2-1-1987

Edge Detection for Object Recognition in Aerial Photographs

Helen Anderson
University of Pennsylvania

Recommended Citation

Helen Anderson, "Edge Detection for Object Recognition in Aerial Photographs", . February 1987.

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-87-14.

Edge Detection for Object Recognition in Aerial Photographs

Abstract

An important objective in computer vision research is the automatic understanding of aerial photographs of urban and suburban locations. Several systems have been developed to begin to recognize man-made objects in these scenes. A brief review of these systems is presented.

This paper introduces the Pennsylvania Landscan recognition system. It is performing recognition of a scale model of the University of Pennsylvania campus. The LandScan recognition system uses features such as shape and height to identify objects such as sidewalks and buildings.

Also, this work includes extensive study of edge detection for object recognition Two statistics, edge pixel density and average edge extent, are developed to differentiate between object border edges, texture edges and noise edges. The Quantizer Votes edge detection algorithm is developed to find high intensity, high frequency edges.

Future research directions concerning recognition system development, and edge qualities and statistics are motivated by the results of this research.

Disciplines

Computer Sciences

Comments

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-87-14.

EDGE DETECTION FOR OBJECT RECOGNITION IN AERIAL PHOTOGRAPHS

Helen Lillias Anderson

MS-CIS-87-14
GRASP LAB 96

Department Of Computer and Information Science
School of Engineering and Applied Science
University of Pennsylvania
Philadelphia, PA 19104-6389

February 1987

Acknowledgements: This work was supported in part by DARPA/ONR grant NOOO14-85-K-0807, NSF grants DCR-8410771, NSF-CER grant MCS-8219196 Air Force grant AFOSR F49620-85-0018, Army/DAAG-29-84-K-0061, DAA29-84-9-0027, NSF-CER/DCR82-19196 Ao2, NIH grant NS-10939 -11 as part of Cerebro Vascular Research Center, NIH 1-RO1-NS-23636-01, NSF INT85-14199, NSF DMC85-17315, ARPA NOO14-85-K-0807, by DEC Corp., IBM Corp. and LORD Corp.

UNIVERSITY OF PENNSYLVANIA

SCHOOL OF ENGINEERING AND APPLIED SCIENCE

MOORE SCHOOL OF ELECTRICAL ENGINEERING

EDGE DETECTION FOR OBJECT RECOGNITION IN AERIAL PHOTOGRAPHS

Helen Lillias Anderson

Philadelphia, Pennsylvania
December 1986

A thesis presented to the Faculty of Engineering and Applied Science of the University of Pennsylvania in partial fulfillment of the requirements for the degree of Master of Science in Engineering for graduate work in the Department of Computer and Information Science

(original signed by)
Dr. Max Mintz

(original signed by)
Dr. Peter Buneman

Abstract

An important objective in computer vision research is the automatic understanding of aerial photographs of urban and suburban locations. Several systems have been developed to begin to recognize man-made objects in these scenes. A brief review of these systems is presented.

This paper introduces the Pennsylvania LandScan recognition system. It is performing recognition of a scale model of the University of Pennsylvania campus. The LandScan recognition system uses features such as shape and height to identify objects such as sidewalks and buildings.

Also, this work includes extensive study of edge detection for object recognition. Two statistics, edge pixel density and average edge extent, are developed to differentiate between object border edges, texture edges and noise edges. The Quantizer Votes edge detection algorithm is developed to find high intensity, high frequency edges.

Future research directions concerning recognition system development, and edge qualities and statistics are motivated by the results of this research.

Acknowledgement: This work was in part supported by: DARPA/ONR grant N0014-85-K-0807, NSF grant DCR-8410771, Air Force grant AFOSR F49620-85-K-0018, Army/DAAG-29-84-K-0061, NSF-CER/DCR82-19196 Ao2, NIH grant NS-10939 -11 as part of Cerebro Vascular Research Center, NIH 1-RO1-NS-23636-01,NSF INT85-14199,NSF DMC85-17315, ARPA N0014-85-K-0807, by DEC Corp., IBM Corp. and LORD Corp.

Contents

1 Introduction

An important objective in computer vision research is the automatic understanding of aerial photographs. The Pennsylvania LandScan (Language Driven Scene Analyzer) system identifies buildings, roads and other man-made objects in urban scenes. LandScan consists of modules which perform edge detection, segmentation, stereo disparity calculation, feature identification, recognition, and query generation. In this work, the initial implementation of the LandScan system is presented, except for the query generation module.

The initial implementation of the LandScan system uses pictures of a model of the University of Pennsylvania campus as shown in Figure 1. In addition, the system is being expanded to perform recognition on a large class of aerial photographs. It is difficult to extend a recognition system which performs recognition on a known picture to perform recognition on a completely different picture such as the one in Figure 2. No single threshold used within a module will be at the right level for every possible picture. For example, in the edge detection module, a certain minimum edge strength is employed. Below this strength threshold, edges are not counted. This eliminates edges which are part of the texture of the campus model's materials. However, the edge strength threshold which applies to the campus model may not be applicable to the edges in a picture of a suburban town on a winter afternoon. Before recognition can be done on a large class of photos, edge strength thresholds, size thresholds and other specific decisions about picture characteristics must be identified and understood. After understanding how these decisions effect recognition, the thresholds can be modified as necessary to perform

Figure 1: University of Pennsylvania model

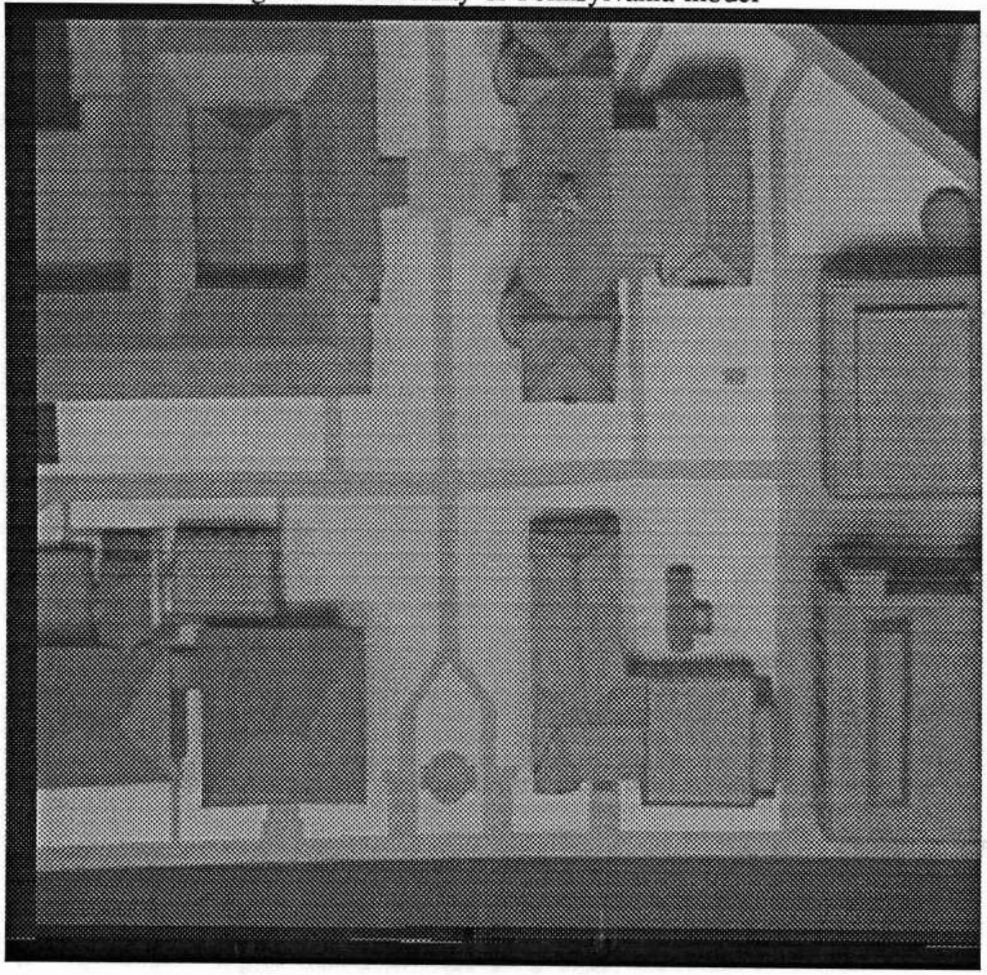
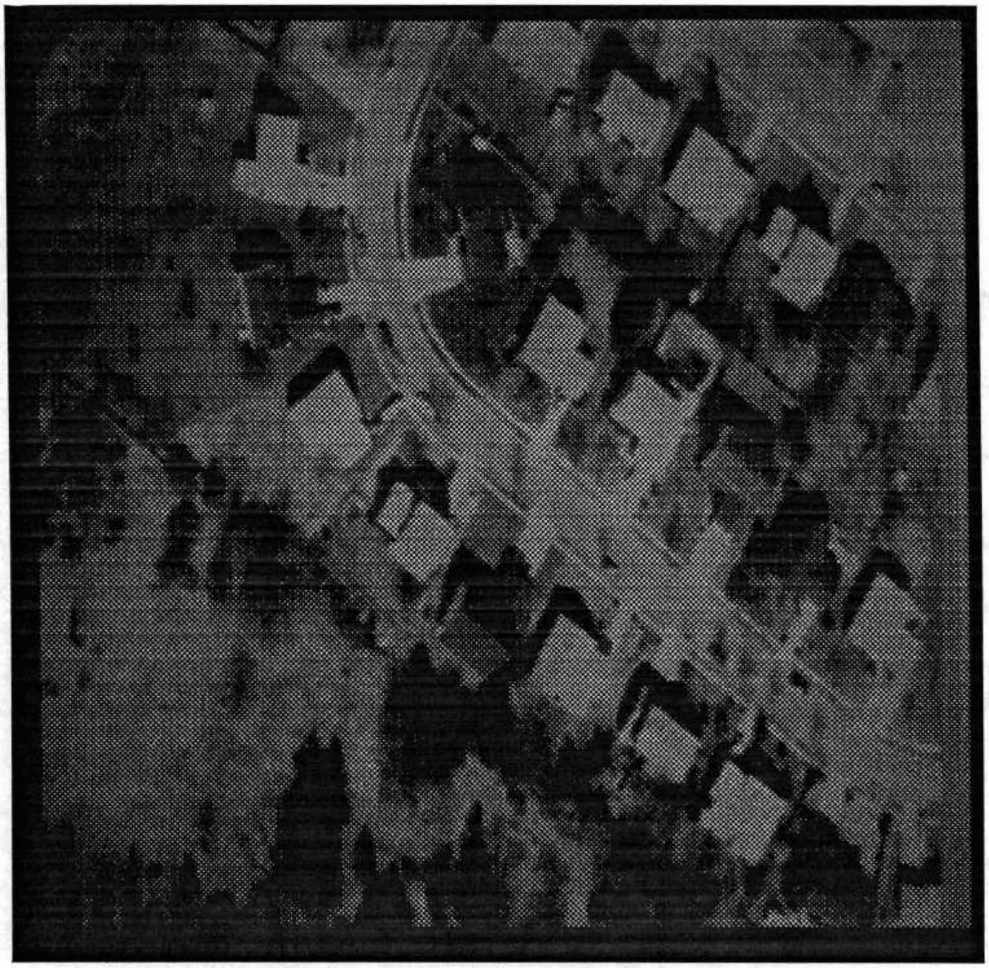


Figure 2: Aerial Photo of Mt. Laurel, NJ



recognition automatically on many other photographs.

This work includes extensive study of the edge detection module. Using different methods of edge detection and line finding, a large number of edge pixels is generated. The proper edges are usually found, along with many edges that a person would not use to represent the image. It is possible to choose one threshold after another until a good set of edges is obtained, and then to proceed with "automatic" scene analysis. In this work, we show approximately how many edge pixels are needed and why. Elimination of the extra edge pixels is studied. Statistics are presented which indicate whether the extra edge pixels were removed.

Even when the edge detection module is truly automatic, it will not work on all aerial photos. For example, in a picture which includes an entire city, buildings cannot all be recognized separately. No object which is represented by a few pixels can be recognized. The resolution of the picture limits the abilities of the recognition system. This problem is discussed, and the analysis bounds the practical expectations for recognition systems.

Finally, an edge detection algorithm is presented which finds appropriate edges for recognizing man-made objects in aerial photographs. This algorithm will find sharp edges, which are characteristic of buildings, road edges and other man-made and man-altered objects. It eliminates fuzzy edges, which are characteristic of trees and many natural objects.

2 Automatic Recognition System Design

2.1 Design Philosophy

The most important decisions in recognition system design are the early ones. First, the goals of the system must be set. Second, the features which lead to recognition must be considered. Third, the structure of the overall system must be designed.

Before setting the goals, it is worthwhile to consider the philosophy of the recognition system. Is this an attempt to perform a subset of human vision? Or, should the system be based strictly on the kind of tasks computers do well now?

Human recognition of complex objects is not a well understood process. Some psychologists suggest that things are recognized as whole objects, for example, as faces rather than as collections of edges or collections of colored regions [MW 39]. Indeed, people can imagine that they see faces in many places where no real faces exist. On the other hand, certain elemental features of a scene are sent from the eye to the brain. Familiarity with the exact shape of object borders does help people recognize complex objects in complex images. In [BLK 80] it was demonstrated that highly detailed representations are unnecessary for target identification training, but the use of line drawings made a significant improvement in target identification performance. People convert the brightness levels and the edges which they see into faces, buildings, and other objects. This conversion of discrete and local image attributes into continuous and global objects is a major problem in the field of perceptual psychobiology [WRU 83]. The nature of this conversion, and of visual knowledge representation in the brain, is not known.

To date, there is no machine system which performs general vision. Berthold Horn suggests that we can "address ourselves either to systems that perform a particular task in a controlled environment or to modules that could eventually become part of a general-purpose system" [BKPH 86]. Systems are in use today which perform particular tasks in controlled environments .

