# U.S. DEPARTMENT OF THE INTERIOR
# U.S. GEOLOGICAL SURVEY

## An Intelligent Systems Approach to Automated Object Recognition
## A Preliminary Study

by

Brian G. Maddox[1], Casey L. Swadley[2]

Open-File Report 02-461

2002

[1] Computer Scientist, USGS Mid-Continent Mapping Center, Rolla, MO, 65401
[2] Student Computer Trainee, USGS Mid-Continent Mapping Center, Rolla, MO 65401

## CONTENTS



## ILLUSTRATIONS

**KEY WORDS**
Object Recognition, Feature Extraction, Computational Intelligence

**ABSTRACT**
Attempts at fully automated object recognition systems have met with varying levels of success over the years. However, none of the systems have achieved high enough accuracy rates to be run unattended. One of the reasons for this may be that they are designed from the computer's point of view and rely mainly on image-processing methods. A better solution to this problem may be to make use of modern advances in computational intelligence and distributed processing to try to mimic how the human brain is thought to recognize objects. As humans combine cognitive processes with detection techniques, such a system would combine traditional image-processing techniques with computer-based intelligence to determine the identity of various objects in a scene.

## INTRODUCTION

The ability to have a computer "look" at an object and identify it has been long sought in computer science. From product defect identification to computer vision, there are thousands of applications for this type of technology. Many attempts at computer recognition have been made over the years and have met with limited success. Unfortunately, none of these attempts have yet reached the 100-percent mark for unattended operation.

Visual pattern recognition is a difficult and complicated process. Problems such as color differences, changing points of view, and occlusion further compound the problem by changing the appearance of the object depending on the observation frame. Recognition is an easy task for humans, as we have brains that have evolved to function in a three-dimensional environment and have cognitive abilities that enable us to make sense of the visual inputs. For a computer, however, this is an extremely difficult problem, because it does not have any built in cognitive capabilities.

As part of the fiscal year (FY)# 2002 National Mapping Discipline (NMD) Prospectus project, *A Parallel Processing Approach to Computing for the Geographic Sciences*, work began at the Mid-Continent Mapping Center (MCMC) to try to develop a system that would be able to simplistically mimic both the biological and the cognitive processes that humans use in object recognition. This method would combine traditional image-processing methods with some computational intelligence concepts to determine the identity of an object.

## BACKGROUND AND DISCUSSION

One of the problems with the historical methods of object recognition is that they perhaps focus solely on the capabilities that a computer might use to recognize something. Typically, they approach the problem solely from an image-processing perspective. These techniques have ranged from brute force pattern recognition, where a known object is distorted in many different ways to try to match an image, to more advanced techniques that use neural networks to train on the presence of certain features. They typically involve pattern matching and are designed so that they can run on a single computer. These techniques have had some successes in automated operation, but usually still require some form of manual intervention.

Optical Character Recognition (OCR) provides one of the best examples of methods that are currently used for object recognition. Initially, OCR simply tried to use a brute force method for matching letters. It later overcame the recognition deficiency by broadening the approach of looking at individual characters to looking at more contextual cues. OCR document recognition

techniques use a variety of strategies, such as dictionary checking, font identification, word recognition, page layout, and sentence structure. Modern OCR systems also incorporate decision trees to identify of individual characters.

Other methods that OCR uses that are representative of past and present computer visual recognition techniques are layout segmentation, character building, and font identification. Layout segmentation breaks the document down into segments, such as pictures, text paragraphs, lines, and words. Character building identifies characters on the basis of their shape and location in the document. For example, 9 and g have very similar shapes, but location on the line and context would aid in the proper identification. After an initial guess as to the character identification, the systems have ways to check their accuracy and use backtracking methods if a dictionary check fails to identify a word. Some OCR systems use font identification to aid in character identification, and others look for patterns rather than working inside a known set of fonts (Belaïd). With all of these techniques, typed characters had an identification rate of nearly 100 percent per word. However, handwritten documents had a recognition rate of only about 50 percent per word, meaning that human correction was needed for every other word (Comay, 2002). Fixed-font typefaces are much easier to detect as they follow set patterns. Human handwriting, however, follows no such fixed pattern and is thus much harder to identify.

