome these limitations. At each node a scheme or frame is proposed to model an abstract object. This system provides for procedural knowledge at each node to guide the search for instances of an object.

The system by [Nagao and Matsuyama, 1980] is a vision system for multispectral data of suburban scenes. Regions are characterized by spectral and spatial features such as color, size, shape and texture. Regions called characteristic regions are a focal point of the analysis. Examples would be large homogeneous, elongated, shadow, water and vegetated regions. A production system is used to incorporate knowledge in the system. It features production rules and image cues for the image extracted at a low-level. In a later paper the group introduces methods for analyzing texture patterns. Texture elements are extracted by a region growing method [Nagao and Matsuyama, 1980]. An aerial scene analysis system called SIGMA is described in [Matsuyama, 1987]. The system utilizes frames. In [Levine and Shaheen, 1981] a system applied to natural scenes is described. The modules of the system are low-level processes, measure analyzer, hypothesis initializer, hypothesis verifier, focus of attention, and scheduler. These processes communicate through long-term memory and short-term memory. The data are arranged in a relational database. The system is implemented as a rule based system. In a later paper [Nazif and Levine, 1984] a rule based system for low-level segmentation is described. In [Binfor, 1982], computer vision systems are surveyed. The ACRONYM system developed by the authors of the paper is described. Points are made in this article that most systems: operate on two-dimensional data, use fairly simple world models; have limited segmentation procedures and use only weak descriptors of shape and texture. The ACRONYM system uses generalized cylinders for described 3-D shape. There are many objects which occur in aerial scenes which do not lend themselves to this description. The recognition strategies bottom-up. In [McKeown, 1985], a rule based system for aerial scenes is described. Airport scenes are considered. The rule based system interprets the scene by building interpretations based upon an initial segmentation which is produced by a region growing program. Region properties are extracted to determine an association between regions and airport features. An initial confidence is calculated to determine how well a region fits the feature description. Rules are organized into classes. Initialization rules give the goal states, map database, class expectations, and low-level segmentation. Region-to-interpretation rules create an initial hypothesis for each region. Local evaluation rules are used to enlarge regions. Consistency rules apply spatial and context constraints to modify the confidence of initial fragment hypothesis. Functional area rules recognize when fragment interpretations can be grouped into functional areas such as runways, taxiways and tarmac. Goal generation rules recognize situations inconsistent with airport structure in order to prune weak fragments from further consideration. Model generation rules assemble functional areas into a model for the airport scene.

Another rule-based vision system for aerial scenes is discussed in [Perkins, Laffey and Nguyen, 1985]. In this system, the image is first segmented using an edge based

segmentation technique. Attributes are then computed for each patch. The attributes include intrinsic properties such as area, average intensity and texture strength. Other attributes are based upon context such as adjacent to a specified region. The patches and their attributes form the data base for the rule based expert system. A semantic network (SN) represents declarative knowledge in a system described in [Nicolin and Gabler, 1987]. An entity of the SN consists of property slots and relation slots. The SN is structured by two hierarchies of relations. The first hierarchy is the generalization and specialization relation. Inheritance occurs through specialization. The other hierarchy is by the composition and decomposition relation. This hierarchy gives the structure of complex objects from less complex objects. The system provides for long-term memory (LTM) and short-term memory (STM).

In summary, many systems use rules for aerial scene analysis. There is a reliance upon low-level segmentation and low-level image cues. The image cues extracted are of the simple variety as typified by edge detectors, thresholding and region growing based upon gray level differences. It is easily recognized that these methods need to be improved and generalized in order to provide more reliable cues and a richer descriptive vocabulary. Vision systems have progressed so that the inferencing mechanisms are more complex. The operators used with the vision systems for the most part have changed little. There has been improvement in obtaining interpretation of the outputs of these operators with better inferencing systems. It is clear that to achieve the analysis of complex aerial scenes it is required to develop substantially more complex systems.

1.7 Useful Properties of Vision Systems

In the previous sections, we mentioned a number of issues important in a vision system. One point is to make certain hierarchies explicit in the vision system. This included knowledge hierarchies and descriptive hierarchies. Another point is the utility of a frame based system which facilitates hierarchical structures and the incorporation of procedural and declarative knowledge. Contextual information tends to be declarative information since it indicates facts about objects and not how to determine the facts. Procedural information is concerned with the method in which information is obtained. The use of declarative information removes the reliance upon training samples and statistics which often is not available. The gestalt and temporal rules used by experts to make interpretations of the imagery must also be utilized. It will be important to incorporate knowledge about how surface materials appear in scene, sometimes called reasoning from first principles [Ferrante et. al., 1984]. A belief maintenance system is needed to combine the information obtained from the different sources. The results obtained from any single analysis will be imprecise and subject to error. Several belief maintenance systems have been proposed. Some are heuristic and some have a theoretical basis. Fuzzy logic and the Dempster-Shafer [Shafer, 1976] theory are examples of systems with a theoretical base.