Vision systems for parts inspection applications perform particular tasks in a controlled situation. There are two aspects of the control. First, this kind of system is provided with a complete set of objects in advance. Like a template, the object shapes must match within a close tolerance or be rejected. Second, the objects are automatically and consistently arranged to be compared to the template. The objects may not always be at the same orientation angle, but they are at approximately the same distance. The objects can even be marked with features specifically to make comparison easy. The lighting is controlled, too. Segmentation into object versus background is done the same way each time [EWK 86]. This is not a recognition problem, but a template matching problem.

The problem of recognition in aerial photographs is not as controlled as the parts inspection problem. First, the objects vary continuously within predictable ranges. For example, road width may vary from 1 lane to 16 lanes. Actual lane width varies some, too, and road shoulders add even more variability. Also, road curvature varies. However, curvature is limited to the performance abilities of cars and trucks. Exact limitations of curvature can be found in highway design manuals. Thus the characteristics are available, but every instance of every road cannot be provided for comparison. Second, in an aerial photo, objects are not arranged for the convenience or performance repeatability of the recognition system. Contrast varies, brightness varies, texture varies and object distance from the camera varies.

The aerial photo recognition problem has some other limits, too. The types of objects and their viewing directions are limited. Buildings, roads, statues, people and cars are expected, whereas flying hamburgers can reasonably be ruled out. Viewing angle does vary, but it will be approximately perpendicular. The aerial photo recognition problem is more difficult than a carefully controlled task but less difficult than a general-purpose task. It is a *limited* task.

The techniques used here do not necessarily apply to a general-purpose system, though they may turn out to be very useful. Also, the techniques need not be copies of human visual techniques, though they may turn out to be. However, we can look to human vision as an existing implementation of a vision system which may be copied whenever it is convenient to do so. If detailed study of edges helps the human recognition system work, we should consider using edge details if they will help our system work.

Beyond the questions about how human vision works, and whether to copy it, we must make many decisions about the smaller goals. We must define exactly what we want to do, and get a general idea how we can do it. The basic questions and choices which define the goals for vision system design are presented in Appendix A.

2.2 Related Recognition System Work

See Table 1 for a description of several kinds of vision systems. It includes the system characteristics of human and machine recognition systems. The human recognition processes which have been studied are used to provide guidance for what a machine may accomplish.

In the literature, there are several machine systems specifically designed to work with objects in aerial photographs. Some systems are designed to extract three-dimensional shape information about the objects, but do not address the problem of recognition of the objects themselves. On the other hand, some systems begin with sketches or maps, where the objects are presented in a consistent manner, then the objects are recognized using a feature comparison.

One system, the 3-D MOSAIC system [MH 84], extracts connected edges of build-

Table 1: Vision System Characteristics

	General Vision	Identify Tanks	Penna. LandScan Design	Complex Machine Vision	Simple Machine Vision	Random Dot Stereo
Computer	human	human	machine	machine	machine	human
Object Set Size	$\sim 10^6$	< 10	< 20	$< 10^3$	< 10	1
Example Objects	all things	AMX30 tank	building, road	military targets	machine parts	dotted cube
Object Set Variance	high	all known low	in ranges medium	in ranges low	identical none	different patterns-low
Instances Per View	< 200*	< 10	< 200	1 – 200*	1 – 10	1
Understand context	yes	yes	yes	sometimes	no	no
Response time	real time	real time	minutes per image	variable	real time	real time
Image size (pixels/ frame)	10^7	10^7	10^6	$10^4 - 10^{10}$	10^4	10^2
Context Set Size	large	small	small	small	1	1
Reference	[BRB 77]	[BLK 80]		[BRB 77,WBS 84]	[BRB 77]	[BLK 80]

* unsubstantiated guess

ings, or junctions, from multiple aerial views of urban scenes. This system matches junctions from two pictures and obtain heights of edges. Horizontal faces are inferred from L-shaped junctions. Vertical faces are dropped from the edges. There is not enough information to find all the junctions from a single pair of pictures. Herman *et al.* are experimenting with multiple views to find the missing L junctions which will show the complete buildings.

This system does not require junctions to be at right angles. After the missing junctions are found or inferred, this system will make a good reproduction of the above-ground portion of the scene. The emphasis of this work to date is on making a 3-dimensional model and display of the buildings rather than image interpretation.

A second system, the SPAM image interpretation system [DMM 84], uses maps to guide the recognition of objects in airports. The knowledge from maps allows the system to used rules with cartographic coordinates, such as elevation and real distances, rather than "the runway has area 12000 pixels." A major problem with this work is the difficulty of region-based segmentation. The objects which appear to be broken into arbitrary fragments are typical of state-of-the-art region-based segmentation on real aerial images. It is extremely difficult to identify these fragmented objects.

A third system, by A. Huertas and R. Nevatia [AH 83] looks for buildings with sharp corners. Lines and corners are extracted from real aerial images to trace boundaries of possible buildings. Next, adjacent shadows are extracted along the direction of illumination. If a candidate building's corners can be paired with its shadow's corners, then the system decides that the object is indeed a building. The shadows are used to confirm that the building is higher than its surroundings. The criteria for identifying buildings are sharp corners and adjacent matching shadows. These criteria find L-shaped buildings, and do not find parked airplanes to be buildings.

This strategy works well for buildings. It is limited by illumination constraints, but stereo could improve it dramatically. Nevatia states that a stereo module is being developed to enhance the system [RN 86]. The intention of this system is to find buildings, not to do exhaustive identification.

A fourth system by D. Rosenthal and R. Bajcsy recognizes 12 objects (car, bus, street, building, median, etc.) using a production system [DAR 84]. Queries drive the system, which searches for queried objects only in highly probable areas. For example, a query to find a median strip would generate a query to find a street then look on the street for the median strip. Region growing was used in this system. The shapes of the buildings and other objects recognized are basic, and the system uses adjacency and containment relationships very effectively. A powerful texture descriptor, homogeneity, is included in the system.

To be recognized as a road's median strip, an object must be on a street, lighter in gray value than the street, within a range of widths, shaped like a rectangle, homogeneous, and almost as long as the street. This system only recognizes rectangular buildings, too. This is an example of a system which works very well on one picture, but would need significant modifications for more general application.

A fifth system, called MAPSEE, was implemented on sketch maps by Mulder and Mackworth [JAM 85]. It recognizes roads, coastlines, bridges and other objects. Sketch maps were used to completely avoid the problem of segmentation and texture. They exhaustively identified all of the elements in their sketches, but they also created all of those same elements. Therefore, the features probably would not be useful for real scenes, but the recognition algorithm will be.

The MAPSEE system uses the method of least commitment for the identification. The principle underlying the method is to stick to the most abstract interpretation until evidence forces more specific interpretation. This was introduced by Marr and Nishihara [DM 82]. It begins with the assumption that the object in question is an object (or part of an object) in the set of all possible objects. There are two subsets which divide the set of all possible objects. One test will determine whether the object is one subset or the other. This reduces the number of possible interpretations of the object, or reduces the ambiguity about it. Each succeeding test further reduces ambiguity about the object until it is identified as a member of a set containing one type of object. It is desirable for the early tests to divide the sets into approximately equal subsets and to be inexpensive to perform.

This approach to the inference problem can be described with a tree structure. Nodes are sets of objects. The root is the set of all possible objects. Each leaf should be a single object. A region starts in the root node interpretation, and moves down into smaller and smaller subsets of possible interpretations. The edges are constraints which allow objects to be disambiguated. An alternative view of the tree is that a node is simply any region satisfying constraints on the unique path between that node and the root. At the leaf level, enough tests have been made to identify the region as one particular object.

The tree arrangement is based on a priori knowledge of the scene. Correctness and efficiency of the tree is crucial to the success of the process. Objects may be located at more that one leaf; however, each leaf has a unique testing process which leads to it. The edges leading to a leaf will become the plan to find the object located at that leaf.

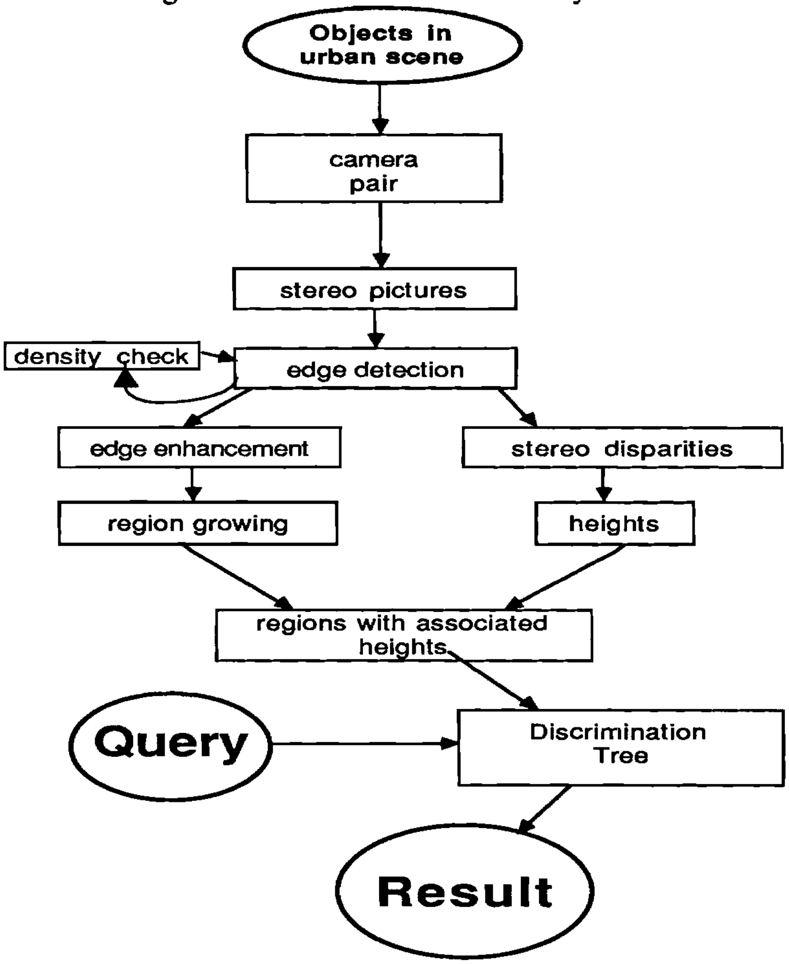
3 Pennsylvania LandScan

When people look at an aerial image of a complex urban scene, they can recognize buildings, roads and other objects with little difficulty. It is clear that there is sufficient information in an aerial image to recognize objects, but for an automatic system, it is very difficult to extract the useful information and then make the identifications. We begin with a long list of gray scale values and wonder how to proceed to do what we know is possible.

The goal of the LandScan system is to identify objects in a complex urban scene from aerial images. We want the capability to identify all the objects in each scene, so that we can consider the context of the objects. While identification of every object in the original image is interesting, it is not efficient if only one particular type of object is sought for further analysis. In the LandScan system, queries are used to guide the search to the appropriate tests. The tests which are used to identify a building may not be the tests which are used to identify a sidewalk, for example. However, some tests, such as edge detection, must be done on the entire image.

The LandScan system is intended to be modular, so that separate pieces can be improved while the system still functions. Figure 3 shows the overall system scheme.

Figure 3: LandScan Identification System



In addition, more interaction between modules is planned. For example, in the initial implementation, edge detection cannot be redone on the basis of the query. However, this and other interactions are planned as future enhancements to the system.

LandScan uses the Pennsylvania Active Camera System [FF 86]. The camera system includes two cameras, and it has capabilities for movement similar to the human head. The camera platform controller lets the cameras converge, tilt, pan and move horizontally and vertically. The lens controller provides control of the focus, zoom and aperture opening. Each camera chip is a black and white CCD array (Fairchild CCD222) with a resolution of 488 lines per frame by 380 elements per line. A real-time digitizer and frame buffer (Ikonas RDS3000) acquires 512 by 512 8-bit image data from each channel. In addition, a single Sony XC-39 CCD camera can be used with the digitizer and frame buffer, but with manual platform and lens control.

LandScan is now working with images of a scale model of the University of Pennsylvania campus made of white, gray and black cardboard. This limits the variations of scale and texture present in the image. Pictures of the model are easily obtainable under different conditions of lighting and noise. Also, statistics about it can be checked easily. We can recognize most objects in the scale model now. However, we want to apply our recognition system to other images. We are trying to increase our flexibility so that we can do this. Future plans include recognizing objects in many different images, regardless of scale, noise level and other variables, and recognizing more types of objects.

Gray value is the most obvious feature of objects in aerial photographs. However, the gray values of buildings and roads depend on the construction materials, the lighting and many other variables. Shape is much more consistent. Roads are long and narrow (relative to their length) whether they are concrete or asphalt, superhighway or alley, well-lit or barely visible. Buildings have more variety in their shapes, but they are rarely as long as roads. Shape is an important factor in human recognition of distant objects [BLK 80].

In the LandScan system, shape is the most important feature used for recognition. Shape is obtained from grouped pixels, or regions. The regions are formed from edge pixels and local similarity of pixel gray values. The edge detection/region growing paradigm is discussed in [RB 86]. In this region-growing algorithm, edges are used as barriers through which regions cannot cross.

Stereo information is very useful in image understanding for aerial photos. It is more reliable than shadow information for differentiating raised structures from ground structures. Stereo allows adjacent objects to be joined for analysis as a group. In this system, stereo information is obtained from a stereo matcher developed by D. Smitley [DLS 85]. To date, the LandScan system attaches the stereo information to regions, so that one height is associated with each region. A planned development of the system is interaction between the stereo information and the region growing process.