A more recent development that may be helpful in geospatial object recognition is the progress in handwritten document recognition. Advanced Character Recognition (ACR) allows the reading of handwritten characters with no supporting information in various document types, forms, and unstructured notes at a recognition rate of about 98 percent in most cases. One thing that makes ACR so different from OCR is the learning procedure that is applied after every recognition procedure. This allows the system to adapt to each individual's writing style. It is this continuous learning that caused most of the recognition success. After a fraction of the document has been processed, all of the words recognized in a dictionary are used as training for the others. However, that is not the only difference between ACR and OCR. ACR also incorporates a sophisticated voting scheme, which allows multiple engines to vote among each other. More than one engine is used whenever possible to ensure the best results. These ACR-type techniques can be used in feature extraction. For example, although some features in cities are easier to identify owing to clear grids and less vegetation, this methodology is key to object identification in the unstructured world outside of urban areas (Comay, 2002).

A better method of visual pattern recognition might entail trying to make a computer mimic the human brain in recognizing objects. Human perception is a very complex process that starts with biological specialization and ends with thought processes that categorize inputs on the basis of environmental cues and historical knowledge. The cognitive processes give us the ability to recognize objects in various conditions. Without the addition of intelligence, we would be

much less successful in judging what something is when we see it in the environment.

At the biological level, the first area of interest is the ganglion cells that lie on the path between the eyes and the brain. These cells are divided into the X and Y cells, where the X cells respond more strongly to patterns and the Y cells respond more strongly to motion (Goldstein, 1989). Note that these cells always send some response. However, the response is much stronger when the cell encounters the specific stimuli for which it is looking.

Another biological area involved is in the visual cortex. The visual cortex is "the part of the cerebral cortex of the brain primarily responsible for interpreting signals from the eye" (vision1to1.com). The visual cortex is made up of six layers that contain approximately 100 million cells (Goldstein, 1989). Cells within the cortex also respond to certain stimuli. For example, the complex cortical cells respond best to bars of light that have a certain orientation. The hypercomplex cortical cells respond to moving lines of a certain length or to certain angles.

After the biological components of perception come the thought processes involved. These processes involve looking not only at the object that is being identified, but also at the environment in which it exists. For example, when trying to determine the identity of a chair, the thought processes could involve questions such as "I'm sitting in a house, so it makes sense to have a chair here" or "I'm standing out in a field, so it's probably not a chair since it does not make sense in this context."

Trying to mimic the human perception process is a very complex endeavor, and is one that would not be practical with a single machine. However, distributed processing clusters may well provide the capabilities necessary to simplistically mimic human perception. Research at MCMC in FY# 2002 involved developing the design and idea behind such a distributed system. This distributed approach to object recognition is presented here.

The overall method is to use a loosely coupled series of stages as part of the computer recognition system that mimics the human perception process. This system would run over a distributed processing cluster, where certain nodes can be dedicated to perform various tasks, such as edge detection and categorization. The system would use some techniques from neural networks and fuzzy logic to perform the categorizations. For example, fuzzy logic is necessary since it will allow the system to determine categorization on the basis of degrees of membership between various sets instead of absolute and discreet values. Neural network techniques will allow each node to "fire" only when a certain threshold of input strength is met. The nodes in each stage will broadcast their findings to the entire cluster so that any node in another stage can react if it is listening for a specific item. The layers are decoupled so that additional stages

can be added or nodes in a certain stage can be reassigned as necessary. The following diagram illustrates a high-level overview of this approach.

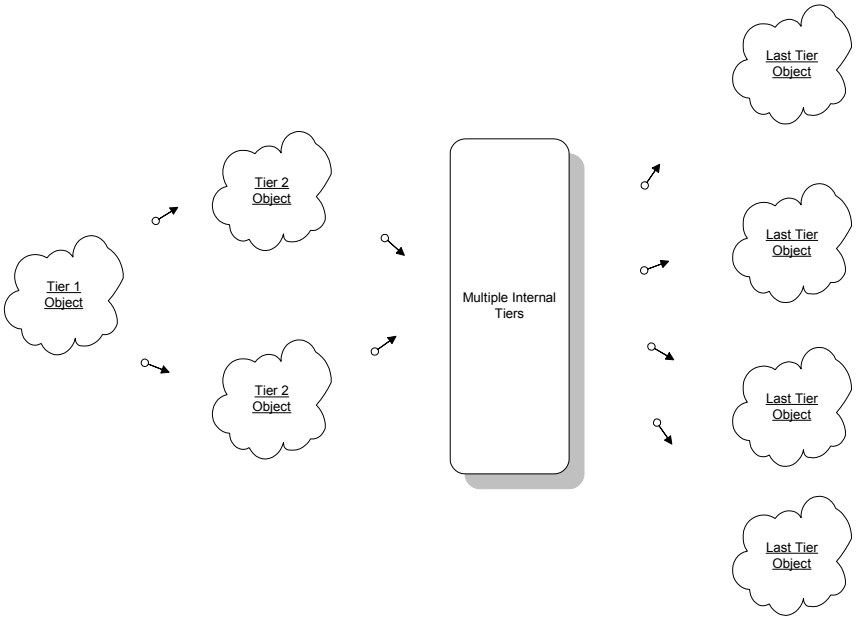


Figure 1. Visual pattern recognition stages.