The image analysis operators are a vital part of any vision system since they extract the image cues from the data upon which all inferences and results are based. It is important to develop operators that reliably return image cues which measure some perceptual property. These operators should have a certain robustness. An example would be operators that preserve perceptual similarity. If one used these operators to extract image cues, then one would extract measures from regions which are close in measurement space when the areas are close in perceptual space. Thus, the vision system should not make severe errors such as calling a residential area some disparate label such as commercial. It is also important to develop operators which extract image cues at intermediate levels in the scene descriptive hierarchy. This will relieve the inferencing system from undue dependence upon low-level image cues. One also needs operators with a rich descriptive vocabulary so that image cues with meaningful perceptual terms can drive the vision system. These descriptions should convey more information than the presence of uniform gray level in a patch or the presence of an edge.

1.8 Knowledge Sources Relevant to Vision Systems

The use of world knowledge is critical to the development of robust systems. The manner in which knowledge is structured and utilized in the analysis of the images is an important consideration. At the present time we designate these knowledge sources as discipline knowledge, general knowledge, specific knowledge, and object knowledge. In this section we indicate some ways in which knowledge can be used to guide a vision system. We will consider analysis of remotely sensed data as an example system to be considered.

1.8.1 Discipline Knowledge

Having knowledge in a particular discipline additionally increases the ability to accurately identify specific features. Experts in botany, forestry, or biology would better interpret vegetation communities. For example, if forests are to be mapped, then a general photointerpreter could do the job. If detailed communities within the forest are to be mapped, a forestry background would be required. A person familiar with temporal changes of a specific feature, would first utilize photography acquired at the time of year where the maximum contrast between features occurs. Second, the interpreter knows what can and cannot be discerned considering seasonal ground conditions. Considering forest communities, certain desired classes would need to be merged or separated depending on the season of image acquisition. Forest communities also are generally located in specific spatial arrangements. Certain communities are along shorelines and are early invasion species while others are located in older mature or stable areas. These few examples demonstrate the need for discipline knowledge to make determinations of basic features.

1.8.2 General Knowledge

General knowledge includes information from two sources: the position of the image (latitude and longitude) and the date. By combining lat/long with global environmental and political maps, information such as nation and biome (large ecological communities, such as tropical forest, desert, or alpine forest) could easily be determined. Knowing the nation is important. For example, grain silos in the United States are round, while those in Canada are square. Further, the distance to the nearest urban area would give an indication of whether the scene was part of a natural or human landscape. Proximity to other cultural and natural features can also be determined from location data. For example, the lat/long of many geographic features in the world are known. Examples would be major cities, railway stations, lakes, mountains, canyons, and world ports. For ports ancillary data such as harbor size and type, maximum vessel size, types of cranes (fixed, mobile, or floating), and access to railways is known. This information would greatly simplify identifying a port. The date can also provide useful information, since the objects one expects to find in a scene are often dependent on the time of year. For example, it would be futile to use the attribute "contains leaves" to identify deciduous trees if the image was obtained during the winter. Further, winter scenes could contain snow and ice, whereas summer scenes would not. Thus, knowledge of the season is important for determining object lists as well as attribute lists.

1.8.3 Specific Knowledge

Although similar environments contain many similar objects, many objects are region dependent. For example, a marsh in Louisiana or California could be expected to have oil wells, whereas this would not be expected in a New England marsh. This information would shorten the object list by generating a list of those objects actually found in that specific region, rather than relying on a generic list for that general environmental class.

Low resolution land use data already exists for much of the world. This data can be used to provide the specific context for a scene. Further, the frequency distribution of objects could help determine which objects were most likely or least likely to occur. This would substantially shorten search times. Statistical information is useful for locating "conspicuous objects". An object is conspicuous if it occurs in an unexpected location. Therefore conspicuousness is related to context. A large building is not conspicuous in an urban but is conspicuous in a forest.

Other forms of statistical information can be utilized besides frequency distributions. Statistical relationships may exist for an object and its distance to some feature. For example, an oil processing facility on the coast is much more likely to be located adjacent to a canal than in the middle of a marsh. If one was searching for such a facility, then the search should first focus on areas near canals. Similarly, an object located greater than 100m from a canal would probably not be an oil facility. Density distributions can also yield valuable information. For example, a forest on a steep grade with low fertility would be expected to have a lower tree density than a forest on a level, fertile piece of land. In addition, the density of commercial buildings is different for suburban and urban areas. An area near an interstate exit would be expected to have a high density of fast food restaurants and motels. By utilizing these spatial relationships, features can be identified more efficiently.