The basic internal shape representation is the surface patch, which is a connected group of pixels (region) with associated height. This is a form of $2\frac{1}{2}$-D sketch, which is useful because it makes explicit information about the image in a form which is closely matched to what the simplest image processing algorithms can deliver [DM 82]. Urban scenes viewed from above with stereo are described naturally with this structure. Few objects have sides which are not vertical, and few have bottoms above ground. This shape representation does not completely describe bridges or elevated highways, but it retains all the information available from the stereo aerial image.

These surface patches are found in two steps. First, regions are found on one gray scale picture from a stereo pair by a combination of region growing and edge detection [RB 86]. Then, point-based stereo matching is done on the stereo pair [DLS 85]. The stereo disparities are associated with the appropriate regions, and the median heights of the regions are calculated.

The system uses a discrimination tree similar to Mulder and Mackworth's, but operating on surface patches instead of sketch elements. The discrimination tree is particularly well suited to a query directed system because each object at a leaf has a unique path from the root. Strategy to find a particular object is automatically generated by following the path of tests from root to leaf. Also, it is efficient to mark partial identifications for the regions which do not satisfy constraints on the path leading to the desired object in case other identifications are requested later in the session. The discrimination tree which has been implemented is shown in Figure 4.

Requests to the system include the type of object to be identified and the picture to be examined. The object sought would have to match a node of the identification tree. Output from the system is a picture of the requested objects.

The intermediate results from the image in Figure 1 and its stereo paired image are shown in Figure 5. Recognized buildings are shown in Figure 6 and sidewalks in Figure 7.

Figure 4: LandScan Discrimination Tree

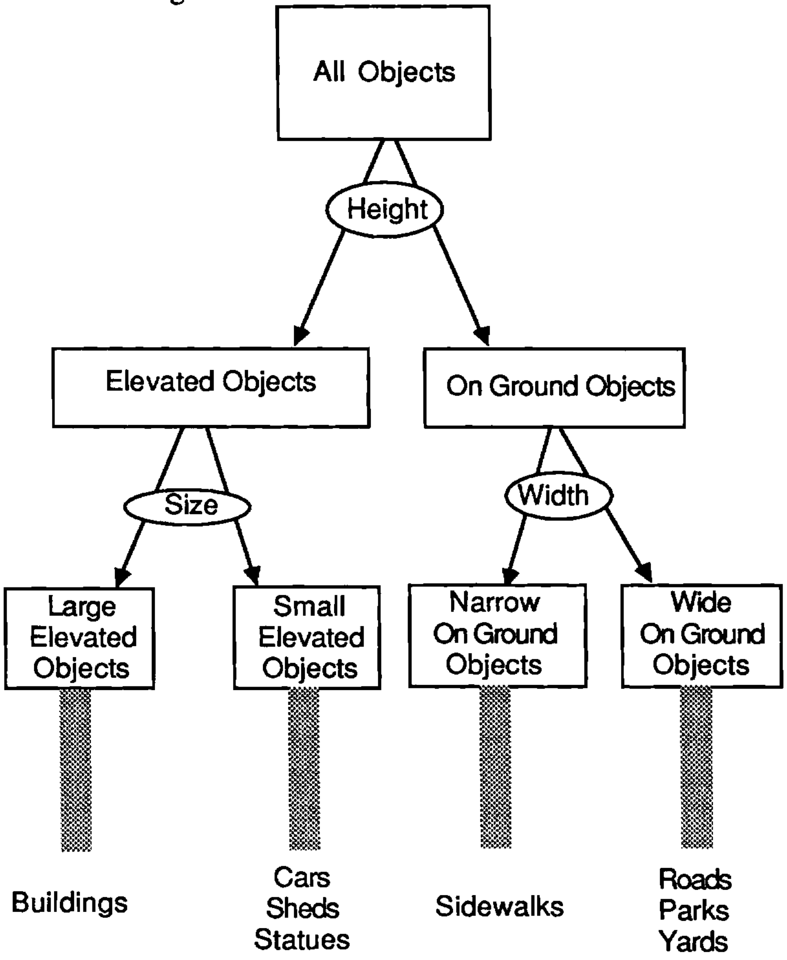


Figure 5: Regions Found to be Above Ground with Stereo Module

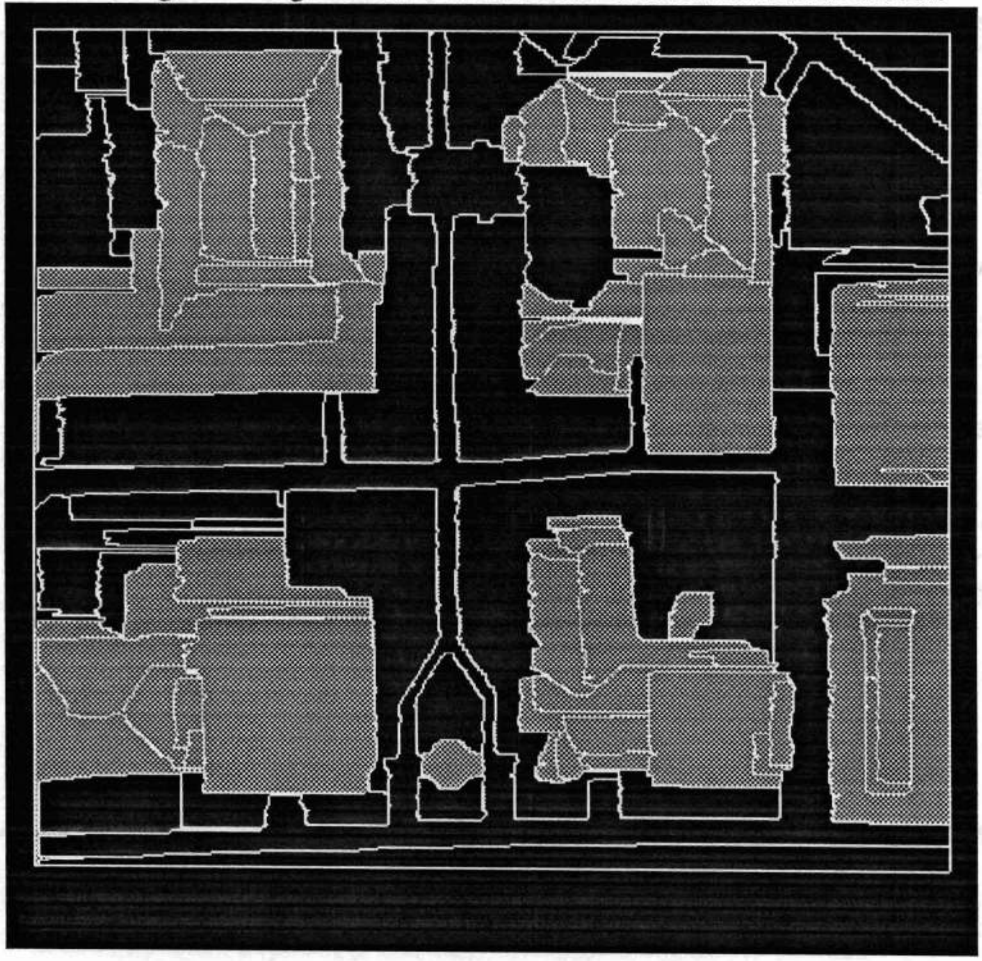


Figure 6: Buildings

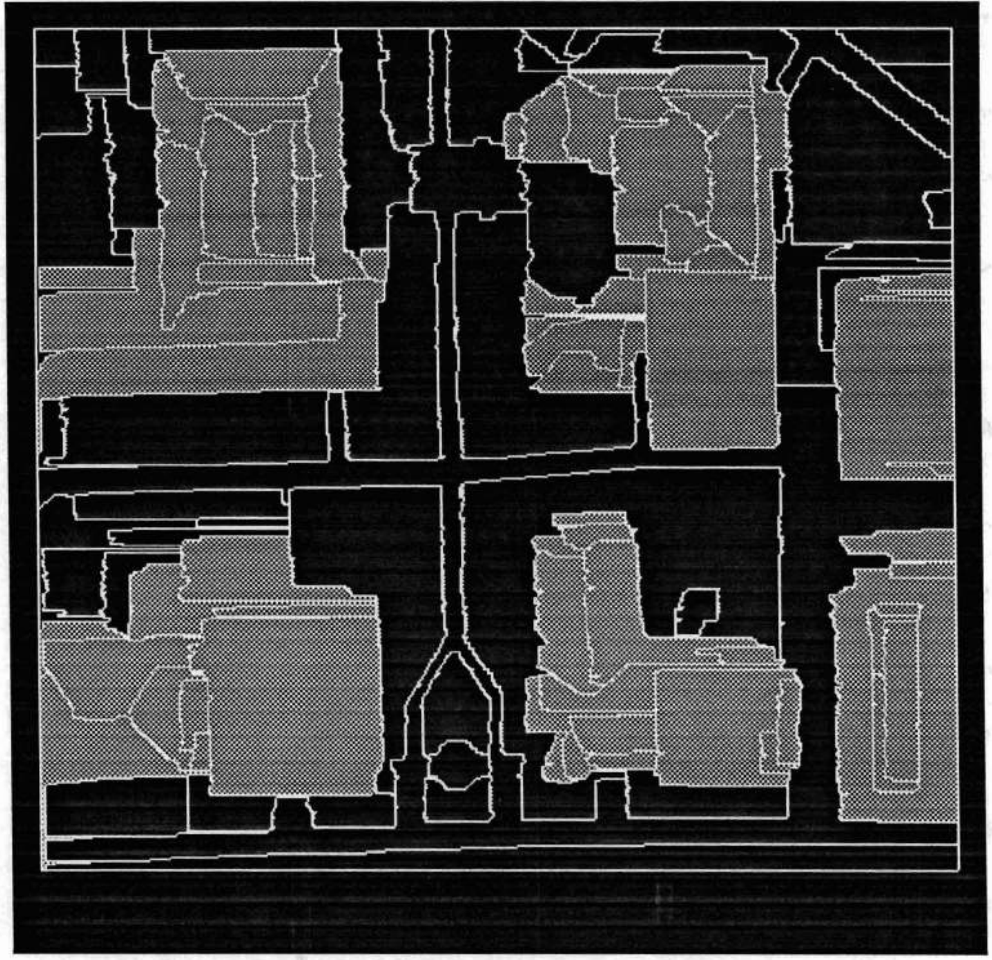
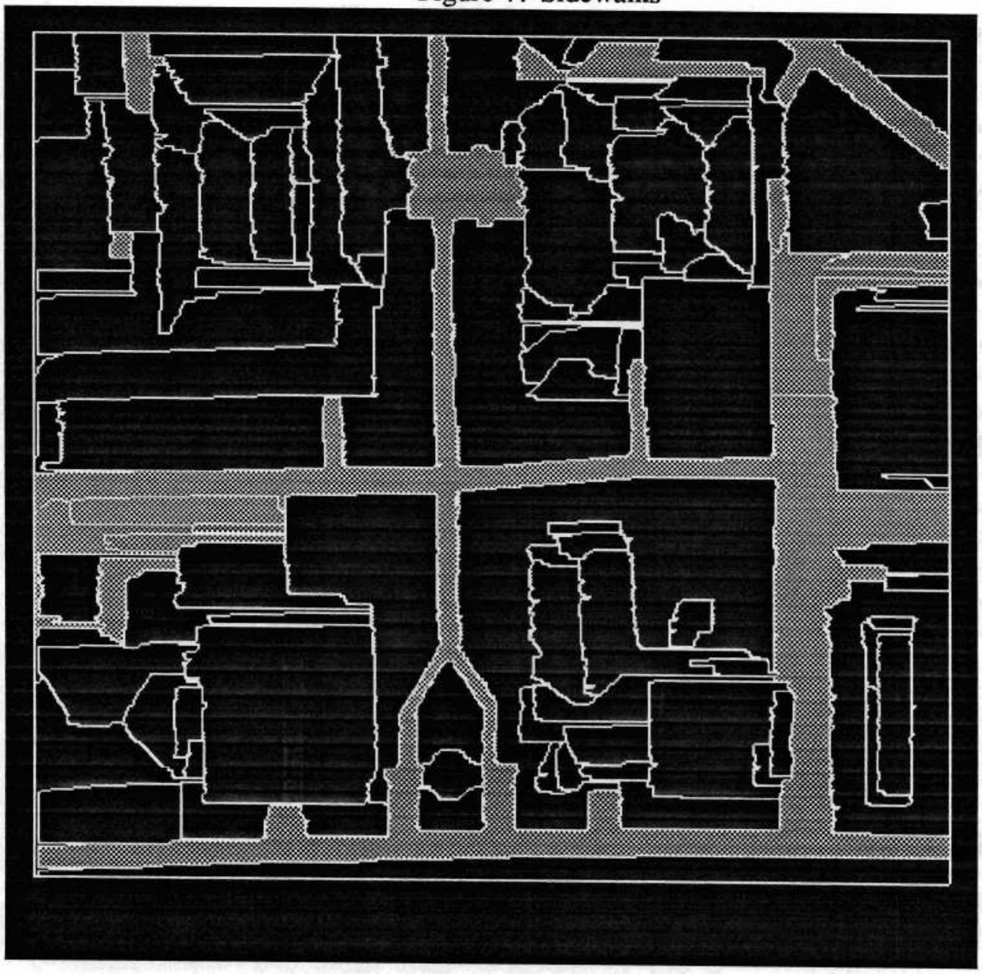


Figure 7: Sidewalks



In an automatic image understanding system, it is important that only the useful information from an image be extracted for analysis. Simple tests should be sufficient to identify objects if the right information has been extracted. To find the useful information, a geometric model of the scene is required. A geometric model is a concept of the 3-dimensional structure of the scene. For example, one possible model of a scene could be a summation of cubes in different locations in space. This is a voxel, or volume element, model. The voxel geometric model is a 3-dimensional model of a 3-dimensional scene. In an urban scene viewed from above, the tops of objects are visible, and vertical sides ordinarily connect raised objects to the ground or to adjacent raised objects. Therefore, the scene can be modeled using regions with heights. If there is a disparity between the height on one region and an adjoining region, a vertical connection is assumed. This is the surface patch geometric model, which is a $2\frac{1}{2}$-dimensional model. It has substantial 3-D information, but not complete 3-D information. The LandScan system uses the characteristic vertical sides in an urban scene to reduce greatly the amount of information to store and analyze.