In this approach, the first stage, or tier, would mimic the biological processes of perception. This would use signal-processing techniques, such as edge detection, to find the edges between various objects in the image. These edges would be converted to a series of small line segments through the application of techniques such as the Hough Transform (CERN). The locations of these segments would then be broadcast to the rest of the system.

The next layer would then be responsible for converting the line segments into basic shapes. This stage would convert them into straight lines, circles, arcs, and so on. This layer would also need to apply some signal processing techniques to smooth out the segment endpoints into meaningful shapes. Once these larger shapes are identified, they are converted into a series of vectors and broadcast to the next level.

The middle stages are responsible for trying to perform some basic feature identification. Some nodes may look for parallel lines, such as roads, while others may look for irregular polygons to find possible locations for features such as lakes or fields. Other nodes in the next few stages could be responsible for finding compound shapes, such as a field made up of tree-shaped objects.

The last layer in the system represents the cognitive stages of perception. This layer listens for the complex or compound object messages sent from the lower levels of the system. This stage will then look at environmental and contextual cues to perform the recognition. Some of these cues will involve texture, color values, reflectance values, and even cues from surrounding areas. This stage may also be done generically, where a node takes all of the information about an object and performs a database lookup to see what matches this object. This is one of the cases where the use of fuzzy logic techniques might be important, as the match is more likely to return varying degrees of membership in certain sets instead of an exact identification.

Some of the preliminary research for FY# 2002 focused on studying edge detectors and various strategies for using them in this type of approach. As the layers are decoupled and respond to messages, multiple edge detectors could be used simultaneously over the cluster. There are many different types of edge detectors and many techniques for manipulating their findings. Some initial work was done to gain experience in implementing edge-detection algorithms in this type of system. Other research went into different techniques to apply during the edge-detection process.

Edge detection is a process that works to highlight changes in intensity within an image. Although there are many different implementations, most can be categorized into the surface-fitting or transformation/filter-based methods. The surface methods work by trying to fit a two-dimensional surface over the image. Detectors in this category typically differ in their use of mathematical models, such as hyperbolic surfaces or derivative models. The filter-based methods apply some form of matrix filter operator to compute an image gradient vector for a given point in the image.

One of the first things tried was experimentation with running edge detectors on the entire image and on subsets of the image. The idea was to determine if the smaller areas would have a significant impact on the mathematical models behind the detectors. Smaller subsets may allow the detectors to perform better by reducing the number of gradients in the local area. The subset method did perform better in some tests. However, overall it did not consistently perform any better than running the edge detector over the entire image.

Research also went into which edge detectors performed better on the types of imagery that would be used for visual pattern recognition for the NMD. A sample image of a Digital Orthophoto Quadrangle was chosen that had a combination of some small urban areas and heavily forested areas with roads running through them. Several edge detectors were examined during the FY# 2002 study to determine their performance with the input imagery. A small clip of the imagery is shown in the following figure.

Figure 2. Sample imagery.

This sample illustrates some features that may be common in input imagery that is fed to an automatic feature extraction system. Dense and sparse tree groupings are present in the image. Roads are present in the image and are also occluded by trees and their shadows in several spots. An irregularly shaped lake is on the left side of the image. Small fields are also present in the image. Two common edge detectors, Canny and Sobel, have been applied to the image and are presented in the next two figures.