1.8.4 Object Knowledge

In a system designed to automatically identify objects from aerial imagery, the system's performance is determined by the complexity of the scene and the characteristics of the object. Any information that can be used to reduce the number of possible objects will improve the performance of the system. For example, one might determine object attributes such as height, size, brightness, texture, shape, etc. The number of round objects that are found in aerial photographs is large probably on the order of hundreds to thousands. Some of the round objects that might be encountered are grain silos, water towers, salt domes, oil storage tanks, traffic circles, radar domes, and buildings. Thus the attribute "round" is not sufficient for a unique identification. World knowledge can provide the system with the context of a scene. Knowing the context reduces the number of expected objects, and also provides a more logical framework from which the search can be executed. Suppose it was known that the scene came from an urban area. Of the seven round objects listed above, only three might be expected in a typical urban area (water towers, traffic circles, and buildings). On the other hand, only two of these objects might be expected in an agricultural region (grain silos and water towers). Thus knowledge immediately reduces the number of possible round objects. Another factor is that without knowledge we know nothing about the expected frequencies of different

objects. Thus, it must be assumed by default that these objects are equi-probable. We then have no basis for selectively testing the most likely candidate. If it was known, however, that a building is the most likely round object in an urban region and a traffic circle the least likely, then the search could be streamlined by first testing for the most probable candidate. A particular object can have many spatial attributes. However, some of these attributes are more powerful for identification purposes than others. For example, shape is not a particularly useful attribute for identifying the World Trade Center, whereas height is. It is therefore extremely useful to know the "most distinguishing characteristic" (MDC) of an object. This MDC is not absolute, however, but depends on context. For example, height would not be the MDC for a tall building in a city full of skyscrapers. The MDC must be derived given the context of the specific area.

The manner in which the function of an object can be used to locate and identify the object. Many objects have a wide variety of possible forms or shapes in which they may appear in imagery which makes it difficult to characterize them by shape or other properties. Road networks have widely varying geometries and sizes for example. The purpose of the road network is to provide access to cities, residential areas etc. A similar statement is true for canals that provide access to petroleum exploration platforms. Buildings and plants have widely varying shapes. If the plant produces electricity or petroleum products, there will be basic differences in the two facilities due to their function. The function of an object will be studied in order to gain additional insights into determining distinguishing characteristics of objects and the manner in which these characteristics can be used by the vision system. Another factor related to objects is that certain features add more to our understanding of a scene. Locating these objects initially therefore aids in subsequent identification. For example, in a scene with motels, amusement parks, jetties, and piers, the object that would add most to our contextual understanding might be the beach.

1.9 References

- Argialas, D., and Harlow, C. A. (1990). "Computational Image Interpretation Models: An Overview and a Perspective." *Photogrammetric Engineering and Remote Sensing*, Vol. 56, No. 6, June 1990, pp. 871-886.
- Ballard, D. B., and Brown, C. M. (1982). *Computer Vision*, Prentice Hall, Englewood Cliffs, N. J.
- Binford, T. (1982). "Survey of Model-Based Image Analysis Systems." *International Journal of Robotics Research*, Vol. 1, No. 1., pp. 18-64.
- Brooks, M. J. (1978). "Rationalizing Edge Detectors: Some Comparisons." *Computer Graphics and Image Processing*, Vol. 8, 1978, pp. 277-285.
- Chen, P., and Pavlidis, T. (1979). "Segmentation by Texture Using a Cooccurrence Matrix and Split-and-Merge Algorithm." *Computer Graphics and Image Processing*, Vol. 10, 1979, pp. 172-183.
- Chen, P., and Pavlidis, T. (1983). "Segmentation by Texture Using Correlation." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-5, No. 1, 1983, pp. 64-69.
- Clowes, M. (1971). "On Seeing Things." *Artificial Intelligence*, Vol. 2, No. 1, 1971, pp. 79-112.
- Ferrante, R. D., Carlotto, M. J., Pomarede, J.-M., and Paul W. Baum. "Multispectral Image Analysis System." *Proceedings First Conference on Artificial Intelligence Applications*, Denver, Col., pp.357-363.
- Fischler, M. A., and Firschein, O. (1987). *Intelligence: The Eye, the Brain, and the Computer*, Addison-Wesley Publishing Company.
- Gonzalez, R. C., and Woods, R. E. (1992). *Digital Image Processing*, Addison Wesley.
- Haralick, R. M., and Shapiro, L. G. (1985). "Survey: Image Segmentation Techniques." *Computer Vision, Graphics, and Image Processing*, 29, 100-132.
- Harlow, C. A. (1973). "Image Analysis and Graphs." *Computer Graphics and Image Processing*, Vol. 2, No.1, August 1973, pp.60-72.
- Harlow, C. A., Dwyer, S. J., and Lodwick, G. S. (1976). "On Radiographic Image Analysis." *Topics in Applied Physics*, Vol. 11., Springer-Verlag,.
- Harlow, C. A., and Eisenbeis, S. A. (1973). "The Analysis of Radiographic Images." *IEEE Transactions on Computers*, June 1973., pp. 678-689.
- Harlow, C. A., Trivedi, M. M., and Conners, R. W. (1986). "Scene Analysis of High Resolution Aerial Scenes." *Optical Engineering*, Vol. 25. No. 3, March 1986., pp. 347-355.
- Havens, W., and Mackworth, A. (1983). "Representing Knowledge of the Visual World." *Computer*, August, 1983, pp.90-96.
- Horn, B. K. P. (1986). *Robot Vision*, MIT Press, Cambridge, Massachusetts,.
- Hwang, V. S.-S., Davis, L. S., and Matsuyama, T. (1986). "Hypothesis Integration in Image Understanding Systems." *Computer Vision, Graphics, and Image Processing*, Vol. 36, pp. 321-371.
- Jain, R., Kasturi, R., and Schunck, B. G. (1995). *Machine Vision*, McGraw-Hill Inc.
- Kanade, T. "Model Representations and Control Structures in Image Understanding." *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, pp. 1074-1082.