It is possible to make a geometric model of an urban scene which is useful for displays but difficult to test for object identification. Conversely, it is possible to make clever tests for identification using information which is extremely difficult to obtain from aerial images. Since the LandScan project has object identification in aerial photos as its primary goal, the identification tests are being developed together with the geometric model. Useful descriptive information about features from the original image are generated along with the model.

A problem in scene understanding occurs when only the characteristics of the image under study are used to make the identification tests. The image under study is understood well, but the system fails on almost any other image. A robust system must start by understanding one type of scene, but the system development must not stop after its first success. The design must allow for enhancement, adding more objects to recognize, improving the modules, and allowing more knowledge-based interaction between the modules. The robust system must accomodate the complete range of object presentations, yet differentiate other objects which may be quite similar. This is much more difficult.

A robust system for scene understanding must have identification tests which are applicable to a wide variety of scenes. High-level tests should recognize curved roads as well as straight roads. Buildings with a variety of shapes should be recognized. However, a system cannot recognize all possible objects. Underground buildings will not be visible and will not be identified, but elevated highways are visible and should be identified.

The low-level processing such as edge detection must work on images of different quality. The system should be adjustable to account for different image scales and noise levels. Of course, there will be a limit of minimum image quality for which the system works, but that limit should be found along with an explanation for the limit.

The LandScan system is successfully working on one type of image, but the difficult task of making the system work on many types of images lies ahead. The system will be tested on new views of the Penn campus scale model and on real aerial images. Failures in identification will be studied, and enhancements to the system will be made. The goal of the error analysis and the study of enhancements is to produce a system which will work on many different images. The enhancements are directed toward automatic adjustments to the modules. However, before this can be done, exact error analysis will be required.

In LandScan, the image analysis modules are designed on the basis of the following imperatives:

emphasize shape rather than absolute gray value. Shape is a more consistent feature of objects than gray value, since the gray value depends on the lighting of the scene.

match the abilities of the scene analysis processor. The characteristics of the features extracted from the images must be useful for the reasoning system.

correspond to interesting aspects of aerial urban and suburban scenes. Edge detection should select sharp, strong edges while removing weak, gradual edges.

preserve excellent edge location while removing low amplitude edges. Edges should be sharp and accurate.

Knowledge about the answers desired at higher levels should be included at every level of image analysis. For example, if the recognition system cannot differentiate between very small objects, the region grower should not produce many small objects. Instead, the region grower should merge similar regions to present good candidates for identification. However, the knowledge included at the low level should be general enough that the scene understanding system works on a large class of pictures.

4 Edges for Recognition

Digitizing a picture produces a value for the intensity at each element (pixel) of a 512 by 512 array. This can be described as a 2-dimensional intensity function. If there is a significant change in local intensity, the pixel at the location of the greatest change is called an edge pixel. Edges pixels are found using a convolution of two or more masks with a two-dimensional intensity function. The convolution produces some form of the directional derivative. The value of the convolution of the intensity function with the mask at a given pixel is the gradient of the intensity function near that pixel. A high gradient means that there is a large intensity change at that pixel, so a high gradient is a strong edge. The value of the gradient will be referred to as the strength of the edge.

In the LandScan system, two methods are used to find edge pixels, Canny's method and a new method, quantizer votes. Canny's method [JC 83,DT 85] consists of convolution in 2 directions, using masks which are the first derivatives of a Gaussian normal function (of width σ), then directional non-maximum suppression.

The quantizer votes method uses a combination of edges from Canny's method obtained from several quantized versions of the original intensity function. This method is discussed further in Section 5.

Edge scale is related to the width of the Gaussian function. Since gradient values are rates of change in intensity with distance, it is important to consider the distance over which the changes occur. The Gaussian filter has the desirable effect of reducing noise by averaging several adjacent pixels. At the same time, the picture resolution is reduced, much like defocusing. As a result, convolution with a narrow Gaussian filter (small σ), yields small and large steps of intensity along with large gradual changes of intensity. Using a broad Gaussian filter (large σ), only large changes of intensity are found, whether they are sharp or gradual. A narrow filter gives good localization of edges; however, it also responds to noise. A broad filter is more robust in the presence of noise, but does not give good localization. Further discussion of edge scales and mask types is included in Section 4.4.

4.1 Border Edges

An important class of edges in an aerial photo is the borders of the objects in the photo. These edges are usually long, fairly straight, and sparse. Typical border edges are curbs, building corners and river banks. Edges which are not borders may be short, wavy, and very close together. Typical non-border edges are brightness changes due to differing vegetation, or waves on a lake. The non-border edges are useful for texture identification. However, for edge-based segmentation, it is desirable to find border edges and eliminate non-border edges.

Borders are useful because they provide region boundaries which separate objects. Also, details of border shapes are important. For example, consider a concrete sidewalk adjoining a concrete street. If the border between the sidewalk and the street is missing, the whole sidewalk plus street will look like a street. However, with the borders present, they can be treated separately. Both are long, but the sidewalk is probably narrower. Width may be the best feature to separate the class of sidewalks from the class of streets. On the other hand, if the sidewalk is divided into separate concrete sections, the objects will look like small squares. The roof of an adjacent building may also have square shingles. The shapes of these objects may be the same. In addition, the recognition system may be overwhelmed with thousands of objects to recognize. The key to recognition by shape is finding the borders of the particular objects that the system is designed to recognize.

To find the best edges it is possible to look for the best edge detection procedure. On the other hand, it is also worthwhile to study the edges themselves. What is it that makes an edge picture good for segmentation and for finding the right border shapes? Other than in machine perception, where are edge pictures used?

4.2 Coloring Book Edges and Edge Statistics

There is another situation, completely different from automatic scene understanding, in which border edges are very important—children's coloring books. Some insight into border edges in aerial photos may be gained from understanding border edges in coloring books. The coloring problem for children is related to the segmentation problem in image understanding. How should areas of the picture be grouped together as regions?

In coloring books for pre-school children, edges are provided as borders of regions. Ordinarily, each region is intended to be all one color, and adjacent regions are different colors. These border edges are usually long, fairly straight, and sparse, just like the edges we seek in the aerial photo. The most interesting thing about the coloring book edges is that the edges are long and sparse with statistical parameters almost **independent of the contents of the pictures** [CB-1].

These descriptions of edges in coloring books and aerial photos lead to statistics which may be applied on all edge pictures. The observed sparseness leads to *edge pixel frequency*, which is the number of edge pixels per thousand pixels. Edge pixel frequency simply indicates how crowded an edge picture looks. The observed edge length leads to *average edge extent*. Average edge extent is like edge length, except that an edges can go in any direction and split. This quantity is estimated by finding all the sets of 8-connected edge pixels, and giving the average set size, in pixels. While there are other possible ways to measure length, this method has the advantage that it favors connected lines and is independent of line curvature. However, it has the disadvantages that it has units of area rather than length and that it favors edges which are not thinned.

When these statistics are applied to coloring books in a qualitative study, it is clear that there are differences in average edge extent and edge pixel frequency among coloring books [CB-1,CB-2,CB-3]. The differences appear to be a function of the age of the child for whom the book is intended. Average edge extent decreases and frequency increases with increasing age of the child. There are many possible explanations for this, but it is clear that an older child has greater ability and/or willingness to make a finer segmentation of a coloring book picture while a 2 year old may happily color the entire picture red. This description leads to the concept of designing the edge detection module to match the design of the segmentation and recognition processing elements of the system.

4.3 Results of Edge Statistics

When edges are produced to form regions for a recognition system, most edges should be boundaries of recognizable objects. The regions should be large enough to have recognizable shapes, and the recognition system should not be asked to work with objects beyond the optical resolution. Thus the regions should be fairly large. Since most edges are boundaries of regions, and the regions are large, the edges should be sparse. For example, if there are 200 regions in a 512 by 512 image, the average area of a region is 1311 pixels. If all of these regions were square, there would be 55 edge pixels per thousand pixels in the image. Of course, most pictures are not that simple, but there are practical limits on the perimeter-to-area ratio of typical regions. Since the scale of an image can be calculated from camera geometry, the approximate size of the objects to be recognized can be calculated. Based on knowledge about the objects, approximate perimeter to area ratios of expected objects in urban scenes can be estimated. These ratios are the basis for analysis of edge detection results using edge pixel frequency.

One system limitation listed in table 1 is *the maximum number of object instances per view*. In most automatic recognition systems, this limitation is not explicitly considered. It can provide useful guidance for boundary edge detection and the recognition system in general. A limit on the maximum and minimum number of objects can prevent the machine system from working on objects too small to recognize or objects which are partially occluded. Segmentation programs normally limit the minimum size of regions, which gives an upper limit to the number of regions possible, but this upper limit is often much larger than the number of objects which the system could be expected to recognize.

The perimeter-to-area ratio of a region is a very useful statistic about the region. In the range of distances between the camera and the objects in which the recognition system is expected to operate, the perimeter-to-area ratio is approximately constant [EWK 86]. If perimeter of a region is measured by counting the exterior pixels and area is measured by counting all pixels, then the perimeter-to-area ratio is inversely proportional to distance. In [WAP 86], "compactness," which is Area/(Perimeter ** 2), is used as an attribute of regions for recognition to avoid this distance variation. However, both perimeter-to-area ratio and compactness are useful ways to compare region shapes.

The number of regions and the average perimeter to area ratio of the regions determine $\mathcal{B}(0)$, the total number of border edge pixels. Let P be the average perimeter of the regions in pixels, A the average area of the regions in pixels, A_t the total number of pixels in the image, and r the total number of regions. When adjacent regions share border pixels,

$$\mathcal{B}(0) = \frac{rP}{2} = \frac{PA_t}{2A}$$

If a strength threshold is used to remove weak edges, then the number of edge pixels can be expressed as a function of strength threshold, $\mathcal{B}(s)$. $\mathcal{B}(s)$ is the total number of edge pixels whose strength exceeds s. As s increases, $\mathcal{B}(s)$ decreases.

Expectations about $\mathcal{B}(s)$ can be used by an automatic system such as Perkins [WAP 86], or by a cartographer operating an interactive system to set edge strength thresholds. Then edges can be used to construct regions. However, the edge threshold will influence the attributes of the resulting regions. For example, a forest in an aerial photo may produce edges at the forest boundary and fainter individual tree edges. If all the edges are used, the resulting regions will be individual trees. The size, texture, shape and compactness of the tree regions may not match the attributes of the expected forest. Knowledge about the scale of the picture could be used to know that the tree borders make too many edge pixels when forest-sized objects are sought. Then a threshold of edge strength could remove most tree edges leaving the forest boundaries in the edge picture.

Besides borders of desired regions, there are other sources of edges. In the tree vs. forest example, the trees and grasses are real, but they cannot be recognized by the system. Let the number of edge pixels they produce be $\mathcal{T}(s)$. In addition, the camera system itself will produce noise, and that noise will produce edge pixels, $\mathcal{N}(s)$.

If we know about the behavior of $\mathcal{B}(s), \mathcal{T}(s)$ and $\mathcal{N}(s)$, then we will know whether a threshold of edge strength is useful. The total number of edge pixels, $\mathcal{E}(s) = \mathcal{B}(s) + \mathcal{T}(s) + \mathcal{N}(s)$. Will a threshold of edge strength remove the texture and noise edges and preserve the borders? This is a decision which is normally made by the system operator and tailored to individual images. However, global knowledge about images can help with this decision. The total number of border edge pixels in the image, $\mathcal{B}(0)$, is known approximately, from the number of regions and the perimeter to area ratio of the objects.

Since the total number of edge pixels exceeding s is $\mathcal{E}(s)$, this can be expressed as

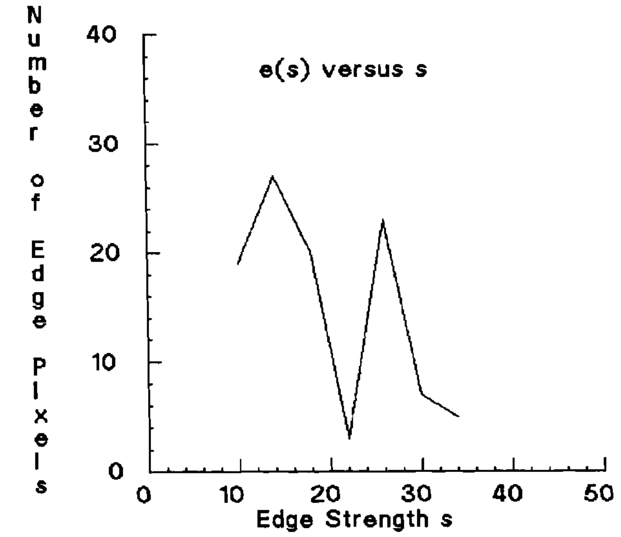
$$\mathcal{E}(s) = \int_s^\infty e(s)\,ds$$

where $e(s)$ is the number of edge pixels at a given strength value s.

$$e(s) = b(s) + t(s) + n(s)$$

To separate border edges, $b(s)$, from texture edges and noise edges, $t(s) + n(s)$, consider