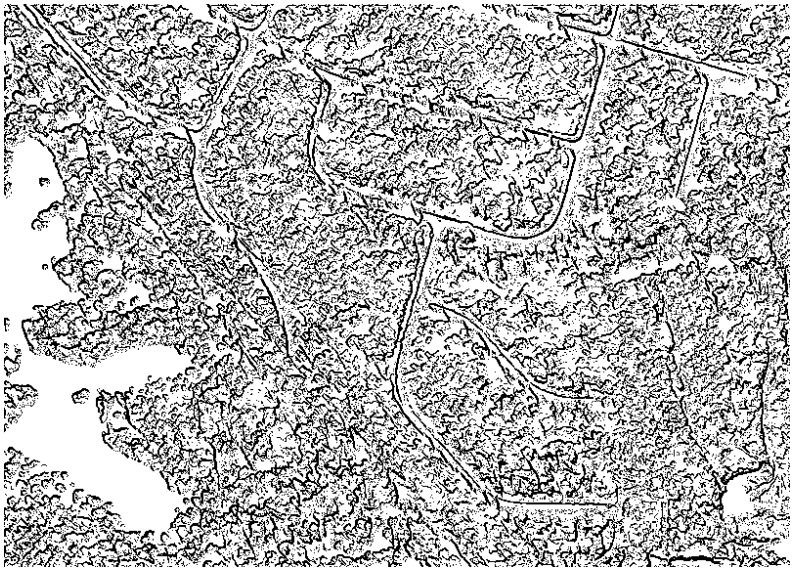

Figure 3. Canny Edge Detector Example.

Figure 4. Sobel edge detector example.

The above figures illustrate the differences between these two common edge detectors. Canny works by first performing Gaussian convolution to smooth the image and then applying a simple two-dimensional first-derivative operator to highlight regions of the image with high first spatial derivatives. Canny also features a non-maximal suppression of the local gradient magnitude (*Feature Detectors - Canny*). The Sobel edge detector uses a 3x3 convolution kernel to perform a two-dimensional spatial gradient measurement of the input to emphasize regions of high spatial frequency that correspond to edges (*Feature Detectors - Sobel*). Canny detected more edges in the input imagery than did Sobel. However, Sobel produces more discreet edges at the expense of discarding some of the weaker edges in the image. The roads in the Sobel image, for example, are easier to identify than in the Canny version. They also suffer less from shadow occlusion in the Sobel version than in the Canny. However, the Canny output also has more edges that could be used to detect individual or groups of trees. This demonstrates how having multiple edge detectors in the first stage of the system can be beneficial in providing the maximum amount of information for the later stages.

The next step of this research involved methods that would generate more discreet edges than are commonly output from the various edge detectors. This is crucial, as it would greatly assist in the generation of line segment vectors from the edges. The example figures illustrate the problem of multiple edges merging together into a single, larger edge. This is a result of how the various detectors operate, since detection based on color gradients can result in thicker bands that surround the various areas. In the Sobel case, larger contiguous features are

easier to detect. However, smaller individual features, such as trees, can get lost in the noise.

Thinning is “a morphological operation that is used to remove selected foreground pixels from binary images” (*Morphology*). It can be used to reduce the output of edge detection to thinner, and hopefully more discreet, lines. It is usually applied to the output of an edge detector and the algorithm checks that it does not break up any continuous lines. This process is generally repeated until no further changes can be detected in the image. However, applying it separately from the actual edge-detection process does present the possibility that some smaller edges will be discarded in favor of larger contiguous ones. Some work went into applying the thinning algorithm during the edge-detection process itself instead of afterwards. It was hoped that it would produce smaller and more discreet lines without sacrificing too many small objects. The results of this method are presented here.

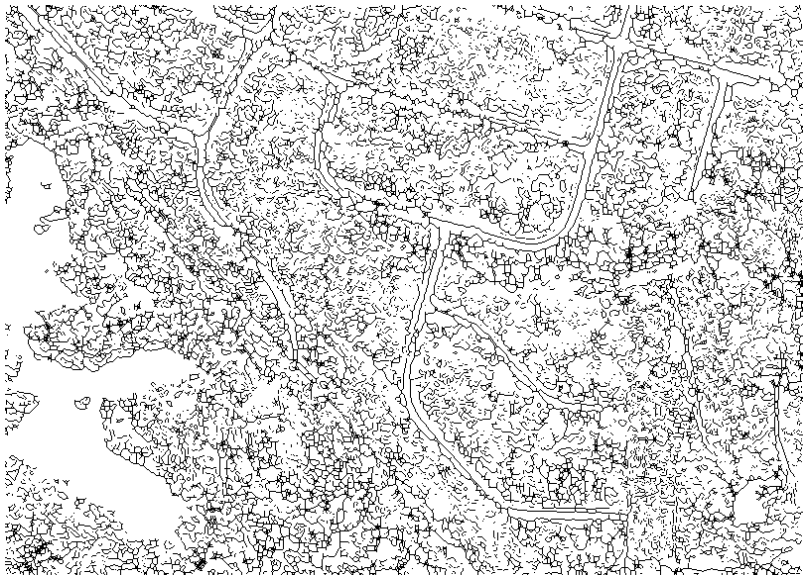

Figure 5. Canny edge detector with thinning.