- Levine, M. D. (1985). *Vision in Man and Machine*, McGraw-Hill Book Company.
- Matsuyama, T. (1987). "Knowledge-Based Aerial Image Understanding Systems and Expert Systems for Image Processing." *IEEE Transactions on Geoscience and Remote Sensing*, GE-25, No. 3, May 1987, 305-316.
- Matsuyama, T. (1989). "Expert Systems for Image Processing: Knowledge-Based Composition of Image Analysis Processes." *Computer Vision, Graphics, and Image Processing*, 48, 1989, 22-49.
- Nagao, N., and Matsuyama, T. (1980). *Structural Analysis of Complex Aerial Photographs*, Plenum Press.
- Nazif, A. M., and Levine, M. (1984). "Low Level Image Segmentation: An Expert System." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI 6, No.5, Sept. 1984, pp. 555-577.
- Ng, W. S., and Lee, C. K. (1996). "Comment on Using the Uniformity Measure for Performance Measure in Image Segmentation." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, VOL. 18, NO. 9, September 1996, pp. 933-934.
- Nicolin, B., and Gabler, R. (1987). "A Knowledge-Based System for the Analysis of Aerial Images." *IEEE Transactions on Geosciences and Remote Sensing*, Vol. GE-25, No. 3, May 1987, pp. 317-329.
- Ohlander, R. K., K., K. P., and D. R. Reddy. (1978). "Picture Segmentation Using a Recursive Region Splitting Method." *Computer Graphics and Image Processing*, Vol. 8, 1978, pp. 313-333.
- Pavlidis, T. (1977). *Structural Pattern Recognition*, Springer-Verlag.
- Pavlidis, T. (1982). *Algorithms for Graphics and Image Processing*, Computer Science Press, Inc.
- Pavlidis, T. (1982). *Algorithms for Graphics and Image Processing*, Computer Science Press, Inc.
- Peli, T., and Malah, D. (1982). "A Study of Edge Detection Algorithms." *Computer Graphics and Image Processing*, Vol. 20, 1982, pp. 1-22.
- Perkins, W. A., Laffey, T. J., and Nguyen, T. A. (1985). "Rule-Based Interpretation of Aerial Photographs Using LES." *SPIE Vol. 548, Applications of Artificial Intelligence II*, 1985, pp. 138-146.
- Rich, E. (1981). *Artificial Intelligence*, McGraw-Hill.
- T. Binford, T. (1982). "Survey of Model-Based Image Analysis Systems." *International Journal of Robotics Research*, Vol. 1, No. 1, 1982, pp. 18-64.
- Trivedi, M. M., and Charles A. Harlow. (1985). "Identification of Unique Objects in High Resolution Aerial Scenes." *Optical Engineering*, Vol. 24, No. 3, May/June 1985, pp.502-506.
- Wang, F. (1991). "Relational-Linear Quadtree Approach for Two-Dimensional Spatial Representation and Manipulation." *IEEE Transactions on Knowledge and Data Engineering*, Vol. 3, No. 1, March 1991, pp. 118-122.
- Wechsler, H. (1990). *Computational Vision*, Academic Press.
- Weeks, A. R., Jr. (1996). *Fundamentals of Electronic Image Processing*, SPIE Optical Engineering Press.
- Winston, P. H. (1992). *Artificial Intelligence*, Addison-Wesley.
- Zhu, S. C., and Yuille, A. (1996). "Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation." *IEEE Transactions on*

Pattern Analysis and Machine Intelligence, VOL. 18, NO. 9, September 1996, pp. 884-900.

Zucker, S. W. (1976). " Survey of Region Growing: Childhood and Adolescence." *Computer Graphics and Image Processing*, Vol.5, 1976, pp. 382-399.