Figure 8: Number of Edge Pixels vs. Edge Strength for the Image in Figure 2

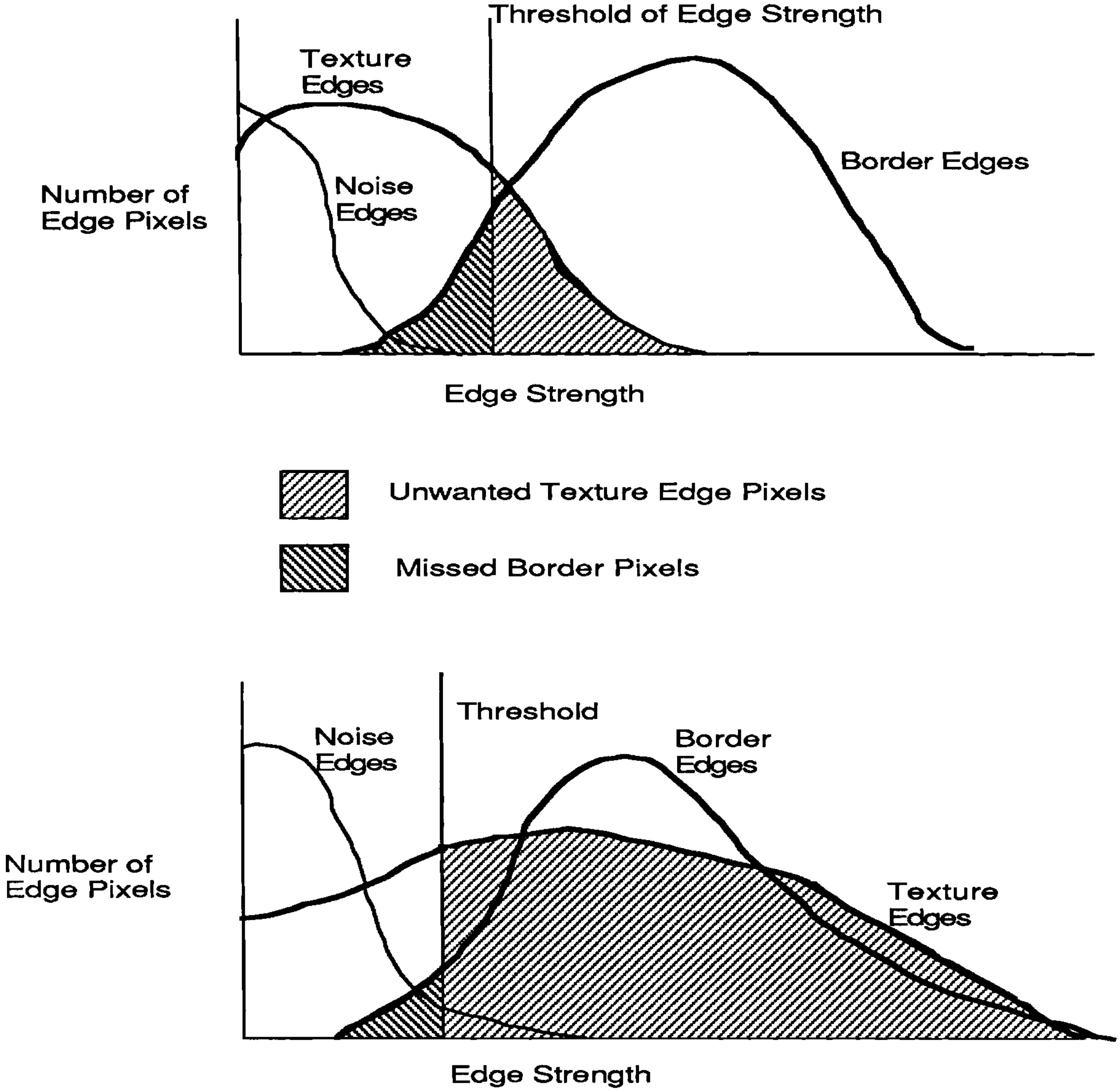


the graph of $e(s)$ versus s in Figure 8. If the border edges have higher contrast, on average, than texture edges and noise edges, then they can be separated with a threshold of edge strength. When s is the local minimum in the bimodal distribution:

$$\mathcal{E}(s) \approx \mathcal{B}(0)$$

However, in some pictures, texture edges have higher strength values than border edges. For example, if the object sought in a photo was a city block of buildings, then the building to building edges within the block may be just as strong or stronger than the building front to sidewalk edges. Figure 9 illustrates the problems edge strength thresholding. On the first example, a threshold of edge strength eliminates most texture edge pixels and few border edge pixels. This is typical of pictures where the texture has lower contrast than the object boundaries, such as Figures 1 and 2. However, in the second example, a threshold of edge strength is not useful. This is typical of pictures like Figure 10 where

Figure 9: Edge Separation using a Strength Threshold



building-to-building edges compete with the city block edges.

4.4 Spatial Frequency of Edges

Edge shapes change when they are found with different scale edge detectors (*i.e.* different values of σ in the Canny edge detection mask). At a larger scale, edges from texture may be removed, but the accuracy of edge location is reduced. In [JC 83], the tradeoff between accuracy of edge location and signal to noise ratio is discussed. In aerial photos, this means that removal of texture (convolution with large Gaussian) produces changes in the shapes of the remaining edges. Building corners become rounded. Lane stripes are widened and the gaps in between them are filled. Figure 11 shows how the shapes of road lane stripes change when viewed at different scales.

Witkin [APW 84] and Bergholm [FB 86] have attempted to track edges across different scales. Witkin uses different scales to fully describe the edges in the intensity function. Bergholm finds edges in large scale, then looks at small scales to give better localization of these edges.

In an urban scene, sampled with pixels spaced at, say, 3 foot intervals, borders are generally man-made or man-altered. The desired edges are the object borders, which are normally long and straight. Urban streets are long and straight. Buildings have straight sides, interrupted at intervals by sharp corners. Oil storage tanks and highways have long edges with specific types of curvature. All of these edges are sharp. The intensity changes can be found at high spatial frequencies in the original image. Gradual changes are found in tree edges, shadows on grass and other natural objects. We want do not

Figure 10: Image at a Scale (approximately 10 feet per pixel) at which High Contrast Edges come from Objects Too Small To Recognize. Photograph courtesy of Prof. R.J. Woodham, University of British Columbia, digitization at the University of Pennsylvania GRASP Laboratory.

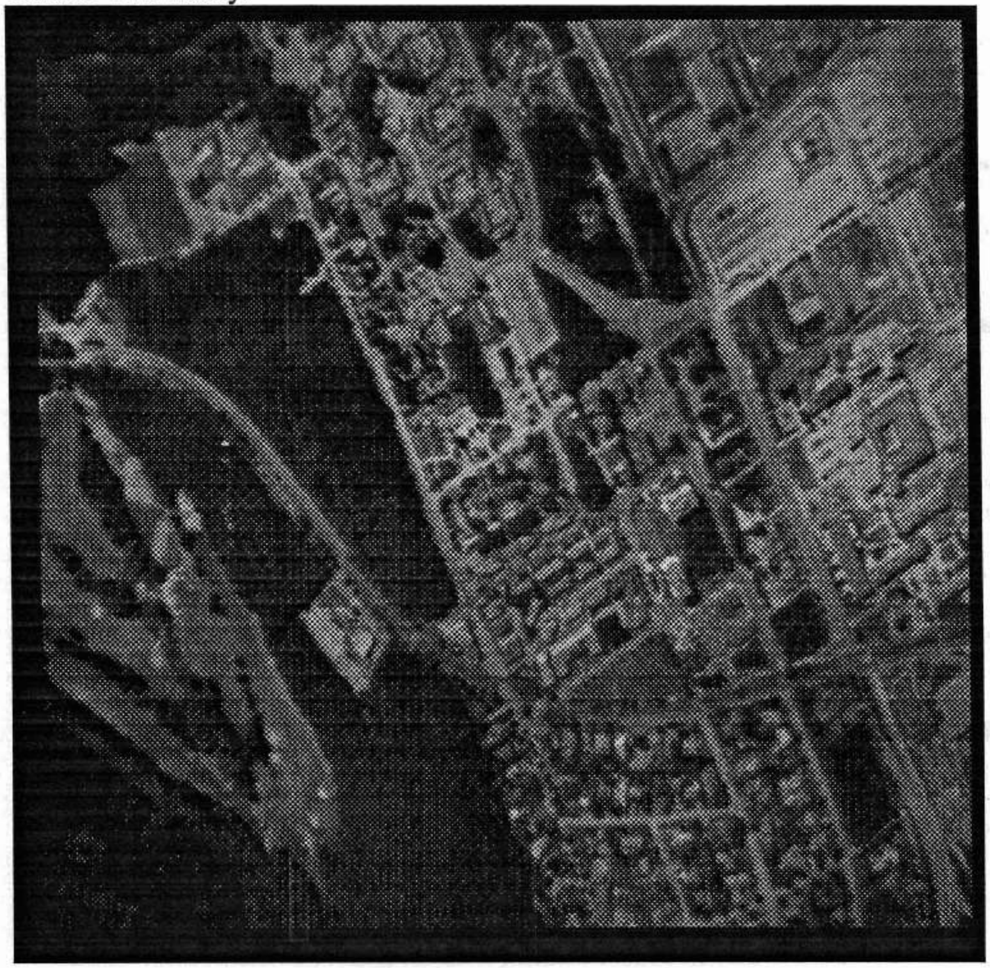
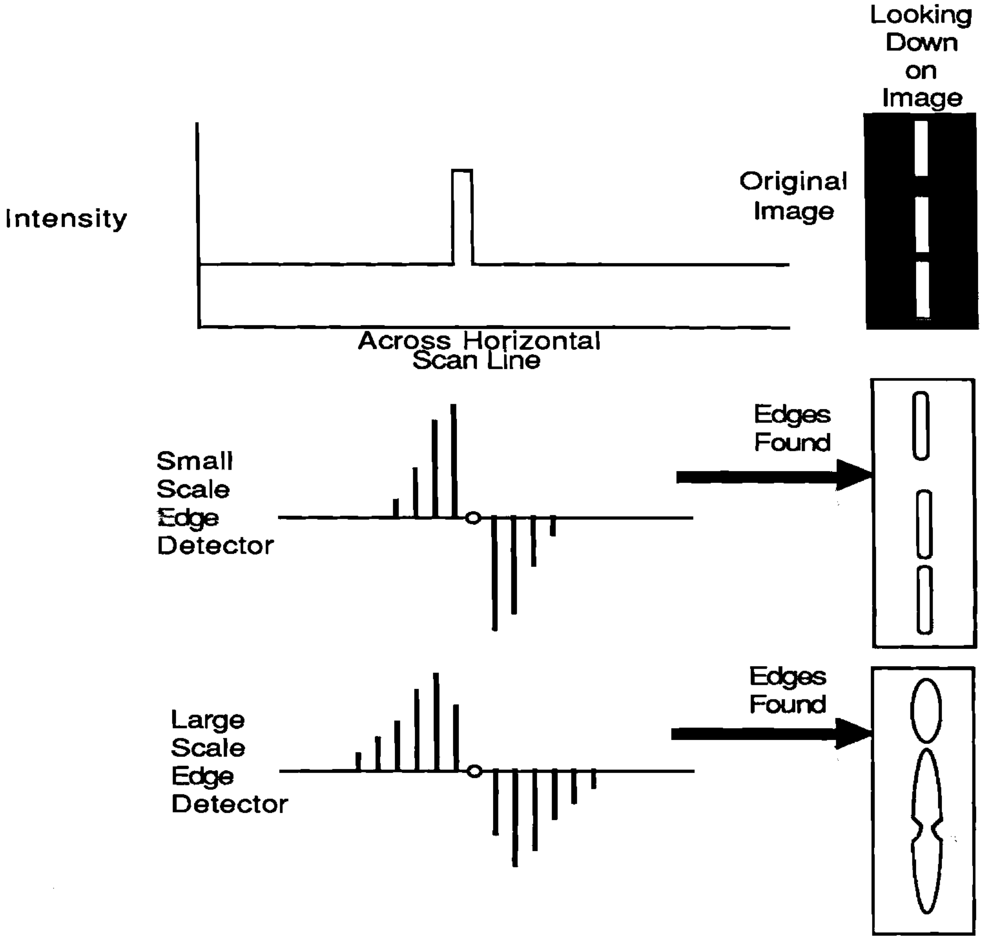


Figure 11: Edges at Different Scales


Looking Down on Image
Intensity
Original Image
Across Horizontal Scan Line
Small Scale Edge Detector
Edges Found
Large Scale Edge Detector
Edges Found

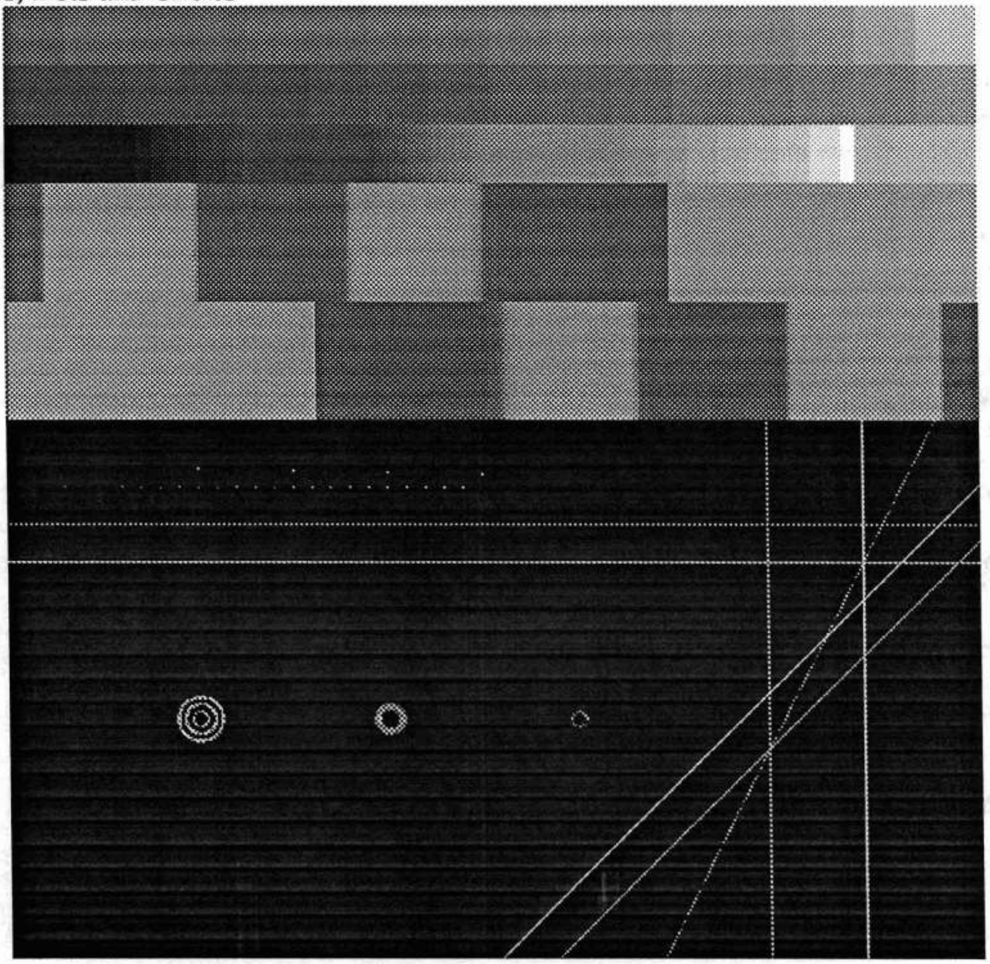
want to recognize trees, but to group trees together and recognize forests. Therefore, we do not want gradual edges which have only low frequency components. However, the edges found at high spatial frequencies include both small and large changes in intensity. We want only the edges with large intensity change. For an unknown image, we do not know how large a change we want until we see the edges.