Thinning applied with the Canny detector does discard some of the smaller edges as noise. However, it also makes some of the longer contiguous features, such as roads, much more discreet. It also helps with some of the shadow occlusion of the road surfaces by cleaning up the noise in their local areas.

Another method that is used to reduce the number of edges is pruning. Pruning is actually another application of the thinning algorithms. Pruning differs in its method of application, as well as in the number of iterations that are applied to the image. It is mainly applied to reduce the number of “spurs” in an image. A spur is an edge that is “produced by the small irregularities in the boundary of the original object” (Morphology). Pruning was also studied to see how it could be applied during the detection process instead of afterwards. This was also run

with a traditional thinning algorithm during the actual detection process to see how well it would work for detecting more continuous edges.

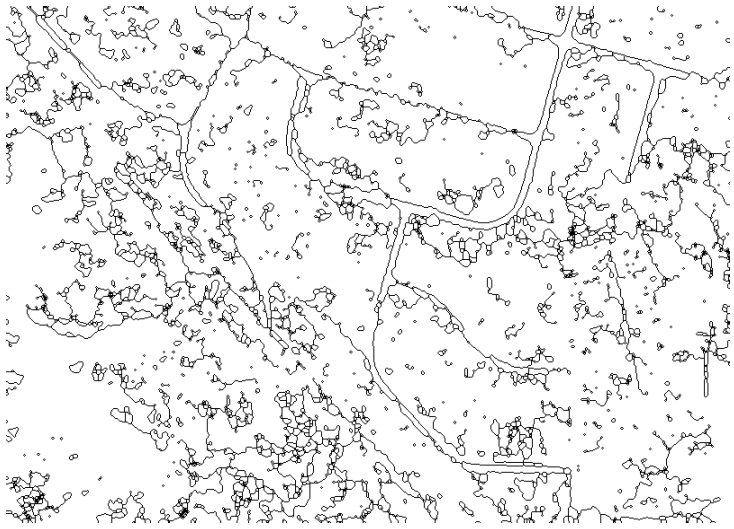

Figure 6. Canny with thinning and pruning.

Thinning and pruning do reduce the number of edges and spurs present in the image. However, they also severely limit the amount of usable information that can be gained. For example, roads are much harder to identify in some areas with pruning applied than when thinning alone was applied. However, some areas also produce a much more discreet line element that could be detected with application of the Hough Transform.

The results of the different edge detectors and different cleaning algorithms highlight the utility of running them simultaneously during the vectorization process that would take place in the first stage of the proposed system. Different detectors and reduction methods can be used to look for different things. This type of functionality would not be reasonable on an individual computer system. However, this is where the true power of a distributed processing system comes into play. Because it is made up of individual nodes with separate memory and processors, a distributed system can easily run several of these algorithms simultaneously. This would enable the first stage of this system to detect far more edges than would be possible if it were limited to a single method.

Some research was also done in FY# 2002 on applying the field of topology to object identification. Topology is “the study of the properties of geometric figures or solids that are not changed by homeomorphisms, such as stretching or bending” (*Dictionary.com/Topology)*. Topology, however, is not the same field as geometry. Geometry is “the mathematics of the properties, measurement, and relationships of points, lines, angles, surfaces, and solids” (*Dictionary.com/*

*Geometry)*. Topology studies qualitative properties that are invariant of shape. For example, a donut and a rectangle are topologically equivalent, because a rectangle can be “stretched” to form a donut and vice versa.

What contributions could topology offer to object identification in geospatial imagery? For many years, mathematicians, particularly in the area of geometry, have studied quantitative (measurable) characteristics of objects. Topologists are mathematicians who study qualitative questions about geometrical structures. This can help to determine invariant relationships between objects, or between various features of the same object. As an example, consider the following London Underground Map.


Figure 7. Map of the London Underground System.

This will not reliably tell you how far it is from Kings Cross to Paddington, or even the compass direction from one to the other. It will, however, tell you how the lines connect up between them. In other words, it gives topological rather than geometric information (Strickland).