Emphasis on sharp edges is characteristic of human vision, too. A small change in gray value is visible when it is presented as a step edge, while a much larger change in grey value is difficult to see as a ramp. On a monitor with 256 gray values, this is easy to show. On a printed page, the quality of a 16 level image cannot be distinguished from a 256 level image [TP 82]. Whether there are 256 or 16 gray levels, machine representations of ramps are made up of small steps. However, a gradual ramp looks like a series of steps when printed with 16 gray levels. A steep ramp is more difficult to localize. Human vision enhances sharp edges, but does not enhance gradual edges. This is shown in the synthetic picture which was used for algorithm testing. It is shown in Figure 12.

5 Quantizer Votes

Quantizer votes is a nonlinear algorithm which eliminates edges with low intensity change and edges with only low spatial frequency. That is, subtle edges and gradual edges are eliminated while sharp edges are retained. The algorithm is adjustable such that it can eliminate low intensity and low frequency edges up to different levels, depending on the content and quality of the picture. This is done without loss of localization accuracy.

Figure 12: Synthetic Picture of 2-Dimensional Intensity Steps, Ramps, Blurred Edges, Lines, Dots and Circles



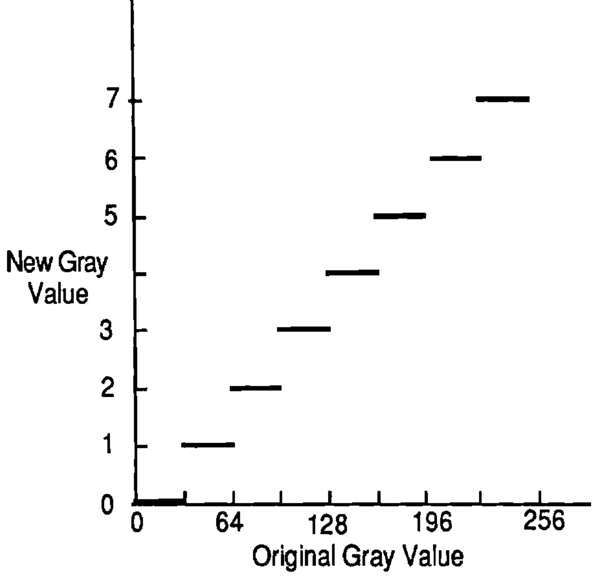
There is no guarantee that the edges which are eliminated are uninteresting. However, several statistics about the resulting edge pictures are presented, and they can be used to rate the effectiveness of the algorithm at different levels on aerial photos.

5.1 Quantizer Votes Algorithm

The algorithm uses a combination of several images of the local maxima of the Gaussian smoothed gradient on quantized versions of the original image. The Gaussian smoothed gradient of an image depends on the standard deviation, σ, of the Gaussian filter. At a larger convolution scale, edges from texture may be removed, but the accuracy of edge location is reduced. The shapes of the remaining edges are changed. Rather than a combination of edges using different spatial frequencies, our algorithm combines edges in the amplitude (intensity) domain.

Since a 16 gray level image looks the same as a 256 gray level image on the printed page, all 256 gray levels are not required to identify objects in an image. An 8 gray level picture will look almost as good as a 256 gray level image with the addition of *dither*. Dither is the addition of noise generated by a random process. It tends to break up contours produced by the coarse quantization of an image [TP 82]. Dithering a coarsely quantized image of a face makes the face more pleasant to look at, since sharp edges are not expected in a face. The re-quantization done in the quantizer votes algorithm restores the contours which may be removed by dithering. Sharpened edges are useful for recognition based on edge shape. For this purpose, the best re-quantization is the minimum representation which preserves the information necessary for object recognition.

Figure 13: Re-quantization Mapping Function



A quantization function maps the input gray levels to the re-quantized gray levels. This mapping can be considered a method of estimating the gray value in the noisy original image [RM 86]. A symmetric step function is the simplest mapping. A step mapping function is shown in Figure 13.

The technique consists of three parts: repeated quantizing and gradient detection, then combination of results. The technique can be used at various levels, depending on the amplitude of edge desired, but a specific set of thresholds is described below.

First, the image is quantized from a 256 gray scale image to a 16 gray scale image. The 16 remaining gray values correspond to the original 0-15, 16-31, 32-47,... Then gradient detection, using the Canny operator, is done on the quantized image.

Next, the image is re-quantized to a 16 gray scale image, using a smaller first bin and larger last bin. The 16 gray scales correspond to the original scale 0-14, 15-29, 30-45, 46-62... Again, gradient detection is performed.

The re-quantizing and gradient detection is repeated until the first bin is vanishes. (An alternative way to consider the repeated re-quantizing is this: add a bias of intensity 2 to every pixel, then re-quantize using the original quantization mapping. This process is repeated until the size of the bias equals the quantization step size.)

Finally, edge pixels from the different re-quantized images are combined. Some pixels are found to be edges in all of the re-quantized images, some are edges in none, and some in between. There are 8 different re-quantizations, so a pixel can get counted as an edge, or "voted," up to 8 times. The intensity change weak edge is in the original image is proportional to the number of votes the pixel has received. A strong edge, with intensity change greater than the step size in the mapping function, will receive the maximum number of votes. Votes are summed for each pixel to produce a combination image. This is demonstrated for a single scan line in Figure 14. A threshold of gradient value is used to prevent extremely weak edges from counting toward the combined image. The gradient threshold removes most of the edge pixels due to noise. Gradual edges become steps when re-quantized. The steps produce edges which are artifacts due to re-quantization. However, *these artifacts move as the re-quantization mapping function changes*. When a vote threshold is used *at each pixel*, no single pixel has enough votes to pass the vote threshold.

Edges found using two different step size thresholds are shown in Figures 15 and 16.

Figure 14: Edge Pixel Votes on a Single Scan Line

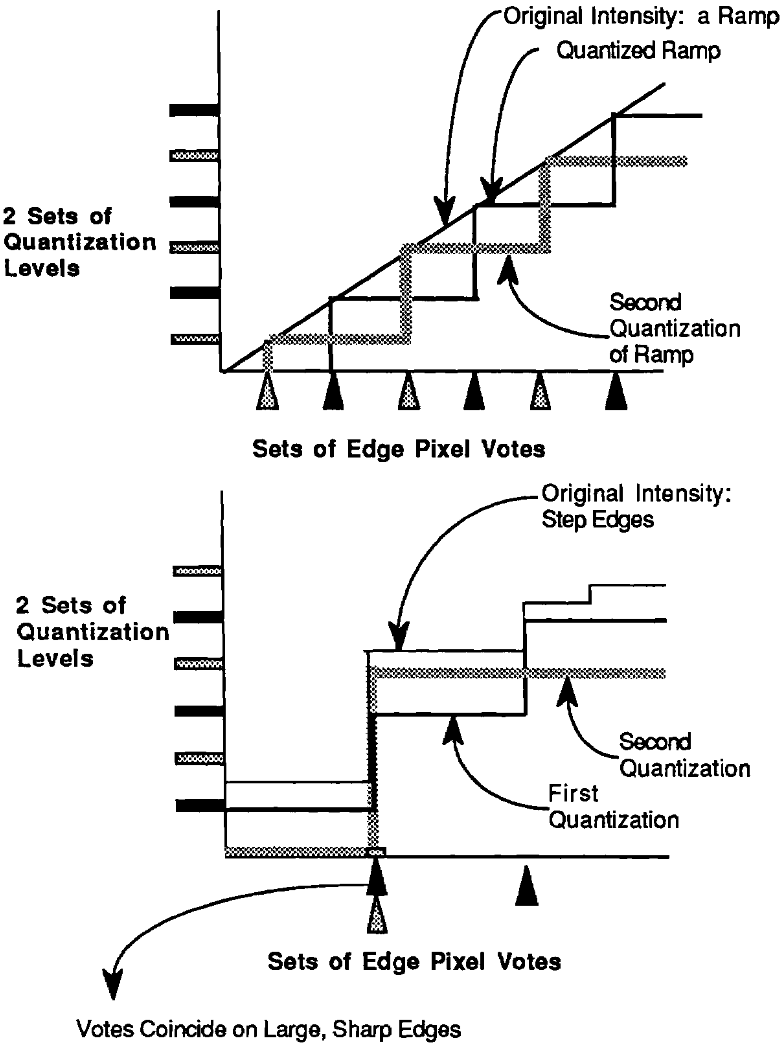

Original Intensity: a Ramp
Quantized Ramp
2 Sets of Quantization Levels
Second Quantization of Ramp
Sets of Edge Pixel Votes
Original Intensity: Step Edges
2 Sets of Quantization Levels
Second Quantization
First Quantization
Sets of Edge Pixel Votes
Votes Coincide on Large, Sharp Edges

Figure 15: Edges Found for Figure 2 Above Step Size 9

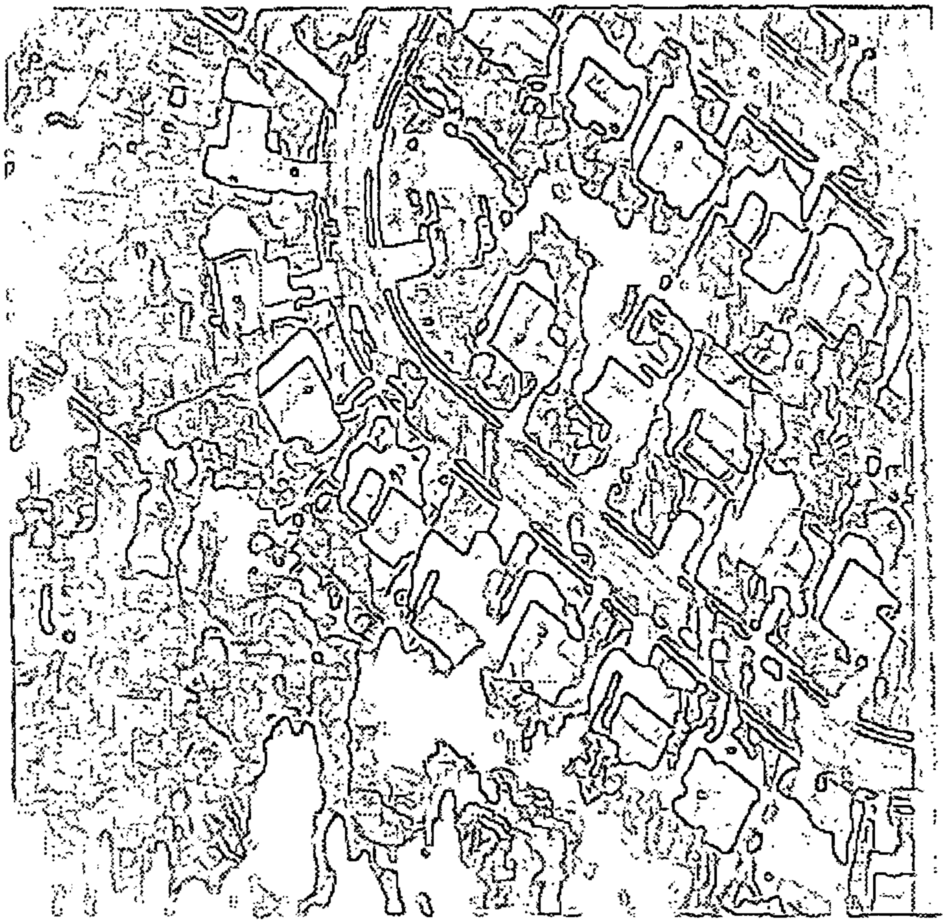
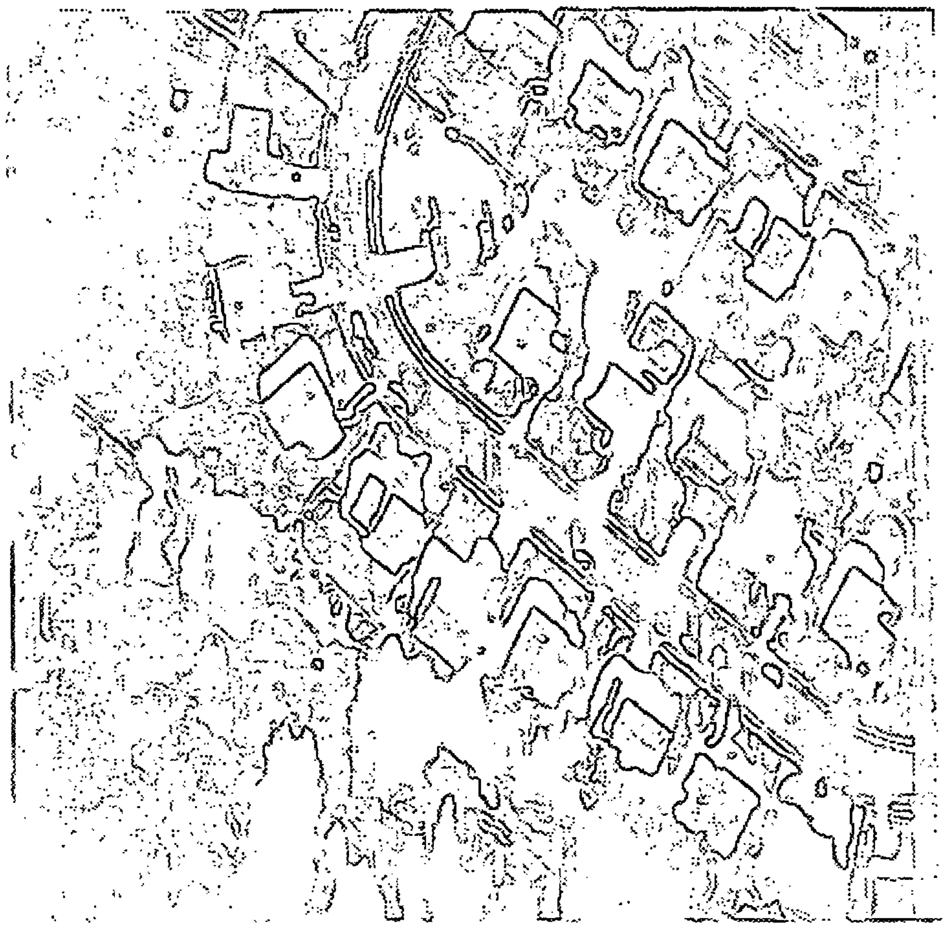


Figure 16: Edges Found for Figure 2 Above Step Size 18



5.2 Results and Discussion of Quantizer Votes Edges

Certain types of edges are attenuated by the quantizer votes process. A ramp edge is made into a series of steps by quantizing. After shifting, the step locations move. Thus, no single location on the ramp edge gets enough votes to pass thresholding. All gradual, or low spatial frequency, changes in intensity are attenuated by this process.

High frequency edges do appear in the shifted image. In the final combination image, a two dimensional step edge (further discussed below) will appear if it exceeds ss.

$$\frac{(v_t - 1)q}{V} < ss \leq \frac{v_t q}{V}$$

$v_t =$ number of votes required to pass threshold

$V =$ total number of votes

$q = \frac{original\ gray\ scale}{re-quantization\ gray\ scale} =$ quantization ratio

Between these two values, the edge may appear, depending on how the edge fell in the quantization. The probability of an edge appearing increases linearly from 0 to 1 over the range.

Thus, the amplitude of edge to be attenuated varies with the quantization ratio and the level of thresholding. The process works well with a quantization rate of $8 \leq q \leq 64$. Below 8, the quantization has no effect. Above 64, too many edges are lost. The threshold to total vote ratio, v_t/V from 2/8 to 5/8 give reasonable results on the test pictures, with the best contours usually at 3/8 and 4/8. In the experiment, the total range of maximum step edge to be attenuated can vary from 4 to 32 on a 256 gray scale (8 bit) image.

The step edge discussed above is a 1-dimensional step, repeated for several scan lines, so the total edge is 2-dimensional. The gradient detection uses information from adjacent scan lines to improve the edge detection of 2-dimensional edges. Taking a threshold of the gradient value before voting removes steps which only appear in 1 scan line. This is useful for eliminating stuck pixels and some other local area noise. Using the Gaussian-smoothed gradient detection method is better than local differencing, which would be sensitive to single pixel noise. It also removes the noise of real objects which are too small to classify.

Convolution is a linear process, but non-maximum suppression is nonlinear, so the combination of these processes is a nonlinear process. The combination process improves the signal to noise ratio without reduction in edge localization, until noise exceeds the maximum attenuation for the filter settings. When that occurs, the process makes the signal to noise ratio worse.

For example, the quantizer votes algorithm has been tested finding step edges in Gaussian white noise. Pixel intensities range from 0 to 255. If the value of the noise is ≤ 6, and a quantization ratio of 32 is used, up to 2 votes out of 8 may be changed by the noise. Using a vote threshold of 3, this will have no effect on the votes combination picture. However, if the value of the noise is > 15, and a size 32 quantization is used, 4 or more votes will be changed. This will cause the noise pulse to be falsely classified as an edge in the final picture.

Experiments on temporal noise in the Pennsylvania Active Camera System indicate that usually $\sigma^2 < 16$ [RM 86] for a maximum signal intensity of 256.

To test the effect of noise on the edge detection, Gaussian white noise with $\sigma^2 = 16$ was added to the synthetic picture. Approximately 95% of the additive noise will be

$\leq |8|$ on the original gray scale. This is individual pixel noise, and that is not likely to produce a 2-dimensional step edge. The Gaussian-smoothed gradient detection process reduces the effect of Gaussian white noise. Therefore, a pre-vote threshold on the Canny gradient values which eliminates step edges ≤ 8 will prevent almost all false detection of edge pixels. However, noise at edges causes a more serious problem that noise in areas of constant intensity. Noise at edges may cause inaccurate edge localization.

To compare the quantizer votes filter using the Canny process with the Canny process alone, a synthetic test picture was used. The test picture includes a series of step edges of different intensities, ramp edges, cumulative Gaussian edges, lines of different orientations and intensities, single pixel dots of various intensities, and several circles. The gradient values of pixels on either side of a perfect step edge are equal, so the Canny non-maximum suppression removes neither. Therefore, the 2-dimensional step edge shows as a double line in the result. However, when a little noise is added, the gradient at one pixel or the other becomes a local maximum, and the edges become erratic line, as in Figure 17. Quantizer votes removes the effect of small changes in local intensity and the same noisy edge is a straight double line. In real pictures, double edge lines are more likely to occur using quantizer votes than using Canny alone because the number of possible intensity values is much lower after quantization. Thus edges are more likely of be symmetric. This can be beneficial when edges are used for region growing because they are more likely to be closed over long distances; however, for other applications double lines may be a problem.

Synthetic intensity ramps appear as very wide edges using gradient detection. They

Figure 17: Effect of Gaussian White Noise of Variance 1 on Canny's Method and Quantizer Votes Edge Detection Algorithms

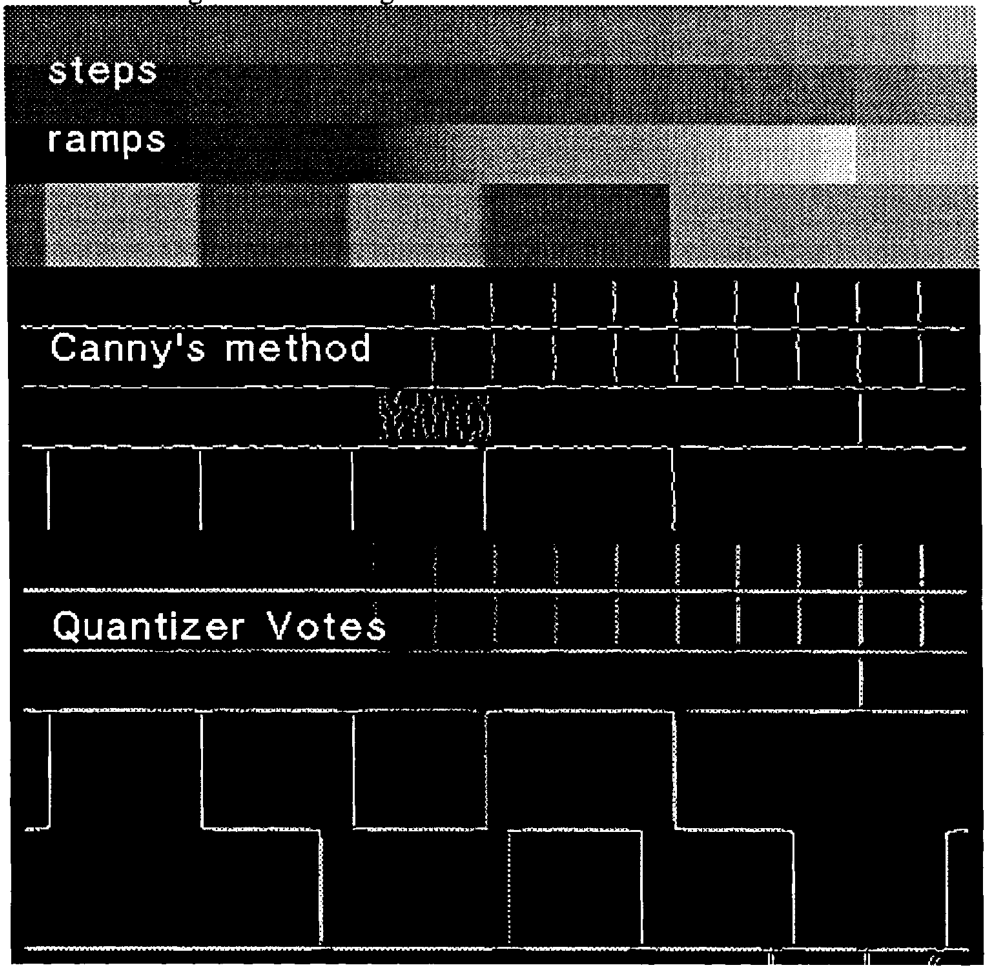
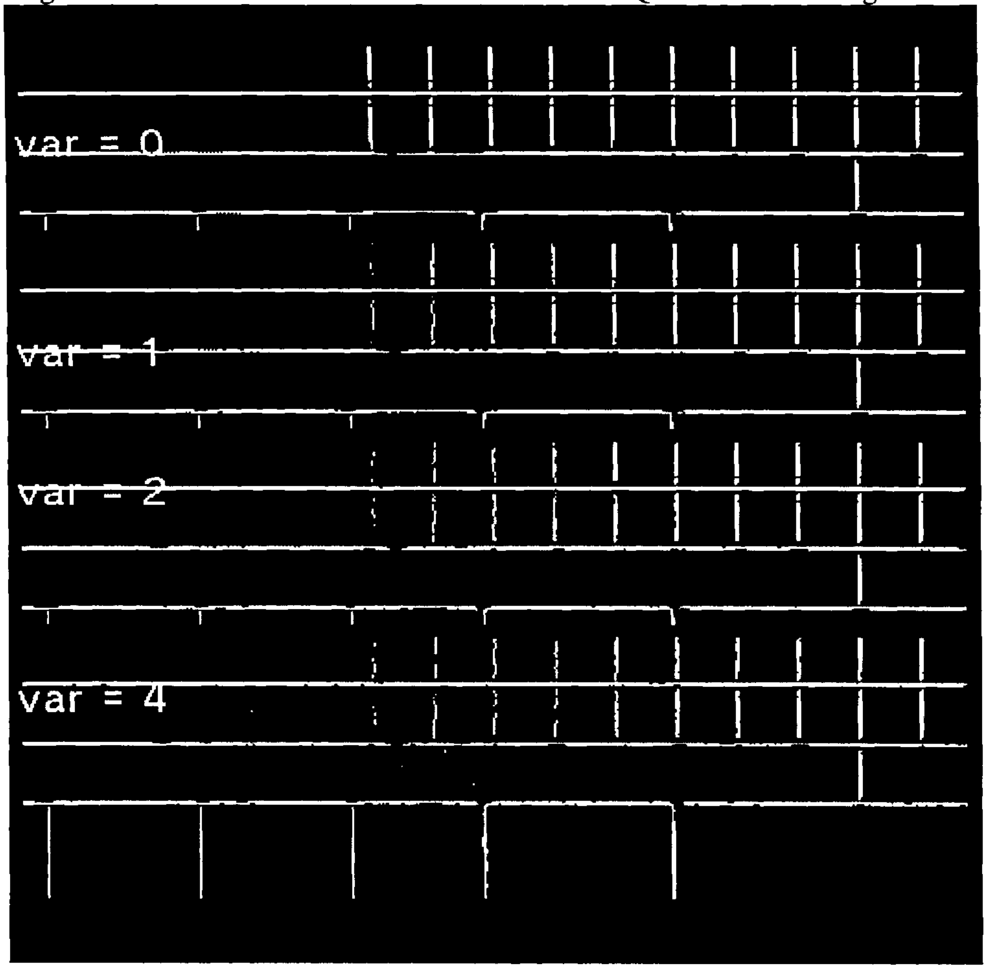


Figure 18: Effect of Different Levels of Noise on Quantizer Votes Algorithm



are not thinned by non-maximum suppression, because the gradient values are identical. With Gaussian white noise added, there are small jumps in intensity and small flat levels of intensity, so the ramp becomes an area of short, erratic edges. The quantizer votes algorithm eliminates these edges until the noise is very strong. This effect is shown in Figure 18.

The quantizer votes process works well using a convolution with a mask which is a directional derivative of a Gaussian function. However, the exact nature of the convolution mask is not specified by the process. Quantizer votes and other amplitude domain processing, may work with a large variety of convolution masks. In principle, quantizer votes could work using local differencing instead of convolution (and non-maximum suppression in either case), but we expect a degradation of performance from the lack of good pre-vote thresholding.

The quantizer votes algorithm could be implemented as a parallel process. The quantizations could be done in parallel, the edge detection could be done in parallel and the non-maximum suppression could be done in parallel. No global information is used until the edge information is completed.

There is no a priori distinction between noise, texture and useful edges. If noise exceeds the amplitude of attenuated edges, that noise will be passed through the process. Conversely, if the amplitude of useful edges is less than the attenuation amplitude, those useful edges will be lost. However, some analysis of the resulting edges will indicate whether the process was successful.