The study of invariant features is extremely important to object identification. The same object can appear in many different ways, depending on point of view, lighting conditions, and other environmental factors. One of the problems with the brute force approach to object identification is that it attempts to match solely on the basis of quantitative factors. It can be impossible to warp the known object to get a close match to the input object. Topology provides the ability to examine an object on the basis of qualitative means. This methodology can succeed where other methods might fail, because it simply studies relationships between objects instead of relying on exact measurements. For example, trying to identify intersections between two roads geometrically relies on an infinite number of angle combinations, but topologically the only necessary condition is that the two lines intersect. In identifying structures, rather than looking for all possible shapes with a variety of sizes, topology would identify the characteristics that all buildings have and use that for the initial detection. Using topological characteristics, we can create object classifications. An idea might incorporate the first tier, described earlier, to place features in classes by specific topological characteristics. Later runs would then look into more detail for the exact object identification. After some objects are identified with high probability, the system

can extract information from those objects to aid in the identification of others, similar to the approach of ACR.

## FUTURE WORK

Several items of future work are required for this system. The upper level categorization methods must be further refined. Communication and synchronization between the various layers and nodes per layer must also be resolved, especially if the top layer works in the more generic manner previously described. With such a decoupled system, synchronization techniques become that much harder because the nodes do not directly communicate with each other. The layer definitions must also be refined so that the functionality of each layer is more discreet. The lower level signal-processing work must also be better developed. More edge detection and vectorization research are necessary to fulfill the requirements of the system.

Additionally, we should examine how elevation and spectral signatures can be used to aid in the decision logic. Man-made objects will exhibit different spectral signatures than naturally occurring phenomena. Accurate elevation data, such as LIDAR, can also be used as inputs to help differentiate various types of objects. For example, a road will not only have a different spectral signature than a surrounding field but will also exhibit different elevation characteristics than the natural area. These inputs would then help the detection logic make the final decision as to the identity of the object found by the lower levels.

## CONCLUSION

Distributed processing technologies may pave the way for the development of visual pattern recognition systems that attempt to mimic the human brain's processes. Multiple computers in a distributed system can be allocated to perform certain functions, ranging from mimicking the brain's biological processes to the very cognitive processes that are lacking in current attempts to construct such a system.

This paper proposes a system that uses a series of loosely coupled stages run across a distributed cluster environment. The lowest levels of this system are analogous to the biological specialization that exists in the brain, and the upper levels attempt to mimic the cognitive stages of perception in a simplistic manner. This system is still very much in the theoretical stages, however. In the short term, more research will be done on developing the signal processing techniques that the lower levels of the system will use to perform the initial stages of perception.

Mathematical techniques can also be used to aid the recognition of objects in a visual environment. Topological studies can be used to attempt to match objects based on qualitative measures, while statistical analysis techniques can be used to determine what degree of set membership must be met before an object can be identified. These techniques, when combined with certain elements of computational intelligence, may be able to produce a system that can visually recognize an object without human intervention.

## REFERENCES


Belaïd, Abdel, *OCR: Print*. <http://cslu.cse.ogi.edu/HLTsurvey/SECTION23>.

CERN - European Laboratory for Particle Physics, *Hough Transform*. <http://rkb.home.cern.ch/rkb/AN16pp/node122.html>.

Comay, Ofer, Dr. Paz Kahana, 2002, *Advanced Character Recognition (ACR) Redefines What Automated Recognition is Capable of Reading:* CharacTell, Ltd. <http://www.ngtvoice.com/downloads/whitepaper.pdf>.

Fisher, R., Perkins, S., Walker, A., and Wolfhart, E., *Feature Detectors – Canny Edge Detector:* The University of Edinburgh. <http://www.dai.ed.ac.uk/HIPR2/canny.htm>.

Fisher, R., Perkins, S., Walker, A., and Wolfhart, E., *Morphology – Thinning:* The University of Edinburgh. <http://www.dai.ed.ac.uk/HIPR2/thin.htm#1>.

Fisher, R., Perkins, S., Walker, A., and Wolfhart, E., *Feature Detectors – Sobel Edge Detector:* The University of Edinburgh. < http://www.dai.ed.ac.uk/HIPR2/sobel.htm>.

Goldstein, E. Bruce, 1989, *Sensation and Perception*. 3rd ed. Belmont, California: Wadsworth Publishing Company.

Lexico, LLC, *Dictionary.com/Topology*. <http://www.dictionary.com/search?q=topology>.

Lexico, LLC, *Dictionary.com/geometry.* < http://dictionary.reference.com/search?q=geometry>.

Vision1to1.com, 2002, *Glossary*. <http://www.vision1to1.com/EN/search/glossary/Glossary.asp?Letter=v>.

Strickland, Neil, *What is Topology?* The University of Sheffield. <http://www.shef.ac.uk/~pm1nps/Wurble.html>.