5.3 Checking Edge Results

Several statistics have been developed which measure whether the border edges were separated from the noise and texture edges. One statistic tests whether the *number* of edge pixels *found* is the *number* of border edge pixels *expected.* The other statistic tests whether the *average length* of the groups of edge pixels *found* has the *average length* of the groups of edge pixels *expected.* These methods will not prove that the edge pixels found are actually the desired border edges. That can only be done subjectively. However, these methods provide indications that the edge pixels are the border edge pixels.

One statistic is the the *edge pixel frequency* introduced in Section 4.2. Edge pixel frequency is the average number of thinned edge pixels found per thousand pixels in the original image. If the graph of the number of edge pixels $e(s)$ versus edge strength s has a bimodal distribution (Figure 8), the local minimum is used as the strength threshold s_t. If the total number of edge pixels $\mathcal{E}(s_t)$ at or above s_t falls into a range of expected values for the number of border edge pixels $\mathcal{B}(0)$, then there is a good chance that most of the border edge pixels are found. This method was used along with subjective evaluations to establish a theoretical basis for edge strength thresholding.

This method is slow for the quantizer votes algorithm because the quantization and edge detection must be done repeatedly for different edge strengths thresholds. The same statistic can be used with Canny's method of edge detection, since a histogram of strength values can be used directly. However, in our tests, edges found using Canny's method did not produce a bimodal distribution.

Another way to confirm that noise and texture edges are separable from border edges uses the *average edge extent* statistic which was introduced in Section 4.2. Average edge extent is a measure of average length of 8-connected edge pixels, but the connected pixels need not be in a line. If there is a jump in average edge extent at the threshold s when $\mathcal{E}(s)$ is in the expected range for $\mathcal{B}(0)$, then this is a strong indication that the border edges were separated. However, if some noise and texture edges exceeded the strength threshold, they will reduce the average edge extent. For example, the image in Figure 2 did not have a sharp increase in average edge extent when the Quantizer Votes algorithm was used for edge detection. When Canny's method was used for edge detection, there was a gradual increase in average edge extent with increasing edge strength threshold.

A faster way to confirm separation of border edge pixels is to choose a strength threshold, then calculate the average edge length and edge pixel frequency. If pixel frequency and length are within acceptable limits, then the threshold is acceptable. If not, use the pixel frequency to choose whether to increase or decrease the strength threshold. More experiments need to be done in this area to find a quick, reliable process.

6 Suggestions for Further Study

More work is needed to enhance machine object recognition capabilities. The new edge thresholding work should be integrated with edge enhancement and region growing processes, and the resulting regions should be checked with the expectations about regions. Knowledge-based interaction between these modules may be a substantial contribution to the field of object recognition.

Quantizer votes can be compared with another algorithm which would produce edges with similar spatial frequency characteristics. This algorithm would combine edges with high edge strength at a narrow scale with high edge strength at a wide scale. This method may give similar results using pyramid image processing rather than parallel image processing.

The edge extent, edge pixel frequency and total number of object instances per view statistics need to be studied further to confirm their usefulness in general vision. In particular, the range of border edge densities found in different types of scenes, not just urban and suburban aerial photos, needs to be studied.

7 Conclusions

The LandScan object identification system for use on aerial photos of urban scenes has been described. The LandScan system performs automatic recognition on images of a model of the University of Pennsylvania campus. Current versions are being enhanced to recognize objects from digitized aerial photographs on a larger set of scenes.

The LandScan recognition system is based on identification of region shapes. It is a modular system, and the modules are being studied and enhanced individually. In this work, several approaches to the edge detection module have been studied. A new concept based on the quantizer votes algorithm has been developed. The quantizer votes algorithm is an edge detection method which produces appropriate edges for recognition of man-made objects. This method is based on re-quantization of the original image and repeated application of Canny's method for gradient detection and non-maximum suppression. This method has been tested and discussed.

Several new statistics about edge pixels have been introduced and studied. These are *average edge extent* and *edge pixel frequency*. Average edge extent is the average length of connected, but not necessarily straight, edges in the image. Edge pixel frequency is the number of edge pixels per 1000 pixels in the image. These statistics were tested on several images and there are indications that they may be useful for general application to edge detection. In addition, a limit on the *total number of object instances per view* has been proposed, based on human perception and machine capabilities. This limit, along with object area to perimeter ratios, will provide information necessary to calculate the edge pixel frequency.

The goal of this work has been twofold. First, a modular automatic object recognition system has been built. Second, that system has been enhanced to use more knowledge about kinds of edges and characteristics of edges to improve recognition on the basis of shape.

A Appendix: Recognition System Design Questions

The design of a recognition system is driven by the answers to the following questions:

- What objects are the system supposed to recognize?

  understand context: Will the system try to recognize everything in the scene, or will part of the scene be considered background?

recognition vs. discrimination vs. inspection: Is the recognition a division into classes, then finer distinctions? Is the object assumed to be in a particular class, and the system is to discriminate between instances? Is the system to compare objects to specific patterns or to a range of features?

range of variability: How much change within object classes will there be? Do object classes ever overlap?

- For each object, what are the important diagnostic features upon which to base recognition?

  1. How will raw information be obtained?
  2. What types of feature information can and should be obtained from the raw data?
  3. How will feature information be combined?
  4. Will recognition of some objects in the scene influence identification of other objects?

- In what situations is the system supposed to work?

  1. Will the same features be available in all situations?
  2. How will feature information change in different situations?
  3. How will the system recognize what the situation is?
  4. Under what situation is the system not expected to work, and what should its response be then?

- How will the system be controlled?
    1. How will the task begin?
    2. How will control pass within the system?
        (a) Can processing be done in parallel?

        (b) Is there internal feed back, self-checking, relaxation labeling, or other repetitious processing toward a goal? What is that goal?
- What is the nature of the result?
    1. What are the termination criteria?
    2. In what form is the result?
    3. How is the result delivered?

Once the overall system design is complete, there are many different kinds of algorithms to choose from. The recognition decisions can be made using a fixed rules or fuzzy sets. The initial identifications can be the final identifications, or they can be updated using a set of rules about inter-relationships. Recognition algorithms should be chosen on the basis of the expected objects' characteristics. For example, if the set of objects to identify is fairly consistent, then a fixed set of rules should produce reasonable results on the first iteration. On the other hand, if objects have overlapping characteristics, a fuzzy set of rules would be more useful. The recognition portion of the system can be modular, so that the recognition algorithm can be improved if it is not adequate. In this way, the system can begin with a relatively simple recognition algorithm. When the limits of the algorithm are reached, it can be replaced without redesigning the whole system.

Features for recognition need to be extracted from the pixels. There are many kinds of features which can be used, and there are various algorithms used to find them. Again, this aspect of recognition can be modularized such that the feature extraction algorithm can be changed if it is not adequate. The exact feature value ranges in the recognition module would have to be changed, but the rest of the recognition module should still function.

References

[RB 86-2] R. Bajcsy, E. Krotkov, M. Mintz. "Models of Errors and Mistakes in Machine Perception," University of Pennsylvania GRASP Laboratory Technical Report 26, 1986.

[RB 86] R. Bajcsy, M. Mintz, E. Liebman. "A Common Framework for Edge Detection and Region Growing," University of Pennsylvania GRASP Laboratory Technical Report 61, February, 1986.

[DHB 82] D.H. Ballard and C.M. Brown. *Computer Vision* Prentice-Hall, Englewood Cliffs, New Jersey, 1982.

[FB 86] F. Bergholm. "Edge Focusing" in *Proceedings of the International Conference on Pattern Recognition* Paris, Oct.27-31, 1986.

[BRB 77] B.R. Bullock. "The Necessity for a Theory of Specialized Vision," Machine Vision Conference paper, University of Massachusetts, June 1-3, 1977.

[JC 83] J. Canny. "Finding Edges and Lines in Images," MIT AI Laboratory Technical Report 720, June, 1983.

[ERD 86] E.R. Davies. "Constraints on the design of template masks for edge detection" in *Pattern Recognition Letters* Volume 4 Number 2, pp. 111-120, April, 1986.

[FF 86] F. Fuma, E. Krotkov, J. Summers. "The Pennsylvania Active Camera System," University of Pennsylvania GRASP Laboratory Technical Report Number 62, February, 1986.

[DJG 81] D.J. Granrath. "The Role of Human Visual Models in Image Processing" in *Proceedings of the IEEE* Volume 69 Number 5, pp. 552-561, May, 1981.

[AKG 73] A.K. Griffith. "Edge Detection in Simple Scenes Using A Priori Information," in *IEEE Transactions on Computers* Volume C-22 Number 4, pp. 371-381, April, 1973.

[MH 84] M. Herman, T. Kanade, S. Kuroe. "Incremental Acquisition of a Three-Dimensional Scene Model from Images," in *IEEE Transactions on Pattern Matching and Machine Intelligence* Volume PAMI-6 Number 3, pp. 331-340, May, 1984.

[BKPH 86] B.K.P. Horn. *Robot Vision,* MIT Press, Cambridge, Massachusetts, 1986.

[AH 83] A. Huertas and R. Nevatia. "Detection of Buildings in Aerial Images Using Shape and Shadows," Proc. of 8th International Joint Conference on AI, Aug 8-12, 1983.

[EWK 86] E.W. Kent and M.O. Shneier. "Eyes for Automatons" in *IEEE Spectrum* Volume 23 Number 3, pp. 37-45, March, 1986.

[BLK 80] B.L. Kottas and D.W. Bessemer. "Comparison of Potential Critical Feature Sets for Simulator-based Target Identification Training," U.S. Army Research Institute for the Behavioral and Social Sciences Technical Report 510, September, 1980.

[EK 85] E. Krotkov. "Results in Finding Edges and Corners in Images Using the First Directional Derivative" University of Pennsylvania GRASP Laboratory Technical Report 37, March, 1985.

[DM 82] D. Marr. *Vision,* W.H. Freeman, San Francisco, 1982.

[RM 86] R. McKendall, M. Mintz, "Models of Sensor Noise and Optimal Algorithms for Estimation and Quantization in Vision Systems," GRASP Laboratory Report Number 80, 1986.

[DMM 84] D.M. McKeown, Jr., W.A. Harvey, J. McDermott. "Rule Based Interpretation of Aerial Imagery" in *IEEE Proceedings of the Workshop on Principles of Knowledge-Based Systems*, pp. 145-157, 1984.

[AMN 84] A.M. Nazif and M.D. Levine. "Low Level Image Segmentation: An Expert System" in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-6, No. 5, pp. 555-577, September, 1984.

[JAM 85] J.A. Mulder. "Using Discrimination Graphs to Represent Visual Knowledge," unpublished Ph.D. dissertation, University of British Columbia, 1985.

[RN 86] R. Nevatia. report at the 1986 Defense Mapping Agency Research Review, St. Louis, Mo., December 4, 1986.

[TP 82] T. Pavlidis. *Algorithms for Graphics and Image Processing* Computer Science Press, Rockville, Maryland, 1982.

[WAP 86] W.A. Perkins, T.J. Laffey, T.A. Nguyen. "Rule-based interpretation of aerial photographs using the Lockheed Expert System," in *Optical Engineering* Volume 25 Number 3, pp. 356-362, March, 1986.

[IP 86] I. Pitas, A Venetsanopoulos. "Nonlinear Mean Filters in Image Processing" in *IEEE Transactions on Acoustics, Speech and Signal Processing* Volume ASSP-34 Number 3, June, 1986.

[DAR 84] D.A. Rosenthal, R. Bajcsy. "Visual and Conceptual Hierarchy: A Paradigm for Studies of Automated Generation and Recognition Studies" in *IEEE Transactions on Pattern Analysis and Machine Intelligence* Volume PAMI-6 Number 3, May, 1984.

[WBS 84] W.B. Schaming, L.E. Toombs. "Multifeature methods for target detection" in *RCA Engineer* Volume 29 Number 6, pp.56-59, November/December, 1984.

[DLS 85] D.L. Smitley. "The Design and Analysis of a Stereo Vision Algorithm," University of Pennsylvania GRASP Laboratory Technical Report 42, May, 1985.

[DT 85] D. Talton. "Implementation of a Gaussian-Smoothing Gradient-Edge Detector," University of Pennsylvania GRASP Laboratory Technical Report 35, January, 1985.

[WRU 83] W.R. Uttal. *Visual Form Detection in 3-Dimensional Space,* Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1983.

[MW 39] M. Wertheimer. "Laws of Organization in Perceptual Forms," in *A Source Book of Gestalt Psychology,* W.D. Ellis, ed., Harcourt, Brace and Company, New York, 1939.

[APW 84] A.P. Witkin. "Scale space filtering: a new approach to multi-scale description", Chapter 3 in *Image Understanding 1984* S. Ullman and W. Richards, Ablex Publishing Co, 1984.

[CB-1] **Pre-School Level Coloring Books:** *Sesame Street A to Z* (1976), Golden Coloring Book, *Color Me Happy* (1985), *A Coloring Book: ABC* (1983), Golden Pre-school Coloring Books, Western Publishing Co., Racine, Wisconsin.

[CB-2] **Kindergarten and First Grade Level Coloring Books:** I. Forte (1982) *Read About It: Activities for Teaching Basic Reading Skills* Incentive Publications, Nashville, Tenneessee. K. Barabas and Joanne Ryder (1984) *My Family* Golden Step-Ahead, M. Cohen (1985) *Grow Safe* Golden Learn About Living, Western Publishing Co., Racine Wisconsin.

[CB-3] **Above First Grade Level Coloring Books:** T. Ho, illus., and M. Williams *The Velveteen Rabbit Coloring Book* Simon and Schuster, New York. J. Tenniel, illus., and L. Carroll *Alice in Wonderland Coloring Book* Dover Publications, New York, illustrations enlarged and text abridged from *Alice's Adventures in Wonderland* (1865) Macmillan and Company, London.